UNIVERSITY OF SOUTHAMPTON
Faculty of Natural and Environmental Sciences
School of Chemistry

A Multi-Stakeholder Investigation into the Effect of the Web on Scholarly Communication in Chemistry

by

Richard William Fyson

Thesis for the degree of Doctor of Philosophy

February 2017

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES
School of Chemistry

Doctor of Philosophy

A MULTI-STAKEHOLDER INVESTIGATION INTO THE EFFECT OF THE WEB ON SCHOLARLY COMMUNICATION IN CHEMISTRY

by Richard William Fyson

As an open, inexpensive, collaborative platform, the Web is ideal for facilitating communication among scholars, enabling near free access to knowledge. The Web's potential goes further still however, allowing researchers to utilise the digital, networked medium to publish more of their research data in comparison to paper-based journal articles, and to publish them in context sensitive formats enabling wider access and visibility, increased discoverability and potential for further use including through e-science techniques. Yet this potential has not been fully realised. Whilst much research is made freely available via open access, this remains a contentious subject, along with other facets of scholarly communication such as peer review and journal impact factors. The range of functions fulfilled by the processes of academic publishing ultimately inhibits the Web's ability to instigate change. This thesis presents an all-encompassing study examining the roles of the major stakeholders in the process of capturing research and making it publicly available, to understand why the full potential of the application of the Web in scholarly discourse has not yet been fully realised. Through doing so a new approach to scholarly discourse, termed *disintermediation*, is formulated whereby the researcher is placed in a central role, using the Web to communicate research findings directly with their peers. An evaluation of *disintermediation* via a socio-economic analysis of researchers' behaviour finds that publisher-enabled mechanisms of reward and recognition are crucial in driving scholarly dissemination. Therefore a more pragmatic approach to instigating change in academic publishing is then investigated. This leads to the development of a prototype Web service to facilitate a process termed *disaggregation*, the breaking down of conventional publications so that their constituent elements may be disseminated and used more freely. An evaluation of *disaggregation* via expert interviews, seeds a final proposal that builds on the disaggregation approach. Whilst not immediately implementable, this proposal recognises the need for a significant cultural shift and engagement from all the stakeholders of scholarly discourse, if it is to truly benefit from the affordances of the Web.

Contents

List of Figures

Declaration of Authorship

I, Richard William Fyson, declare that the thesis entitled *A Multi-Stakeholder Investigation into the Effect of the Web on Scholarly Communication in Chemistry* and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- none of this work has been published before submission
- parts of this work have been published as: Fyson (2011), Fyson et al. (2012) and Fyson et al. (2013)

Signed:.....

Date:.....

Acknowledgements

Many thanks must first go to my primary supervisor Simon Coles, whose support and feedback throughout the whole PhD has been invaluable. Our meetings have always been thought-provoking and enjoyable and I should have arranged to have many more of them! Thank you also to Les Carr for not only providing excellent feedback at key points throughout the PhD, but also for the whole Web Science opportunity.

I would also like to thank my friends in the Web Science CDT, both those who are still studying and those who have long since finished! Special thanks must go to my housemate Chris Phethean for listening to me talk about my work and often giving valuable insight in response, and to Richard Gomer who generously allowed me to develop my Disaggregator prototype using his PHP framework. The CDT experience has made the last few years incredibly enjoyable and fun and I would like to thank everyone who made it all possible, staff and students alike. Thank you also to all my friends throughout my time at university, with special thanks to Chris Asquith whose questioning of my work over the years has no doubt helped to improve it.

Thanks must also go to the EPrints Services team who have kindly employed me whilst I went part-time on my PhD, and have been so accommodating in helping me balance my work and PhD commitments. Special thanks to Justin Bradley also for arranging some much needed technical support. Similarly I would also like to thank Jeremy Frey and his group for letting me work in the safety of their office amongst all the other chemists.

Another round of thanks goes to my interview participants, who took part so enthusiastically. I would also like to thank the EPSRC Digital Economy Programme for funding the Web Science CDT, thus making it all possible to begin with!

Finally I would like to thank my family for their support, especially my Mum.

Chapter 1

Introduction

Initially conceived as a tool for facilitating academic collaboration at CERN (Berners-Lee, 1989), the Web has since gone on to be shaped by both social and technical forces into a number of different directions. The Web's unique properties of openness and neutrality has created a tool that has been utilised to revolutionise a range of industries and has changed the way many of us go about our daily lives, with its reach extending far beyond issues of academic collaboration and communication. The Web provides the capacity to not only allow ideas to be communicated to an audience of any size, but to allow audiences to form around the content they seek, and as such it is a powerful tool for dissemination.

Presented in this fashion, the Web appears to be an ideal platform for academia; an environment driven by the exchange of ideas, data and knowledge. Yet the benefits the Web affords to dissemination have been inhibited. For example, while the Web allows near free access to information which can be reproduced at no marginal cost, much research is only accessible through paywalls. Institutions spend large sums of money subscribing to journals, as illustrated by freedom of information requests made by Lawson and Meghreblian (2015) and Gowers (2014), with subscriptions tending to rise above the rate of inflation (Research Libraries UK, 2014); and individuals not affiliated with a subscribing institution can expect to pay a fee for each article they wish to access.

Furthermore, web technologies mean that research data need not necessarily go unpublished – the technology affords researchers the ability to share all of their work. Before such a rapid and reproducible means of information dissemination was possible, research findings needed to be presented and distributed as articles in journals, a faster and more concise publishing than other mediums such as monographs. However academia now has an opportunity to embrace "transparency, openness and reproducibility", which are "recognized as vital features of science" and yet fails to do so (Nosek et al., 2015). An issue spanning multiple academic disciplines, unpublished data may include negative results (Fanelli, 2011; Franco et al., 2014) or simply "low priority" results and may not

be published for a variety of reasons (Scherer et al., 2015; Franco et al., 2014). As illustrated by the example of clinical trials, greater access may produce significant benefits (Doshi et al., 2013); and as modern research practises develop, conventional publishing approaches may not be sufficient when researchers are faced with the "data deluge" (The Royal Society, 2012; Hey and Trefethen, 2003) and begin to adopt the "Fourth Paradigm" using data to drive new methods of research (Hey et al., 2009).

Whilst problems persist in academic publishing, the Web has nevertheless had a transformative effect on the domain. The open access (OA) movement, conceptualised with the Budapest Open Access Initiative (BOAI) proposed using the Web to provide "free and unrestricted access" to "the world-wide electronic distribution of the peer-reviewed journal literature" (Chan et al., 2002) and has gone on to influence the policies and actions of research funders (Research Councils UK, 2013; Stebbins, 2013), academic institutions (Russell Group, 2014), and publishers (Nature Publishing Group, 2016; Elsevier) alike, leading to an overall improvement to the free and open access of research articles (Harnad et al., 2008; Gargouri et al., 2012). However, whilst OA has been widely adopted over the years (Gargouri et al., 2012), to achieve these aims has been difficult – new policies are being devised to ensure OA in 2020 (Council of the European Union, 2016), 18 years since the BOAI proposed OA and 31 years since the Web was conceived (Berners-Lee, 1989).

Therefore, this thesis aims to critically examine the unique impact the Web has had on academic dissemination, an impact that has failed to have a disruptive or transformational effect on academic publishing and instead simply allows the traditional processes to continue at a faster pace. An examination of the history of academic publishing will look at the various functions it has beyond the dissemination of research, including how the community measures the quality of research outputs and the impact of researchers' and institutions' contributions. A look at the Web's relationship with academic publishing will also highlight its potential to revolutionise scholarly discourse through enabling e-science techniques and presenting alternative approaches to established procedures such as peer review. Building upon this examination, the thesis shall then go on to propose two original contributions to this domain. First the concept of disintermediation and how the Web allows it to be applied to scholarly discourse to present a paradigm shift in the way in which research outputs are communicated when the interactions and relationships between academia's stakeholders are reconfigured. Such a disruptive change has not materialised however and so this thesis will then go on to evaluate why disintermediation is ultimately an unrealistic prospect. The second significant contribution learns from this evaluation to present a new approach for researchers to communicate their research data called disaggregation. Disaggregation examines how research data may be extracted from publications to help the author enhance their reputation in the community, and help the community through access to more research data. This contribution draws on expert interviews with key stakeholders in academic publishing, from

researchers through to publishers, to provide greater insight into the machinations of the scholarly knowledge cycle and what lessons need to be learnt to propose both pragmatic and beneficiary changes to scholarly discourse.

1.1 Scope of the Research

The dissemination of research outputs is an issue that crosses international boundaries, creating an area of study for this thesis which is ultimately too expansive in its scope. Therefore to examine the issues concerning academic publishing and the effect of the Web demands this scope be restricted to prevent the topic from becoming too unwieldy. To this end, the thesis will focus on the dissemination of research outputs from the perspective of stakeholders in the UK's research landscape, the academic publishing context with which the author is most familiar and the one that presents the most opportunities to learn more about developments in this area by attending local conferences and talking to those who work in this sector. However the UK context will only serve as an initial insight into academic publishing and will not form be fundamental to any proposals concerning scholarly discourse or academic practise which should be applicable within an international scholarly context.

Similarly, whilst academic publishing is an international issue, it is also a multidisciplinary one. Different disciplines are disposed to dissemination in various manners: some, such as the natural sciences, place significance upon academic journal articles, whereas the comparatively younger discipline of computer science presents much of its research at conferences for example, some of which have attained equivalent levels of status to prestigious journal publications. Various properties of a discipline shape the way in which its knowledge is shared and communicated, with both the epistemological and social structures and processes influencing dissemination.

Thus to prevent the subject of this thesis from becoming overshadowed by a study of disciplinary differences and quirks, this thesis will focus on the subject of chemistry as a case study. Chemistry provides an ideal setting to experiment with new approaches to scholarly discourse. It is a conservative discipline, as evidenced by a reluctance to adopt an OA model (Björk et al., 2010); whilst a majority of those in chemistry do believe OA will bring benefits to their field, the proportion of those who do is smaller than many other disciplines (Dallmeier-Tiessen et al., 2011). Not only is chemistry a conservative field, but it is also a comparatively commercial one, with the confidentiality of data and its commercial value being an important factor (Long and Schonfeld, 2013). Finally, chemistry is primarily data-driven, a discipline that lends itself to ideas of autonomy (Bernal and Daza, 2010) and big-data driven "e-Science" potential (Hey and Trefethen, 2003), and yet is also a discipline that takes place on a small scale, with small contributions by individual scientists.

As a result chemistry provides an interesting and important case study for examining the relationship between academic publishing and the Web, for if chemists can be encouraged to disseminate their research results in a manner that befits the medium of the Web, then there is much to be gained. Any advancements in understanding within this context may provide a range of potential benefits, from making a researcher's life more convenient by allowing them to spend more time in the lab doing research rather than writing up results, through to broadening the range of knowledge and data available to them, through the publication of vast quantities of data, only made manageable through use of the Web and data mining techniques. And with more information comes the potential for greater transparency and efficiencies, further facilitating reproducibility, a tenet of the scientific method that underpins all chemistry research.

1.2 Thesis Outline

Starting with an examination of the history of academic communication, chapter 2 will identify the significant stakeholders and processes of scholarly communication, establishing the concept of the scholarly knowledge cycle. Alongside an examination of academic publishing will come an introduction to the Web, with a particular focus applied to the success of the open access movement in its ability to change the way in which research stakeholders consider disseminating research outputs.

Having examined the literature around academic publishing and the Web, and having learnt the lessons of those who have contributed to that body of literature, chapter 3 proposes a new approach to scholarly discourse based upon the Web and the interactions of the stakeholders in scholarly discourse. Termed "Disintermediation", the theory outlined in chapter 3, which highlights the researcher-centric nature of academic publishing, is informed by a novel perspective of the subject area, presenting an alternative system for scholarly discourse that achieves many of the same roles, whilst also eliminating many of the issues.

A paradigm shift in the way in which research outputs are disseminated is ultimately unlikely however, and so chapter 4 evaluates why a new method of publishing which may be beneficial to so many stakeholders ultimately fails to materialise. To do so a second literature review is used to investigate a number of theories for explaining behaviour in social systems, and to look at the results from past studies that have sought to better understand or encourage particular behaviour amongst publishing researchers. Through proposing and evaluating the idealistic, yet ultimately flawed approach that is Disintermediation, the entrenchment of conventional publishing practises is highlighted along with the importance of reward and recognition within the research community as a currency for measuring success.

Thus chapter 5 shifts the focus of the thesis on to a new approach for disseminating research, one which strikes a balance between the potential that disintermediation poses, the powerful and influential legacy of conventional publishing and pragmatism. "Disaggregation" proposes that researchers can be encouraged to share their research outputs more openly via the means of breaking down the conventional artefacts of academic publishing such as journal articles, theses and conference posters, into their smaller, aggregate parts. By reducing the problem to smaller transactions, researchers are then given more control over the manner in which they make them available to their peers, alleviating concerns they may have. To explore this proposal, a software prototype is developed within the context of a tool that aims to facilitate research within chemistry. The development process itself helps to inform the disaggregation proposal, a process which is documented throughout the course of chapter 5. The chapter then concludes with a brief presentation of the software prototype and a discussion as to how best to use it to evaluate the concept of disaggregation.

Chapter 6 details a study design to evaluate disaggregation. A number of expert interviews are conducted to elicit the opinions of those scholarly discourse stakeholders identified in previous chapters. The interviews reveal a number of key themes for scholarly discourse in chemistry, highlighting the importance of cultural norms and the role of stakeholders in academic publishing, and suggesting pros and cons of disaggregation and the way in which it has been implemented as described by chapter 5. The lessons learnt in chapter 6 are then used to form the basis of chapter 7 where a new approach to disaggregation is proposed and discussed.

Concluding with chapter 8, the lessons learnt from evaluating disaggregation and the overall discussion of the scholarly knowledge cycle that forms the thesis, are synthesised to generate a number of original proposals for how the work presented may be developed further.

Chapter 2

Background

Before engaging in a deeper discussion of the issues pervading academic publishing and what possible steps may exist on a pathway to resolving those issues, it is important to develop an understanding as to why such complications have arisen at all and to arrive at some definitions for terms that are frequently referred to over the course of this thesis.

To do so requires a brief look at the history of scholarly discourse and the role it plays, demonstrating how the dissemination of research outputs hinges upon a number of different models and frameworks, derives and results in numerous processes, and sits among a complex network of stakeholders each with their own methods and motivations.

2.1 Knowledge Creation

To discuss the dissemination of knowledge, first demands a discussion as to the nature of knowledge itself, what it means and how it is created. Epistemology is the branch of philosophy that primarily examines how knowledge is acquired, presenting two approaches to understanding knowledge. Knowledge may be either *a priori*, deriving from "rational insight into abstract objects"; or it may be *a posteriori*, based upon empirical observations (Landesman, 1997), with both approaches being employed for different needs within a discipline. How these ideas manifest themselves in reality however varies among disciplines, as commented upon by Miller et al. (2008), "different disciplines carry with them different epistemologies, or theories of knowledge" and as such "have a different conception of what constitutes knowledge, how it is produced, and how it should be applied". Quoting MacMynowski (2007), Miller et al. (2008) emphasises the importance of understanding epistemologies, describing how they "drive presumptions about the "relationship between the researcher and the system/object of study and the modelling of system processes" [...] They shape how researchers answer questions regarding the validity of knowledge [...], the legitimacy of methods to produce knowledge

(experimentation, induction, hypothesis testing, etc.) and the assumptions inherent in partcular conceptualizations of the object of study and certain methodologies".

For example, in the natural sciences, knowledge is typically created through the processes associated with the scientific method, with experiments being planned and then repeated to verify the robustness of the results (Repko, 2008). van Gigch (2002a) describes the "Knowledge characteristics" of the physical sciences further still, exploring how knowledge is acquired "on the basis of two different worlds", "*the empirical world*" and the "*the mathematical world*"; verified by an objective approach to study that aims "to keep the object under the study [...] separate and independent from the individual making the observation". Conversely, the social sciences employ "a position diametrically opposed" to the physical sciences, using a "constructivist epistemology" where knowledge is "'constructed' through acts of cognition, such as representation, imagination, understanding, intuition and so on" van Gigch (2002b). Miller et al. (2008) attempts to summarise approaches to knowledge via three metaphors: "knowledge as mechanistic", the "objective, replicable" approach; "knowledge as contingent" emphasising the "importance of agent and context" with knowledge relying on "behavior, variability, and relation to socially held norms"; and finally "knowledge as narrative" which is "interpretive and critical" and where "knowledge is inherent to object and represents values that may be shared or individually held".

Establishing knowledge is not just a matter of philosophy however, it is also a social conundrum. As stated by Walsham (1995) summing up Latour (1987), "science in action, as opposed to accepted science, is a controversial and contested social process of fact construction, in which those who are trying to establish certain "facts" need to enrol and mobilise a broad community of supporters sufficient to overcome particular opposition", thus presenting a competitive community, which whilst no doubt is necessary to ensure the best interpretation of the world around us emerges, may also have the effect of altering the way in which the community at large operates. As commented upon by Walsham (1995) and observed by Latour (1987) the process of creating a community or network which supports certain facts results in people tending "to either ignore new ideas, or take them up but *transform* them" thus changing the way in which the knowledge is constructed or represented and introducing an element of conservatism in the way in which the academic community approaches innovative, new ideas.

Thus both the philosophical approaches to knowledge and the construction of the scientific community determine the way in which new knowledge is created and affirmed, with facts building upon one another, to either expand the overall corpus or to change perceptions on facts that have come before. As such the process of research is inherently iterative, consisting of one step after another, each building upon the last in some way, with each step only being made by the collective weight of a whole community. Latour (1987) asserts that for this iterative process to work a "machine" is required, "a stratagem, a kind of cunning, where borrowed forces keep one another in check so that

none can fly apart from the group". Walsham (1995) develops this further still highlighting the journal as a "powerful "machine"", but it is whole "networks of people, and their organization" that are "critical in understanding the process of fact construction in science". The scholarly knowledge cycle is an example of such a machine, one which has developed over hundreds of years to shape the way in which research takes place.

2.2 Modern Research Environment

An awareness of epistemologies and philosophy provides an insight into the procurement of knowledge. However, another side of knowledge creation exists which concerns a social science perspective and an understanding of the interactions that take place within and around research institutions and how research operates within society. The last few centuries have shown how research makes leaps forward often as a result of the findings of just a handful of notable scientists, whose contributions have been so significant that whole branches or concepts within their disciplines are named after them. But as knowledge and technology becomes increasingly advanced, so too does its impact on society as a whole and the more it becomes "central for successful development" (Kaul et al., 1999). Thus the approaches taken to progress our understanding further still, begin to necessitate more collaborative and resource intensive techniques. Similarly as the demand for solutions to the problems and challenges facing society grows, mechanisms for producing research are required that take into account its qualities as a public good (Kaul et al., 1999).

To use the UK as a case study, much research is publicly funded, with the funds allocated to universities: government ministers decide upon the "overall size of the funding for science and research and its distribution between the Research Councils" and the allocation of money within the research councils is guided by the Haldane Principle meaning "decisions on individual research proposals are best taken by researchers themselves through peer review" (Department for Business Innovation & Skills, 2010). The Higher Education Funding Council for England (HEFCE) provides a "dual support" system, providing both "annual funding for English institutions in the form of a 'block grant', with a number of separate research councils providing separate funding for "specific research projects and programmes" (Higher Education Funding Council for England, 2016). HEFCE allocate "'Quality-related' research funding" where funds are distributed "on the basis of research quality, and take into account the volume and relative cost of research in different areas". How funds are then allocated to individual institutions is then decided upon through mechanisms such as the *Research Excellence Framework* (REF), through which institutions endeavour to report back the impact of their research through the creation of "impact case studies" which aim to report on "changes and benefits to the economy, society, culture, policy, health, the environment and quality of life that have arisen from research" (King's College London and Digital

Science, 2015) – an extensive range of possible impacts. The programme specific funding of the research councils is administered by a number of individual organisations, divided by disciplines: Arts and Humanities, Biotechnology and Biological Sciences, Engineering and Physical Sciences, Economic and Social, Medical, Natural Environment, and finally Science and Technology Facilities; which collectively act together as the RCUK.

Similar approaches to public funding of research are adopted elsewhere around the world. In the US, the National Science Foundation (NSF) funds "research and education in most fields of science engineering" through the allocation of "grants and cooperative agreements", with approximately 11,000 of 40,000 proposals submitted to the NSF receiving funds (National Science Foundation). In Europe, Horizon 2020 is the "biggest EU Research and Innovation programme ever" and welcomes proposals from researchers, whilst structuring funding into a number of key sections including "Excellent Science", "Industrial Leadership" and "Societal Challenges" (Commission, 2016).

It is through "Societal Challenges" – complex problems faced by society that demand answers to maintain our way of life or avoid bigger issues in the future – that indicate how the infrastructure of research has not only developed over time but also come to influence research itself through the manner in which research questions are formulated. Researchers are now faced with the pressing problems society faces such as global food security and energy, that demand interdisciplinary approaches. The research council infrastructure allows the different funding bodies that represent different disciplines to work together in this way, with the RCUK funding cross-council research (RCUK, 2014) to tackle such problems. However this presents new challenges for researchers: as previously discussed epistemological differences exist between disciplines and so years of academic study have resulted in the "siloing" of disciplines which need to be reconciled to prevent the "disciplinary strutucre of knowledge itself" from inhibiting interdisciplinary research (Miller et al., 2008).

Structuring research in this manner, turning it into an industry in which researchers are given facilities and funding for research and yet are also accountable gives greater prominence to the research life cycle. As discussed previously, the generation of new knowledge is an inherently iterative process, which also demands a certain degree of rigour, thus ultimately a cyclical approach emerges: research builds upon previous research yet must also comply to requirements such as being reproducible and being subject to peer review – a research life cycle facilitates this whole process, whilst mapping the ideas behind knowledge creation on to the real world and modern research infrastructure. As emphasised by the *Patterns of information use and exchange* report (Research Information Network, 2009), through studying a number of case studies, a life cycle emerges that consists of various activities which can be broadly mapped on to the *Knowledge Transfer Cycle* described by Charles Humphrey (Figure 2.1).

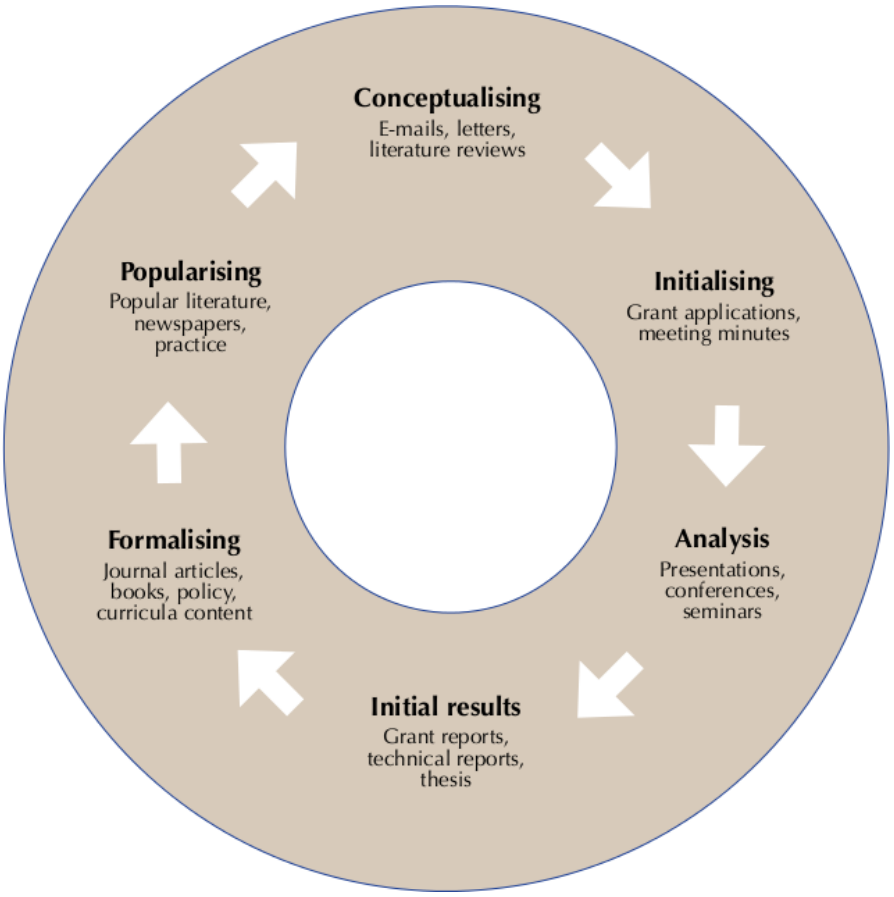

FIGURE 2.1: Charles Humphrey's Knowledge Transfer Cycle describing the various stages that a research project goes through, from "conceptualising" a new idea and gaining support and resources to carry it out, before going on to "formalise" and eventually "popularise" the results so that others may build on the work (Research Information Network, 2009).

The cycle consists of a number of stages that highlight the extent to which the research sector has become a broad and complex place that involves not only the sharing of data and collaboration amongst peers, as well as the "popularising" of results once a research project comes towards its end, but also the competitive nature of applying for grants and presenting at conferences, as well as stages such as "conceptualising" where the research itself is rather more nebulous and not yet fully formed.

Systematising research in this way, may lead to better understanding of knowledge creation and thus lead to new efficiencies and opportunities in the research sector, but it also begins to create a network of stakeholders, and demands for resources and services. When knowledge creation takes place at such a vast scale, how new findings are communicated becomes crucial. This is a problem that researchers and research institutions have never needed to immediately address however as an answer was already present; the academic publishing industry is an old one with a long and influential history.

2.3 A Brief History of Academic Publishing

Society grows and matures as new ideas, theories, works and technologies are developed, whether they pertain to social theories, philosophy, art and culture, or industry and science. For this process to be sustainable and self-perpetuating then a record of knowledge needs to be stored and accessible to those who wish to build upon it, and it should be noted here that accessibility does not simply mean physical access to such a record, but the ability to be able to actually use the record and glean meaning and understanding from it. In short knowledge needs to be communicated not only amongst people, but also through time in a reliable manner which preserves the original intent and findings.

There are a couple of immediately obvious historical developments that have helped with this process. Firstly the development of reading and writing, a most fundamental aspect of learning. Secondly the development of the printing press, with Gutenberg's mechanical model in 1450 often considered to be a key information technology development (Harnad, 1991; Eisenstein, 1980). This allowed a much more rapid approach to reproducing the written word compared to previous methods, which in turn permits a greater spread of writing and a more robust record to be made, simply by allowing numerous copies of knowledge to be created and distributed.

However, whilst these developments can largely be taken for granted now, the beginning of the modern era of knowledge dissemination began in 1665 with the publication of the first academic journal, *Philosophical Transactions of the Royal Society* (Kronick, 1991), to facilitate the dissemination of the many discoveries being made at the time by a growing community of scientists (with early volumes featuring the works of Isaac Newton (Newton, 1671) and Robert Boyle (Boyle, 1665) among many other notable luminaries). The purpose of the journal was established in its introduction, written by founder Henry Oldenburg:

"*To the end, that such Productions being clearly and truly communicated, desires after solid and usefull knowledge may be further entertained, ingenious Endeavours and Undertakings cherished, and those, addicted to and conversant in such matters, may be invited and encouraged to search, try and find out new things, impart their knowledge to one another, and contribute what they can to the Grand design of improving Natural knowledge, and perfecting all Philosophical Arts, and Sciences. All for the Glory of God, the Honour and Advantage of these Kingdoms, and the Universal Good of Mankind.*"

Thus the goal of the journal is made clear from the start and in turn so are the expectations placed on scientists; to create a repository of knowledge that all can contribute to, whilst employing measures to ensure such contributions are valid and should be repeatable, all for the greater good of expanding the boundaries of knowledge - all of which are concerns and priorities that are still at the forefront of scientific communities today.

The inclusion in the list of benefactors from this endeavour of "the Universal Good of Mankind", implies that the knowledge presented within should be exploitable by all and for all, foreshadowing the idea of knowledge as a public good (Kaul et al., 1999) (although how one might envision the "Universal Good of Mankind" may be subjective). For Oldenburg, who worked as a publishing agent of Robert Boyle, the motivation that lay behind the formation of the journal was partly a matter of convenience: he hoped it would "promote the [Royal] Society's activity, save him the labour of copying the same news to dozens of correspondents in half a dozen languages and enable him to earn a decent living" (Moxham, 2015); whilst Oldenburg certainly established an impressive legacy, with this particular journal still in print today, ultimately it failed to furnish him with an independent living (Moxham, 2015).

Whilst this may represent the beginning of the modern journal publication system that is still in use today, the form contributions published in these early volumes took is quite different. Articles are presented in the format of letters that describe in detail the work carried out by the author along with their observations. This format was ideally suited for the manner in which science was taking place at the time, allowing scientists to concisely describe their experiments and helped enable the necessity of reproducibility that is required by the scientific method. However, over time the nature of research has developed as it begins to tackle increasingly large and complex problems and as such becomes increasingly expensive and requires more sophisticated infrastructure to carry out, making it increasingly difficult to convey all the information required in the short form permitted by a journal article.

Thus academic journals were born out of the requirements of the natural sciences community and the need to publish the discoveries of this domain. As scientific enquiry continued to progress and grow more complex, more journals began to emerge to focus on increasingly specific fields of study as well as serving the needs of other disciplines. With journals essentially having a monopoly as the medium through which an author could quickly disseminate their ideas to interested audiences, over time the journals and their publishers have become increasingly significant stakeholders in the processes of academia, research and knowledge creation, which has resulted in a complex legacy.

2.4 The Impact of Academic Journals

Academia is comprised of many different players, with journal publishers being just one. Yet their actions can have an effect that ripples across the network and influences the actions and motivations of other stakeholders. Similarly, the presence of the journals themselves, over time can begin to accrue a status and gravity that may impact the work of those who find themselves striving to become published among its pages. As a result, publishers have found themselves in a position of influence, able to manage and

manipulate the academic community's perception of research dissemination, resulting in other aspects of the research process to become tightly entangled with the actions of the publishers themselves.

2.4.1 Measuring Impact

One such area in which journals have demonstrated their influence is the measurement of impact; that is to say the extent to which an individual or organisation has influenced the work of others in their academic community, any economic success their work may have produced, or the how they may influenced society in some way. With publishers possessing a vast body of work, along with information as to how research is accessed publishers are in a position to provide researchers with data that tells them the affect and reach their work is having. The degree of impact that either a journal or an author can possess is predominantly based upon citations, with the number of times the article has been cited by other articles giving an indication as to how useful it has been as a seed for future research - arguably a sensible proposition due to the cyclical nature of knowledge creation. This data is frequently summarised into a more concise and understandable form as an impact factor, a figure associated with a journal calculated annually by Thomson Reuters. The impact factor is found by dividing the number of times articles published in the preceding two years have been cited, by the number of articles published in total in those two years (Thomson Reuters, 2012). However, intrinsic to this calculation is the grouping of a number of articles together, and in doing so the validity of the impact factor is lessened; it only represents an average article within the journal, an average that no one article may actually reflect. As commented upon by Woodside (2009), summarising the findings of Maffulli (1995) and Vinkler (2005) "the most cited half of the articles are cited, on average 10 times as the least cited half [...] Even the uncited articles are then given full credit for the impact of the few highly cited articles that predominantly determine the value of the journal impact factor".

Other measures of citation impact are also available, such as the Hirsch index, an index proposed in Hirsch's 2005 paper, that aims to quantify the "cumulative impact and relevance of an individual's scientific research output" (Hirsch, 2005). In his paper, Hirsch sets out the index simply: "A scientist has index h if h of his or her N_P papers have at least h citations and the other $(N_P - h)$ papers have $\leq h$ citations each" (Hirsch, 2005). This has effect of only rewarding a scientist for their cited works and not punishing them for their infrequently cited papers; however disadvantages also exist, with the rewarding element restricted: "once a paper belongs to the h-core, it does not matter how many more citations it will receive" alongside numerous other problems as illustrated by Egghe (2010). The upshot is that articles are judged by their authors rather than the company they keep in journals, but even with this change, a researcher's impact on their scientific community, and thus their reputation, is only measured by the articles

they have authored and fails to consider the broader range of academic outputs they may have contributed towards.

Whilst these metrics do provide some insight concerning the contributions a researcher or an institution may have to a scholarly community, they are not without their flaws. For example there is a time component to be considered with some impact not necessarily being realised immediately. The conflict between types of impact and the time it may take to emerge is illustrated by Smith (2001): "scientists would think of the original work on apoptosis (programmed cell death) as high quality, but 30 years after it was discovered there has been no measurable impact on health. In contrast, research that is unlikely to be judged as high quality by scientists - say, on the cost effectiveness of different incontinence pads - may have immediate and important social benefits". The use of citations to measure impact is flawed further still when considering that a citation is unable to convey context. For example Fleischmann and Pons' paper on 'cold fusion' has generated many citations (1,548 listed on Google Scholar at the time of writing), however no other laboratories have been able to reproduce the work, which as highlighted by Moed (2005) demonstrates "that citation impact on the one hand and research 'quality' or contribution to scientific progress on the other, are distinct concepts".

Most significantly however, the use of citations to indicate the impact of research is limited by only considering how other researchers use the work – a measure of one researcher's influence on their peers and successors. But, as identified by the Research Councils UK (RCUK), the body responsible for funding much research in the UK, there are numerous "pathways to impact", some of which are less concerned with academic contributions, instead placing an emphasis on a wide range of activities that fall beyond being published such as "public engagement" and the identification of "relevant users of research and stakeholders" (Research Councils UK, 2014a). The RCUK redeveloped their approach to reporting following the "RCUK Outcomes Harmonisation Project", an initiative born out of the RCUK's "responsibility to demonstrate the value and impact of research supported through public funding" and the need to fulfil various roles enabled via the collection of research outputs (Research Councils UK, 2014c). One outcome of this project was the "Common Question Set", which outlines the full range of research outputs that can be reported upon, ranging from the traditional, such as "Publications", through to other common research activities such as "Collaborations & Partnerships" or "Engagement Activities" to potentially less frequent, but nevertheless significant results such as "Spin Outs" or "Intellectual Property & Licensing" (Research Councils UK, 2014b). The extent to which researchers will report on different outcomes will likely rely on their discipline and area of research, yet the RCUK Outcomes Harmonisation Project's own outcomes demonstrates a shifting nature as to how funders assess the way to which their funds are put to use. One further result of the RCUK Outcomes Harmonisation Project, is the introduction of "Researchfish", an "online facility

that enables research funders and Research Organisations to track the impacts of their investments, and researchers to log the outputs, outcomes and impacts of their work", (Research Councils UK, 2014d), which principal investigators (PIs) will be required to report to. Similarly, other research councils, such as the US National Science Foundation (NSF) are also seeking out impact beyond traditional publication citation impact, as highlighted by Piwowar (2013), stating how the NSF "asks a principal investigator to list his or her "products" rather than "publications" when applying for grants, showing how "to the NSF, a scientist's worth is not dependent solely on publications. Data sets, software and other non-traditional research products will count too".

Yet despite research funders explicitly stating what they value as impact and the disconnect between their definition and the answers provided by journal based indices, it has been noted that researchers have a tendency to become overly concerned with publishing in high impact journals, a behaviour dubbed "impactitis" (Van Diest et al., 2001). Institutions make use of the citation based metrics when selecting work to submit to the REF process (and the similar Research Assessment Exercise (RAE) that came before it). As observed by Trevorrow and Volmer (2012), "it is no secret that department heads have advised researchers to publish in high impact factor journals" and, citing Corbyn (2008), claims "it was 'common knowledge' that universities had used ranked lists of journals to decide which papers to submit to the RAE", even though the REF and RAE are supposedly free from the influence of journal impact factors (Trevorrow and Volmer, 2012). Conversely the use of impact factors during the REF process may help the implementation of what has become a costly exercise. As outlined by Martin (2011), impact itself is a difficult concept to define which means different things to individuals form different disciplines and a "complex assessment system would be very costly to set up and maintain" – "there would be not just the direct financial costs, but also the elaborate bureaucracy to maintain and operate the system". Thus impact factors provide an efficient shortcut in this elaborate, but necessary requirement of funders, a facet of impact factors identified by Casadevall and Fang (2014) who states "relying on publication in highly selective journals as a surrogate measure of quality provides a convenient, if intellectually lazy". Casadevall and Fang (2014)'s reasons for an overt fondness for impact factors are numerous and ultimately concludes "scientists act rationally in their own self-interests despite the detrimental consequences of their actions on the overall scientific enterprise" and alludes to the "tragedy of the commons".

Processes such as the REF are important however; the research community needs to be aware of the impact that various research projects are having so decisions can be made as to how limited resources and funding can be allocated in future. If such judgements are influenced by impact factors, future allocation of funds may be skewed such that quicker or cheaper projects that may yield useful results may miss out on funding opportunities for not being sufficiently complex or exciting. Similarly researchers who make valuable contributions to their field through channels beyond publishing journal articles may not

be adequately recognised for their efforts. In a similar manner to how an education system may not cater for some learners if it fails to consider different learning styles, the tendency for researchers and institutions to aim for high impact factor journal articles may leave some researchers going unfairly unrecognised. It should be noted however that simply because journal publishers provide the information on how articles are cited and used, it does not make them accountable for the way in which it is used and any issues that may arise from its use.

2.4.2 Peer Review

Another area in which journals have influenced scholarly discourse, is that of quality control. There is a need to ensure that what does get published is scientifically sound and can be verified so that any disciplinary corpus does not get swamped by incorrect, unreliable or unoriginal research outputs. With the majority of scholarly dissemination requiring distribution via a journal, the journal publishers naturally occupy the role of gatekeeper being able to choose what gets published, a function carried out both by the journal's editor and the peer review process. The aforementioned Philosophical Transactions of the Royal Society, introduced the notion of peer review and the Royal Society went on to establish a "Committee on Papers" in 1752 to "review all articles considered for publication" (Michaels, 2006). For a long time the use of peer review remained "the exception rather than the rule" (Michaels, 2006), with the "institutionalization of the process" happening in the last century to meet the demands imposed upon editors by ever widening fields of study and increasingly specialised expertise (Burnham, 1990), to ultimately become a firmly entrenched aspect of scholarly discourse, considered to be essential in ensuring that only valid and innovative research is published. Thus, the current sequence of events that typifies the peer review process is as follows: a researcher submits an article to a journal where the editor then acts as an intermediary, first identifying experts to review the article and then passing their comments back to the original author who will have a chance to act on them if appropriate, and a decision will then ultimately be made by the reviewers as to whether or not the article should be accepted into the journal (but the process varies from journal to journal (Smith, 2010)). As the name suggests, the reviewers are other members of the author's academic community who are deemed sufficiently knowledgeable to be able to comment on the veracity of the author's contributions, and as such it should be noted are not employees of the publisher, but instead academic volunteers, for reviewers are typically not paid for their time reviewing.

As outlined by Smith (2010), peer review is used when "deciding which grant proposals will be funded, which papers will be published, who will be promoted, and who will receive a Nobel prize", yet Smith (2010) also notes, "Yet not only do scientists know little about the evidence on peer review but most continue to believe in peer review, thinking

it essential for the progress of science". Like the impact factors discussed previously, peer review is not without its flaws, flaws which become increasingly pronounced as the scale of research dissemination becomes ever greater. The issues that pertain to peer review are of a different nature however, primarily involving social issues, which will inevitably arise when organising a process that involves human judgement.

The most immediately evident problems are that peer review is both slow and expensive (Smith, 2010), a natural consequence of a process which is facilitated by time-poor academics. However problems extend beyond these, such as numerous cases where mistakes in the reviewing process have lead to erroneous conclusions being published, or instances in which professionals in some fields claiming the presence of "cliques" are suppressing some works and supporting some inferior ones (Henderson, 2010). There is also evidence to suggest that not only do peer reviewers favour particular cliques, but that wider societal problems may also arise, such as sexism and nepotism (Wennerás and Wold, 1997) (see (Bernstein, 2015) for a more recent example). The natural propensity for human error or bias in the peer review process may be exaggerated further still when faced with the challenge of reviewing interdisciplinary research, when reviewers may be called upon to comment on disciplines they may not be so familiar with (Lee et al., 2013), thus highlighting how modern research problems require modern approaches across all the facets of scholarly discourse.

A similar problem, although one born out of prestige as opposed to cliques, was demonstrated by one study which resubmitted twelve already published psychology papers to a number of journals (Peters and Ceci, 1982). The papers, authored by researchers from prestigious departments and previously published in highly regarded journals, were resubmitted under fictitious authors and departments but sent to the same journals as before. Upon resubmission, only three were detected as being resubmissions and eight of the remaining nine papers were rejected (Peters and Ceci, 1982). Reasons given for rejection were typically "serious methodological flaws" and with the papers having been previously accepted, this provides an indication towards a bias in favour of recognised and respected names and departments as opposed to methodologically sound and original results (Peters and Ceci, 1982).

Another issue facing peer review, is that of confirmation bias: reviewers are much more likely to agree with arguments that support commonly held beliefs and theories, making it much trickier for new, disruptive ideas to be published and gain traction (Mahoney, 1977). Mahoney (1977) recommends training academics to be better peer reviewers may help to alleviate the problem, but admits it would not be a perfect solution, and it becomes weakened further still when considering the far greater number of papers being published today than compared to over 30 years ago; training all peer reviewers would be a costly exercise and presents no short term solution. And, as Mahoney (1977) observes, a consensus may still be prejudiced if the pool of researchers share the same "ideological or methodological biases". This may be an unavoidable consequence of the

way in which knowledge is generated: as has been discussed previously, for knowledge to be accepted it has be popularised, yet this may stifle more radical and disruptive ideas, and while this may be an inevitable aspect of knowledge creation, leaving decisions as to which findings are accepted into a wider corpus to a small handful of reviewers each time provides an additional obstacle to what is already a challenging and demanding procedure.

The extent of these issues and the concerns they pose to academic communities can be evidenced by the fact that events are still being held today that aim to deal with these problems. Whilst, many of the papers referenced here were authored decades ago, the Royal Society's "The Future of Scholarly Scientific Communication" event held on the 20th April 2015, hosted a debate entitled "Is peer review fit for purpose" which focused on many of the issues (The Royal Society, 2015). The ongoing discussions surrounding the important yet imperfect nature of peer review serve to highlight the high regard in which the academic community at large holds certain aspects of the system they operate within and the inherent difficulty involved when trying to shift well established cultures to better suit new contexts and technologies.

2.4.3 Serials Crisis

Thus it becomes clear that the position within scholarly discourse that the journal publishers occupy, places them in a natural position to facilitate certain services demanded by their target market. By fulfilling these services the journal publishers are adding value to scholarly discourse, allowing researchers to better allocate their funds, identify the best work to build upon and ensure that only academically sound research is published. But the journal publishers have one last impact, one of a financial nature, typically referred to as the "serials crisis" (Butler, 1999). As outlined by Shapiro (2013), the serials crisis is a result of a number of trends in academic publishing. Firstly in the 1980s when "new journals flooded an already crowded market and greater and greater numbers of papers were published", the result was a "declining number of subscribers, on average". As a consequence many institutions had to cancel some of their subscriptions and so the commercially driven publishers increased their prices to account for this shortfall in their income (Shapiro, 2013). Simultaneously "major commercial publishers" have controlled "over 60% of the market for peer reviewed journals", with mergers resulting in "consolidated control within a few hands" (Shapiro, 2013). The oligopoly that has a formed as a result has lead to market failure, with a "correlation between this increased merger activity and journal price inflation" (Shapiro, 2013) that has been left to escalate until even some of the most affluent institutions are in a position where they are now no longer able to provide their researchers with access to the literature they require (Harvard University, 2012).

This problem is foreshadowed by a market failure that arises when dealing with knowledge transfer, for which the normal market mechanisms of supply and demand fail to work efficiently. A potential buyer of knowledge is uncertain of its true value until the knowledge itself has been revealed, upon which moment the buyer no longer has any reason to pay for it (Buckley and Casson, 1976). Whilst the problems of information asymmetries are mitigated by the use of freely available abstracts that present a brief overview of an article's content, journal publishers can nevertheless take advantage of this trait to push on to academic institutions the knowledge for which there is little demand: institutional libraries can subscribe to journal bundles, in which access to a group of journals is purchased in a single transaction, with some of the bundled journals possessing much more value to their researchers than others.

Furthermore, the academic publishers can exploit the imperfect competition market structure that has arisen in the market for academic journals. The market possesses traits of both an oligopoly with only a small handful of publishers possessing the brand value that in turn encourages researchers to publish in their journals and thus necessitates the demand for institutions to subscribe to them; and of monopolistic competition with the different journals each specialising in particular strands of research such that few are directly competing against one another with respect to the content they publish. The influence this unique market structure lends the journal publishers allows them to ask institutional libraries to sign non-disclosure agreements when subscribing to journals (Van Noorden, 2013), which the libraries typically agree to with the understanding that this may result in a lower price Cornell University Library (2015). However the libraries and those analysing the academic publishing market cannot be certain, due to the very nature of the non-disclosure agreements. A study by Bergstrom et al. (2014), acquiring pricing information through the Freedom of Information Act found that journal publishers charged from 3 to 10 times the price per citation compared to equivalent non-profit publishers, although this disparity is reduced for smaller institutions with correspondingly smaller budgets. A further finding of Bergstrom et al. (2014) indicates that this lack of transparency between the buyers and sellers, is resulting in "bargaining efforts [which] will frequently fail to reach efficient outcomes" and thus the "publishers lose revenue that they might have gained from more moderate offers" (and researchers lose access to publications which may have aided their research efforts).

2.4.4 A Complicated Legacy

Journals and their publishers have contributed a lot to scholarly discourse throughout their history and represent a crucial component of academia and its unique processes and methods – Figure 2.2 presents an overview of the various aspects of the scholarly knowledge cycle that publishers have an influence on by controlling the dissemination of research outputs.

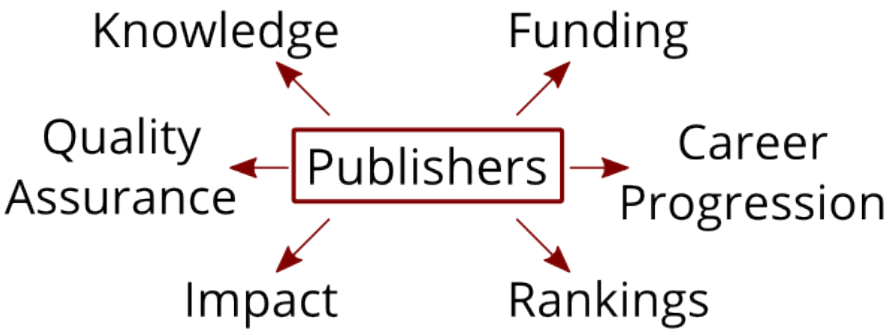

FIGURE 2.2: The many facets of academic publishing. The roles of journal publishers have become increasingly influential over time, to the extent that there are implications for other aspects of the research life cycle. The wide reach of publishers along with their monopolistic characteristics in the academic publishing market, may result in varying levels of efficiency or effectiveness for each of the facets.

Yet they have also taken a lot out of the research sector through their costly subscription fees and through the obfuscation of their flaws be it deliberately or through an unwillingness of their primary users to notice or take action. However they have remained largely unchanged since their inception, whereas the way in which research takes place and the demands that those engaged with research impose upon the infrastructure responsible for disseminating research outcomes has changed drastically. Fortunately researchers are no longer constrained to disseminating their research findings, engaging with their community, and performing the functions required of scholarly discourse via the medium of the printed paper: perhaps the biggest revolution in communication since Gutenberg's printing press is now available to them, the World Wide Web.

2.5 The World Wide Web

Having established how approaches to research and knowledge dissemination have developed over time, it is possible to place one of the more modern and most influential communication technologies, the World Wide Web, within this context. However it would be almost impossible not to consider the Web in this knowledge dissemination context when considering its genesis as an information management system that emerged from CERN (The European Organization for Nuclear Research). Tim Berners-Lee's original proposal concerned "the problems of loss of information about complex evolving systems", a problem which Berners-Lee aimed to tackle using a "distributed hypertext system" (Berners-Lee, 1989). The issues described by Berners-Lee represent a much more fluid and synchronous approach to discourse than has been previously facilitated by academic journals, but sits in a slightly different context. The problems encountered at CERN were similar to that of any organisation that involves staff specialising in different parts of a process or hierarchy and who take that information with them when they leave (Berners-Lee, 1989) (Figure 2.3), a well-documented problem for many

organisations (Cabrera and Cabrera, 2002); whereas academic journals are designed as a venue to publish the results of research once a well-defined portion of the work has been described and completed with some results and a conclusion.

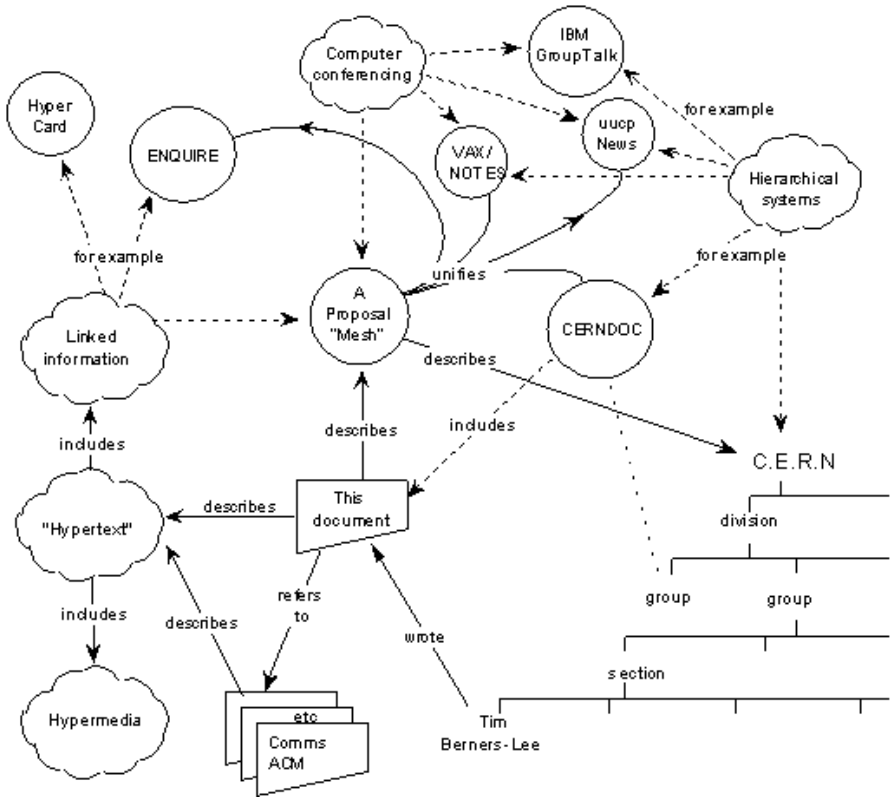

FIGURE 2.3: Tim Berners-Lee's hypertext proposal that went on to become the World Wide Web. The proposal aimed to solve issues of "losing information" by allowing individuals to exchange information in a manner that reflects the non-hierarchical observed communication patterns at CERN and in such a way that allowed the documentation of CERN's "constantly changing ideas" – a "big book" would be insufficient (Berners-Lee, 1989). These initial requirements behind the proposal also reflect those required by academia at large.

Of course the Web has gone on to have a vast impact, well beyond its initial scope as outlined in Berners-Lee's proposal (Berners-Lee, 1989), changing nearly every aspect of people's day-to-day lives, revolutionising the way in which people communicate by allowing its users to receive and broadcast, affording many-to-many communication, at a minimal cost (Shirky, 2009a). The open nature of the Web allows anyone with an Internet connection and sufficient digital literacy to publish what they want and to respond to the communications of others. This idea has so much potential that the Web becomes a complex entity to understand and elicits all manner of social phenomena that can be challenging to predict. The Web means different things to different people. The co-constitutional nature of the Web (Halford et al., 2010) legitimises these differing perspectives, because if the Web is an "unfolding, enacted process" (Halford et al., 2010) then any snapshot of the Web will be derived by the players and the stage where they are playing. For example for those working in the area of knowledge representation,

their view of the Web may place greater emphasis on the semantic web, whereas those looking at cybercrime may choose to place greater significance on the social structures and hierarchies that materialise around the sites of the "Dark Web". Whilst many of the technologies may be shared across these different domains of study within Web Science and the same affordances of the Web may enable the processes that take place, the view of the Web from economic, social and technological perspectives nevertheless differs.

Thus a perspective of the Web is required to investigate how it may relate to research and scholarly discourse. In the context of research, the Web can be viewed as an enabling tool that presents the research community with not only a new, efficient way of communicating with one another, but also a new resource that enables new methods and practices. These developments typically fall under the umbrella terms of e-Science or e-Research, comprising of three strands "the sharing of computational resources, distributed access to massive datasets, and the use of digital platforms for collaboration and communication" (Wouters, 2006), each of which presents new challenges. For example, the problem of the "data deluge", whereby sensors can now provide so much data new standards for data collection, management and metadata are required so that this data can be used in a meaningful and purposeful manner (Hey et al., 2009; Hey and Trefethen, 2003).

This in turn leads to new opportunities for expanding the methods used by scientists. If data is sufficiently well described in a standardised fashion, automatic analysis becomes possible that allows researchers to instantly identify the most interesting aspects of the data that has been created (Hey and Trefethen, 2003). Similarly, in-silico experiments allow hypotheses to be investigated without the prerequisite of a laboratory (Hey, 2005). Thus a whole new market opens up in academia, to provide researchers with services and databases that help them to take advantage of new technologies and train researchers with the necessary skills. And with the new services comes new information that can be disseminated among the community, such as workflows that can be reused and reworked (Goble et al., 2010) and provenance data that provides more detailed insights as to how research results came to be (Borkum et al., 2010; Zhao et al., 2004). Table 2.1 summarises the potential benefits from adopting e-Science techniques.

With regards to scholarly discourse, it is pertinent to frame the Web as an open, neutral and (almost) free to use communication platform where ideas may be published, consumed and responded to. These processes may incorporate any number of web technologies or new social norms that the Web has instigated, but it is also important to consider the overall simplicity, flexibility, and open and collaborative fundamental nature of the Web when probing the issues of academic publishing and the dissemination of research. Once the Web has been framed in this context it reveals an array of possibilities, presenting researchers with new ways in which the collaborate with one another and share resources. As a result any thinking concerning how the Web can transform

Provenance	Enhance trust in results and easier reproducibility Borkum et al. (2010); Zhao et al. (2004)
Collaborations	Collaborations among people and the sharing of equipment and resources Hey and Trefethen (2003), "with the aim to improve efficiency and stimulate greater collaboration" EPSRC (2016).
Data-Driven Research	Big data insight enabled by grid computing, sensors and improved processing power Hey and Trefethen (2003) and *in silico* experiments Goble et al. (2003); Hey (2005)
Crowd Sourcing	Citizen science projects present researchers with an opportunity to engage the general public in the data collection process Bonney et al. (2009); Silvertown (2009)
Dissemination	Opportunities to disseminate more research outputs including workflows Goble et al. (2010) and negative results Matosin et al. (2014)

TABLE 2.1: A brief summary of e-Science techniques, enabled or facilitated by the Web.

scholarly discourse should keep in mind the extra affordances that come with new means of communication.

2.6 Academic Publishing and the Web

Having gained a perspective on what the Web has to offer, it is fair to assume that the Web is therefore in a position to improve academic publishing. With academic publishing involving many stakeholders however, the notion of how the Web may make it better can be observed from many different perspectives. For the publisher for example, the Web may be seen as a tool for ultimately improving profit margins as the costs of distributing publications changes. For institutions and researchers the Web may better scholarly discourse by making research openly available and by conveying more relevant information with which researchers can assess and validate previous results. For funders it may be in its ability to better convey the impact of the work that has been funded.

Academic publishing is already being transformed by the Web, but the extent to which it has been improved is less certain. The Open Access movement (to be discussed in more detail later) has changed the way in which many publications are made accessible (Harnad et al., 2008), a change reinforced by the mandates of research funders (Research Councils UK, 2013; Stebbins, 2013) and has ultimately made access to research more open. However changes have often been more of a process of optimising existing procedures rather than a disruptive one as has been the case for so many other industries. For example the music industry has changed from one based around the distribution of music to a focus on live performances (Weller, 2011), changing consumers relationship

with music as a result, with music streaming services making access to music much easier. The newspaper industry has also been unable to evade the disruptive influence of the Web, with the "traditional rules of thumb such as "price equal to marginal cost"" breaking down for "digital information goods" (Smith et al., 2001), not only must the newspapers find new sources of revenue (Weller, 2011), but their relationship with their readers has also changed substantially (Hermida and Thurman, 2008). However, within academia, the publishing model that was first instantiated in 1665 exists to this day, largely unchanged by the many technical and social upheavals that its publications have conveyed over the years. Whilst it previously permitted a flourishing of research and research communities before the Web, the same model could be argued to be constricting the flow of research now the Web has arrived and pushed the upper boundaries as to what is possible when dealing with matters of research dissemination, new techniques for communicating research data are required (The Royal Society, 2012; Hey, 2005). It is therefore possible to think about the impact of the Web on scholarly discourse in two different ways: how it effects academic publishing, and how it effects the communication of ideas and research outcomes.

As has already been discussed, journal publications provide a service to researchers and institutions, which allows publishers to charge a subscription fee to cover the costs of publishing a journal. Whilst many of a journals costs are not borne by the publishers themselves, such as the production of the journal's content and the process of peer review; journal publishing nevertheless gives rise to a variety of costs. Fixed costs include items such as editorial costs and manuscript management; variable costs cover paper, licensing and distribution; and finally there ares overhead costs such as premises and management costs (The Wellcome Trust, 2004). Running a reputable journal also leads to higher costs, the "more effort a publisher invests in each paper, and the more articles a journal rejects after peer review, the more costly is each accepted article to publish" (Van Noorden, 2013). However the Web presents publishers with an opportunity to rethink the way in which they produce a journal, with many of the variable costs being eliminated. Naturally many of the fixed costs are still present when publishing journals electronically, and whilst some variable costs are removed, others are added; namely the cost of hosting articles and maintaining servers. In a report commissioned by The Wellcome Trust to investigate different business models (The Wellcome Trust, 2004) it was concluded that these changes only resulted in very slight savings. Conversely a report published by the International Network for the Availability of Scientific Publications (INASP) states that electronic publishing can reduce costs although this will be nullified if print editions continue alongside - a product that most readers still demand (Morris, 2006).

The INASP report highlights a number of benefits to electronic journal publishing which may make it seem like a more attractive offer to publishers, although it is worth considering that the report is now almost a decade old and it would be hard to imagine any publishers not taking advantage of the online medium now. Nevertheless benefits

listed included an international reach; opening up the demand side of the market at little extra expense to supply; greater speed of publication, allowing articles to be available as soon as they are ready; additional functionality with basic extras such as linking to other articles to improve the reading experience; and finally the creation of new opportunities and potential for new business models, such as article-by-article publication (Morris, 2006). To capitalise on these advantages, the journal publishers must convey the benefits to consumers too and studies have found that electronic journals do provide a number of benefits for libraries and researchers. Provision of electronic journals leads to a broadening of the range of journals read by students, less time spent identifying and locating articles and are cheaper on a per use basis (Hansen Montgomery and King, 2002).

But the Web also presents an issue for journal publishers. As with all digital goods, supplying a product online results in negligible marginal costs (the cost of producing one extra unit of the good) and thus the cost of purchasing a unit should decrease to reflect this, as only the fixed costs of producing the initial good need to be covered. However this creates a market failure - there is no incentive to create the first instance of the good as one cannot reasonably sell at a price to recoup the start-up costs (Quah, 2003), or in instances where a market has already been established, this introduces difficulties in maintaining existing profit margins. The power of the Web to massively reduce the marginal cost of supply for a product is a phenomenon that has appeared in numerous sectors and has had a dramatic effect on each, forcing numerous industries to adapt their business models. As observed by Kreider (2013), "The Internet seems like capitalism's ultimate feat of self-destructive genius, an economic doomsday device rendering it impossible for anyone to ever make a profit out of anything digitized again". Yet within academic publishing this disruptive force appears to be muted as demonstrated by earlier discussion of the serials crisis and hinted at by academic publisher's profits, with one of the largest publishers, Elsevier, earning an adjusted operating profit totalling £760m in 2015, 37% of revenue (RELX Group, 2015). Van Noorden (2013) also highlights how "commercial publishers are widely acknowledged to make larger profits than organizations run by academic institutions" with profit margins "at 20% for society publishers, 25% for university publishers and 35% for commercial publishers", noting the "irritant" effect it has on researcher with money going to shareholders "rather than being ploughed back into science or education".

Nevertheless, the Web has become the primary communication medium for scholarly discourse, with its ability to allow its users to communicate with one another instantly, at any scale, at a very low cost being of a natural benefit to academia, but these advantages are typically only being employed for two purposes: either informal, often unstructured communication between scholars (The Royal Society, 2012; Long and Schonfeld, 2013) or for distributing journal articles and accelerating article discovery (Renear and Palmer, 2009), in a manner that has befitted the scholarly business model for the last few hundred

years. This may not seem immediately problematic, but with the underlying business model of the journal publisher persisting alongside the Web, large swathes of cutting-edge research are being closed off to all except those who can afford the journal subscription fees, and even those who can afford access suffer an inevitable trade-off.

Not only is the persistence of this form of publishing costly, it perpetuates an outmoded style of communication that is largely unsuitable for modern research methods, having been devised at a time when science was conducted in a radically different manner, which itself would seem significantly out of date to most modern scholars. The anachronous nature of the scientific publishing in the information age has become clear to many scholars however, with many taking action - most notably those who support Open Access (OA). Some disciplines have embraced OA; for example the arXiv preprint server, popular amongst "physicists, mathematicians and computer scientists" attained one million articles in 2014 (Van Noorden, 2014b) and continues to observe steady growth in the monthly submission rate, reaching approximately 9,000 articles submitted in August 2016 ArXiv (2016). In other areas, researchers have taken action against organisations within the existing system, such as those who have signed the Elsevier boycott (Gowers, 2012). Yet ultimately the overall impact of such attempts may have been a little inhibited as a result of the established nature of academic publishing in the wider academic community and how solutions proposed are not necessarily suitable for all disciplines (for example the monograph favouring disciplines of the humanities find OA a problematic approach to dissemination (Suber, 2004)).

It is only by understanding publishing as a multi-faceted process, that possess different consequences for different stakeholders that it becomes possible to fully understand the benefits and changes the Web may bring to scholarly discourse and to engineer an approach that changes publishing in such a way that is compatible with each individual stakeholder's motivations and ambitions. The discussion thus far has lead to an indication as to what the academic publishing process involves, but it is worth considering also what publishing might mean to the scholars of the future. It is clear that everything that publishing encompasses at present is still required by the research community, but the means by which processes such as quality control and peer assessment are conducted need not remain an ossified part of academic publishing, with the Web providing other avenues to satisfy academics' demands. Whilst the Web's ability is principally applied by changing the medium of publishing from paper and matters such as peer review are orthogonal to the method of distribution, the two are intrinsically linked, as illustrated by Marshall McLuhan's phrase, "The medium is the message" (McLuhan, 1994). This concept is further illustrated by Renear and Palmer (2009), summarising how the Web alters the ways in which researchers interact with the research literature, as a network of articles emerges with researchers "chaining references backward and citations forward", creating an every growing body of literature and ultimately affecting the manner by which researchers make judgements of "relevance, impact and quality".

Ideally, with the Web allowing fast, scalable and inexpensive communication, academic publishing should extend to allow dissemination of all research outputs in such a manner that allows them to be acknowledged and valuable in a variety of contexts, with results not just exposed via what is published in journals. Rather than a small number of peer reviewers selected by an editor deciding which research is of significant interest to be made more widely available, instead the wider research community could serve as their own gatekeepers, bringing attention to those research projects which are of most interest and relevance. The Web already provides a near perfect environment for markets based around the concept of the long tail to flourish (O'Reilly, 2007) and thus seems ideal for the dissemination of research where there are so many niche specialities. The publication of more research should also allow for greater reproducibility in science and a more efficient allocation of resources, which in turn should ultimately lead to fewer research projects being needlessly duplicated – a concern for many scientists with "selective reporting" and a "pressure to publish" often identified as causal factors (Baker, 2016). Similarly, by removing elements of arbitrariness from publishing, peer review also has an opportunity to be parted from the publishing process, by allowing the community as a whole to verify scientific soundness rather than a select few reviewers. And with much research being simply disseminated as articles that have followed the same format for decades (if not longer), academic publishing has been presented with the chance to exploit the digital medium offered by the Web to allow research outputs of any format to be not just accessible but also recognisable as useful contributions to the community. Figure 2.4 gives an indication as to the potential of web-based publishing where the digital medium is taken advantage of to offer services beyond the abilities of conventional journals.

Funders now recognise the potential of the Web in expanding publishing opportunities. As already touched upon, the RCUK has sought to reassess the impact of funds through the Outcomes Harmonisation Project. Similarly, researchers are also expected to provide data management plans when applying for grants to ensure that access to the data is made available. In the US, the National Science Foundation (NSF) expect all research projects to make best efforts to publish resulting data where appropriate (National Science Foundation, 2010), and the RCUK's "Concordat On Open Research Data" recognises the importance of research data as "a valuable result in itself of the research process" (RCUK, 2015). The concordat makes a compelling case for making research data more widely and openly available, citing and the "societal benefits from making research data open", including "economic growth, increased resource efficiency, securing public support for research funding and increasing public trust in research" (RCUK, 2015). The concordat goes on to outline a number of principles including, "Open access to research data is an enabler of high quality research, a facilitator of innovation and safeguards good research practice" and "Good data management is fundamental to all stages of the research process and should be established at the outset" amongst many others, which together help form a number of expectations of the research community

FIGURE 2.4: The affordances of the Web present a number of opportunities to resolve existing issues and improve upon a number of facets of academic publishing. Many of these areas are currently left either under-served or highlight the flaws in current approaches to publishing. By fully taking advantage of the digital medium of the Web, it should be possible to remedy many of these issues, to create a publishing environment that better serves researchers and the future of knowledge creation.

including the notion that "researchers will make their research data open and usable within a short-well-defined period" and that research institutions "will seek to provide appropriate access to infrastructure systems and services to enable their researchers to make research data open and usable" (RCUK, 2015).

Whilst it is clear that the Web provides many opportunities to rethink facets of academic publishing, it also provides a resolution to perhaps the most pressing issue that requires addressing; that of the serials crisis and costs. The future of publishing has the chance to be a much more open process, free of many financial barriers. Issues of cost are complex, providing infrastructure to disseminate research is not cheap; however a moral argument remains: where research is funded by public funds for the benefit of society, a case can be made that such research should be open for any who wish to use it or build upon it to further the public good that it enables.

The image of publishing in the future is one of openness and flexibility that removes restrictions and allows researchers and the wider community to use the outputs of research as they see fit. It represents a more distributed structuring of the various processes, allowing functions such as the ranking of peers to sit outside of the publishing system, not built into it where their presence distorts the flow of knowledge. However whilst it is easy to imagine a future of academic publishing where many of the existing issues are resolved and new value-adding services and methods have been woven into scholarly discourse, it is important to consider the nature of this publishing environment, which comprises of a multitude of different disciplines, each of which has its own nuances and

approaches to dissemination. As with publishing pre-dating the Web, it is likely that one size does not fit all and new systems will need to be wary of the context in which they operate. It should also be noted, that when discussing a future academic publishing landscape, it is easy to forget that many initiatives and innovations are already in motion, meaning any analysis needs to consider not only the legacy of long standing stakeholders, but also the modern history of academic publishing.

2.7 Open Access

To downplay the impact of the Web upon academic publishing as merely maintaining the status quo, whilst allowing it to play out at a much greater pace and scale would be to devalue the actions of those behind the Open Access (OA) movement and those who support and promote it; who have recognised the Web's potential to ease and optimise the dissemination of research and acted upon it. The Web combined with scholars' immutable desire to push the boundaries of their fields should allow for the production of an "unprecedented public good" - the intent of the Budapest Open Access Initiative (BOAI), which envisaged a two-fold approach to opening academic discourse through author self-archiving after conventional publishing, or via author-pays OA journals (Chan et al., 2002). The BOAI is based on two fundamental principles: "the willingness of scientists and scholars to publish the fruits of their research in scholarly journals without payment, for the sake of inquiry and knowledge" and that the Internet allows "the world-wide electronic distribution of the peer-reviewed journal literature and completely free and unrestricted access by all scientists, scholars, teachers, students, and other curious minds" (Chan et al., 2002). With these two principles in mind the BOAI comes to the conclusions that with "tools and assistance" scholars can deposit their "refereed journal articles in open electronic archives", and to complement this activity, "scholars need the means to launch a new generation of journals committed to open access, and to help existing journals that elect to make the transition to open access" (Chan et al., 2002).

As mentioned above, there are two approaches to open access, "gold" OA whereby an author publishes their article directly in an OA journal, and "green" OA, which involves publishing an article in a non-OA journal, but then self-archiving it in an OA archive (Harnad et al., 2004). Since the original BOAI proposal, both routes to OA have witnessed significant developments, fundamentally changing the way in which researchers access, and consequently cite articles, with OA articles being recognised as offering a citation impact advantage (Swan, 2010; Björk and Solomon, 2012; Harnad and Brody, 2004). Both of the routes have their own benefits and concerns however, which have been reflected in the way in which OA has been adopted by different stakeholders in academia.

Gold OA has been accomplished through the arrival of OA journals – academic journals that researchers pay an article processing charge (APC) to publish in, with some of the most significant OA journals originating in biomedicine and coming from new publishers BioMedCentral and Public Library of Science (PLoS) (Björk and Solomon, 2012), although as noted by Björk and Solomon (2012) the "importance of the APC business model for funding OA journals has grown rapidly" with established publishers adopting the practice. Gold OA has the advantage of providing an immediate solution to the access problem and can easily be integrated within the existing publishing system; the existing journal publishers use their infrastructure to provide open access to an article straight away for a fee – deemed a much "better value" proposition than subscriptions (Van Noorden, 2013). However, this fee is often still expensive for many institutions and researchers (Solomon and Björk, 2012; Björk and Solomon, 2012) and when combined with the need for universities to continue subscribing to journals and articles which are not available via OA, results in a financially unsustainable situation of "double dipping", whereby "a publisher gaining from two income streams, APCs and subscriptions, in a way that its overall income from the same customer rises" (Pinfield et al., 2015).

Gold OA has also given rise to the issue of "predatory publishing" (Beall, 2013), arising from the gold OA journal's conflict of interest "the more papers a journal accepts, the more money it makes", an issue which Beall (2013) believes may threaten scholary societies. Beall (2013) tackles this problem with a "black list" of predatory publishers and a further investigation by Bohannon (2013) has helped the Directory of Open Access Journals (DOAJ) to filter out such journals. Berger and Cirasella (2015) however recognising that "low quality publishing" is not just the domain of OA journals and that Beall's list may compromise OA, highlights the role librarians can take in steering authors towards the most appropriate journal for their work, and through doing so "counter-act the misconceptions and alarmism that stymie the acceptance of OA" – an effect that can also be achieved by drawing researchers attention towards repositories and self-archiving.

Self-archiving and repositories represent the green OA route; a path which allows academic institutions to support their researchers in making their articles open, with a number of software providers such as EPrints, DSpace and Fedora, each contributing towards a competitive repository market. Green OA also possesses a number of advantages. Firstly it builds upon the existing approach to academic publishing. Researchers still publish in the usual manner and free access to literature comes at no direct cost to the author or reader as per gold OA or conventional publishing, enabling rapid adoption of open access with over 90% of journals allowing the green route to OA (Harnad et al., 2008).

However free access can only be achieved by the author depositing a version of the manuscript in a repository – green OA demands a robust infrastructure. Such an infrastructure has been built upon the work of the Open Archives Initiative (OAI), an

organisation with the aim of creating a framework supporting "federation and interoperability" so as to enable "increasingly multi-disciplinary" "scholarly endeavours" and ultimately present a "low-barrier solution" (Lagoze and Sompel, 2001). The OAI also firmly established the need for openness "from the architectural perspective" with the two most popularly deployed repositories, DSpace and EPrints (University of Southampton, 2015), adopting open source models , recognising its ability to allow users to "modify or extend" repositories to "meet local needs", and to help "institutions and organizations with minimal resources" (Smith et al., 2003). Through the work of the OAI and the repository community, there are now 3093 repositories recognised by the repository authority OpenDOAR, with 84.5% of these being institutional repositories (OpenDOAR, 2016b), a number achieved through steady growth enabled by the repository community (Figure 2.5).

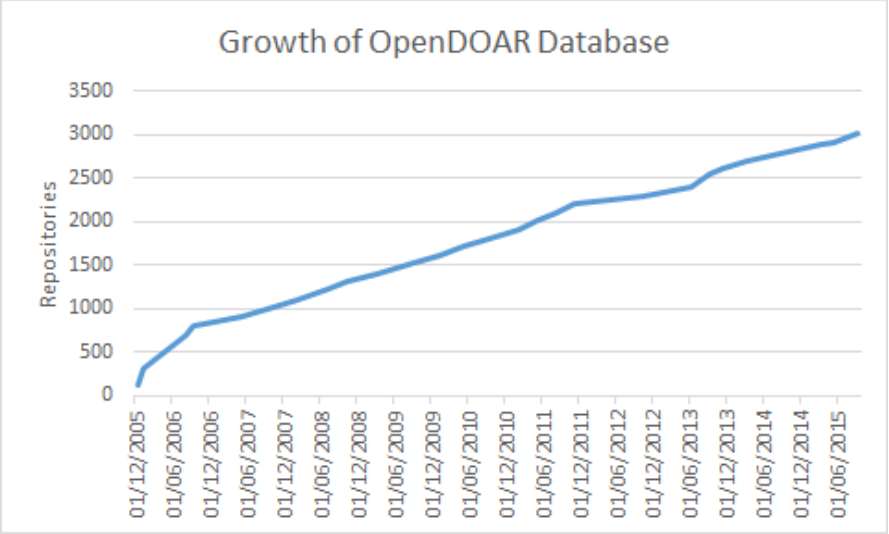

FIGURE 2.5: The growth of repositories registered in the OpenDOAR database, showing a steady growth over the last 10 years (OpenDOAR, 2016a). As noted by OpenDOAR (2016a), this only "reflects the work of OpenDOAR rather than the growth of the number of actual repositories" indicating the actual number of repositories may be much greater.

It is through the development of the institutional repositories (IRs) that the, the Web's ability to change the way in which research is disseminated, can lead to value added services. Whilst an IR may cost an institution through its development and maintenance, it also provides a platform from which the institution may establish "digital-asset management" plans (Chan, 2004), researchers can distribute preprints with their peers (Gibbons, 2009), and by curating "intellectual property" in one place rather than it being "diffused through thousands of scholarly journals" institutions and researchers alike can demonstrate their "scientific, social and financial value" (Crow, 2002). Not only are there benefits for the hosting institutions, but, as per the intentions of the OAI, "metadata [...] about resources available via a well-defined protocol", "allows other organizations (service providers) to harvest it" enabling "external services" such as "cross-repository searching, current-awareness, and reference linking" (Lagoze and Van de Sompel, 2003).

IRs are now recognised as "essential infrastructure for scholarship in the digital world" (Lynch and Lippincott, 2005).

However green OA is not without its issues. Firstly it is dependent on the "sustainability of the existing journal system" (Craig et al., 2007). The sustainability of the existing system with the introduction of green OA has been called in to question; the Association of Learned and Professional Society Publishers (ALPSP) highlighted this in their response to the RCUK proposing that "fundees should self-archive [...] their final drafts of journal articles" (Berners Lee et al., 2005), claiming it would "inevitably lead to the destruction of journals". However, as highlighted by Berners Lee et al. (2005), if the publishers continue to add value that would not be present in the researcher's pre- or post-print deposit (Berners Lee et al., 2005), for example by facilitating the peer review process, "self-archiving and journal publication can and do continue to *co-exist peacefully*". Nevertheless, as indicated by Björk et al. (2014), "many authors can be assumed to knowingly break copyright rules by posting the actual published articles", creating concerns with "both the nonsystematic storage and breach of copyright" making "long-term availability uncertain". Similarly, Björk et al. (2014) identifies "author behaviour" as "barrier to green OA" simply because "many authors are unaware of what they can do and lack of time and other priorities also prevent frequent uploading", with Rowlands and Nicholas (2005)'s study backing up this claim where 38% of survey participants were unwilling to deposit their articles. This in turn leads to low uptake of green OA, despite its potential, with Björk et al. (2014) finding "green OA coverage of all published journal articles" to be "approximately 12%".

The second concern for green OA is the effect of publisher imposed embargo periods, which limit the availability of articles via an OA route. Laakso and Björk (2013)'s survey investigating delayed open access, found this practise to be most common with high-impact journals, which typically used embargo periods of up 12 months. With such policies common amongst publishers, the repositories community has looked to the history of academic discourse to provide a solution (Sale et al., 2014), building into repositories a mechanism that allows authors to share preprints "under the provisions of fair dealing in the world's Copyright Acts". The 'Request-a-copy' button, presented to users trying access embargoed articles on a repository, will email the author of the paper who has the opportunity to accept or reject the request, and whilst initial surveys indicate approximately half of authors ignore the emails (Sale et al., 2014), it nevertheless indicates ways in which social pressures, leveraged by technological developments, can encourage openness in academia.

Thus with OA being implemented via different means, different stakeholders in academia have varying perspectives of how it should be achieved. The last few years has seen a number of key developments for the OA agenda in the UK with the publication of the Finch report representing a significant milestone. The Finch report (Finch, 2012),

entitled "Accessibility, sustainability, excellence: how to expand access to research publications", chaired by Dame Janet Finch, was charged with "the important question of how to achieve better, faster access to research publications for anyone who wants to read or use them", and ultimately came to the conclusion of supporting a predominantly Gold OA approach, but nevertheless allows researchers to pursue either option (Figure 2.6). This outcome met with a mixed response, with publishers such as Elsevier welcoming the changes proposed (Wise, 2012) whereas the Director General of the Russell Group, Dr Wendy Piatt speaking on behalf of a number of the UK's leading research universities, would have preferred to see "more consideration to the possibilities of the 'green' model" (Piatt, 2012). A year after the publication of the Finch report, a further report was released which declared the "government was mistaken in focusing on Gold as a route to full open access" (Business Innovation and Skills Committee, 2013), highlighting the financial consequences of simultaneously asking institutions to continue with their subscriptions to journals in order to maintain access to much of the existing literature, and to pay OA fees that the Gold OA model demands. A much more recent review of the consequences of the policies, carried out by the RCUK found that the although both green and gold OA routes are compliant with the policy, "the fact that the policy states a clear preference towards the gold model has caused confusion" and there are indications that "the policy's strong preference for gold is creating a two-tier system with those favouring the green route following a perceived inferior route" (Burgess, 2015). The report also describes how comments from the publishing sector "highlight the use of funds for green as not being a proper use of funds" (Burgess, 2015), a point picked up on by Research Libraries UK's (RLUK) David Prosser, who highlights the unusual nature in which "providers of a service (the publishers) are trying to dictate to funders and institutions what is and what is not a proper use of the funds that those funders provide to the institutions", going to note "It is hard to think of any other sector in which this would tolerated and it is rather odd that the publishers' comments are repeated in the Review" (Prosser, 2015). The presence of these comments in the review do serve to stress the unique scenario of academic publishing and the complex relationships that exist among its stakeholders.

Not all OA developments are supportive of the Gold route however. The policy for the post-2014 Research Excellence Framework establishes that "to be eligible for submission to the post-2014 REF, authors' outputs must have been deposited in an institutional or subject repository" (Higher Education Funding Council for England et al., 2015), providing a strong incentive for institutions to ensure that their best research is openly available.

FIGURE 2.6: The Finch report's recommendation as to which OA routes researchers should follow showing a preference for gold OA where possible and green OA as an alternative should gold not be possible. Image: (Research Councils UK, 2013)

2.7.1 A Critique of Open Access

OA has done much to improve the accessibility of academic publications, but its ultimate goal has been to take the disruptive nature of the Web and apply it to improving the accessibility of research, all whilst simultaneously maintaining the status quo. Making access to research more open to all as defined by the BOAI is naturally a worthwhile pursuit, yet the conventional approach to OA fails to grasp the full potential of the Web; a not unreasonable expectation considering it pre-dates so many of the Web's developments and the ever evolving nature of its use that has emerged since 2002.

As such, some have attempted to develop the ideas of OA further still. Recent developments have seen the arrival of a third type of OA, typically called "Diamond" OA (Gowers, 2013), which turns its focus to more than just issues of accessibility. Diamond OA professes to offer a free to read and free to publish approach, via the medium of an "overlay journal", such as the service offered by the Episciences project which proposes to "promote the emergence of epijournals, namely open access electronic journals taking their contents from preprints deposited in open archives [...] that have not been published elsewhere" (Centre pour la Communication Scientifique Directe, 2013). The method for securing free and open access here is to utilise software that facilitates the traditional publishing process such that the cost to publishing and reviewing academics is no greater than that imposed by their interactions with conventional publishers (Gowers, 2013).

Therefore Diamond OA goes beyond merely dealing with issues of accessibility, as noted by Steven Harnad, who disagrees with prepending new "colours or precious metals" to OA, stating that OA is "not about cost-recovery models (nor about peer-review models); it is about research access" (Harnad, 2013). Ultimately Harnad comes to the conclusion that overlay journals are simply Gold OA journals funded by either subsidies or author fees, when the subsidies are eventually exhausted (Harnad, 2013). Yet his comments on some of the wider aspects of the publishing process and how they do not pertain to OA, or rather how OA does not concern these other aspects, reveal why OA alone may not be a sufficiently thorough proposal, albeit perhaps a suitably subversive one (Harnad, 1995), when considering the potential of the Web to disrupt the established system. Diamond OA begins to take advantage of some of the Web's properties with it incorporating elements of the publishing process such as peer review (Centre pour la Communication Scientifique Directe, 2013). Nevertheless, the Web could be used to shake up the academic publishing system further still, with the potential existing to allow a more agile research landscape, able to combat some of the other problems that afflict scholarly discourse.

It should be noted, that whilst OA does primarily concern "research access" rather than "cost-recovery models" (Harnad, 2013), the impact of widespread adoption would have repercussions for the other stakeholders in the academic publishing ecosystem, primarily the learned societies that accompany many disciplines and depend on publishing revenues to fund other activities which are themselves often identified as public goods and recognised as contributing towards the "international standing of UK research" (Gardner, 2013). This concern has been noted by Dame Janet Finch, chair of the aforementioned Finch report, who has stated it is "important to give learned societies, which are often heavily dependent on income from their publishing arms, 'time to adjust"' (Jump, 2013). The role of learned societies is not one academic communities would wish to see diminished, yet shifts towards OA threaten their business model. Thus whilst OA may not concern the financial aspect of scholarly discourse (or many of its other facets), it is closely connected and as such it is a necessary angle to consider when debating the merits of different OA flavours and reflecting on its continuing impact and legacy. The potential for negative consequences from adopting OA, does not mean it should not be pursued however, as with every new approach to scholarly discourse that is developed, there lies the opportunity for new sources of revenue and business models to materialise which can in turn support organisations such as the learned societies in what is a persistently "uncertain publication environment" (Gardner, 2013).

This brief overview of Open Access, describing the different ways in which it may be implemented, the varying preferences held by different stakeholders and the publications of recent reports that aim to navigate the politics that has emerged around academic publishing all serve to further illustrate the unique relationship amongst its stakeholders. OA as described by the BOAI should facilitate a scholarly utopia where the motivations

of researchers to produce and share their findings ("for the sake of inquiry and knowledge" (Chan et al., 2002)) are sufficient to realise the end goals of OA. So why are all articles not freely accessible? Similarly why was the Finch report required eleven years after the Budapest Open Access Initiative? The answer to these question may lie in the mutually dependent relationship that exists between researchers and journal publishers. Assuming that researchers are in fact self-interested, economic agents, their actions will reflect what is best for them; which in the instance of the publishing researcher would be to secure a publication in a journal perceived to have high impact, thus enhancing their own and their institution's reputation and status in the academic community. Such developments may in turn lead to increased pay or chance for promotion, but simultaneously provide a motivation for keeping knowledge to one's self until the moment when publishing the research can yield the greatest benefit.

A counter to this argument would be to suggest that an OA article may, through being open to a greater readership, receive more citations when compared to a non-OA article (Swan, 2010) and in doing so will broaden and enhance the recognition of an author. But ultimately it is often the power of a journal's brand and its history that have a tendency to influence the choices of a publishing academic. The advantage of the OA journal may not also be immediately perceptible when the often expensive costs of publishing in a gold OA journal are their first experiences of this new approach; an approach, that when placed next to the mechanisms of career progression in academia, has an undesirable air of vanity publishing.

OA has had a profound impact on academic publishing, changing the attitudes and policies of all stakeholders (Research Councils UK, 2013; Stebbins, 2013; Russell Group, 2014; Nature Publishing Group, 2016; Elsevier). Yet, as remarked by David Prosser during the Repositories Fringe 2015 keynote speech, it's impact needs to be questioned: "There was the idea that open access and repositories would revolutionise scholarly communications. There was the idea that we would solve librarians financial problems concerning the serials crisis. There was the idea that publications would be free for all to read including SMEs and members of the public.". Speaking to an audience of mostly institutional repository managers, Prosser goes on to say "We have failed to engage researchers adequately and I think that the busyness of academics is an insufficient reason to explain that. So why have we failed to engage and to get academics to see this as something they should do on a daily basis?". It becomes clear that OA alone is insufficient to instigate a significant upheaval of scholarly discourse, although it may nevertheless be a crucial component. Ultimately it may be the ideas of reward and recognition which are driving forces in scholarly dissemination and their influence should not go underestimated. It is this aspect of the dissemination process however that perhaps reveals the thinking behind numerous new innovations in this field that fall outside the realms of conventional academic publishing.

2.7.2 Open Access to Data

As has already been touched upon, researchers are increasingly expected to make their research data "open wherever possible" (RCUK, 2015), with funders (the NSF (National Science Foundation, 2010), or ESRC (Economic and Social Research Council, 2015) for example) and publishers (Nature's *Scientific Data* journal "mandates the release of datasets" Springer Nature (2016)) both having been known to require researchers submit their data to a repository. This introduces a new dimension to open access; the opportunity to publish research data, whilst being of great benefit both within and outside of the research community, presents new dilemmas as to how this should be implemented.

As noted by Klump et al. (2006) data publication, like research articles, also needs to "comply with the principles of Open Access" meeting "criteria of accessibility, persistent identification and long-term availability". More recently Lawrence et al. (2011) notes that formally publishing data is not a simple process and demands procedures to be "put in place for peer review and formal citation of datasets". Kratz and Strasser (2014) summarises the debate about data publication and comes to the consensus that published data "is publicly available now and for the indefinite future" and that "published data must be adequately documented such that, at a minimum a researcher in the same field could reproduce or reuse it" – an aspect of data publication Kratz and Strasser (2014) describes as being "frequently the most laborious step for researchers in taking data data from useable within the lab to useable by others" [*sic*]. To help ensure that data remains "discoverable", "open", "linked", "useful" and "safe" (Parsons et al., 2011), numerous data repositories have been established to support the requirements of data publication. Whereas the wider OA community has seen the rise of institutional repositories, and whilst many institutions do run their own data repository, many data repositories are organised along disciplinary boundaries (Springer Nature, 2016), with the "highly heterogeneous and distributed" nature of research data demanding "simpler" and "flexible" technological solutions (Parsons et al., 2011).

However whilst the benefits of openly publishing data are well grounded, and the scholarly infrastructure is largely present to facilitate such processes researchers are not always keen to engage with such practises, with Key Perspective Ltd (2010) highlighting how data management policies are "viewed by researchers as bureaucratic obstacles to be negotiated in the quest for research funding" and that at present "compliance with data-related funding policies [...] is not especially well monitored and sanctions for non-compliance are rarely applied". Parsons et al. (2011) also highlights the "cultural diversity" of those who collect their data, citing this, as opposed to the volume of data" as being one of the "greatest challenges" of data publishing. The issues are further compounded when considering the need for research data "to be integrated with reference and community data" (Parsons et al., 2011) and with publication repositories being

grouped by institution and data repositories based around disciplinary boundaries a disconnect between the two facets begins to emerge, with publications ultimately taking priority from the perspective of the researcher Key Perspective Ltd (2010).

2.8 Academia 2.0

The Web has permeated many aspects of many people's lives, not only changing the manner in which they consume content, but also giving new opportunities for interactions, and avenues for producing and broadcasting content (in a multitude of mediums). It is therefore no surprise that the Web is beginning to make an impact upon the day-to-day business of research, in areas beyond the conventional research dissemination process that the OA movement largely concerns itself with. As a result a number of new platforms have emerged, alongside some of the more established ones, as enterprising individuals find new ways to add value and provide services specifically for researchers.

A number of social media platforms have been co-opted for academic use. For example, (Veletsianos, 2012) highlights scholarly usage of Twitter, categorising it under a number of themes such as: "Information, resource, and media sharing", "Expanding learning opportunities beyond the confines of the classroom", "Requesting assistance offering suggestion", "Digital identity and impression management" and "Connecting and networking". Similarly Bik and Goldstein (2013) outlines a number of "benefits from an online presence" including "a powerful way for scientists to boost their professional profile", "a way to share journal articles, advertise their thoughts and scientific opinions, post updates from conferences and meetings, and circulate information about professional opportunities and upcoming events." Gu and Widén-Wulff (2011) describes how "Social media provides technical support for scholarly writing" through mediums such as "blogs and wikis", which are "suitable to cultivate the habit of online scholarly writing and collective writing" and Kirkup (2010) describes how blogging may help researchers develop a "multiphrenic" identity, one that is "not only created out of a variety of narratives, but performed and presented through a variety of media", helping academics in the "continuous development of a professional 'self'.". The use of social media in academia centres around the concepts of identity, discovery and engagement, helping a researcher to establish themselves as part of the wider academic community; an idea promoted by some academic societies. For example, the Royal Society of Chemistry's (RSC) road show events run a session entitled "How can I use Social Media to raise the profile of my research and to connect to the scientific community?" (Royal Society of Chemistry, 2014), during which attendees are suggested to "use social media to raise the profile of your research", to engage on blogs and become virtually involved in academic conferences.

Identifying the influence and potential of social media on scholarly communication, some have sought to establish social media sites specifically designed for academic purposes. Van Noorden (2014a) discusses the way in which ResearchGate and Academia.edu, two academic social networking sites – "a scholarly version of Facebook or LinkedIn" – are changing the various facets of publishing. Describing ResearchGate's co-founder, Ijad Madisch's vision for the site, Van Noorden (2014a) quotes "a key venue for scientists wanting to engage in collaborative discussion, peer review papers, share negative results that might never otherwise be published, and even upload raw data sets", emphasising ResearchGate's position to provide an innovative, new service. Similarly Van Noorden (2014a) describes "San Francisco-based competitor Academia.edu says that it has 11 million users. "The goal of the company is to rebuild science publishing from the ground up," declares chief executive Richard Price". However Van Noorden (2014a) finds that users' motivations may be more conventional: "The most-selected activity on both ResearchGate and Academia.edu was simply maintaining a profile in case someone wanted to get in touch – suggesting that many researchers regard their profiles as a way to boost their professional presence online", suggesting that the sites simply complement existing dissemination practises.

Using social media, both multi-purpose platforms such as Twitter, as well as those designed for a scholarly audience, is not without its issues however. For example, with regards to Twitter and blogging, Veletsianos (2012) notes how this can lead to a situation "where personal and professional identities blend". Researchers may also seek to share their articles via these platforms, but as evidenced by Elsevier's insistence that Academia.edu remove papers which authors have published through them (The Economist, 2014), this presents legal complications, and thus the benefits of scholarly social media use may be inhibited by the existing frameworks of academic publishing. Where OA has been developed on top of established practises, scholarly social media is an emergent practise that academic stakeholders need to respond to. As indicated by Veletsianos (2012) "online practices may not be valued or understood by peers and academic institutions even though scholars themselves may have found scholarly value in participating in online spaces", presenting a conflicting scenario for a would be user of these services.

However, as discussed earlier, funders are adapting their approaches to impact assessment, looking for products of research beyond publications (Piwowar, 2013). Thus the demand for "altmetrics" (Priem et al., 2011) has arisen in an attempt to track and measure new forms of impact that scholarly social media facilitates. As outlined in the altmetrics "manifesto", the ever growing use of online reference managers and social media channels such as Twitter, presents an opportunity to improve upon and complement some of the functions of dissemination that the traditional filters of peer review and citation based indices are struggling to support with the ever growing body of research outputs, and allow us to measure the influence of the "hallway conversation",

which has now "moved to blogs and social networks" (Priem et al., 2011). The notion of altmetrics has been implemented by organisations such as Altmetric (Altmetric, 2015) and ImpactStory (Impactstory, 2015) which provide users with a range of quantitative measurements that give an impression of how far their work has spread, typically giving values as to how widely user's work has been 'tweeted' or shared across a variety of sites that host research outputs. As outlined in the altmetrics manifesto, there is a danger here that should be considered: do "altmetrics really reflect impact, or just empty buzz" Priem et al. (2011). With scholarly discourse on the Web actually representing a vast network of human interactions, such metrics may often lack the vital context that makes them truly meaningful. For example, even something as simple as a 'retweet' can be interpreted in different ways which have different consequences, ultimately making it very tricky to establish what the influence on the scholarly community has actually been, and providing even further complications if one were to take a "Big Data" approach to this analysis (Tinati et al., 2014) in a bid to understand the aggregate impact of contributions made in these spheres.

Whilst the impact of academic social media use may be difficult to measure, it nevertheless illustrates Marshall McLuhan's concept of "The medium is the message" (McLuhan, 1994). As has already been discussed, the medium of the journal article affects the way in which the various stakeholders, communicate and measure the impact of research. Now too, the Web is changing and influencing the way in which researchers communicate with one another and share their research outputs, with new services and metrics arriving to help facilitate this change, and it is allowing funders to interact with and assess the researchers they fund in new ways which should hopefully go on to result in more efficient funding allocations in future (Research Councils UK, 2014c). However with so many different platforms emerging, offering different facilities and expectations, combined with the potential ephemeral nature of online services, researchers may be left in an anxious state: researchers may be faced with the possibility of repeatedly reporting on their research outcomes to the different audiences that different sites reach, along with the possibility that a service they use will one day no longer be available (not to mention the sophisticated level of digital literacy required to properly understand the effects and consequences of using the Web in this fashion). Thus it becomes clear that enacting a change to the established habits and norms that dominate scholarly discourse is not an easy task in this multi-stakeholder, multi-faceted environment; especially when researchers' careers and the allocation of public funds are dependent on the outcomes.

2.9 Conclusion

Examining the history of academic publishing and the socio-technical impacts of the Web (both actual and potential), makes it clear that the realm of scholarly discourse

is a complex landscape. The scenario at present depicts a market which has been established over hundreds of years, by a number of stakeholders, each with their own roles and agendas. Thus to define what is meant by the oft-mentioned phrase 'scholarly discourse' in this thesis, it refers to the overall system in which researchers share with one another the results of their work, for either purposes of citing and building upon results, reviewing and critiquing, or to demonstrate skills, knowledge and impact. The primary transaction that takes place in the system is the exchange of journal articles, typically brief summaries of a researcher's work published in an academic journal which may exist either offline or online (and commonly in both), assembled and published by journal publishers. 'Scholarly discourse' does however include processes which feature other stakeholders, namely research institutes and their associated libraries who fund researchers and subscribe to journals respectively, so that researchers have all the resources required to perform their work. Finally 'scholarly discourse' may also include the less formal communications that occur between scholars themselves in the sharing and doing of research. In short 'scholarly discourse' can be used to refer to any communication that involves a researcher relating to their work, or any communication made between two parties, where one is acting on behalf of the researcher.

Having defined scholarly discourse, a number of concerns are revealed. Firstly there is the problem of the 'serials crisis' (Butler, 1999), an issue that affects research institutions around the world. Access to much research literature is not cheap and yet to distribute research does require infrastructure and does require administration, thus the process itself is costly (The Wellcome Trust, 2004). The extent to which the cost of subscription pays for the cost of distribution is now questionable however with reduced marginal costs, as is the structure of this market: should publicly funded institutions be paying for access to publicly funded research? The legacy of traditional academic publishing does not just have financial implications however, with peer review (Wennerás and Wold, 1997; Henderson, 2010; Lee et al., 2013; Peters and Ceci, 1982; Mahoney, 1977), reproducibility (Baker, 2016) and the misplaced importance of citation impact (Casadevall and Fang, 2014; Van Diest et al., 2001), all presenting issues which should be addressed. Publishing results through journal articles is ultimately seen to be an outmoded method of communication, useful for briefly communicating findings, but using a format not suitable for the scale and interdisciplinary nature of much modern research, and via a business model that results in uncomfortable trade-offs for research institutions.

Not only are there issues that need to be addressed, but the Web presents new opportunities. Whilst the success of the OA movement is responsible for many beneficial developments, its characterisation of bringing openness to academia as being an issue of access to papers fails to realise much of the Web's potential. The Web can enable a much more open culture with the potential for all data and results to be shared, a culture which, if enthusiastically adopted by the research community, would allow for not only greater access to research findings by more people and organisations, but also

brand new opportunities for adding value and preventing the unnecessary duplication of work that may have failed and gone unpublished previously.

To fully take advantage of the opportunities that are presented by new Web technologies and the social changes they have wrought (and vice versa) requires thinking about the problem not simply in providing the necessary means via technology, but also the wider system in which these technologies reside and are expected to operate; it necessitates an interdisciplinary approach. Open Access provides an illustrative case study. OA approaches the well defined problem of access to research and presents a simple solution of either creating technology to enable access (repositories) or a social change to enable it (pay to publish, open access journals), yet in doing so implementing open access from both a technological and policy perspective has been a slow, and at times contentious, process due to the way in which it has implications for a wide range of stakeholders and processes. The holistic view of scholarly discourse reveals there is a complex system of institutions and organisations comprising of a myriad of individuals all of whom need to demonstrate their own impact and value – the more ambitious of whom will want to use this information to not only expand the boundaries and advance the sum of human knowledge, but also, expand the boundaries of their own careers and the reach of their affiliated organisations. Taking such a view will provide further insight from which to inform future scholarly discourse developments.

Chapter 3

Disintermediation

3.1 Identifying a Social Framework for Scholarly Discourse

An analysis of the literature and listening to and observing the open knowledge and open science communities, highlights a number of issues with scholarly publishing. As such, lots of research has taken place to document and define the issues, such as that surrounding the serials crisis for example (Butler, 1999), as well as studies that endeavour to promote the advantages of adopting more open approaches, from citation advantages that benefit researchers (Swan, 2010) to paradigm shifts that would advantage whole communities (The Royal Society, 2012). The reasons that lie behind this constant need for pushing and promoting more open and modern means of communication are less well understood however.

It can also be noted that the arrival of the Web has had a peculiar affect upon academic publishing. The most significant upshot of the Web on the dissemination of research is that it allows the old processes of academic publishing to happen at a faster rate and on a bigger scale, but it has failed to disrupt them with few significant innovations or changes to the overall infrastructure of academic publishing. A small proportion of the academic community push for change and new approaches are slowly being introduced but as a whole the system has become ossified.

Thus, the interconnected nature of scholarly discourse is revealed, with the literature highlighting the numerous interactions and roles taken by those who act together to form parts of a much larger system. This in turn can inform the epistemology and methodologies required to study the subject further, in an attempt to better understand how the research community may be encouraged to take advantage of the affordances of the Web to resolve some of the aforementioned issues or to produce innovative approaches to disseminating research online.

Looking at the system and infrastructure that is scholarly discourse through the roles and interactions of the individual actors that make it reflects and recalls ideas concerning structural functionalism, as proposed by Talcott Parsons when discussing the social system and the role of individuals in achieving societies overall goals. As summarised by Turner (1991), Parsons argued against the assumptions of classical economics "that economic actors are rational and egoistic" theorising that if this were true, society would break down into an "atomistic" free for all and that economics own arguments for solving this problem are contradictory and "incoherent" (Turner, 1991). Parsons came to a different solution: "a culture of common values" which promotes individuals working cooperatively, creating "a realm of relatively autonomous social values which cannot be reduced to material interests or environment" (Turner, 1991). This is described more concisely by Kingsbury and Scanzoni (1993) who describes ""Actor's" conformity to a set of preexisting standards that promotes the greater good of the larger whole to which Actor [sic] belongs" and is elaborated on further still by Alexander (1990) who describes the "moralism" of Parson's theory, "His [Parsons] actors are imbued with a desire to be good, and they are understood as trying to conform with principles that express this aspiration. Because moral principles are reference points points for human action, human beings want their institutions to bear an appropriate relation to them". Thus a connection between the choices people make and larger social constructs emerges, with ideas of moralism and choices seeming particularly pertinent when considering the motives and actions of those researchers who are enthusiastic supporters of openness in scholarly discourse, and presents sociologists with an opportunity to measure how such functionalism works by examining the choices made by individuals.

Parson's work has been the subject of much criticism however, mostly for its "hyper-abstract, logically faulted" perspective and its almost naive approach in discounting the "role of material forces, technology and social class in shaping the social structure" (Turner, 1991), significant criticisms when considering an attempt to apply structural functionalism to explain the role of the Web in scholarly discourse. However, as would be expected of any academic community, the ideas behind structural functionalism have been built upon, with Merton using "theories of the middle range" (Merton, 1968) to derive a more nuanced approach at examining social systems. Middle-range theory exposes the flaws of structural functionalism, that "it is hopeless and quixotic to try to determine the overarching independent variable that would operate in all social processes, or to determine the essential feature of the social structure" (Boudon, 1991) and instead attempts to derive theories "that lie between the minor but necessary working hypotheses that evolve in abundance during day-to-day research and the all-inclusive systematic efforts to develop a unified theory that will explain all the observed uniformities of social behaviour, social organization and social change" (Merton, 1968) (or more succinctly "theories [..] that involve some abstraction but are still closely linked to observations" (Carroll and Swatman, 2000)).

When considering social systems Merton therefore devised his own "paradigm for functional analysis" (Merton, 1968), which criticised the assumptions required of functional analysis, "functional analysis assumes or explicitly operates with some conception of the motivation of individuals in a social system", and introduces new types of functions to allow for a more nuanced perspective of the actions of individuals and the manner in which these create a social system. The first is "Manifest functions", those functions which are "intended and recognized by the participants in the system"; with the second being "Latent functions [...] which are neither intended or recognized" (Merton, 1968). Other elements of Merton's paradigm include the notion of a "unit" to explore the notion that actions may be either "functional for some individuals and subgroups and dysfunctional for others", and "functional alternatives, equivalents or substitutes" used to introduce variations to the manner in which a function is performed (Merton, 1968).

Scholarly discourse represents a social system wide in scope and scale and to apply Merton's framework for functional analysis to it, an interpretivist approach will be taken to explore this area. As described by Schwandt (1994), "The constructivist or interpretivist believes that to understand this world of meaning one must interpret it. The inquirer must elucidate the process of meaning construction and clarify what and how meanings are embodied in the language and actions of social actors. To prepare an interpretation is itself to construct a reading of these meanings; it is to offer the inquirer's construction of the constructions of the actors one studies.". This represents a significant contrast to positivist approaches "where it is assumed that the "objective" data collected by the research can be used to test prior hypotheses or theories" (Walsham, 1995). In the process of trying to understand why scholarly discourse has been affected by the Web in the way it has, interpretivism presents several advantages, permitting a holistic approach that allows systems as a whole to be interpreted and by allowing findings and ideas to emerge over the course of the analysis, and for ideas to be evaluated within their specific contexts.

3.2 A Functional Analysis

With consequences of scholarly discourse involving many more issues than just the dissemination of research outputs, and the arrival at the conclusion of needing to take a holistic perspective when considering academic publishing, it becomes necessary to identify the interactions in the process in more detail, and to identify the function of each actor. Merton's framework for functional analysis permits the necessary degree of freedom and flexibility for exploring the functions that constitute scholarly discourse, however it should be noted that in this instance the actors in question are not all humans. Actor network theory permits treating different entities, be they either social or

technical, as "inseparable" and should be "analyzed with the same conceptual apparatus" (Walsham, 1997) – an important yet necessary perspective when conducting any Web Science research with its inherent mix of the social and technical (Halford et al., 2010).

By identifying the various stakeholders, their motivations for engaging in the system and their concerns; it then becomes possible to formulate a picture as to how they interact and once these connections have been drawn, how they may be rewired when considering the social and technical framework in which they exist, which is to say with a consideration of the social norms of academia and the capabilities of the Web.

Table 3.1 presents a list of the stakeholders and endeavours to list what they provide and expect from scholarly discourse so as to illustrate their functions. The stakeholders have been identified as a result of the background literature; the currency of scholarly discourse is research outputs in the form of journal articles and theses, and as such the stakeholders described here are all affected by this currency in some way. This list is by no means exhaustive, especially as it fails to reflect unique roles that may only be found in particular disciplines, but it aims to include all of those stakeholders that are found across disciplines. It should also be noted that whilst some of the stakeholders may seem tangential, with research councils looking for research impact to have societal and economic gains as well as academic ones (Research Councils UK, 2014a), the reach of academia becomes quite wide.

Stakeholder	**Example**	**Goal**	**Provide**	**Expect**
Researcher	PhD students, Professors	Do research, Enhance status in community	Research findings	Research findings, Reputation/Status
Institution	Universities	Maximise output, Increase ranking	Organise research, Provide facilities and services	Researchers to secure grants, Funding from govt., Students
Discipline	Chemistry	Understanding	Epistemologies	Disciplinary corpus, Quality assurance
Research Council	NERC, EPSRC	Report academic, societal and economic impacts	Allocate funds, Report back impact	Demonstrable academic, economic and/or societal impact

Continued on next page

Table 3.1 – continued from previous page

Stakeholder	**Example**	**Goal**	**Provide**	**Expect**
Government	Department for Business Innovation & Skills	Achieve academic, economic and societal impacts	Funding to research councils	Academic, economic and/or societal impact, Growth
Journal Publisher	Nature Publishing Group	Profit	Research findings, Organise peer review, Formatting, Impact factors, Source of recognition	Subscribers, Readers, Profits, Publishable findings
Institutional Library	University Libraries	Provide researchers with access to research	Manage journal subscriptions and access	Funding from institution
Learned society	Royal Society of Chemistry	Sustainable income to support service members	Publishing outlets, Conference organisation, Accreditation, Source of recognition	Funding via membership, Funding via subscriptions
Third party database	The Cambrdige Structural Database	Sustainable income/profit	Scholarly resource, Outlet for research data, Value adding services	Non-open access databases require subscriptions, Free, crowd-sourced databases require user contributions
Third party metadata	Web of Knowledge, Scopus	Profit	Metadata aggregators, discovery and metrics	Subscriptions
Student	Undergraduates	Education, Employment	Source of funding, Future generation of researchers	Teaching, Research led learning environment
Industry R&D	GlaxoSmithKline	Develop new products and services, Profit	Collaborations with universities, Research findings	New products and services, Profit

Continued on next page

Table 3.1 – continued from previous page

Stakeholder	Example	Goal	Provide	Expect
Research User	Hospitals, SMEs	Improve products/services	Research Questions, Resources to conduct research, Use cases	Access to research to enhance products/services

Table 3.1: The actors present in scholarly discourse, illustrating their ultimate goals or motivation, what they provide to the system and what they expect in return.

As demonstrated in Table 3.1, the stakeholders of scholarly discourse are a varied collective, that together represent a market; with government on the consumer side of the market, demanding research takes place so that it might benefit from its findings and elicit technological or societal progress, and with researchers and the various infrastructure that surrounds them emerging to serve the supply side of the market. This scholarly market differs from a more economically conventional consumer, producer, and sometimes intermediary, set-up that is more often observed however, with many of the stakeholders described above not easily placed on one side of the "transaction" or the other. Researchers are the agents responsible for both producing and consuming research and thus many of the main transactions that take place in the provision of research occur within in this group. The majority of the other stakeholders (perhaps with the exception of "Industry") are all present to either enable or facilitate the work of the researchers, some of whom are more vital than others. For example, stakeholders that enable researchers to do their work include disciplines, which are required to provide structure and epistemologies for researchers to follow; society which provide public funds so researchers can afford to their work; and institutions which are responsible for providing a venue and resources with which research can take place (although it could be argued that institutions are not strictly vital to research process, they simply make it more economical, structured and feasible).

Some stakeholders provide useful services, but are however wholly dependent on what other stakeholders provide or expect from the system: research councils allocate funds from one stakeholder to another, journal publishers and learned societies distribute research among the researchers, often via institutional libraries, and third party organisations fill a gap in the market to facilitate the work of researchers by providing or

aggregating data from other sources. Whilst these stakeholders all provide useful services, in some instances they may add little original value; it is these stakeholders that represent the intermediaries in this market.

3.3 Transactions of Scholarly Discourse

Table 3.1 explains how each of the actors have their own, individual role to play in the process of scholarly discourse, but the interactions between the actors themselves is not so clear. Figure 3.1 attempts to illustrate these interactions, describing what services the stakeholders provide to one another, which in turn reveals two different goals that academic publishing ultimately contributes towards.

When considering academic publishing from the perspectives of the different stakeholders, two separate markets start to emerge. First there is the "macro" level in which society demands research to be conducted in order to achieve technological and social progress, improving overall quality of life and well being. In an attempt to satisfy this demand for research, a second "micro" level market arises, in which transactions take place chiefly among the researchers. In this market, the good that is being traded is research outputs, necessary for the continued generation of yet more research, putting researchers in the rather unusual position of both being the suppliers and consumers of the good that they are exchanging, a relationship which can be identified by the loop in Figure 3.1 where research stems from the researcher, through to a journal publisher or learned society, on further towards the institutional library before it is then passed back to the researcher.

At first inspection there should be no reason why researchers cannot conduct these transactions among themselves, but the sharing of knowledge presents a complex conundrum. Not only do the typical market mechanisms fail when exchanging knowledge, due to information asymmetries (Buckley and Casson, 1976), but there are externalities associated with this exchange of research. By making one's research available to the wider community, it is put under scrutiny and so peer review takes place. Similarly, it provides a means by which researchers can allocate reputation and status upon individuals and institutions, which ultimately has an affect on the way other interactions among the stakeholders take place (for example how money is distributed by the research councils).

It now becomes clear why a holistic view of scholarly discourse is required, with the decisions stakeholders make and act upon in one part of network having potentially widespread and possibly unforeseen consequences on other parts. The entangled mesh of transactions may also help serve to explain why the arrival of new technologies and behaviours has had difficulty in disrupting the traditional processes of academic publishing. As the stakeholders are pulled in different directions by varying motives and

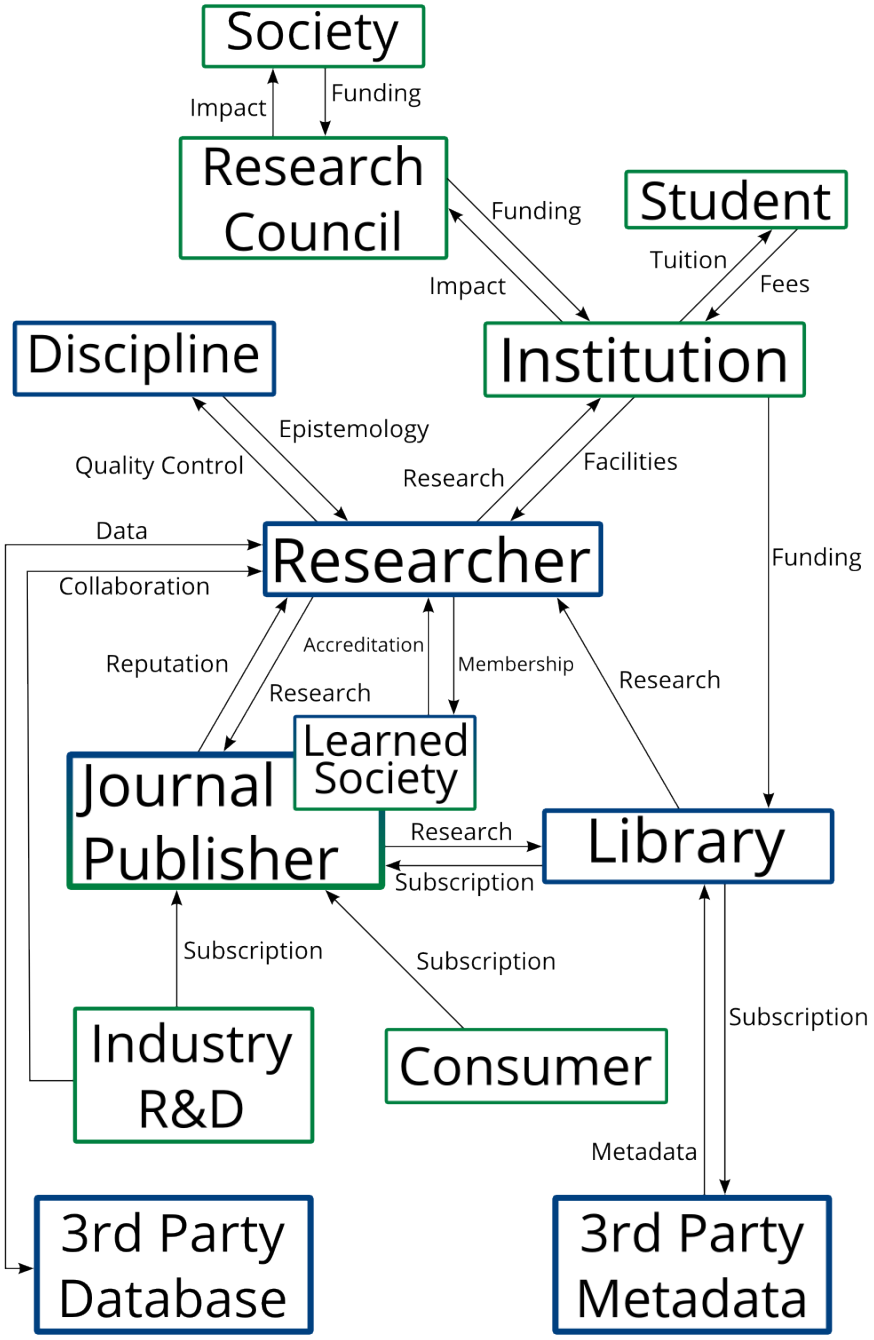

FIGURE 3.1: An illustration providing an indication of the transactions that take place between some of the major stakeholders in the dissemination of research outcomes. The arrows connecting each stakeholder indicate what service or resource one stakeholder provides to another. The stakeholders depicted with a green border are those who take part in the market in which society demands and funds research. The stakeholders depicted with blue boxes represent those who take part in the market in which researchers demand access to research. The publisher and learned society operate in both of these markets.

agendas, their freedom and potential to try new approaches to certain tasks is inhibited, introducing significant risks to progress towards their own, personal objectives.

By taking this broad, holistic overview of scholarly discourse it is possible to identify some quirks that are apparent in Figure 3.1. As already discussed there is a cycle in which research flows from researchers, to researchers, via the journal publishers and libraries. The number of roles that a researcher fulfils also becomes apparent, responsible for not only producing research, but also for providing impact to their institutions and ensuring the work produced is original and valid. Another peculiarity that emerges from the diagram is the role of the research councils which seemingly act as intermediaries between research institutes and the society that funds them. Through identifying these special relationships in scholarly discourse, it possible to begin thinking about how and where the dissemination of research outputs may be pliable to change, particularly change in the form of disintermediation.

3.4 Background to Disintermediation

In many instances where individuals interact with one another, a third party is often required who acts as an intermediary, facilitating the interaction between the other two actors. The demand for an intermediary typically arises when one individual lacks the necessary skills, knowledge or resources to take part in the transaction in such a manner that is sufficiently desirable and efficient for them, and thus another actor is required to lend their expertise for the transaction to go ahead. In doing so, whilst the intermediary allows the transaction to run smoothly, they are also presented with an opportunity to add value to whatever good or service is being provided, contributing to the overall system. This relationship typically manifests whenever an individual needs someone to act on their behalf; travel agents or insurance brokers making for prototypical examples.

However, the Web has presided over a shake up of many of these traditional relationships which has seen the roles of intermediaries redefined, or in some cases, removed completely. This process can be defined as disintermediation: "the displacement or elimination of market intermediaries, enabling direct trade with buyers and consumers without agents" (Wigand, 1997). The travel agents example nicely illustrates how the Web has affected intermediaries, with disintermediation in this instance taking place in two phases, an "initial" phase in which the Web assisted travel agents with their job, followed by a "terminal" phase in which the need for travel agents at all is surpassed (Tucker, 2010) as Web users and services become sufficiently adept at arranging the travel requirements themselves. The Web is renowned for its ability to break down barriers to communication and by doing so, lessens the demand for certain intermediaries, with consumers and buyers coming to realise that they can quickly access and use the knowledge that the intermediaries previously took advantage of; resulting in pressure on the intermediaries' part to demonstrate they can still add value. Thus, intermediaries often need to find new roles in the market, falling back on sub-functions they may offer to prevent themselves from becoming irrelevant, with the only route to

a sustainable business model being able to understand the other stakeholders' perceptions of value adding services (Tay and Chelliah, 2011). This transformative phase for intermediaries is often only short lived however, with disintermediation also creating an opportunity for "reintermediation": the disruptive changes to a market imposed by the Web can create complex new electronic markets that massively increase the number of buyers and consumers available, potentially resulting in the need for new intermediaries that are in a position to negotiate this new market (Wigand, 1997). Whether or not new intermediaries are the same actors as the old intermediaries is not always certain however.

Disintermediation is not the only Web enabled process that threatens to shake up the traditional roles and interactions that take place in many markets. A similar disruption, named "apomediation", may result in increased prominence given to existing players, with "influential peers and opinion leaders" becoming the "primary conveyors of trust and credibility" (Eysenbach, 2007). Apomediaries operate in a similar fashion to intermediaries, but direct potential consumers "through the onslaught of information", giving "additional credibility cues" and providing "metainformation" rather than following the approach of traditional intermediaries and carefully selecting information to be presented to consumers (Eysenbach, 2007), a role only made feasible by the rise of blogs and social networks such as Facebook (O'Connor, 2009). As discussed by Eysenbach (2007), whilst apomediaries are seen as more "equitable and democratic" approaches to providing value added services; conversely influential "hubs" may emerge and develop increasingly powerful positions, and the traditional intermediaries may often find themselves in an optimum position to establish themselves as "credibility hubs" thus eroding any democratic aspect of apomediation. When such hubs emerge, so too emerges the possibility of adverse side affects, either as a result of those who wish to take advantage of a hub's influential position or through a hub misappropriating its influence either deliberately or inadvertently.

Another example of disintermediation can be observed via the Web's transformation of news provision, with citizen journalists frequently breaking and then disseminating news online through sites such as Twitter (Kwak et al., 2010), showing that such a process is possible even when intermediaries perform an important role. However it is important to make a distinction here, as Shirky (2009b) does between form and function. The form, in this instance newspapers, is changing to a (typically) free to distribute digital format: websites. However the function that the newspapers and now websites act as an outlet for, is journalism; it is the effect of the Web on the function that is most significant, not the form. Disintermediation may result in a deficit of investigative journalism, an expensive task that requires experience, and one which citizen journalists are unlikely able to fulfil, although they may still be able to assist by both providing information and used for crowd sourcing tasks such as trawling through large data sources (Houston, 2010).

The same distinction can be made in the context of academic publications: journal articles are simply the form scholarly discourse has taken, but the function of research is much more important and does not hinge upon journal articles. It does however depend upon the flow of knowledge and information for future research to build upon. It is not impossible other forms of communication may still provide this function, eroding the dominant position that journal articles currently possess within scholarly communication and in turn within the academic community, concerning matters of reputation and funding. The case for disintermediation in scholarly discourse is strengthened further still when considering a key difference between research and journalism; the source of value-adding intermediaries. Investigative journalism requires news agencies that are prepared to fund expensive investigations and acquire sources and contacts - all paid for by readers and advertisers. In comparison to scholarly discourse, whilst undertaking research can also be very costly, these expenses are met by research councils and institutions providing funding and resources, and not typically via the route of selling research outcomes.

Thus the process of disintermediation (or apomediation) can have mixed results which depends on a number of factors. Firstly to what extent does the intermediary add value? As highlighted by Dunleavy (2010) writing about disintermediation in government, it is seen as a key to "strip out layers of redundant or non-value-adding process and bureaucracies from service delivery", which stands in contrast to the health sector for example where disintermediation has lead to conflicts between physicians and patients (Eysenbach, 2007). Secondly disintermediation can arise where the intermediary's ability to fulfil their roles has been superseded by an alternative; as demonstrated by the Web's potential to turn anyone into a citizen journalist. At a glance it would seem that many of these principles apply to scholarly discourse, but the extent to which the Web is able to transform this sector is playing out at a comparatively slow pace compared to some of the aforementioned examples, suggesting the processes of disintermediation is not so easy to apply within this context.

3.5 Disintermediation in Scholarly Discourse

Whilst the role of disintermediation has been applied to a number of different domains previously, with Web users as producers and consumers ("produsage") being a key theme of the Web (Bruns, 2008), it has yet to be applied to the scholarly knowledge cycle, with its potential to disrupt academic publishing remaining unexplored. To examine the role of disintermediation in scholarly discourse, requires rewiring some of the interactions between the stakeholders that were identified previously, to take advantage of the opportunities for disintermediation, and thus efficiencies, in both the "macro" and "micro" scale markets. The extent to which disintermediation might be pursued however

depends how sufficiently the affordances of the Web allow the other stakeholders to adequately fulfil the roles originally conducted by the intermediaries and how such changes impact on a stakeholder's relationships elsewhere. Figure 3.2 shows a new configuration of the relationships between the stakeholders once disintermediation and apomediation have taken place, representing the two markets where these intermediary-diminishing processes have had a significant effect (an effect that is quite ambitious and indicates significant changes which would be impossible to realise within the scope of this thesis).

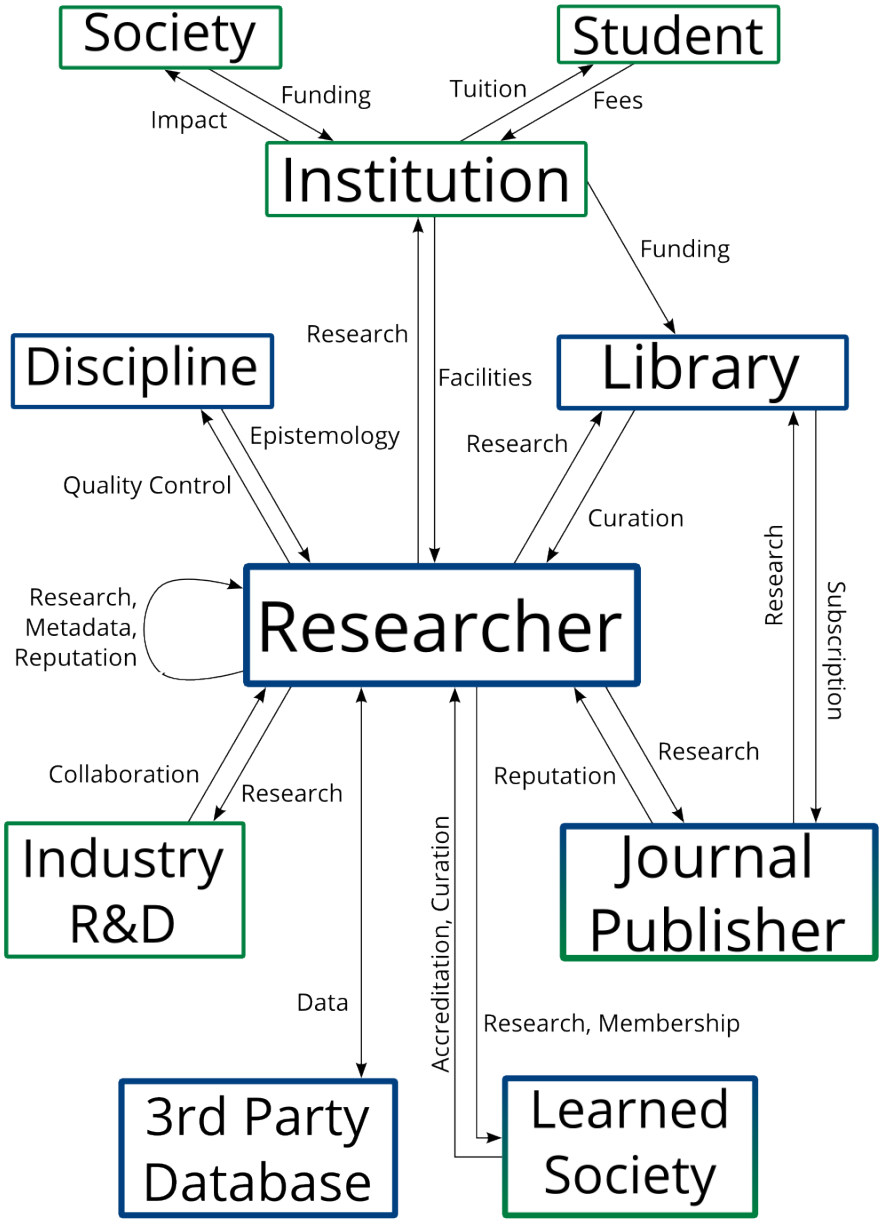

FIGURE 3.2: The interactions between the main stakeholders once disintermediation has taken place. The researcher now plays a much more central role, whereas other stakeholders have been removed. It should be noted that this figure illustrates an exaggerated view of the effects of disintermediation rather than a pragmatic one. As before, the green boxes represent the stakeholders in the macro level market for research, and the blue those who take part in the micro level market.

As illustrated by Figure 3.2 there are numerous opportunities for disintermediation and apomediation in the micro level market, where the researcher now plays a much more central role in academic publishing. The transactions that occur at this level facilitate the exchange of knowledge from one researcher or group of researchers to another in return for credit or recognition. Not only are they responsible for providing research to one another, but for generating their own metadata and allocating reputation amongst themselves, as symbolised by the looping arrow that no longer passes through any of the the stakeholders.

It is clear that the Web is capable of allowing people to freely exchange information (although its ability to easily exchange knowledge may be a more ambiguous notion) and this has already been observed in the publishing industry, with researchers moving to accessing papers online (Long and Schonfeld, 2013). What is perhaps less clear however is how efficiently and effectively the Web is able to facilitate the allocation of recognition and the process of establishing a reputation among one's peers – an unavoidable concern when working in academia (Van Diest et al., 2001). This function of academic publishing is not so easily disintermediated; it should be safe to assume that a researcher cannot be responsible for determining the level of one's own abilities and worthiness of contributions, but that the wider community is capable of evaluating the efforts of an individual researcher. At present this process is achieved via a combination of the peer review process and the influence of journal brands, however these approaches have known flaws (Peters and Ceci, 1982; Priem and Hemminger, 2010) and scope for improvement is available with all the information required to evaluate a researcher's value ultimately being provided by the research community itself.

For example, it should be possible to implement a Web service which allows the research community to organise the peer review process; one could imagine an online platform whereby researchers join, state their area of expertise and affiliations and upload papers and articles that need reviewing. The site could then match papers to potential reviewers and the process of peer review could be conducted in an open, reliable and hopefully efficient manner. Yet such an approach has seemingly not emerged – the research community highly values the peer review aspect of publishing (Ware, 2008) and it is perhaps this value that makes it too precious to experiment with, thus engendering an ironic element of conservatism in a community which is relied upon to be innovative.

Should the research community be unable to be self-organising, the roles played by publishers and learned societies may ultimately be retained, albeit as apomediaries rather than intermediaries, assisting researchers in making these decisions, whilst not holding so much influence or restricting access to results, so that it may be possible to achieve the free and open exchange of knowledge and the further opportunities for disintermediation and value-adding services a more open environment may present. For the potential for disintermediation extends beyond just the distribution of research outcomes and recognition. At present, institutional libraries pay not just for access to journal publications,

but also to metadata provided by third party organisations such as ISI Web of Knowledge or Scopus. If research outputs were made openly available, then the tools that provide this aggregated information on research could also be made openly, either by the community or via open source business models, with new tools introducing an element of competition to the market which in turn may lead to more competitive pricing models from existing services, or incentives to find new ways of adding value. Whichever route taken, research institutions and the researchers they cater for would stand to be advantaged.

Not only can existing, tertiary services be provided via the use of openly published research and the data that becomes available as a result, but there would also be new opportunities for adding value; a virtue that has been embraced by the Open Data community which sees great value in data being made available wherever possible, with data from government departments being made openly available with the aim of improving public services being one example (Berners-Lee and Shadbolt). For example, whilst it is unlikely (and undesirable) for all publishers to be dinstermediated, many of the administrative roles that are conducted by intermediaries such as by administrators at research institutions for example, may be streamlined by tools built upon an open approach to research publications. One such use case may be for a researcher to easily pull together their year's research outputs to submit a report to their institution which wishes to monitor the output of its academics as part of a performance review.

Ultimately a more open scholarly environment might permit a more agile approach to research, in which researchers can more accurately assess each other's progress and easily identify opportunities for collaboration. A rigorous case for openness in academia is made by the *Science as an open enterprise* report commissioned by the Royal Society, which makes a number of recommendations for the future of scientific publishing to "improve the conduct of science, respond to changing public expectations and political culture and enable researchers to maximise the impact of their research" (The Royal Society, 2012). The report emphasises the role that should be played by numerous stakeholders, such as universities or learned societies, in supporting researchers in being more open with their research outputs; in effect making a case for disintermediation not taken to its full extent where researchers are left to communicate findings themselves, but to a degree whereby the role of potentially inhibitive stakeholders is lessened and those non-profit stakeholders who are present to serve researchers, have their roles expanded upon (The Royal Society, 2012).

With disintermediation and apomediation possible in a number of the transactions that permeate the micro level market in academic publishing, there is then further potential for these changes to initiate a trickle up effect to the macro level with the openness of research making the functions of the intermediaries in this market more easily replicated by other means. For example, in the macro level market the roles of research councils and funders may be able to undergo disintermediation. They take funds from one stakeholder

and allocate it to another. Their role is necessary for two immediately clear reasons: firstly the source of much of the funds, the taxpayer, does not possess the knowledge to know how to most efficiently or sensibly allocate the funds available (in a market where the market forces approaches fails to operate), and secondly it can be assumed that researchers would more often than not desire their project receives funding over another and so some mechanism is required to allocate funds at this level.

Thus research councils provide a useful role in overcoming these problems. However the process by which funds are allocated is typically administered by other researchers, which in the UK takes the form of the REF exercise. If a form of peer review which operates via disintermediation could be expanded upon to work on a national level, it may therefore be possible to reduce the extent of the roles of research funders too, eliminating the problems associated with mechanisms such as the REF procedure (Trevorrow and Volmer, 2012). Avoiding institutional biases or gaming of the system might be difficult however and would require reviewing that was both independent and carefully regulated. However, whilst the vision of a research community which could allocate funds efficiently amongst its own community is an ambitious one, it is nevertheless one which should be considered when thinking about the potential for disintermediation and its ultimate scope. The Web is a powerful tool for communicating with others both at scale and across vast distances; it would be possible to organise researchers on an international level to judge the efforts of their peers in different countries for example, thus allowing a fair reviewing system whilst also hoping to remove any preferential treatment from more local ties (whilst being careful to avoid diplomatic or political ones!).

Such a solution would certainly not materialise in the short term and, as is the case with many of the ideas suggested in this thesis on the topic of disintermediation, to adopt these radically different approaches to scholarly communication would require a significant cultural shift from researchers and many of the other stakeholders. It is clear that the possibilities of disintermedation and apomediation pervade both of the two main markets in scholarly discourse and, as seen in other industries and new markets that have emerged online, the technology is available to make much of this disintermediation possible and thus introduce efficiencies and value into the market. Yet in scholarly publishing this desire for change is seemingly lacking among the community as a whole. A deficiency of progressiveness from the community may be a key factor in the lack of disruption in this market that has led to scholarly publishing becoming stuck in the "initial" phase of disintermediation (Tucker, 2010). The Web is asserting its functionality in academic publishing, but only to facilitate many of the existing processes. Despite the possibilities made available by the technology, an impetus to embrace change is missing from some of the major stakeholders, namely the researchers who are largely content with the workings of the current system which for many has served them well thus far, and the publishers who continue to profit from the present system and so have comparatively little external pressure to change their business model. As a result, the

"termination" phase (Tucker, 2010) is never reached and thus many of the problems of scholarly publishing are yet to be "terminated". Naturally some stakeholders will not wish the termination phase to begin, with it not being in their best interests, but if the market is to move ahead and take advantage of the Web, it may be an inevitability. However, as discussed previously, this does not necessitate the end of their participation, just a shifting of roles.

3.6 Disintermediation in Practice

An examination of the stakeholders and their interactions has led to two main conclusions: first that researchers are the major stakeholder and a large majority of the interactions that occur between any two stakeholders largely depend upon the work or contributions of the researcher at some point; and secondly that researchers act as both producers and consumers in the market of scholarly discourse. The extent to which disintermediation can take place is reliant on how effectively the Web can facilitate the roles of old intermediaries, with a change in the role of intermediary that is the journal publisher offering the biggest potential for change.

Journal publishers' function within scholarly discourse can not be reduced to one single role however. They are responsible for the distribution of research, the provision of editorial functions (both with respect to content and formatting) and the administration of a reviewing system for quality control that impacts upon not only what work is and is not accepted for publication, but also the reputations of individual researchers and research institutions alike. Not only do these functions need to continue in some form, but the needs and wants of the researchers themselves need to be adequately fulfilled for new systems to develop and remain sustainable. Thus if disintermediation were to work in practice a number of technologies and resources would need to exist to enable all these functions to continue, alongside a shift to a more decentralised culture among the research community.

3.6.1 The Tools that Enable Disintermediation

Whilst academic publishing fulfils many roles, which in a disintermediated environment would still need to be carried out, the Web has demonstrated its ability to provide different services in innovative new ways, which may either directly parallel those functions found in academic publishing or may be applied after adaptation.

3.6.1.1 Researcher Portfolios

To substitute the first function of disseminating research and allocating reputation, the Web can allow researchers to host their work on a server from which it is then made available to a much wider audience; a page which demonstrates a researcher's contribution to their field and is able of conveying their achievements and capabilities to whomever is interested. Such a page should essentially work for the researcher, helping them to progress with their career, be it furthering their status in academia or applying for jobs in industry, ideally presenting the researcher's work in the context that best suits the current viewing audience.

However, such a solution is far from being simple: the work needs to be disseminated in some way such that other researchers are able to find it and it would be useful if there was an indication as to the level of quality of the work presented and its significance to the research community, information that journals manage to convey via metrics such as impact factors, even if these may be flawed or misappropriated. Infrastructure and support would also be required to make such a service operational that takes into account not just technical requirements but also researchers' digital literacy. Nevertheless services that fulfil many of these requirements have been known to flourish on the Web: Web users do find relevant, high quality content thanks to search engines and innovations such as Google's PageRank algorithm (Page et al., 1999); and have been known to work in a collective and organised manner. Mandavilli (2011) for example highlights the potential of social media services such as Twitter to be used to "weed out sloppy work" on top of their ability to act as a channel for debate. Finally, whilst facilities would need to be provided, the success of Web 2.0 shows that many users are familiar with creating and maintaining personal profiles.

Not only does such a service need to fulfil these criteria, but it would also need to offer a convenient experience to users, with any attempt to collate a researcher's work being a streamlined solution that integrates with pre-existing tasks (Cabrera and Cabrera, 2002), preferably harvesting content as it is generated (via integration with tools such as electronic lab notebooks) and as it is being formulated into more traditional forms of scholarly discourse (journal articles, conference proceedings, theses). It is fair to assume that researchers do not want to be tasked with more administrative roles and would instead rather be able to spend a larger proportion of their time conducting research; and so a system which respects this preference will more likely be favoured by the academic community. Ultimately it is hoped that such a portfolio of research outputs would act as the authoritative representation of the researcher on the Web, in a sense that might mimic proposals suggested by the semantic web community when real world resources cannot be adequately dereferenced from their URI (Sauermann et al., 2007) and users are instead redirected to a page about the resource they are looking for.

However, whilst such a service demonstrates how disintermediation may operate to serve certain stakeholders in academic publishing, it is unclear how it would be implemented. Possible routes of service provision are numerous, yet require new burdens to be placed on existing stakeholders or involve services which are currently available but often perceived to be ineffective compared to traditional, more embedded processes, for example the use of a research-centric social networking site. A service would need to fulfil the dual requirements of being both a convenient site to use which does not impose further administrative duties on researchers, and also carry an air of authority and credibility. Numerous sites already exist that allow a researcher to develop an online profile, but these are often time consuming to cultivate and curate, and it could be argued none have yet reached a critical mass of users that lends them the legitimacy required to be properly acknowledged by academic communities. Thus two potential stakeholders would likely be appropriate facilitators of this role: academic institutions or learned societies.

In the case of the former, any researcher is almost certainly already going to be affiliated with an institution and as a result will likely possess some sort of profile on the institution's website, which itself will likely also have a repository to store its research outputs. However in most instances, a researcher's profile is likely to be no more than a brief biography and a collection of links to further work, rather than a complete collection of their work to date and detailed indication as to their contributions and influence upon their field. Whilst it may attempt to act as a homepage for a researcher, it will likely not be the sole resource a researcher would direct those who are interested in their work to. Institutional repositories themselves can be disadvantaged by being unwieldy to use (Kim, 2006), a finding that originates in the earlier days of many institutional repositories, but may nevertheless still hold true for many researchers today.

The potential for a much more flexible and informative page describing a researcher and their work, coupled with the institutional environment presents an opportunity to provide a number of benefits for the individual and the institution alike. Not only can an individual's research activity be collated, but so too can information pertaining to projects that are run within the institution that the researcher is involved with. The processes of harvesting research activities could also be tied in to administrative procedures required by the institution saving researchers' time and introducing administrative efficiencies for the institution. In the short term however this may impose significant costs on the institution that would need to develop it, and in the long term to maintain it. Other problems arise when considering that an institution may not wish to be responsible for hosting a page that presents a researcher's achievements outside of their position at the institution; and the complexities and inefficiencies that may arise from different institutions implementing different solutions. Whilst the competition aspect this introduces may ultimately result in the provision of better services and greater

levels of innovation, it begins to move away from the decentralised, disintermediation proposal initially set forward and as such new problems begin to emerge.

The other stakeholder in a position to implement such a service would be the learned societies that are typically associated with a particular discipline. By engaging with this new form of scholarly discourse, the learned societies find themselves with an opportunity to build upon their status within the community and develop a new business model in a market which is currently under going changes, imposed by the Web and the proliferation of open access (an even greater concern in the context of the findings of the Finch report (Finch, 2012)), which may not be to their financial benefit. They may also be able to use their disciplinary expertise to assist with the development of platforms for communication, allowing them to leverage their previously held positions to add value in a new market (Eysenbach, 2007). However, problems are once again present with this approach also, with the transfer to a disintermediated environment potentially being very costly for learned societies during the development phase, and with the paradigm shift in publishing possibly discouraging those researchers who prefer traditional approaches to publishing, there remains a possibility that researchers may not continue to renew their membership to support such an enterprise. An argument can also be made that this fails to properly represent disintermediation, with a single stakeholder in each disciplinary field being responsible for the the dissemination of scholarly articles, although the concern is perhaps lessened by the non-profit nature of these stakeholders.

Whilst a researcher's portfolio that represents a researcher's findings would be a useful device in achieving disintermediation, it is ultimately unclear how such a service would be provided. However the philosophy of openness that accompanies the ideas behind disintermedation, mean that in practice, the implementation of these pages should not matter. By publishing researchers' findings as open data, the specifics of the infrastructure that make such a page possible are no longer important and instead the interoperability of services becomes much more powerful. For example an institution may wish to provide a service which can take a researcher's data and help the researcher to automatically generate a professional development report in a quick and streamlined manner. Alternatively, a researcher could deploy a feed that keeps track of developments made by certain researchers or in certain topics to keep abreast of the latest research in a field. Such services could only be achieved through the use of APIs and consistent identifiers, but with technologies and services such as ORCID, a 16-digit unique identifier for researchers and accompanying registry run by an open, non-profit organisation (ORCID, 2015), it is not impossible to imagine a system that is capable pulling resources relating to a researcher from a wide number of locations (Fyson, 2015), creating a more decentralised system for dissemination.

Similarly semantic web technologies could be applied to ensure that entities are described in a consistent and usable fashion. The CiTO "Citation Typing Ontology" (Shotton

and Ciccarese, 2013) provides a typical example of a vocabulary that could be used to define the relationships between researchers and the research outcomes they produce and use. The CiTO ontology contains a wide range of properties for describing how different scholarly resources may be connected to one another, but it is ultimately quite generic (although this trait does also make it quite powerful and flexible) and so in many cases it is likely a subject specific ontology would also be required to better describe the picture which emerges as researchers make their work available. These ontologies may be generated by the community as it begins to build a collection of readily available research outputs or may be built on existing ontologies, such as that described by oreChem (Li et al., 2010). It is likely however that the semantic web technologies that underpin any new approach to disseminating research, would need to remain hidden from many users who may be unfamiliar with the technology and ultimately dissuaded from engaging with it if it is so required, further compounding the problems of researcher responsibilities, learning curves and the subsequent need for institutions to provide support.

3.6.1.2 Crowd Sourced Editorial

The editorial role that forms part of the process of producing a publication is two-fold. First journal editors are responsible for choosing which articles will make for appropriate contributions to a specific issue of the journal and allocate reviewers. Secondly copy editors are used to assist with the presentation of an article to ensure that it is fit for publication with respect to its layout, style and formatting. Both are important roles in the production of a journal, valued by the community it aims to serve, yet arguably the value added here does not equate to the financial costs imposed upon institutional libraries.

Thus for scholarly dissemination to follow a broadly similar approach to conventional publishing once disintermediation has taken place, both of these roles will need to be substituted in some way. However, with disintermediation manifesting as a form of decentralisation, where the idea of a gatekeeper is no longer present to highlight those works of interest, the problem as to how interesting and important pieces of research are brought to the attention of the wider community becomes apparent. One route to identifying interesting research outcomes may be to use the many eyes of the Web to highlight items of interest, by allowing viewers to simply rate an article as interesting or not, with those that garner a number of votes slowly rising to top of their respective fields and being seen by more researchers (in a system not dissimilar to the popular social news site Reddit (http://www.reddit.com)). A potential disadvantage of this approach would be the requirement of some central site however, which would be responsible for allowing this service to take place, but this does not need to threaten the decentralised nature of disintermediation; it can merely be a service built on top of the data that has been made available, acting as an "overlay" journal (Centre pour la Communication Scientifique

Directe, 2013) or implemented by existing publishers in a value adding capacity whereby the value added exceeds the cost from the perspective of the researching consumer.

The second role of ensuring appropriate formatting and styling are applied to research outcomes, with the intent of making it easier (and more attractive) for others to make use of them may be a trickier problem to resolve without imposing new costs upon researchers. Whilst strict formatting guidelines may no longer be necessary if a researcher is making their work available via their own, personal means rather than through a published journal, it is likely that for their work to continue to be taken seriously by the wider community, it would need to maintain a sufficient standard of legibility. This may provide an opportunity for learned societies to assert their relevance in the modern research environment, by proposing standards styles and guidelines that researchers should follow for a myriad of the research outputs relevant to their discipline, without which the scholarly corpus may morph into a confusing and inconsistent state. Alternatively publishers could continue to offer a copy-editing service for a fee, allowing them to continue to profit from those areas in which they do add value to research.

The role of editing, in whichever guise it takes, does however have the potential to be outsourced to the "crowds" of the Web and the wider research community. It is possible to assume that the need to compile collections of research outputs on a particular theme, as a journal does, is no longer necessary in a world where researchers can, and increasingly do, search for articles using a search engine rather than perusing journals – one study concerning the research habits of chemists that found "interaction with the scholarly literature is now primarily driven by keyword and other searches" (Long and Schonfeld, 2013). The need for an editor is increasingly unnecessary and perhaps instead is being replaced by a demand for search and efficient and effective methods for tagging research such that it becomes much easier to find and categorise.

3.6.1.3 Publish then Filter Peer Review

The final role that would need to be fulfilled is that of ensuring published work is of a sufficient standard so as to be beneficial to the academic community and has been conducted in a rigorous, sound and correct manner. At present this is achieved via peer review, a process organised by journal publishers but carried out by researchers. Thus with disintermediation it is merely the process of organising peer review that a substitute needs to found for, although should other forms of quality control emerge as a result of disintermediation, these would not be unwelcome.

The very nature of disintermediation and the philosophy that underpins it suggests that a "publish then filter" approach to peer review is desirable. Disintermediation operates under the assumption that all research outputs should be made more widely available as there is a chance they will be useful to someone in the academic community, even

if that audience is very small, or that by having the outputs available it will help with the interpretation of some other form of the research that has been published. Thus only a system which reviews work after it has been published in some manner would be appropriate.

Attempts at post-publication peer review have been tried previously. Allison et al. (2016) describes a group of researchers who collected together weekly lists of articles and contacted the authors or journals where they found errors in the articles. They found that there were a number of recurring errors, but when they contacted authors or journal editors they "seemed unprepared or ill-equipped to investigate, take action or even respond". The problem was compounded further still when publishers either asked for a fee to publish the corrections letter or charged the author of the article to withdraw the paper (with one publisher charging "US$10,000" (Allison et al., 2016)). Ultimately this approach to peer review is an unfamiliar with Allison et al. (2016) noting that "where to send expressions of concern is unclear", "no standard mechanism exists to request raw data" and "informal expressions of concern are overlooked": the post-publication peer review model is not standard and as such fails to gain recognition or acceptance in the wider academic community.

One example where post-publication peer review is expected is the F1000 Research "Open Science publishing platform" F1000Research (2016). As described by Hunter (2012) "articles submitted to F1000 Research are first processed through an in-house sanity check and then, assuming they pass, published immediately". Following publication, the articles are "subjected to formal peer review" where the "referee's reports are published on the site and all referees are named". Hunter (2012) highlight how the reviewers are asked solely to comment on the soundness of the work presented and not its novelty – a necessity as scientific disciplines become increasingly specialised and as the volume of scientific output continues to grow.

However this approach is not always without its issues. Employing the many eyes of the Web may introduce negativity within the scientific community if reviewers are able to comment anonymously (Knoepfler, 2015). Mandavilli (2011) highlights how social media sites may be used to act as a channel of debate; convenient locations as many scholars are already using these sites to discuss their work. However this raises concerns that such a scheme would result in a deluge of publications from which it would be difficult to identify the higher quality pieces of work (Smith, 2010). There would also remain the possibility that many in the research community would not give discussions held on social media sites the same level of significance as to the traditional peer review process, with social media's potential lack of structure or rigour, not to mention that this process would exclude those voices who are not present on social media sites and the problems that can arise from using a channel of which the user is not in control.

The alternative option has already been briefly touched upon and would require the development of an online platform that would help researchers organise the peer review process. Once again, whilst this strays away from the disintermediation idea by resorting to an intermediary to manage a certain process, it nevertheless hones in on an area where previously the roles performed by intermediary stakeholders do add value, allowing these value-adding services to be retained but (hopefully) at a more reasonable price to the research community. Few implementations of journal-free peer review platforms exist at present, but interest in them is growing (Pandelis, 2013), with the Open Scholar C.I.C. community aiming to implement peer review within institutional repositories aiming "to capitalise on the existing infrastructure offered by open access repositories and to enable their conversion into fully-functional evaluation platforms with associated quality metrics" (Open Scholar, 2015). To what extent the research community would embrace a new approach to peer review is unknown however and the success, or lack thereof, of Open Scholar's Peer Review Module for Repositories will provide a useful insight of the extent to which researchers may be inclined to use the Web in innovative ways.

3.6.2 A Model for Disintermediation

Having identified the major roles which would need to be substituted in the absence of a journal publisher, it is clear that a number of alternative solutions exist, each of which places new pressures on certain stakeholders. Perhaps the most pragmatic approach at implementing disintermediation therefore would arise not by decentralising many of the major roles, but ensuring that the processes of publishing occurs in a well-defined manner such that they are interoperable and compatible with those performed by others. To achieve such a goal would require collaboration from a number of different stakeholders and the use of a wide range of technologies as illustrated in Figure 3.3. The model depicted in Figure 3.3 is just one interpretation of how the services provided by the stakeholders of academic publishing could be reconfigured to offer a more open approach to discourse and involves learned societies, who are trusted to act as benevolent representatives for their discipline, taking on a larger role by providing services which would otherwise be left wanting in the absence of traditional journal publishers. These roles include both the provision of copy editing that would assist researchers in producing accessible research outputs and infrastructure that would help standardise research outputs such that new services can be built more efficiently on top of them.

Having established the role of learned societies, the demands placed upon research institutions becomes clearer. Institutions are expected to employ researchers and provide them with the facilities they required to do their work, much as they do in the present approach to scholarly publishing. However in this new model, the institution also provides each researcher with a "Portfolio" site which their ELN (assuming a chemistry context) can push content to. Naturally this would require more infrastructure required

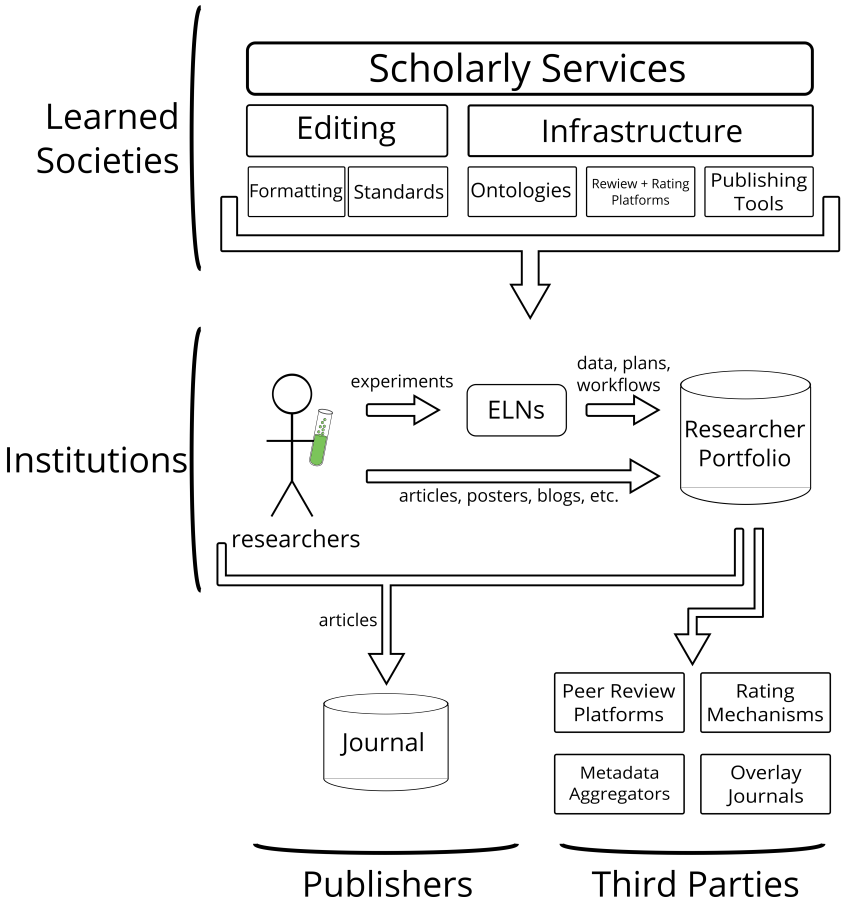

FIGURE 3.3: The roles and services provided by stakeholders in a new scholarly environment. The role of the non-profit learned society has been expanded upon to provide a number of services to the research community that fall under the dual services of editing and infrastructure. Institutions are responsible for providing researchers with the resources they need to continue their work as they presently do, but also for hosting the outputs of their research on a researcher portfolio site. The lower left portion of the diagram represents journals and researchers' continuing practice of submitting articles, and the lower right portion represents new services that can be built upon the open infrastructure that is the collection of researcher portfolio sites.

at the institutional level, although with the widespread use of repositories, much of this is already in place and this approach has the advantage of spreading the burden of hosting research outputs across a distributed network of such sites rather than in one central location (as per the plans of Open Scholar C.I.C. (Open Scholar, 2015)).

As has already been elaborated upon, much of the technology to make this model of scholarly publishing become a reality already exists: open data standards and the application of semantic web technologies could be deployed in such a manner that ensures the research outputs made available can be used effectively (Gurstein, 2011). Many of the typical concerns that are raised about effective data use are not a problem in this instance: the provision of Internet access and services for accessing data is resolved for researchers who have institutions to manage these concerns for them. If researchers

are expected to develop their own portfolios which represent their contributions then a degree of digital literacy would no doubt be required, especially when considering the various requirements that academic publishing demands: "provenance, quality, credit, attributions and methods to provide the reproducibility that enables the validation of results" (Bechhofer et al., 2011). However, these problems should be resolvable through well designed systems and the privileged position institutions are in to educate and develop their researchers to equip them with all the research skills required for them to lead successful careers.

Finally it is worth noting that journal publishers continue to operate and draw the attention of researchers in the new model. As illustrated in Figure 3.3, in the new model for publishing, researchers are able to pull together their research outputs from both their ELN and their "Portfolio" sites when putting together journal article submissions. Complementary to this process is a range of services any third party stakeholders provided by any who wish to build tools and services that operate on the open infrastructure that has a arisen as a result of the portfolio concept, creating a whole new sector which is able to add value in the scholarly publishing market.

3.7 Previous Attempts at Disintermediation

The Web's potential to revolutionise academic publishing by breaking down the barriers that may exist between researchers has not gone unnoticed. Previous work in this area has also looked at the multi-stakeholder situation that occurs in academic publishing, with (Line, 1992) identifying authors, publishers, librarians and consumers as the principle stakeholders that all have dependencies on the publishing medium and highlights an unfortunate trade off that permeates all work on academic publishing: that "none performs very well for all parties". Harnad (1990) phrases the potential impact new communication technologies may have in a different context, bringing attention to the "continuum" of publishing that highlights that the various stages of research each have an opportunity to be revolutionised, placing particular priority on the "prepublication phase" where "most cognitive work is done".

Perhaps the most notable proposition of a new view on academic publishing is that described in Smith (1997)'s "The Deconstructed Journal", which arose from a number of familiar observations: "that e-journals attempt to maintain the style and organisation of paper journals"; "the 'paper' model constrains and inhibits the inherent flexibility of the network medium" and that the "scientific journal has existed almost unchanged for 200 years" indicating its fundamental importance in scientific discourse and the need to maintain the status quo in some form.

The analysis of academic publishing in the proposal of "The Deconstructed Journal" pursues a similar course as described above relating to disintermediation, analysing the

roles that a journal plays with several functions being identified, some more explicit than others. These numerous functions once again suggest that the role of publishing and the medium which facilitates it extends beyond the mere dissemination of work, drawing attention the other roles publishers perform. Some of these functions have been discussed already, such as the need for quality control (both with respect to "content" and "form") and "conferring recognition of work done", but Smith (1997) identifies a few more, such as "subject-" and "community-defining", and "marketing" roles. The classification of some roles as being hidden is also a unique insight and whilst the model of publishing discussed by Smith (1997) does not take into consideration the wider network of stakeholders and the different levels at which the publishing market may operate, by highlighting select publisher roles as being hidden, Smith (1997) draws attention to the fact that some members of the research community are unlikely to possess a thorough understanding of the system of which they form just a small part. Such an insight is important to consider when disintermediation requires the same such stakeholders to adapt the nature of the roles as they currently perform them, and to understand why such a paradigm shift may be desirable or necessary.

Having "deconstructed" traditional journal publications to gain a greater understanding of their place in scholarly discourse, Smith (1997) goes on to "reconstruct" a "New Scientific Journal" (NSJ), one which is designed to much more ably exploit the affordances of the Web and bears some likeness to the disintermediation model proposed above. The NSJ proposed would be a Web service in which the operator does not own the rights to the items which it points towards; can act as a filter between the contents and the user (rather than as a repository of the material); and may or may not be responsible for the quality control stages of publishing (Smith, 1997). Thus the paper presents feasible route to a publishing system that is not entirely unlike the description of disintermediation given previously. Yet actual examples of the NSJ are rare, if not non-existent. The paper touches upon some of the difficulties of introducing a new approach to publishing, focusing on the need for "acceptance by the user community" and reaches the conclusion that acceptance will come when "professional and funding bodies accept publishing in this model" and their ability to treat it on a par with conventional publishing (Smith, 1997). Whilst acceptance by the community is a recognised requisite for a new publishing model, the emphasis here is placed on authoritative bodies. However the paper may be underestimating the greater need for researchers to recognise a need for change, for without this the authoritative bodies will have little impetus to act – the absence of any form of NSJ may thus be a further indication of researchers' insufficient desire for change.

The lack of success for the NSJ can provide a number of useful insights for future attempts of introducing new publishing models. It highlights the difficulties inherent with any paradigm shift, principally the use of certain words and terms that have different

meanings in the old context compared to the new (Smith, 1997). The outcomes of previous attempts in this field do not indicate necessary future failure for similar projects however. The nature of the Web has changed significantly over the last decade, with many researchers being much more technologically savvy individuals who are familiar using the Web for both personal and professional activities. The evolution from Web 1.0, read-only Web to the read-write focused Web 2.0 has resulted in a change in the manner in which anyone, not just researchers, feel able to communicate their ideas with peers and wider audiences (Murugesan, 2007). It has previously been assumed that by simply providing a platform on which researchers can upload their work for all to benefit from, the rewards of opening up scholarly discourse would materialise but it is now becoming clearer, with the rise of influential social media and social networking sites aimed towards academics, that it is a social layer that has been missing from previous attempts to affect change in scholarly discourse. Thus not only is an understanding of the bigger picture of stakeholders required, so too is an understanding of how researchers behave in their environment, how they interact with one another and how they value each other's contributions, such that a successful social layer can be constructed that will facilitate open discourse.

3.8 Conclusion

It is clear that the technology is in place for disintermediation to take place within the dissemination of research, and if implemented could introduce a number of benefits. These benefits can be categorised into one of three themes that emerge from disintermediation: the faciliatation of research; the creation of new opportunities; and the introduction of efficiencies. If disintermediation were to be implemented using some of the ideas proposed above, then over time one might observe a radical shift in the research environment, with free and open dissemination of a wide range of research activities. This would allow researchers to not only lay claim to their contributions, but also to demonstrate their skills and achievements in a more effective manner, that helps to make clear their overall impact upon a range of areas, rather than just those impacts that are typically measured by institutions and academia. Not only do existing research practices benefit, but the open discourse proposed by disintermediation should make it much easier to identify new research areas and possibilities for collaboration and outreach, allowing for flexible and agile methods to be employed by researchers and institutions. Finally, despite some new costs in the short term, over time disintermediation may reduce costs through the elimination of subscription fees and it presents all stakeholders with opportunities to hone their actions to those areas where they most add value and for which the rest of the community are most thankful. When envisaged at its full potential, disintermediation would be able to save on subscription fees to journals and to third party metadata aggregators, whose role could instead be accomplished

via tools and services built upon open alternatives. Researchers would also be able to easily report their progress back to their funders and auditing tasks could be facilitated by tools that can take advantage of the fact that research outputs would be openly published, allowing researchers to spend more time with their research rather than with administrative duties.

However the ideas outlined here, whilst illustrating a number of benefits, nevertheless represent what would be a huge paradigm shift for the academic community; not only in the degree to which many processes are changed, but also in the number of stakeholders such changes would have an impact upon. The lack of readiness to change the nature of scholarly communication is a complex issue. Open access, as an on-going attempt to revolutionise academic publishing, serves to highlight how despite sufficient technology and infrastructure available to make research more openly available, a lack of both consensus and interest from parts of the academic community may have hindered any innovative and disruptive developments that one might otherwise expect. To better understand why disintermediation has failed to take hold in this market, it becomes necessary to look at some of the theories that attempt to explain how and why people act how they do, the nature of researchers and previous work that has been conducted in this area.

Chapter 4

Disintermediation: An Evaluation

4.1 Methodological Approach

Having analysed the various roles and relationships between the different stakeholders in scholarly discourse new ways of rearranging these interactions become apparent, with reconfigurations seemingly inviting a number of benefits for the stakeholders the system deigns to serve. Yet the Web has failed to bring about much in the way of reconfiguration, either at the macro- or micro- scale, even though the technology allows – and, considering the Web's origins at CERN, promotes – such changes to take place. However the theory that researchers want to disseminate their findings ("the willingness of scientists and scholars to publish the fruits of their research [...] without payment, for the sake of inquiry and knowledge" (Chan et al., 2002)), and thus the Web as an open communications tool will disintermediate scholarly discourse and in doing so address some of its issues, is ultimately an ambitious and optimistic theory.

As such Merton's "theories of the middle range" need revisiting: whilst the approach may provide a framework for functional analysis that presents an interesting perspective on scholarly discourse and allows new ideas to be explored, it fails to fully explain why an alternative approach to scholarly discourse which remains technically feasible has failed to materialise. Middle range theory however presents an opportunity to use theories that lie between both grand ideas and day-to-day observations to evaluate ideas derived from functional analysis, which will continue to be explored via an interpretivist approach, building upon previous research and theories which have already been subject to extensive empirical testing.

The framework used thus far is also restricted by the analysis of the motivations that drive actors to take part in society. As touched upon by Parsons, whilst economics would drive individuals to act out of self-interest and that if this were solely true then society would break down, and with Durkheim believing individuals instead act out of "moral

obligation" (Knorr-Cetina and Cicourel, 1981), the idea that individuals act rationally from an economic perspective begins to be worn down. Merton elaborates on actor's motivations further still, by highlighting the confusion that arises from "confining sociological observations to the positive contributions" and "the tendency to confuse the subjective category of motive with objective category of function" which leads on to the ideas of manifest and latent functions (Merton, 1968). However Knorr-Cetina and Cicourel (1981) presents a new approach to understanding an individual's actions: cognitive order, which concerns "a shift [...] towards language use and cognitive processes that represent and interpret the relevance of values and obligations", with participants acting "in terms of tacit knowledge and rules which they know how to apply in specific situations, but which they may not be able to explicate". Thus the determinants of actor's decisions and behaviour may be complex and even unknown to the actor themselves, which as Knorr-Cetina and Cicourel (1981) indicates, tasks the social scientist with identifying the "rules and tacit resources which underlie everyday activities", a perspective that will need to be considered for ongoing, interpretivist evaluation of disintermediation (and future theories such an evaluation may give rise to).

4.2 Socio-economic Factors

To focus on researchers as a central stakeholder who underpin the scholarly discourse processes, requires thinking about the motivations and decision making processes that underpin the way in which they try to navigate their careers and profession. It is at this point that it once again becomes necessary to think about the discipline of chemistry as a particular case study, for the way in which a researcher engages with their discipline and the varying scholarly workflows that are available to them, can alter significantly between disciplines, even if the overall processes and procedures broadly play out in parallel with one another (Meier zu Verl and Horstmann, 2011). Thus to consider a discipline-less researcher would prohibit the successful application of other theories and studies, which have been designed to be applied within a particular, given context.

4.2.1 Game Theory

The products of chemistry research often possess not insignificant commercial value: chemistry has a lot to contribute to profitable industries such as medicine and energy and as such the research environment is quite competitive, with charities and industry also conducting research and patenting their findings (Long and Schonfeld, 2013). With a chemist's success often dependent on the economic impact of the research they produce, a competitive environment may emerge, resulting in some researchers ultimately feeling reluctant to disclose their findings until they can do so under the credit claiming protection provided by a journal. However this protection will come at a cost; the

journal will typically hold the copyright for the piece of work, preventing the original author from distributing it as they see fit.

These attitudes towards sharing one's work emerged from the responses given as part of a preliminary study conducted at the University of Southampton, that took the form of a questionnaire distributed to chemistry department's early career researchers (see Appendix A). When asked about the impacts of publishing research outputs online and not in a journal, participants responded with concerns about IP theft perpetrated by not just the commercial sector, but also their fellow researchers, citing the "novel and commercial" aspect of some research as being the key issue. Similar findings have also been reported in larger scale surveys of the chemistry community, such as those found in the *Advocacy to Benefit from Changes* report, which conducted a nationwide survey of chemists under the auspices of Jisc (Burns et al.), and in the responses of interviewees documented in the Ithaka S+R report, *Supporting the Changing Research Practices of Chemists* (Long and Schonfeld, 2013).

Decisions concerning whether to disseminate research outcomes or not and the risks and benefits from doing so can be modelled using game theory. Game theory aims to simulate what decisions players will make under certain conditions and can be used to determine what factors are affecting certain situations, and can ultimately be applied to lessen or enhance certain aspects of a scenario to encourage desired choices. In the context of chemistry, researchers are faced with a choice as to whether to publish their work and share it with colleagues or keep it private until a later stage where it can be published in its complete form in a journal; in short they may either collaborate with their colleagues or compete against them, and a successful chemist will strike the right balance between the two. This dichotomy is mirrored by the common game theory scenario of the prisoner's dilemma.

The prisoner's dilemma involves two players who have allegedly committed a crime. They are taken to separate rooms (from which they cannot communicate with one another) and asked to either confess or deny to committing the crime. The resulting prison sentences for each player may vary depending on the combination of the responses given, as shown by Table 4.1.

		Player 2	
		Confess	Don't Confess
Player 1	Confess	5, 5	0, 10
	Don't Confess	10, 0	1, 1

TABLE 4.1: Prisoner's dilemma payoff matrix, showing the years in prison for each player based upon their response and that of the other player.

The payoff matrix in Table 4.1 indicates that a player's dominant strategy would be to confess the crime, resulting in them either getting 0 or 5 years in prison (as opposed to the risk of getting 1 or 10 years). However, if both players choose to confess they get 5 years – should both players not confess they will just get 1 year each. The option where both do not confess is the socially optimum outcome (determined by their combined prison sentence), and yet will not emerge if both players take their own personal, dominant strategy. Without knowing what the other player is choosing there is an inherent risk with the "Don't Confess" option, and even if both players can communicate with one another, an issue of trust arises: will one player keep their word or exploit the trust of the other?

Looking at the prisoner's dilemma from the chemistry perspective it is possible to identify a number of parallels that produce the same conflict between making a decision to suit one's own purposes and securing a socially optimum outcome: players can choose to either share their results or withhold them (Table 4.2).

		Player 2	
		Share	Don't Share
Player 1	Share	8, 8	3, 10
	Don't Share	10, 3	2, 2

TABLE 4.2: A payoff matrix for chemists in possession of some data. The numerical values presented here are an abstract representation of some sort of quantifiable benefit of the potential actions (e.g. citations, patents, future work, etc.).

The above payoff matrix has been filled with abstract values which have been chosen to reflect the value of various benefits which may be derived from taking a certain course of action. These differ from those used by the prisoner's dilemma to represent the different context academic publishing represents; with every player benefiitng in some way regardless of their decision. Should both players choose to share their results, then the socially optimum result is achieved; both get to build upon their work, whilst potentially benefiting from the work provided by the findings presented by the other player (potentially also opening up opportunities for feedback and collaborations). However, the payoff is greater for the player who does not share when their counterpart does – they receive the advantage of seeing the other player's findings, but are the singular beneficiaries of their own achievements, giving them access to comparatively more information.

When both players choose to share, neither players will receive any kind of advantage with regards to rewards such as promotion or recognition, as both are in the same position (other than natural biases which may exist, that is to say one player may simply be better at their job than the other). It should also be noted that of the player who chooses to share and the other who withholds information, the former may still

derive some benefit through any opportunities or feedback that may arise from having the work in a more open arena. In the final situation, where neither player chooses to share, the benefits received are merely a product of the player's own work being available to them, thus generating none of the added advantages that may be generated. Nevertheless the personal dominant strategy remains being not to share, thus creating the tension that exists between what is best for the individual and thus what is best for the community as a whole.

A similar problem arises in the world of business, when corporations find themselves in a state of needing to both compete and collaborate with fellow corporations in the market. These conflicting tensions were given the name "co-opetition" by Loebecke et al. (1999) in a paper which makes proficient use of the prisoner's dilemma to illustrate how decisions are made when contradicting ways forward are presented. Sharing knowledge among organisations is a complex issue without any simple solutions, principally because, as already identified in the academic context, the normal market mechanisms of supply and demand fail when dealing with knowledge transfer (Buckley and Casson, 1976). It is therefore important in any knowledge exchange environment, for a preview of the knowledge in question to be made available to lower the risk associated with purchasing something that could be of little use.

Loebecke et al. (1999) use the prisoner's dilemma to try and determine when knowledge will be shared in this environment when traditional economic mechanisms break down. Table 4.3 shows a payoff matrix adapted from Loebecke et al. (1999) that endeavours to depict this more complex "game" that exists in academic publishing, delineating the various factors involved that may influence any decision making.

		Player 2	
		Publish	Don't Publish
Player 1	Publish	2r, 2r	r, 2r + va
	Don't Publish	2r + va, r	r + va, r + va

TABLE 4.3: Payoff matrix adapted from Loebecke et al. (1999) to illustrate the publishing dilemma. (r = basic value of knowledge; va = value added by monopolistic knowledge)

Loebecke et al. (1999) introduces new terms such as "basic value" (the value of the work's content) and "monopolistic value" (the value of withholding the knowledge to the individual who possesses it) to argue that knowledge sharing is only beneficial when there is a comparatively high basic value to low monopolistic value . It could be considered that holding the work to one's self is not deemed worth the potential reciprocation that may follow in exchange for making an initial giving gesture. More concisely, collaborative behaviour arises when $r > va$ and competitive behaviour emerges when $r < va$.

However Table 4.3 was originally designed with a business context in mind – the publishing context may provide further ways in which the various benefits can be broken down, providing more variables to consider and use to tweak scholarly discourse scenarios. For example, the publishing researcher may receive valuable feedback on their work and opportunities to collaborate further, essentially improving its basic value. It is also possible that certain benefits may only be gained when both works are published, allowing them to build upon each other or complement each other in such a way that new knowledge is discovered, creating synergy: "the extent to which cooperation yields additional value from interdependent knowledge sharing beyond the sum of the parties' individual knowledge" (Loebecke et al., 1999). Benefits such as these are depicted in Table 4.4, included as *f*, representing feedback on published work; and *s*, the synergy produced by both works being available together.

		Player 2	
		Publish	Don't Publish
Player 1	Publish	2r + f + s, 2r + f + s	r + f, 2r + va
	Don't Publish	2r + va, r + f	r + va, r + va

TABLE 4.4: A further adaptation of the payoff matrix adapted from Loebecke et al. (1999). (r = basic value of knowledge; va = value added by monopolistic knowledge; f = value of feedback on published work; s = synergy produced by combination of published works)

One limitation of game theory when applied in this manner however, is that it fails to address the reality of the game actually involving far more than two players. The more people there are willing to publish, the greater the multiplier of the value *r*, thus the greater the benefit reaped when others publish. Similarly, we can also assume the more people there are in the academic community, the value of *f* may be multiplied by some value with respect to the total population size, thus incentivising publication. However, more players may also increase the value of *va*: the more researchers there are in a community, the greater the chance that someone else will be able to put this information to use in a faster, or more effective manner, than the initial researcher, placing the originator at a potential disadvantage if they share it. Similarly, by withholding knowledge from a greater number of peers, a researcher may be able to gain greater influence and significance in their field by formally publishing the research in a reputable journal.

Therefore the benefits that are derived from researchers sharing their work in a disintermediated environment as opposed to disseminating via the channel of reputable journals, may swing on the overall balance of decisions taken by all researchers in the field. The more knowledge researchers disseminate however, the greater the increase in the multiplier of *r* and once a certain threshold has been reached, the risk of publishing is diminished as one player knows that other players can be relied upon to contribute to

the shared pool of knowledge and thus will also stand to benefit. This situation mirrors that of the public good dilemma, which whilst identifying the problem at hand, does not provide an easy solution. Rational players follow their dominant strategy of not publishing, thus the "deficient equilibrium" position is reached, at which no player has an incentive to publish as they will instantly be put a disadvantage, a barrier Cabrera and Cabrera (2002) call a "social fence".

Having considered the various scenarios and mapped out the possible outcomes, what decisions are researchers likely to choose? For a single turn of the game it may be assumed that the dominant strategy will be chosen and, if the case of the deficient equilibrium has been reached, then the option of not publishing will be chosen. However, the model does not incorporate an element of time and thus fails to show how decisions may change as time goes by and players have the opportunity to reflect upon past actions. The most successful strategy in an on-going variant of the prisoner's dilemma scenario is tit-for-tat: cooperate on the first move and for all following moves do whatever the other player did (Axelrod, 1984). Tit-for-tat has a number of advantages as a strategy for encouraging cooperation as it is able to foster a trustful environment, for if everyone employs the strategy no one player should defect. However, should a player not act cooperatively, the strategy allows for retaliation, meaning it cannot be exploited by those wishing to seek maximum gain from a system.

Therefore whilst players may be motivated by self-interest, cooperation may still emerge providing players are able to recall previous interactions with other players and are able to see that their interactions will continue for the foreseeable future (Hutchens, 2007). The tit-for-tat model may not sustain cooperation on a long term basis, as the ability to retaliate may atrophy over time: "cooperation can flourish if the public spirited majority can punish freeloaders" (Darwen and Yao, 2002), but without this some individuals will exploit the situation; once again drawing parallels with the public good dilemma. This problem is typically solved by the presence of an authority figure, be it an individual or organisation, and thus the elegance and simplicity of the tit-for-tat system is eroded.

The application of the tit-for-tat strategy to the community of researchers may break down slightly however. Typically one would not know who is accessing one's work until after the fact and if so, there is little guarantee that they would provide anything of relevance in return. It is also impossible for a researcher to restrict the scope of their interactions with others in the community: if a researcher were to choose not to publish some data in retaliation of others not publishing, this would have the effect of penalising everyone else who still was, meaning cooperation which is possible through tit-for-tat immediately breaks down. To take advantage of the collaborative nature of tit-for-tat, a much smaller system for disseminating research that operated on a more local level, allowing for closer, more informed and more controlled interactions between its members would be necessary. For example, the confines of a research group may provide a suitable environment for operating a tit-for-tat style platform for disseminating

research data, presenting an opportunity for researchers to collect and prepare their data for presentation at an earlier stage of the research project, from where it possesses the potential to be distributed more widely at a later date.

4.2.2 Knowledge Management

An alternative approach at looking at scholarly discourse is to view the process as a knowledge management problem. Knowledge management, as defined by Sarvary (1999), is a "business process" through which "firms create and use their institutional or collective knowledge" and can be broken down into three sub processes: "organizational learning", the acquisition of information or knowledge; "knowledge production", the transformation of information into knowledge and its application; and "knowledge distribution", the means by which members of the organisation "access and use the collective knowledge of the firm". And, as outlined by Nonaka (1994), "the ever increasing importance of knowledge in contemporary society calls for a shift in our thinking concerning innovation in large business organizations", thus giving rise to a domain that has been the subject of much research and the creation of an opportunity to think about adapting previously ossified processes. Unavoidably, with discussion of knowledge management comes its original context, the realms of business, industry and knowledge economies. The benefits of knowledge management are well understood and desired by those who run organisations; better knowledge management can result in gaining a competitive advantage, increasing market effectiveness and improving product innovation (Cabrera and Cabrera, 2002). Yet the principles of knowledge management may be effectively applied to the research sector. In summarising core knowledge management principles Piccoli et al. (2000) states "while individuals create knowledge, organizations play a critical role in stimulating knowledge creation, articulating and codifying such knowledge and cultivating a culture of knowledge augmentation and sharing" before drawing a parallel with universities who "have long recognized the value of the expertise of their principal employees – the faculty – and have devised systems to motivate the creation, archiving and sharing of faculty-generated knowledge". The need of organisations for "employees who can quickly contribute novel solutions, communicate and cooperate effectively, independently seek out and evaluate information, and readily share their knowledge in a manner that is intelligible and readily usable by co-workers" (Piccoli et al., 2000) near exactly mirrors the attributes of a good researcher, with only one significant difference: it is not the organisation that demands these skills but the community of researchers as a whole.

In the instance of disintermediation, it is the role of researchers and their participation within a much larger system in which they fulfil a vital part that needs to be considered, and in knowledge management the need to influence individuals to make decisions that

are for the greater good of the organisation of which they are members is not an uncommon problem. As has already been touched upon in previous discussions, knowledge is a public good, possessing two properties, it is both non-excludable and non-rivalrous. It is the non-excludable property that is most relevant within the knowledge management context. When knowledge becomes available, everyone can benefit from it regardless as to whether or not they have made a contribution towards it. This allows for the emergence of free-riders, who may take advantage of the good being provided, without having to contribute towards it, which in turn can lead to the good deteriorating if this behaviour is permitted to continue – and it is likely to with it being the dominant strategy for "rational, self-interested individuals" (Olson, 1965). Naturally a solution needs to be devised to stop the free-rider problem in any given public good situation, or else no one can benefit from the good as there is no longer any incentive to contribute towards it. In some cases, government provides the good, with mandatory contribution towards its provision through the means of taxes. However, in the instance of an organisation wanting its employees to provide their knowledge so that others may benefit, enforced regulation may not be an option, thus necessitating other methods to encourage contributory behaviour, which frequently involve the deployment of a knowledge management system (Cabrera and Cabrera, 2002).

However, knowledge management research, in its attempts to resolve the problem of encouraging individuals to contribute, has uncovered other inhibiting factors that extend beyond an example of a public good failure. Sources that dissuade engagement include the cost of the time and effort involved in contributing an item of knowledge, which is greater than that required by sharing knowledge informally with a colleague (Ardichvili et al., 2003); a lack of understanding of the benefits of knowledge sharing both on a personal and organisational level; and a feeling of intimidation that comes with making personal thoughts public, a fear of being belittled for making trivial contributions (Cabrera and Cabrera, 2002). Whilst these possible reasons for a lack of participation have all been derived from a business environment and thus they may not be completely transferable to an academic setting, it is fair to assume however that the underlying psychology that influences the decisions made by users of knowledge management systems may still help in informing new approaches to dissemination of research outcomes.

Research by Cabrera and Cabrera (2002) indicates that the technical solutions tend to be the most effective way of encouraging people to create, disseminate and exploit organisational knowledge, providing a channel through which users can both contribute and receive data, typically working better than solutions that involve altering organisational structure or implementing new human-resource policies. This implies that many employees when asked to contribute to a knowledge management system, identify the benefits to sharing knowledge, but the cost-benefit ratio is weighted in favour of the costs. A well implemented technological solution should reduce the cost of carrying out the activity it is is supposed to facilitate, thus in this context making it more convenient

to contribute and also pull out relevant knowledge from such a system. Cabrera and Cabrera (2002) indicates that employees cannot integrate knowledge sharing techniques into their everyday activities – the provision of a technical solution that has this goal in mind, may make a drastic change to the extent at which knowledge is provided, highlighting the importance of building this process into pre-existing ones such that the knowledge sharing step becomes a seamless part of the work.

Whilst providing technological channels through which employees can conveniently share their knowledge may help a firm leverage its knowledge potential, a medium for communication alone is not enough to reap the full benefits of knowledge management. The other problem to overcome is that of knowledge hoarding, a result of the "'knowledge is power' syndrome" (du Plessis, 2005), with hoarding typically taking place where there is competition. This is a particularly prominent issue in the academic environment, when pressure is placed upon researchers to release better quality and a greater quantity of publications than their peers, and thus a strong disincentive to the sharing of knowledge outside of a channel through which one can be formally recognised and rewarded. Thus it is necessary for organisations to attempt to encourage a transparent, open and friendly culture among those it wishes to share knowledge. One way through which this might be achieved may be to create communities of practice in which like minded colleagues are brought together to share their work, with such communities often already present in the research environment where postgraduate researchers are managed as parts of a wider research group. One example where such a practise already takes place, within the chemistry domain and on a larger scale than an institution is the crystallography community. As described by Coles and Gale (2012), crystallography frequently "will require recourse to the most highly powered instrumentation available" and as such "cost implications may then require the establishment of shared facilities", achieved in the UK by the National Crystallography Service (NCS). Through doing so the NCS is in a position to generate "large volumes of data from a high turnover of samples", increase "understanding on the part of crystal providers as to the kind of results crystallography can offer" and ultimately conduct research with the aim of "moving the subject forward" (Coles and Gale, 2012).

The open culture approach might struggle to counter a further problem faced by those wishing to implement a knowledge management strategy; the hoarding that occurs as result of "functional silos" within an organisation (du Plessis, 2005). A more effective technique may be to offer more direct incentives for contributing to the knowledge management system, restructuring the manner in which employees are rewarded. Bartol and Srivastava (2002) found that providing extrinsic rewards, either monetary or not, did help to encourage the sharing of knowledge, although the provision of an extrinsic reward did erode intrinsic incentives (the pleasure derived from simply performing the task). Of particular note are the benefits of the rewards provided at a team level, which were found to encourage "cooperation and coordination" and develop a "focus on group

goals and performance" (Bartol and Srivastava, 2002). Ultimately the effect of providing extrinsic rewards within the academic context may be limited when it is already a fundamental part of the researcher's workflow to disseminate the outputs of their work, but suggestions of a change in the way in which the achievements of academics are recognised have been called for previously, with universities and research institutes encouraged to "support an open data culture by: recognising data communication by their researchers as an important criterion for career progression and reward" (The Royal Society, 2012) (and in doing so a distinction is drawn between conventional publishing practices which are expected to occur and more open, innovative approaches which are encouraged via the extrinsic reward).

Similarly, whilst incentives may be required to "push" knowledge into a system, they are also necessary to encourage individuals to "pull" knowledge, an act which may suffer from the "not-invented-here" syndrome (Katz and Allen, 1982), which arises when individuals do not wish to use knowledge sourced from elsewhere for fear of feeling inadequate or for not trusting its validity. Financial motivation tends not to work as successfully when encouraging people to pull knowledge; if not managed carefully it can result in the hoarding of knowledge (Hauschild et al., 2001). Within the academic context however this issue is one that researchers should already be familiar with however: researchers cannot avoid calling upon the knowledge and findings of their peers and should be well equipped to assess the validity of the work they are using (a facet of their day-to-day work research institutions should prepare them for).

As has already been noted, much of knowledge management research falls within the business domain, trying to ensure that a corporation makes the most effective use of its employees and that vital knowledge is not lost when an employee leaves the organisation. Nevertheless, there are parallels between the realms of business and industry and that of academia and as such knowledge management systems, but to simply deploy knowledge management systems to facilitate scholarly discourse would be to oversimplify the problem. The type of knowledge that needs to be handled in academia may differ greatly from that found in business – scientific knowledge is difficult to aggregate as it frequently changes and may require specialised expertise to be understood, and needs to be codified to be stored, a complex task when dealing with cutting-edge research (Bos et al., 2007). Typically researchers also have greater freedom to pursue high risk and high reward ideas than individuals in other professions and many scientific endeavours are cross-institutional, which in turn creates new knowledge management problems (Bos et al., 2007). It is these challenges that academic publishing has evolved to handle, but knowledge management can inform how new systems may be integrated within the existing research environment and how they should be designed to encourage participation, suggesting streamlined, technological systems that may be complemented with extrinsic reward as key to encouraging new forms of discourse.

4.2.3 Collective Behaviour

Whilst game theory can be used to help predict what decisions individuals may make given a number of factors, and how we might engineer a situation to elicit the desired outcome, it is nevertheless restricted in its ability to model n-player games where n is greater than 2; a notable shortcoming when discussing the populous world of academic publishing. For example, it is difficult to measure how penalties and rewards may be distributed among players, when all the other player's decisions also need to be accounted for, resulting in a very large number of possibilities needing to be modelled as more players are added (Yao and Darwen, 1994). Another shortcoming hinges upon the not insignificant assumption that the players of the game are rational.

An alternative approach to looking at the behaviour of many inter-connected individuals is to consider the theories put forward on the subject of collective behaviour. A collective behaviour model can be used when actors have two choices (in this case to disseminate or not to disseminate) and the costs and benefits of the action depend on the number of actors who choose the alternative (Granovetter, 1978). Individuals in the system each have a threshold value: the number or proportion of others who must make one decision before a given actor makes a similar decision, the point where the net benefits exceed the net costs for that particular actor. This model as proposed by Granovetter (1978) highlights the importance of understanding not only individual norms and preferences, but also the aggregate actions of the population in question. Comparing the beliefs of the individual and the actual outcomes of the population often reveals paradoxical effects, as they fail to align with the intentions of the individuals who together are responsible for them. The model goes on to describe various roles that are fulfilled by certain individuals in the population: "instigators", those who will commit even if no one else is doing so; and "conservatives", who possess very high threshold values and may only commit once 80-90% of others do also. Within the context of academic publishing, the idea of instigators can be seen quite clearly when looking at the actions of academics such as Tim Gowers, the instigator of the Elsevier boycott in protest of their rising journal subscription prices (Jha, 2012), with the actions of one going on to spur many more.

Granovetter (1978) identifies several possible reasons for why actions that take place may be as a result of the threshold values people possess rather than their "norms". Norms are the actions which are carried out because of an individual's beliefs and attitudes that affect their decisions should they be in a closed environment without any third party influences. Thus norms are only one "causal influence" on behaviour, with thresholds being the other, with the two influences typically being interrelated. For example, in the publishing context, it is possible to assume that a researcher would be happy to publish their work in an open, accessible format and this belief would have the effect of lowering their threshold for publishing – however, despite wanting to openly publish,

they may never do so if the threshold is not surpassed, because they are concerned about what impact this will have upon them in the wider context of the research community, as understood by the established norms. Typically, factors that may influence one's norms and thresholds include determinants such as social class, education and social position (Granovetter, 1978); in the case of academia these factors may translate to a researcher's career progression (early career researchers may be more protective of their data), influence of peers and supervisors (a supervisor may influence a student's decision on where to disseminate data or not) and occupation (a lab assistant may be less inclined to publish data if it adds a significant burden to their routine tasks for example).

Therefore, when encouraging researchers to be more open with their research data it may be important to first identify those with the lowest thresholds and take advantage of the aggregate effect. This strategy is reflected in the *Advocacy To Benefit From Changes* report (Burns et al.), which suggests producing "local champions" by encouraging those who have little to lose by publishing data openly to take a more active role in advocacy. Care must also be taken to manage the threshold values of individuals in a population such that an insurmountable divide cannot form between the two groups where the conservative threshold is so high it is impossible for a proportion of the population to surpass it. Any system that would take advantage of collective behaviour theory would want to encourage users to not only lower their threshold values, but also help actors become aware of when this value has been reached so that they may act when they see fit. This manipulation of thresholds and norms can only be achieved however by understanding the determinants as described above and by building systems which, as indicated by the paradoxical outcomes of collective action, may differ to what a user might initially specify as expectations and requirements.

4.3 Previous Work

Various theories relating to decision making, knowledge management and behaviour provide an interesting evaluation of the disintermediation ideas put forward in the previous chapter. However, whilst they might give some indication as to why disintermediation, or a similar such approach to disseminating research, has not materialised, merely examining the theories fails to present an opportunity examine the specific context of academic discourse in more detail. Thus to provide further understanding of the interactions and decisions that occur within this domain, demands an analysis that involves studying the stakeholders themselves.

4.3.1 Studies

Having arrived at the conclusion that change within academia hinges upon the reactions of researchers to different ideas, studies that examine researchers' behaviour and motivations can provide valuable insight as to why previous ideas fail to take off successfully or what may need to be designed into a new approach for it to more likely gain traction – as noted by the *Advocacy To Benefit from Changes: Scholarly Communications Discipline-Based Advocacy* report, "embedding new behaviours takes time and needs a detailed knowledge of not only the discipline but also the local culture and infrastructure" (Burns et al.). The *Supporting the Changing Research Practices of Chemists* report (Long and Schonfeld, 2013) provides such an insight, having conducted 60 interviews with research support professionals and academic chemists, with the aim of the report to inspire research support professionals "to develop robust, relevant and innovative research support services". In doing so, the results broke down into three thematic areas, all of which have implications for scholarly discourse: the need for "better support in data management and preservation"; the changing nature of "information discovery" in the "wake of the transition to electronic journals"; and the slow uptake or adoption of "new publishing models" (Long and Schonfeld, 2013).

To arrive at these thematic areas the report investigates the daily life of researchers in chemistry, looking at different aspects of the research cycle from initial phases of discovery and formulation through to actually conducting the work in a collaborative, lab based environment. On the subject of journals for example, the report finds that chemists "feel more pressure to [keep up with the latest research] as the amount of online content they have available to them continues to grow", which has resulted in a transition of how chemists keep up to date with latest results as "interaction with the scholarly literature is now primarily driven by keyword and other searches" as opposed to browsing the journals' contents pages, eliminating any notion of "serendipitously discovering scholarship not previously understood to be related to their needs" (Long and Schonfeld, 2013). This finding is backed up by Renear and Palmer (2009) who state "As indexing, recommending and navigation has become more sophisticated in the online environment [...] strategic reading practices have intensified" describing the researcher's process as "having many different kinds of objectives" and akin to "playing a fast-paced video game" – far removed from the relaxed browsing approach described by the study participants of Long and Schonfeld (2013).

Opinions on journals extend to choosing in which journals to publish, basing decisions upon how the journals "are perceived and what kinds of articles they normally publish and how...the whole peer review process goes", with researchers often consulting with fellow academics about where to publish (Long and Schonfeld, 2013). This finding complements those of the *Advocacy to Benefit from Changes* report (Burns et al.), which in its attempt to establish "advocacy programmes" through interviewing and surveying

chemists and economists, found that "in chemistry, the most influential publishers have a powerful impact on how the scholarly communications process develops" (Burns et al.) and concludes with the "use of local champions from within the field to promote new ways of working has been shown to be the most effective since academics prefer to find out about new developments through colleagues in their department or discipline".

The *Supporting the Changing Research Practices of Chemists* report also goes on to back up the assertions made by Trevorrow and Volmer (2012) and Corbyn (2008) that mechanisms such as the REF place pressure upon researchers to publish in high impact journals. Interviews also suggested that researchers in chemistry were typically "agnostic" towards open access, both because they "do not experience significant problems accessing the content that they need" and because "they see little need for their content on the part of the public" (Long and Schonfeld, 2013). Consequently they only accommodate such processes due to demands placed by research funders (Long and Schonfeld, 2013). With the highly specialised nature of much chemistry research, the argument for open access that the public should have access to research holds little influence amongst chemists who instead perceive open access to be a "very expensive" approach to research dissemination (Long and Schonfeld, 2013). However, chemists were found not to be keen users of the more cost effective approach to open access that is institutional repositories either, with 67% of surveyed respondents indicating they had never deposited any of their publications in a repository, although many went on to say they would be "receptive to the idea of using them", with one researcher highlighting the benefit of a repositories ability to "create a convenient collection of her own papers" (Long and Schonfeld, 2013).

The final aspect of relevance to the discussion about the changing nature of scholarly discourse, touched upon by the report concerns the manner in which chemistry researchers go about their work. Work within chemistry is formulated around lab groups, conducting "complex research projects that require the time and resources of many chemists, students and technicians, and they work in collaboration with other scientists at their own institutions and in other countries" (Long and Schonfeld, 2013). Research is broken down into varying aspects that reflect research life cycles with different people within the group being able to work on different phases within one cycle, with the lab leader typically coming up with research projects, which they let graduate and post-doctoral researchers manage and implement (Long and Schonfeld, 2013). Work can also be delegated to lab technicians who are often chemists with their own research interests but specialise in operating certain equipment.

Naturally, with collaboration organised in this fashion, data management becomes an important aspect of a researcher's work. Data management in this context differs from the knowledge management discussed earlier – with data management concerning how the products of a researcher's day-to-day activities are stored, managed and shared, and knowledge management concerning the final insights derived from the data and how knowledge is shared and retained within a group. The benefits of adopting good data or

knowledge management have previously been discussed and are well known amongst the administrative stakeholders in research such as the funders and research institutions. Yet it is here that the *Supporting the Changing Research Practices of Chemists* is perhaps most damning, stating that chemists' "current data management practices [...] are sub-optimal" (Long and Schonfeld, 2013). The report goes on to elaborate: "They [chemists] are unaware of what they are required to keep, and they do not always create the appropriate metadata to help others interpret their work" (Long and Schonfeld, 2013). This lack of good practice is seemingly partly as a result of different funders varying policies causing confusion amongst researchers, but there also appears to be a lack of "time or inclination to closely monitor data management" in chemists' labs, with data management tools and support often being left as a responsibility for the lab group itself as opposed to the wider institution, and thus in turn is often simply managed by the individuals themselves creating an approach with "little continuity" (Long and Schonfeld, 2013). However in doing so, collaboration within the lab may be hindered and in the longer term, the reproducibility of published results may be hampered. The study also found that when data sharing did occur, it was often "informal sharing between scholars, deposits in subject-specific or institutional repositories, and the publication of data as supplemental materials alongside journal articles on online platforms" and this was typically only upon request as opposed to from the researcher's own initiative (Long and Schonfeld, 2013). Not only does poor data management hinder lab-based collaboration, but it also presents researchers with increased difficulties when complying with funder policies on data sharing. (Long and Schonfeld, 2013) comments how whilst "funders' data retention policies have made chemists more mindful of keeping their data, they are still often unaware of specific requirements and they do not closely monitor compliance", revealing further issues regarding data management.

Ultimately the *Supporting the Changing Research Practices of Chemists* report provides a valuable insight into the thinking of researchers, highlighting their practices and assumptions, which in turn provide some indications as to why disintermediation would not necessarily work. It is clear that for researchers in chemistry, there are very few problems with the current system and that more open approaches either require more work or a financial cost, with very little gain. Data management is already an issue, but it is one researchers have been faced with for some time and only respond to when they are required to do so, typically either by their peers or by their funders; and thus we can identify two strong influences on the actions of researchers: how they are perceived by one another and how they meet their funders requirements. Finally the power of journal brands and notions of their impact are reiterated, with institutions applying pressure on researchers to publish in reputable journals, further emphasising the importance of the role of the journal and thus the impractical nature of the disintermediated research environment.

4.3.2 Questionnaire

For a study conducted as pre-cursor to this thesis, a questionnaire was distributed among the early career researchers at the University of Southampton's chemistry department. The questionnaire (full questions and responses available in Appendix A) was designed to elicit thoughts on three core topics: publishing, motivations for publishing and implementation of publishing systems. The response rate for the questionnaire was low, with approximately a 7% response rate translating to a total of 16 respondents, a result which itself informs about the population sampled, indicating that the topic of dissemination of chemistry research may be of little interest to chemists. Those that did respond however gave some useful insights, with some very much in support of more open, informal dissemination and others against new, disruptive means of communication. However the small amount of data generated by the questionnaire respondents means the findings presented below may ultimately fall victim to subjective bias.

It is clear from the survey that researchers place a lot of importance on the role of journals and the influence that a successful article can have on their reputation within their field. Linking this back to the ideas of disintermediation presented earlier, whilst academics may be in a position to disrupt the market of scholarly discourse, they may have little desire to mimic the actions of consumers in other markets that have been disrupted by Web technology (such as the music industry for example), instead being content with the system currently in place. However this does not mean that a suitable replacement system would not necessarily be unsuccessful, just that the flaws in the current system are not deemed sufficiently problematic to warrant altering by many. Concerns about new publishing models are typically placed with regards to its function in assuring good quality science is conducted and that poor work does not become too prominent and ensuring that researchers do not fall victim to intellectual property theft. The current medium is perceived to perform this function to a suitable level and any changes to the publishing model would be required to meet these needs as sufficiently, if not more so, to build confidence and gain acceptance.

Respondents also made it quite clear that they were keen to retain control over the distribution of their work and that it should not become too readily available too soon, preferring to capitalise on its monopolistic value. Conversely, respondents agreed that being able to access the records of other researchers would be useful and that with more available to view, collaborations would become more likely. Hence most would like the provision of data and yet the value of keeping it to one's self is too high for this to happen, thus all researchers stick to their dominant strategy, depriving the community as a whole from benefiting. This illustrates once again the application of the public good dilemma, and so some instigating force would be required in order to attempt attaining the socially optimum outcome. Use of the theories put forward by collective behaviour and knowledge management may be the solution to this, encouraging researchers to

disseminate their data, but it is likely that a sustained change in attitudes will only be possible by changing the current reward mechanisms; introducing the possibility of new benefits to publishing and deflating the monopolistic value. Thus significant changes are likely to come as a result of mandates from authority figures as opposed to emerge naturally from the community as a whole.

4.3.3 Preliminary Interview Notes

Whilst investigating the area of disintermediation a small number of interviews were conducted within the University of Southampton's chemistry department. Two interview guides were used, one for early career researchers and another for academic support staff. The interview guides focused on the relationship between a researcher's contributions to their field and how they are recognised for it, and how they manage and disseminate their data (the full interview guides can be found in Appendix B). Ultimately the sample size was small however and so the results are only indicative as opposed to significant.

The results from the interviews suggested that early career researchers tend not to manage an online presence, having little interest in developing an online presence and instead are focused on work that will lead to poster presentations and papers. Those who do have an online presence however, aim to develop an online reputation focused on networking, which sits alongside their academic reputation in the field. These two aspects of a researcher's presence within their community of peers can complement one another, for example links on an online CV may direct readers to further information about the researcher's scientific achievements. As highlighted by one interview participant, the modern researcher needs to be able to demonstrate a range of skills, demonstrating both academic and practical capability, as well as networking and having the experience of working across a range of different departments. An online portfolio of a researcher's contributions, an artefact that could be created through disintermediation, would therefore be useful to researchers.

With regards to sharing data with peers, trust plays a large factor with early career researchers, with researchers happy to share data with those they trust. Early career researchers stated that the acceptable degree of openness of research data may change over time, with data that was initially held privately having the potential to be made more openly available once the work had been formally published, although ultimately the decision on what data to share resides with supervisors. Academic support staff suggest however that not all data generated contributes to published articles, and yet there are still implications for making this available due to it's potential to have value at a later time.

Thus interviewing early career researchers and academic support staff reveals that the idea of disintermediation has potential, creating a more flexible approach to the scholarly

knowledge cycle that would help researchers demonstrate their expertise, whilst also helping readers gain access to more data. However it also highlights how the rigid structure of scholarly discourse at present, with its focus on publications and supervisor oversight, restricting the potential to transition towards more agile methods.

4.4 Problems with Disintermediation

Thus an evaluation of the ideas presented by disintermediation shows them to be problematic in a number of key areas. Whilst the technology may exist for disintermediation to take place, and may have been proven to work to this end in other contexts, it is nevertheless lacking within the social construction of science. It is clear that many of the stakeholders in the market of scholarly discourse have little impetus to enact changes, with present business models serving sufficiently well for researchers and publisher alike, whilst embracing change would present too significant a risk for researchers who have their own career goals to consider. However as discussed previously, the model of discourse that has developed over the years prior to the Web is outmoded and costly from a researcher's perspective, and despite these problems it has remained remarkably stable to external, disruptive influences. This stability may be attributed to the fact that whilst it is is imperfect, it is ultimately functional, allowing researchers to communicate their ideas with one another, and in its provision of other useful services such as providing mechanisms for ensuring published work remains of a suitable standard and for both ranking fellow researchers and their institutions. Another possible contribution to the robustness of the system may stem from the network of stakeholders, in which arguably the most important stakeholders, those from which the entire system hangs upon, the researchers themselves, find many of the problems with the system obfuscated by other actors. For example institutional libraries deal with the explicit financial costs of transactions with academic publishers, abstracting the costs away from the researchers.

4.4.1 Reward & Recognition

With researchers having been identified as the key stakeholder in academic publishing, to move on from the idea of disintermediation, it is necessary to understand the researchers, their complex role as both consumers and producers of scholarly discourse, and their motivations and concerns that affect their behaviour. Disintermediation requires researchers to adapt their approaches to scholarly discourse as they become responsible for making more of their outcomes available to the wider research community and so it is understandable that should they feel any changes may impact upon their success in the field, they may be resistant to such disruptions. Therefore it becomes vital to frame any suggested changes in such a way as to fit with researchers' current perceptions, whilst also engineering a degree of manoeuvrability that will allow a shift of the culture that

envelopes scholarly discourse. With current perceptions of reward and achievement in scholarly discourse being heavily weighted towards the prestige associated with publishing in particular journals, for researchers to change the way in which they disseminate their research outputs, it is the measures of success employed within academia that need to be tweaked; a point raised by the Royal Society's *Science as an open enterprise* report which states, "major barriers to widespread adoption of the principles of open data lie in the systems of reward, esteem and promotion in universities and institutes", (The Royal Society, 2012) going on to elaborate "It is crucial that the generation of important datasets, their curation and open and effective communication is recognised, cited and rewarded".

It is often assumed that researchers are motivated solely by their curiosity and desire to push the boundaries of knowledge and the literature concerning the sociology of science does much to back-up this assertion. Most scientists are sufficiently driven by an interest in the subject they study yet as with any institutionalised system, there sits a reward structure, which ultimately impacts upon the behaviour of researchers (Ben-David and Sullivan, 1975). With the aim of science being to create new knowledge, "originality is at a premium" (Merton, 1957) and thus it is for new discoveries that scientists are rewarded. This in turn leads to an emphasis being placed on the priority of discovery and in turn the publishing process as a way of laying claim to one's discovery. As Merton observed, it is with priority and divisive arguments that can arise when there are conflicting claims over who takes priority of a discovery, that the institutional norms of science can be identified, principally the need to be recognised by one's peers, the only "property right" a scientist may reserve, having given up the fruits of their labours to the rest of the community (Merton, 1957). As observed by the Research Information Network (2009), "Above all, they [researchers] see data as a critical part of their 'intellectual capital', generated through a considerable investment of time, effort and skill. In a competitive environment their willingness to share is therefore subject to reservations, in particular as to the control they have over the manner and timing of sharing", indicating how the relationship between research data and reputation creates a culture that may not encourage the sharing of results.

With scientists striving to contribute original knowledge to their corpus, it is only natural that recognition from peers is the currency of reward, with the one of the greatest forms of recognition being eponymy. Yet eponymy by its very nature can ultimately only be applied to a very small number of individuals of the total community (Merton, 1957) (see Figure 4.1 showing the hierarchy of reputation that operates in academia). This scarcity of recognition extends further still with many researchers unlikely to ever be acknowledged for their contributions in any significant manner (Gustin, 1973) and yet still many researchers are driven to participate and contribute their findings. The other popular, incentivising reason for conducting research is that of personal enjoyment, which Gustin develops into the notion of the allure of scientific charisma, "because

through it, a man [or woman] seems to or is thought to come into contact with what is essential in the universe" (Gustin, 1973). Fundamental to both the concept of recognition and charisma is the act of publication: for many researchers "getting things into print becomes a symbolic equivalent to making a significant discovery" (Merton, 1957) and without a publication, a "new discovery is incomplete – it is not a work" (Gustin, 1973). The process of peer review lends further significance to publications, legitimising a researcher's standing in the community and elevating certain members of the community to gate-keeper status, itself a form of reward (Ben-David and Sullivan, 1975).

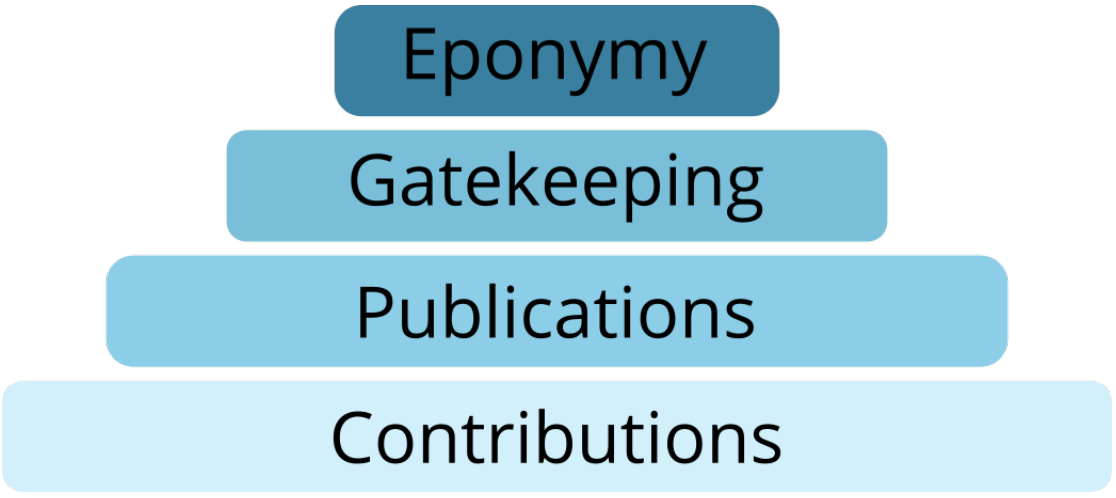

FIGURE 4.1: The varying levels of status and recognition available to a researcher. Recognition is the only premium in academia through which researchers can measure their success. Many will make discoveries and secure publications, and some will go on to become gatekeepers of their academic community, but few will achieve fame and eponymy.

Whilst it should be noted that many of the ideas presented above pre-date the Web and as such scientific practice has moved on in recent years with the arrival of new methods and technologies, the institutional norms of science can still be observed today, with researchers still concerned over issues of priority and maximising the value of their research findings. The pressure to demonstrate novel and valuable results is perhaps greater than ever before, with significant pressure being placed on researchers by funding agencies and institutions to demonstrate the impact of research. The "publish or perish" atmosphere that has been cultivated in academia has resulted in academics publishing to not merely become part of a wider community and be associated with the charisma attached, but to also "gain grants, promotions and tenured positions" (Altman and Melcher, 1983). This has seemingly had an overall effect of muddying the waters of reward and recognition mechanisms in research, with concerns such as an author's h-index influencing a researcher's publishing and citing patterns (Kelly and Jennions, 2006).

The nature of modern research problems has also resulted in a changing "concept of authorship" (Ravetz, 2012) as a result of large collaborative projects and the introduction

of citizen science. Changing attitudes towards the impact of academic research, thinking beyond the metrics based solely upon citations also poses problems for the interlinking of authorship with reward – some individuals may make valuable contributions to their field whilst not authoring papers, and a system of reward that leans on authorship so heavily imposes unnecessary challenges for early career researchers who may need to demonstrate their work and abilities but are yet to have many opportunities to be publishing authors. Simply enumerating a more extensive list of authors may not be the solution however; this can reduce the "visibility of individual role performance" (Merton, 1969) and thus likely lessen the feeling of reward achieved by securing a publication. Developments in the way in which papers are authored may represent that start of a change in the way in which individual researchers perceive their position in their field and how they value what they do; but whilst not significant enough to make discourse more open, may present an opportunity for more modern and open approaches to reward and recognition, and in turn scholarly discourse to be introduced.

Thus with the institutional norms that place great significance on the need to publish, alongside the manner in which recognition acts as reward, there is little incentive for researchers to experiment with the way in which they publish, for fear of jeopardising their potential future careers. It is this aspect of scholarly publishing that the OA movement failed to take into account, considering researchers "to publish [...] without payment" (Chan et al., 2002), whereas Owens (2003) observed, "as long as scientists rely on publishing in high-impact journals to secure funding, there is no incentive to switch to new, open-access models". Whilst the previous statement was made back in 2003, with some OA journals becoming increasingly prestigious since, it nevertheless indicates that the reward and recognition structure in place in scholarly publishing influences the flexibility of the system and the scope with which it can be experimented with. The same paper also raises the issues facing early career researchers: "PhD students have their career to make and should not be sacrificed on some altar of idealism" (Owens, 2003) with the publishing process striking a balance between the need for dissemination and the need for researchers to prove their capability. The interconnected nature of publishing and career progression helps to explain that whilst scholarly dissemination may not be considered to be very efficient and can, at times, hinder every day scientific processes, it is not perceived as significant problem by a majority of the scientific community, most of whom without these mechanisms would not be in the position they are in. These observations can however feed into future experiments, which need to consider ways in which both research data can be made more openly available in more useful formats, whilst also preserving the existing system which researchers depend upon; to suggest career researchers simply abandon one approach to discourse to take up another which may not serve them as well on a personal level, may simply be unrealistic.

4.4.2 Reintermediaries

A final concern for the disintermediation model calls into question the extent to which disintermediation is actually enabled by the Web. Whilst many examples can be given where existing intermediaries have found their roles under pressure from new services and channels, the disruptive effects of the Web have often resulted in the introduction of new intermediaries, those who are in a better position to innovate and exploit the new ways in which many of us now interact with data, information and content. Whilst the Web's ability to allow anyone to produce content as well as consume it, those who do successfully become new intermediaries. For example, within the academic context, some scholars enhance their reputation via blogging (Kirkup, 2010), and through doing so confirm their role as new intermediaries. Similarly, as outlined by Witt et al. (2015) discussing Uber (the 'peer' economy approach to a taxi service), whilst "the narrative of decentralisation is of course correct", a service where users are both providers and consumers "cannot be understood as disintermediation – Uber's search for regulatory legitimacy actually depends upon it demonstrating that it is a better regulator than taxi networks". The disintermediated approach to scholarly discourse suffers from much the same problem, still requiring learned societies and academic institutions to provide a framework within which academic publishing may take place.

Similarly, whilst the concept of decentralisation may well be designed into the fabric of the Internet, and the Web in turn is built upon the Internet with the intent of being open and neutral, whether decentralisation is sustainable as more organisations and industries depend on the Web and place greater demands upon it remains to be seen. This is a concern addressed by those who are working on "decentralization" projects who in October 2015 organised the Redecentralize conference to discuss the transformation of the Web from its original "open, free and decentralized" state, into the "home of vast monopolies", citing how for many their "primary internet experience is limited to facebook" (Redecentralize.org, 2015). Thus the concept of disintermediation is not inherent with the Web, which has become less open and decentralised over time.

The disintermediation model posed previously also contains numerous instances of new intermediaries being introduced instead. For example, to facilitate the key word based searches that would be required to find scholarly content – and which researchers are not typically using as standard (Long and Schonfeld, 2013) – services such as Google Scholar are required that simply act as a different sort of gateway to research compared to publishers. Similarly learned societies would be called upon to help provide infrastructure and support for the high demands that managing one's research outputs would require in the disintermediated environment, and the services provided by organisations such as ORCID would be necessary for a new system to work smoothly (Haak et al., 2012). Whilst the involvement of such intermediaries and non-profit organisations may not pose as many problems as can be found in the current model, and may represent

a more open and distributed infrastructure than is currently possible, to encourage all these stakeholders to work in concert would be a difficult and demanding request.

Disintermediation would also require a high level of digital literacy among researchers burdened with the task of managing peer review and distribution of research articles. However it should be noted that literacy and tools required to facilitate disintermediation would develop over time, as illustrated by the open access movement: green OA represents a disruptive change to scholarly discourse but repository infrastructure has materialised (OpenDOAR, 2016b) nevertheless, and with help from stakeholders such as institutional libraries and funders, the requirements imposed on researchers are kept to a minimum (Chan, 2004).

4.5 Conclusion

Disintermediation of academic publishing should be a viable concept. Researchers are already responsible for all the main functions of scholarly discourse, namely the authoring of content and the reviewing of articles. Technologies such as institutional repositories and the semantic web, identifiers such as ORCIDs, and stakeholders such as academic institutions and learned societies are all in position to facilitate disintermediation, making the process easier for researchers.

However, the theories, ideas and studies described present an evaluation of disintermediation that makes a compelling case for why a new approach to scholarly discourse has not emerged as a result of Web technologies and affordances. Game theory reveals the competitive world of academia where a socially optimum outcome is not attained because each researcher wants to maximise the potential gain from their own data before releasing it to the wider community. Knowledge management practises demonstrate how researchers may have fears in engaging in fully open scholarly discourse, and whilst technological solutions to knowledge management may be best, these have been lacking within the laboratory context (Long and Schonfeld, 2013). Similarly incentives for researchers to engage in better knowledge management practises are also found to be lacking where the main motivation for engaging in scholarly discourse is to better one's reputation via publishing in high impact journals. It is the notion of reward and recognition which ultimately make disintermediation impractical – scholarly discourse may take place openly and freely online, but it needs the actions of the publishers to add an internal meaning to the stakeholders so that researchers and institutions alike may assess one another and so funders can easier allocate finite resources.

Thus socio-economic factors can prevent a community from benefiting from the full potential of Web technologies, highlighting the co-constitutional nature of the Web (Halford et al., 2010). Any approach to encouraging researchers to adopt more open means to

scholarly discourse must consider the important role of journal publications and aim to be complementary to existing processes.

Chapter 5

Disaggregation

5.1 Background

The proposal of disintermediated scholarly discourse reveals the potential for a much more open, data-centric approach to the scholarly knowledge cycle. Not only would this address some of the flaws currently found in academic publishing, such as issues around the reproducibility of previous research and peer review, but it would also allow a more modern approach to the dissemination of research that meets the demands of modern research. Allowing researchers to publish their findings freely on the Web creates a potential where researchers are recognised for all their contributions. However disintermediation is ultimately implausible; it represents a huge paradigm shift, asking researchers to take significant risks, with the socio-economic factors of scholarly discourse making it untenable.

Analysis of disintermediation and the literature and background studies that lead to its proposal reveal that the role and importance of research data has not gone unrecognised in an age of Web-enabled scholarly discourse. As discussed previously funders and publishers alike have introduced policies mandating the need for data to be published and for researchers to present data management plants Springer Nature (2016); Economic and Social Research Council (2015); National Science Foundation (2010); RCUK (2015). Fundamental principles for data publishing have also been outlined, highlighting the need for data to be "discoverable", "linked", "open" and "useful" (Parsons et al., 2011). Yet the evaluation of disintermediation demonstrates the importance of reward and recognition and in this respect "publications reign supreme" when compared to publishing data (Key Perspective Ltd, 2010). With the benefits of transitioning to a data-centric approach to scholarly discourse well founded, and the entrenched nature of academic publishing's stakeholders and processes found to be inhibiting change in the dissemination of research outputs, an approach that satisfies both conditions needs to

be found – one which places equal emphasis on the concepts of reward and recognition derived from publishing, and the potential of the Web alike.

Thus this thesis shall now change its perspective, transitioning from the high-level perspective of stakeholder interactions, down to a perspective that permits a more pragmatic approach to developing the way in which researchers disseminate their results. The need to make more data available to the research community, and conventional publishing's lack of suitability for this, has not gone unnoticed and practical attempts to change this through the introduction of new technologies and services have been explored previously. As noted by Murray-Rust and Rzepa (2006), "Most publicly funded scientific information is never fully published and decays rapidly" and commenting specifically on crystallography services they describe how institutions carry out "hundreds of analyses per year" with each "publishable in its own right, but the majority remain as "dusty files" where the effort required to "write them up" for a full peer-reviewed paper cannot be found". Murray-Rust and Rzepa (2006) proposed a micro-publishing solution to this issue, identifying issues with making data available as journal publication "supplemental information" and instead suggests the publication of "datuments", machine understandable documents that "provide complete information for every component", with the aim of enabling e-science text and data-mining techniques via the "robotic scientific reader" and shifting to a means of discourse that separated the data from the way in which it is presented to the user.

A focus on making data a unit of currency in scholarly discourse alongside the conventional scholarly artefacts of journal articles, conference proceedings and theses, lies behind many data repository initiatives. Coles et al. (2006) cites the benefits of e-science as a justification for adopting the "Open Archive Initiative approach to the dissemination of information", setting up a repository, "eCrystals" to make "available all the raw, derived, and results data from a crystallographic experiment [...] with little further researcher effort after the creation of a normal completed structure in a laboratory archive". Projects such as the Royal Society of Chemistry's (RSC) ChemSpider, provide a "free, online" database by "aggregating data from nearly 400 different data sources and connecting them by means of chemical structure as the primary record in the database" (Pence and Williams, 2010), as well as allowing "registered users to enter information and annotate and curate the records".

In some instances, data repositories have been very successful. For example the Cambridge Crystallographic Data Centre's (CCDC) 'Cambridge Structural Database' (CSD) contains over 800,000 entries, consistently growing year on year (The Cambridge Crystallographic Data Centre, 2016), success it has enjoyed by giving back to its community, conducting original research and developing the "CSD System software" to improve "search capability", "structure visualization" and "data analysis (Groom and Allen, 2014). However Groom and Allen (2014) also attribute their initial success to timing, "the number of published crystal structures was such that a database was of immediate

value, but the number was not large enough to present an insurmountable backlog of existing structures to process". The CSD has had time to grow, both with regards to infrastructure and relevance, as its community has matured also, creating a social and technical environment that leads to success. However many data-repositories do little to take into account the conventions of the scholarly knowledge cycle and the larger system of scholarly discourse of which they form a part. For example "datuments" as proposed by Murray-Rust and Rzepa (2006) "are expressed in XML", and whilst a number of tools are listed for helping researchers generate "datuments", there is no reason why researchers should engage with such practices within the confines of the "publish or perish" nature of scholarly discourse. Whilst Witt (2012) describes how the "data deluge" presents a case for an "expanded role for research libraries in digital data stewardship" via institutional data repositories, the number of data repositories remains significantly less than conventional institutional repositories for storing published works. The Registry of Open Access Repositories (University of Southampton, 2015) only lists 28 "Research Data" repositories as opposed to 2,897 "Research Institutional or Departmental" repositories. Similarly, the eCrystals repository has not seen any crystals deposited since 2011 (University of Southampton, 2016), suggesting a lack of community engagement. The disintermediation perspective highlights the role the reward and recognition in academic publishing along with the socio-economic factors that drive researchers' decisions; factors that the data-centric approaches to discourse described above fail to take into account. Publications remain the dominant currency, because of their externalities, even though they may not be the most suitable medium for scholarly communication in the age of the Web.

Therefore, central to any new approach to scholarly discourse is the notion that any tools or processes must sit comfortably alongside existing tools and processes, complementing them rather than posing a paradigm shift or method too risky for any researcher to ultimately feel able to adopt. Not only must it complement existing methods of dissemination, but it should also appeal to researchers and thus present them with an incentive for using a new approach. Therefore the process of making the research outcomes more openly available should also allow the researcher to be recognised for their contributions. Not only must a new approach provide incentives for it to be used, but it must make efforts to dissuade researcher's fears that may arise when thinking about disseminating their research: will a new approach harm their prospects of achieving (so called) high impact journal publications; or it would it expose their data in a such a way that may allow other researchers to gazump them or infringe upon any agreements made between researchers and their collaborators that specify data must be kept confidential. The importance of communicating data and making it available for e-science must not be at the cost of conventional academic publishing which the researcher and their wider community will continue to need.

Methods of combining a focus on publications, whilst emphasising the importance of

data have been explored previously. As described by Attwood et al. (2010), the "need [...] to rescue the knowledge being systematically isolated in literature and data silos is now widely acknowledged", an issue that has become problematic with a lack of "adequate tools, either to mine the rapidly increasing data - and literature - collections efficiently or to extract useful knowledge from them". The "Unlocking Thesis Data" (UTD) project recognises that PhD theses "make use of newly created data but these data can be trapped in a appendix or DVD - either unavailable or not suited for reuse" (Grace et al., 2015); and speaking at the Repositories Fringe 2015 conference, Stephen Grace of the UTD project, highlighted the potential benefits for different stakeholders, including students, funders, institutions and researchers. The aim of the project is to make data from theses more widely available through the use of identifiers – DOIs for theses and ORCIDs for researchers – an approach to data management that an initial survey has found to be of interest to many higher education institutions (Grace et al., 2015), however the project is still in its infancy and has yet to put many of its recommendations into practice.

Whilst the UTD project is concerned with the infrastructure required to support new approaches to dissemination, other projects have begun to take advantage of the resources already present. For example, the National Compound Collection (NCC) also aimed to "capitalize on the rich legacy of PhD thesis across the UK to create a unique library of accessible and structurally diverse chemical molecules" (Royal Society of Chemistry, 2016). As noted by Royal Society of Chemistry (2016), "since PhD theses are published documents, any IP issues associated with disclosing the structures is avoided" and it allows the outputs of the project to be "openly viewable", with content being sourced from "several hundred academic theses". The NCC used a manual approach to extracting data from the theses, but several projects have tried to either extract data automatically or annotate documents. Jessop et al. (2011) outlines the potential for text-mining, highlighting "that about 10 million chemical syntheses per year are published in the public literature" and have developed OSCAR (Open-Source Chemistry Analysis Routines) to extract chemical data from "natural language narrative". Hawizy et al. (2011) deployed OSCAR in the development of their ChemicalTagger parser which developed upon the text-mining concept further by adding semantics and thus demonstrating that "structured scientific data can be extracted from unstructured scientific literature", with the extraction of both "chemical entities" along with the "relationships between those entities". The Utopia documents project takes a different approach: to ultimately take advantage of the Web to provide a publication with features that a conventional paper journal article simply cannot compete with Attwood et al. (2010) recognises the advantages that "enhanced HTML" can offer, yet comes to the conclusion that "most papers are still read and stored by researchers in personal archives, as PDF files" enumerating the benefits of having a personal PDF collection, such as their permanence once saved locally and their conventional nature. Thus Attwood et al. (2010) implemented a desktop application for annotating PDF documents which could be "configured with

appropriate domain-specific ontologies" and "was designed with two classes of user in mind; the reader, as consumer of published material; and the journal editor, as curator".

5.2 Disaggregation - A Proposal

Following on from disintermediation and previous attempts at encouraging the development of data as a first-class citizen of scholarly discourse, this thesis now proposes the concept of disaggregation, linking data and documents together.

- Conventional publications may be broken down into smaller chunks of data
  - A publication is an aggregated piece of work
  - Finer granularity of research data gives a researcher more control over its dissemination
  - Research data is more useful given a publication context

The proposals above focus on the connection between conventional publications and the importance of data, whilst not undermining the important role of either. Behind the notion of disaggregation lies the idea that a publication is an aggregated piece of work, that is to say it is made up of constituent parts, that one or more individuals may have worked on, an idea in keeping with the findings of the *Supporting the Changing Research Practices of Chemists* report (Long and Schonfeld, 2013). The proposals also invite a number of other areas which may be explored.

- Researchers are motivated to disseminate research data as a means to enhance their reputation within their community
  - Disaggregating data from publications will permit greater dissemination of data whilst minimising the risk to researchers' pursuit of publications
  - An online portfolio of research contributions will assist a researcher in demonstrating their knowledge and experience

Disaggregation as proposed above, differs from the works of previous efforts such as the UTD project or Utopia documents in two key respects. Previous work largely depends on annotating existing literature to provide more detailed information, whereas the approach proposed by disaggregation allows for brand new objects to be born out of the literature that may retain associations to their parent documents, but can exist without them; although naturally provenance metadata connecting any disaggregated components to their source is crucial for creating trust in the data released via such a mechanism. The second area where disaggregation departs from document annotation

is the notion that through disaggregating a document a researcher may enhance the way they are perceived by their scholarly community, allowing a researcher to develop a portfolio of research results by extracting items from their publications and through allowing more complex data structures to be represented via the disaggregation process. A chunk of information extracted from a document does not have to entirely reflect the information found in the document itself with additional information such as the roles of other contributors being added on top of the data. Annotating documents presents innovations around consuming research outputs, whereas disaggregation aims to introduce new methods to dissemination.

5.3 Disaggregator

5.3.1 Disaggregator Development

To investigate the above proposal, a Web based service called the Disaggregator, has been devised to put the ideas into practice, creating a tool that facilitates the disaggregation process. A service that can be presented to the various stakeholders in the research dissemination process becomes necessary to evaluate the ideas that lie behind disaggregation and to inform future directions of research, which hinges upon feedback and comments from those who are directly involved and experience the day-to-day machinations of scholarly discourse. However the process of developing a tool to investigate disaggregation allows for some evaluation to take place before such a prototype has even been deployed and presented to stakeholders. The disaggregation proposal was largely derived from the interpretivist analysis of disintermediation presented in the previous two chapters. Yet the notion of disaggregation is not immediately obvious from the conclusions drawn to on the topic of disintermediation.

5.3.1.1 Image Redactor

The idea of disaggregation began by first developing a simple tool that aimed to present researchers and educators with more opportunities to disseminate their work when previously stopped from doing so by otherwise sensitive items in the work that prohibited sharing the work with a wider audience. The tool, known as the “Image Redactor” was designed for a particular use case; when a work could not be distributed due to copyright that applied to an image, with the tool allowing the user to redact the image and replace it with one licensed under one of the more permissive Creative Common (CC) licenses (as well as allowing the user to embed licenses within the image’s metadata or removing the image completely as alternative approaches to resolving licensing issues). For example, a university lecturer may be able to use copyrighted material when presenting a lecture to a small class of students, but should they wish to make the lecture more

widely available after the lecture so that their students or any other interested parties could learn from it in future, they may be prohibited from doing so because of the presentation's contents. By using the tool, they could redact their presentation, removing infringing images and replacing them with ones which are freely availably. Whilst this task could be achieved manually by the presentation author, it was hoped that the convenience the tool offers would make it more likely that an author would pursue a more open approach to dissemination and it could facilitate the building of a greater corpus of CC licensed material. The tool was a simple prototype built to read the .docx and .pptx Office Open XML files (associated with Microsoft Office 2007 onwards) and then presented the images found within to the user, who could redact them and generate a download link to a new version of the document (Figure 5.1 provides an overview of the redactor features showing different interfaces presented to the user).

The Image Redactor tool also presented an opportunity to reflect on some of the ideas concerning disintermediation, asking to what extent researchers may be able to subsume some of the roles of publishers by preparing their materials for open dissemination; and allowed for an assessment of an app-like approach to replacing some of the roles carried out by different stakeholders, by providing simple tools that provide a service to carry out a singular task in a streamlined manner. However during development of the tool, it became clear the its implementation was hindering some of the potential that the idea of redaction possessed. With knowledge of redacting images came the ability to redact text too, and as research questions formulated around the Image Redactor, such as exploring its potential to create a repository of CC licensed materials for others to use and share (license permitting), it was decided upon to implement a new version of the Redactor. The original version of the tool was developed using JavaScript, a client-side programming language for developing services that run in Web browsers on the user's machine; the new version of the Redactor was instead implemented using PHP, a server side language that offers the potential for users' contributions to be saved and accessed from anywhere online (Figure 5.2 highlights some of the new additions of the second iteration of the Redactor although the significant changes in its implementation are invisible to the user interface).

5.3.1.2 Basic Disaggregator

Once again, the development process introduced new research areas, with the redaction of text showing the greatest potential for changing the way in which researchers may be able to share their research outcomes. If text could be manipulated by the user via a Web interface, it may also be extracted from a document. Thus a new tool was devised that allowed the user to extract text from a document and present it as data that could be presented outside of the document itself, thus giving rise to the idea of disaggregation as proposed above and the research questions that it invites for investigation. The first iteration of the disaggregator was more limited in scope than the

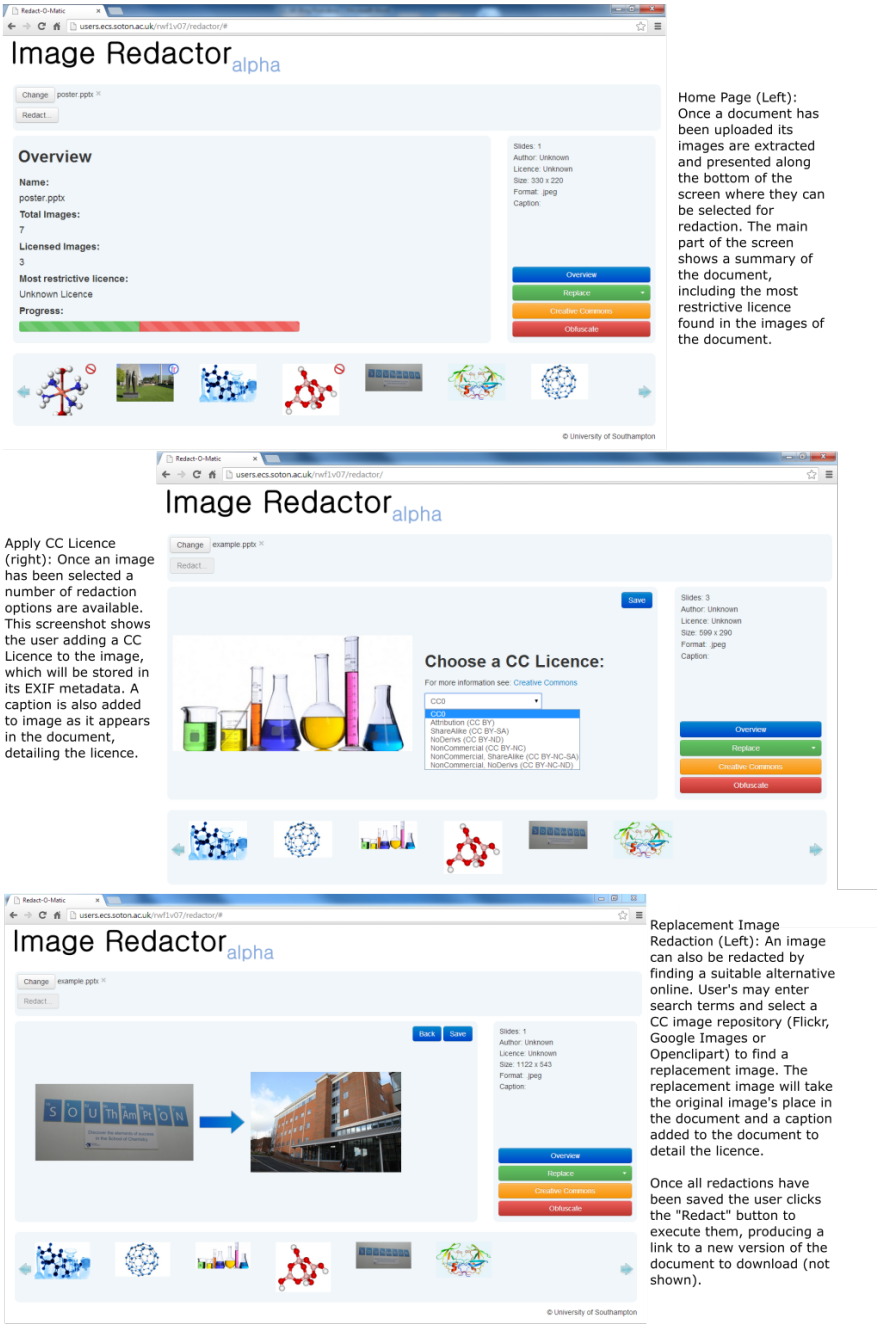

FIGURE 5.1: Some of the main screens presented to the user whilst redacting a document using the Image Redactor. Users are able to upload documents and modify images either by adding a CC licence to the image metadata, replacing the image with a CC licensed alternative or obfuscating the image (by making the image too unfocused to be interpret). Redactions can be saved and once the user has finished selecting them they can be executed, creating a new version of the document. Documents and saved redactions only persist for the duration of the user's session however and are not available to return to later.

final version that is presented for further evaluation however. Developing on from the experience of implementing the Image Redactor, the initial disaggregator prototype no longer followed a single instance workflow process where a resource is input and a new

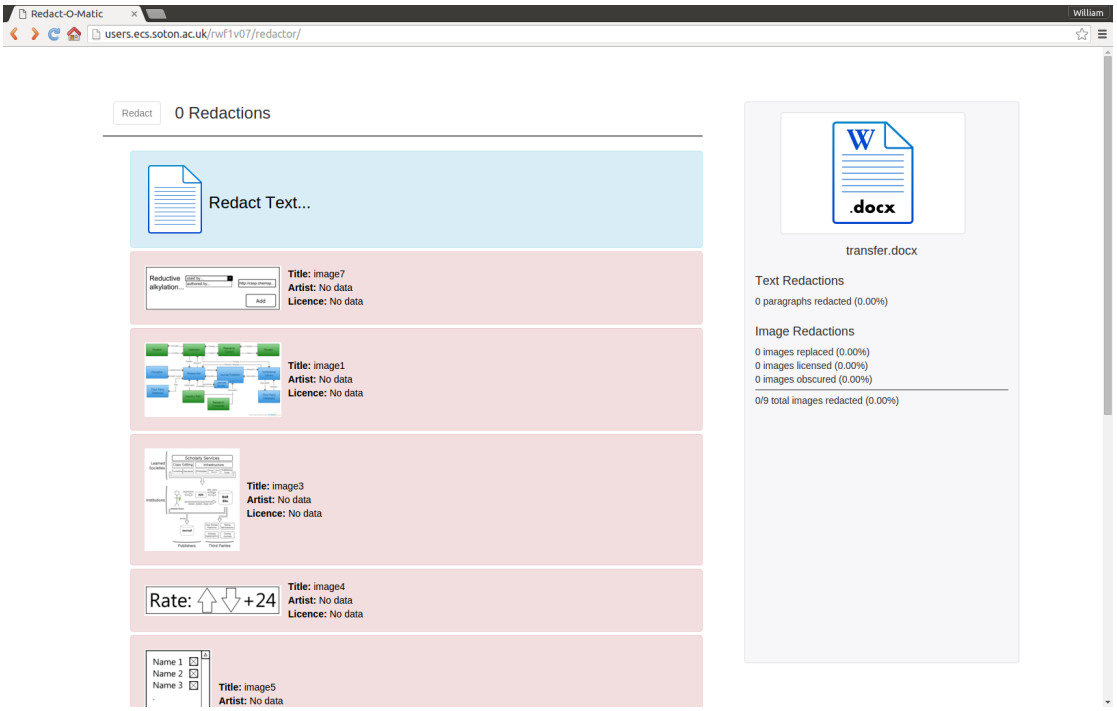

FIGURE 5.2: An updated version of the redactor which supported the disaggregation of text content from .docx documents as used by Microsoft Office Word 2007 and later editions. This version of the redactor also sported a new homepage which showed the user's progress in redacting the document, summarising the changes made. Along with additional functionality, this version of the Image Redactor differed from the former being implemented as a server side application using PHP as opposed to a client side application written in JavaScript.

resource output marking the beginning and end of the user's experience with the tool. Instead the disaggregator prototype provided a system whereby users could register and upload documents, which, along with the items extracted from them, could be saved against the user's account for future retrieval. The initial disaggregator platform was a rigid, proof of concept however, with the items that could be extracted hardcoded in, only allowing "Compounds", "Reactions"and "Spectra" to be extracted, but once these had been derived from an uploaded document they could not be edited again. Nevertheless, whilst the system developed was not as robust as the final iteration designed for future evaluation, it introduced many of the ideas built into the future disaggregator: namely tools for automatically scanning documents for relevant items using text mining techniques and the ability to add data to disaggregated components from other sources beyond the original document, such as via file uploads or external sites such as orcid.org. The hardcoded nature of the items that can be disaggregated along with the lack of flexibility it offered to users with managing their data resulted in a system that would be infeasible to imagine fitting in alongside real-world workflows and as such its potential for research into the idea of disaggregation was ultimately limited (Figure 5.3 shows a number of screenshots from the initial disaggregator concept, highlighting the fixed compounds, reactions and documents tabs and the rigid way in which these concepts

relate to one another).

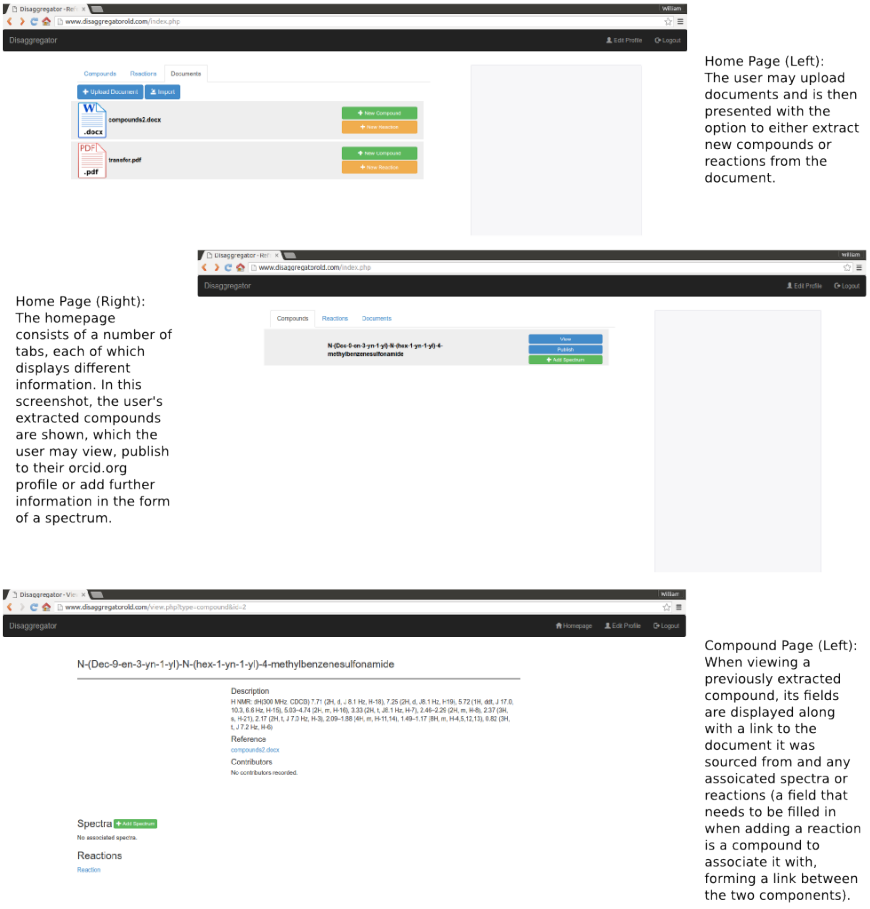

FIGURE 5.3: Screenshots of the initial disaggregator prototype, showing an interface comprising of three tabs to represent compounds, reactions and documents. The document tab borrows heavily from the latest version of the redactor, but new interfaces had to be devised now that the system dealt with types of artefacts beyond documents. However, with the ideas of compounds, reactions and spectra being hardcoded into the system, it soon became evident that building in new features would be unsustainable and the prototypes use as a tool for evaluating disaggregation was limited.

Thus the development processes from the Image Redactor through to the initial prototype of the disaggregator proved to be an exercise in what is required from new research infrastructure. Whilst creating software that represents a single workflow provides some scope for investigation, it limits the extent with which the tool can be used to discuss and evaluate new approaches to scholarly discourse and the capacity for new tools to integrate with existing research infrastructure which are already embedded within the researcher's environment. Developing one feature for investigation also presents opportunities for other avenues of investigation as understanding of the technologies matures and reveals alternative applications. The combination of a greater understanding of

what is possible within the technological scope of the research, and the social and theoretical frameworks within which the research questions lie, ultimately leads to new areas of study.

5.4 Disaggregator Overview

Having outlined the history behind the development process, the origin and inspiration for many of the ideas that lie behind disaggregation and the Disaggregator begin to emerge. The Disaggregator as it exists now is a Web based service, allowing it to be deployed on any server running the required software (a PHP LAMP stack with the appropriate modules and frameworks installed). It has also been designed to be customisable, such that rather being a single site users can log on and use (akin to some academic sites for sharing research outputs such as ResearchGate or Figshare), it is instead designed to be deployed in different instances and in different contexts, akin to configurable institutional repositories such as EPrints. For example an instance may be deployed among a particular research group within a university faculty. The code for the Disaggregator is available at `https://github.com/wfyson/Disaggregator2` (Fyson, 2016) and screenshots for all the features described below can be found in Appendix C.

Once an instance of the Disaggregator has been installed, users need to create an account on their Disaggregator instance, before they can then log in and proceed to upload any Microsoft Office Word or PowerPoint documents (.docx or .pptx files), or PDF files to begin the process of "disaggregation". When creating an account users are required to enter their name along with a username, password and, optionally, an ORCID (to allow information about the user to be queried via a recognised identifier).

A design decision was taken with the Disaggregator to place the documents at the centre of the user experience, making them first-class citizens within the system. It is from the documents that most of the data in the system is sourced and thus where many of the processes begin, thus they are the first thing the user is faced with. However other interfaces are presented along the sites navigation bar to view either "components" the particular instance of the Disaggregator recognises or the user's "portfolio" or "groups" (Figure 5.4).

5.4.1 Components

It is the "components" that are the constituent parts of a publication the Disaggregator may extract from a document and which can then be used to build up the user's portfolio, where the user can share their outputs with the rest of the world. The idea of components builds on the work of Bechhofer et al. (2010) who discuss "the need for mechanisms that support the production of self-contained units of knowledge and

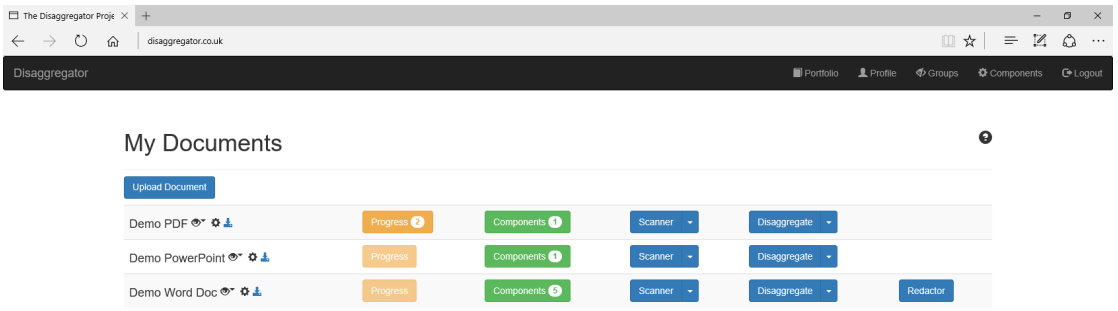

FIGURE 5.4: The Disaggregator homepage, showing the documents a user has uploaded and the various actions that can be taken with those documents. Along the top of the screen lies a navigation bar with links to “Portfolio”, “Profile”, "Groups", “Components” (and “Logout”).

facilitate the publication, sharing and reuse of such entities” and outline a number of principles their “Research Objects” should adhere to. Similarly the idea of “nanopublications” is explored by Groth et al. (2010) who draw a distinction between “a scientific assertion” and the context within which it appears: “the context of a scientific statement is implicit in its immediate environment: the scientific publication”. The Disaggregator draws a similar distinction with documents and components. With all components being extracted from an associated document, this gives rise to a collection of scientific statements or units of knowledge, each with an associated document providing the necessary context and provenance to further assess the validity of the knowledge being presented.

The Disaggregator allows any type of component to be defined by the user, allowing for a level of flexibility as to what a user may choose to share. Components are the system’s metadata, describing the data which is extracted, and can consist of a number of different fields of different types and may be set to contain multiple values, or to be mandatory. Each field as well as being assigned a type (table 5.1) is also given a title to describe it. One field may also be defined as the “preview” field of the component, allowing it to represent instances of the component that have been disaggregated (for example a “name” or “title” field may typically be used as a preview of a component, so that a user may know what the disaggregated component is at a glance).

Following early feedback on the design of the Disaggregator from a publisher, the decision was also made to allow component fields to be specified along with a Resource

Text	For recording textual information such as the name or title of a component that has been disaggregated.
File	For more complex data structures, it is not unreasonable to expect all of the relevant information pertaining to the component to be stored elsewhere. The file field type allows a user to upload other files which can be associated with a component or alternatively select files embedded with documents, such as images.
Component	The component field type recognises that disaggregated objects may relate to one another, and thus allows the user to associate another component that has been defined as a field (which also allows for a recursive data structure as a component may be a field of itself). An example use of the component field may be to describe a chemical reaction which results in a chemical compound, which itself can be defined in terms of text and file fields.
Contributor	A contributor allows a user to define any role another person (or organisation) took with regards to the creation of the disaggregated component. By allowing the user to specify field titles, any type of contribution may be recorded (e.g. supervisor, facilitator, instigator, etc.) and it allows for contributors to be noted even if they have not had a role in the creation of the document.

TABLE 5.1: A component is comprised of a number of fields, each being one of the above field types.

Description Framework (RDF) property to allow the relationships the fields have with respect to the components themselves to be defined in a formal manner. Bechhofer et al. (2010) also highlight how "what is needed in addition to the basic aggregation structure, is metadata that describes the relationships between the resources within the aggregation" and lists a number of ontologies that may be suitable for this purpose. Similarly "Utopia documents" (Attwood et al., 2010) incorporated ontologies into their annotation software to allow their components to "cooperate flexibly" and to allow "sensible relationships and behaviour being inferred at run-time" and it was an approach adopted extensively by Hughes et al. (2004) who devised an ontology to describe the experimental processes in a non-rigid fashion, allowing for flexible storing of experiment information that may "be extended piecemeal and at will without breaking existing applications". Therefore when defining a component the user may choose a namespace, defined by a given URI, and from that namespace choose any properties of the namespace defined by type "rdfs:Class" or "owl:Class" to describe the component. Similarly when defining a field, the user may choose any properties of a namespace of the type "rdf:property", "owl:AnnotationProperty" or "owl:ObjectProperty" to describe the field's relationship with the component. Thus the disaggregator gains the ability to formally describe what a disaggregated component is and how its fields relate to it (for example a field labelled "Title" can be formally described as the title of the component via use of the Dublin Core namespace "title" property.

Ultimately Disaggregator components differ from similar ideas such as the "Research

Objects" of Bechhofer et al. (2010). Whilst the research objects, like Disaggregator components, may also be comprised of various different bits of information such as data, methods and people, and are also semantically enriched, they differ in terms of how the units of knowledge are produced and what is expected of them. The Disaggregator takes an a-priori approach to producing components, with the aim of being compatible with the existing processes and demands of the scholarly knowledge cycle (that is to say publishing high-impact journal papers) along with meeting the needs of the community and research funders through making data openly available in an appropriate manner (Parsons et al., 2011). This approach is ultimately a more relaxed approach to extracting data with the connection to a parent document being key to providing necessary context for the disaggregated component. In contrast Bechhofer et al. (2010) suggests the creation of research objects which are "born digital" and as such the research objects should be "reusable", "repurposable" and "reproducible" (amongst many other requirements) in isolation. Bechhofer et al. (2010) illustrate the requirements of research objects through a number of "stereotypes" suggesting ideas such as a "Publication Object" which are "aggregations that explicitly bring together the presentation of a piece of work – the "paper" – along with the evidence for the conclusions that are being presented". Another example cites the "Method Object" which "reports methodological research" and "exposes the method for easy consumption by other Research/Work Objects".

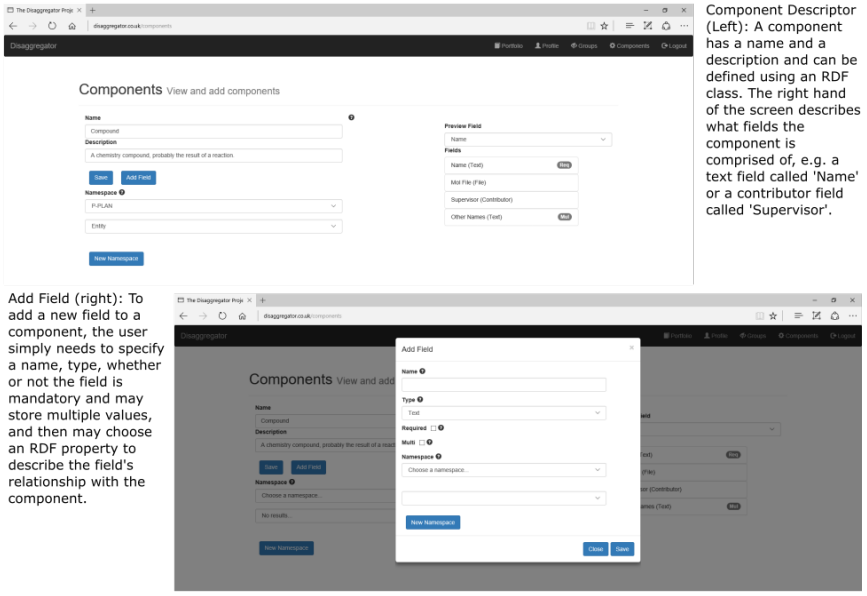

FIGURE 5.5: Screenshots showing a component which describes a 'Compound' and is comprised of several fields, as well as a pop-up form for adding new fields to a component. The fields described here will go on to create a workflow for when disaggregating the component from a document, with a stage showing for each individual field.

Figure 5.6 shows a representation of a component that may be defined in the Disaggregator, showing what fields the component is comprised of. All components receive two fields by default, "Creator" and "Source", automatically populated by the user who is

logged in when disaggregating the component from a document and a reference to the document used in the disaggregation process respectively. The other three fields have been defined via the Disaggregator's interface as described in Figure 5.5. Not all of the fields are related to the component by an RDF property, a trait which has not been made mandatory to allow for instances where a suitable URI either does not exist or is unknown to the user. Using the Disaggregator as a tool for allowing researchers to define new RDF properties or classes where appropriate ones are unavailable may pose an interesting research question for future investigation (but one which falls outside of the scope of this thesis).

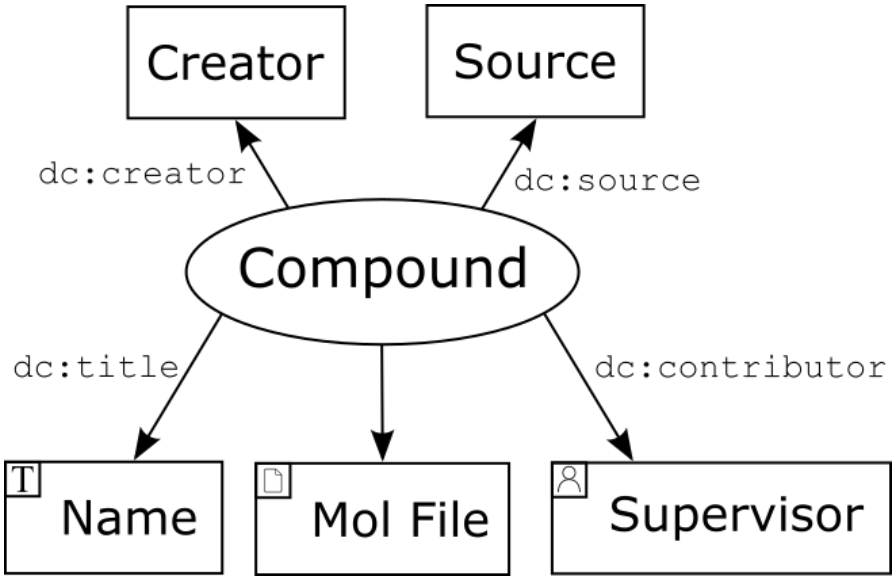

FIGURE 5.6: A component defined in the Disaggregator that describes a "compound". The "compound" is comprised of the five fields. Two are default fields, attached to all components automatically, one to describe the user adding an instance of the component and one to refer to the source document. The other three are custom fields, referring to "Name", "Mol File" and "Supervisor" each with their own type as listed in table 5.1.

5.4.2 Disaggregating

Once a document has been uploaded to the Disaggregator and at least one component has been defined, the user may begin the process of extracting information from documents that describe the component selected. Two approaches to this process have been devised, an automatic method whereby the Disaggregator scans the text of any uploaded documents and makes suggestions for components to disaggregate, and a manual method through which the user can build up components themselves from content found within the document or with data they have simply added manually.

5.4.2.1 Scanners

Scanners within the Disaggregator are used to identify components automatically through the use of text mining techniques, with different scanners being registered in the Disaggregator along with a particular component that they may scan for. The Disaggregator has been designed so that a number of scanners may be written and used, with the ability to provide a suite of text-mining scanners to extract a range of components. However the Disaggregator is also a prototype system, designed to evaluate the notion of disaggregation and the ideas behind it, and as such a fully function plugin system that allows anyone to write their own scanner and register it with the system has not yet been implemented, with new scanners needing to be added manually.

Once a scan has been run over a document the results are presented to the user who may choose to accept or reject each result on a case-by-case basis. By introducing this element of choice, the scanners act as a curated import mechanism, presenting the user with a greater degree of control over what they choose to share via the Disaggregator. Any component that has been extracted is not marked as complete until all of its required fields have been filled and so if the scanner is unable to find results for every field, the user is able to fill in the gaps of the component manually after the scanner's initial attempt. In the prototype, the ChemicalTagger Java library (Hawizy et al., 2011) has been used to search documents for compounds, however the implementation for the Disaggregator is simplistic, merely adapting one of the ChemicalTagger examples (Williamson, 2016) to illustrate the potential for text-mining, but not investigating it further as a research topic.

Through deployment of the text-mining in the form of the scanners, an automatic approach to gathering data is presented. However a degree of autonomy is stripped away from the Disaggregator tool, returning a sense of curatorial oversight to the user, affording the user an opportunity to finesse the outcomes that are returned before they made more widely available, a task which can be achieved via the manual aspect of the Disaggregator.

5.4.2.2 Manual Disaggregation

The manual disaggregation option provides an interface for the user to identify and describe components themselves rather than through the means of one of the scanners. To begin manual disaggregation the author must choose first a document and then a component to disaggregate, or continue working on a component that has been partially disaggregated previously (and thus already has a document and component associated with it). For each field of the component being described the user is presented with a form along with a suitable interface for either extracting data from the chosen document or adding other information such as a contributor, or another extracted component that

has already been extracted and stored in the Disaggregator. For a text field, the user is presented with the document's contents in the browser and may highlight text to select it as the value for the field. Contributor fields can be given a value by assigning either a previously registered user or adding a new contributor by giving their name or ORCID. Component fields list the previously disaggregated components already stored in the system for the type of component being asked for, and fields which require files either allow files to be uploaded from the user's device or may be selected from within a document (if an image from within the document needs to be given for example). Figure 5.7 demonstrates the text, contributor and component disaggregation stages.

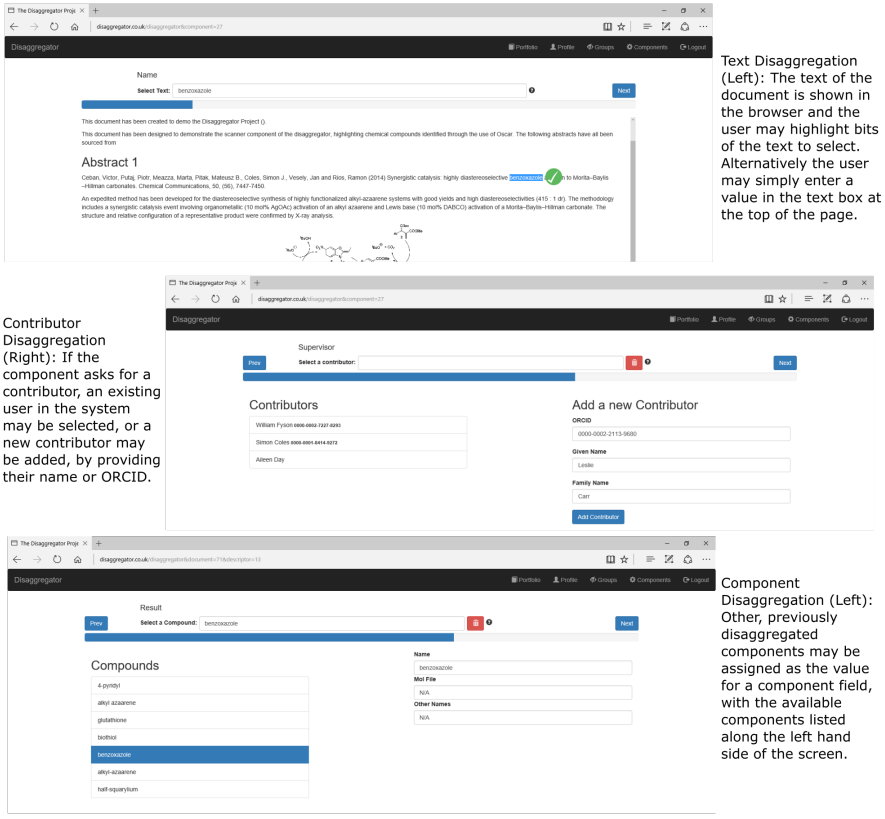

FIGURE 5.7: Three manual disaggregation stages, showing text disaggregation (along with the contents of the document), contributor disaggregation and component disaggregation.

5.4.3 Portfolio

Fundamentally, the process of disaggregation is a mechanism for data to be extracted from documents to be distributed more openly amongst the research community, and it is through the portfolio interface that disaggregated components are exposed to the Web. A user's portfolio page consists of three separate tabs: "Documents", "Components" and "Contributions" which in turn present links to the various items associated with the given user. The documents tab presents the user's documents from which they can be downloaded; the components tab displays all of the components that a user has

extracted from their documents so that other users may view them; and the contributions tab shows all those components in which the user has been described as having made some form of contribution, even if the component belongs to another user. It is via this interface that one of the potential benefits of disaggregation is explored, for the portfolio is comprised of a range of different component types that are defined in the Disaggregator, and specifically those which the author has chosen to share. The potential of disaggregation partly lies behind the increased control a finer degree of granularity gives; for each document and component listed on the home page interface, an option exists for the user to set the particular record's visibility. Users can choose to set items to be either available to all ("Open"), completely hidden ("Closed") or select a particular group of users for the item to be visible to with groups defined from a list of users registered with the Disaggregator instance (see Figure 5.8).

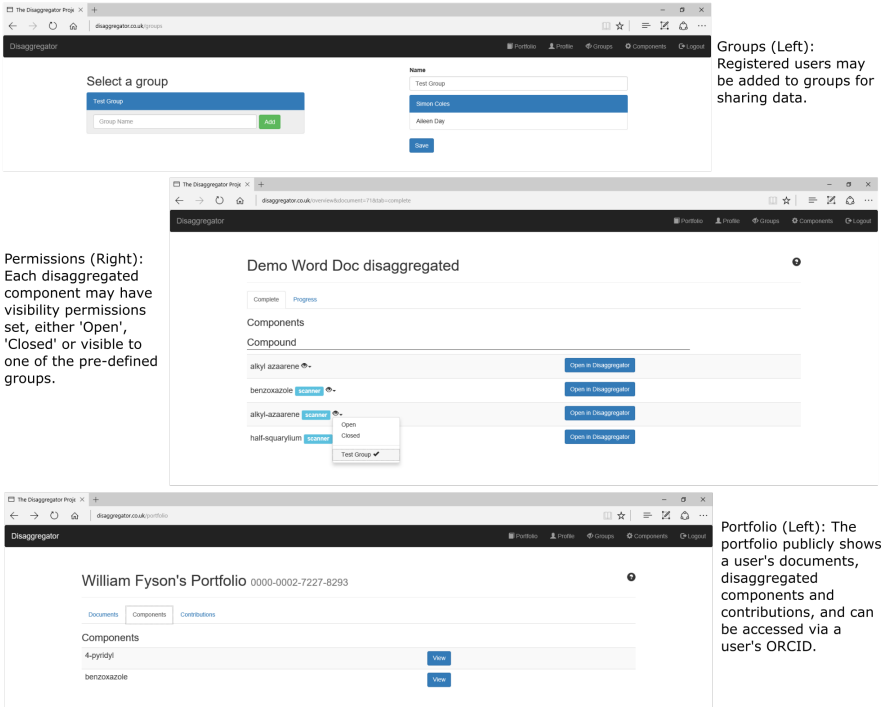

FIGURE 5.8: Users may define groups with which they may share disaggregated components, set the visibility permissions for each disaggregated component individually, and a portfolio page shows the documents, components and contributions a user has set to be publicly available.

For more screenshots showing the various interfaces of the Disaggregator, please see Appendix C.

5.5 Conclusion

Thus with the concept of disaggregation and the tool that presents a possible implementation of such a process, a number of research areas begin to emerge around a particular

context. These facets together provide a framework upon which to hinge an evaluation of the ideas proposed above and the manner in which they relate to the ideas around disintermediation from which they have been developed.

5.5.1 Research Questions

In order to conduct an evaluation of the idea of disaggregation and the suggested context the Disaggregator tool represents, a number of research questions have been identified which will form the basis of further studies.

- How would the process of disaggregation operate alongside conventional processes and workflows already established for the various stakeholders of academic discourse?
- How might disaggregation be implemented with respect to current services, standards and protocols?
- How might disaggregated items be used? How should they be stored and presented to maximise their utility?
- What effect would disaggregation have on the functions of scholarly discourse, including the dissemination and assessment of research outputs, and assignation of reputation?
- What motivates researchers to engage with new approaches to scholarly discourse?

Collectively, these high-level research questions probe many different aspects of disaggregation, ranging from technical and workflow concerns such as how to actually perform the new tasks that are required in disaggregation, through to the social issues that may provide some indication as to what may lie behind disaggregation's success or failure. In endeavouring to answer these questions, not only will it be possible to evaluate the proposal of disaggregation as a means for encouraging a greater degree of academia's research outputs to be made more openly available and to help alleviate some of the issues affecting academia, but it should also further inform previous analysis of the current state of academic publishing discussed in previous chapters. Disaggregation stands in the context of being a response to an evaluation of disintermediation along with new ideas introduced through the development process necessitated by the programming of tools to explore new approaches to encouraging dissemination of research outputs. As such when evaluating it, an open minded approach should be adopted that will not only allow an evaluation to comment upon the manner in which disaggregation has been proposed, but will also provide a foundation upon which disaggregation may be reformulated or through which new theories may be developed, opening up the evaluation to

the possibility of being part of a grounded theory approach. Thus the methodology and implementation of disaggregation's evaluation will demand careful consideration.

Chapter 6

Disaggregation Evaluation

6.1 Methodological Approach

Disaggregation as a new approach to scholarly discourse is born out of an interpretivist evaluation of disintermediation through the lens of different theories and studies; with disintermediation itself arising as a result of taking a holistic view of academic stakeholders in an attempt to consider a wide range of different influences and motivations. Therefore to evaluate the ideas that lie behind disaggregation an approach is required which will build upon these previous studies by directly involving the stakeholders in question, namely researchers, academics and publishers, and asking them to comment upon the ideas and the Disaggregator prototype described previously.

Methods for eliciting feedback and thoughts from study participants are numerous, with both quantitative and qualitative methods each possessing their pros and cons. The feedback required for this research however demands a thoughtful and considered approach; the ideas that study participants are being asked to reflect and comment upon are complex and original, requiring an understanding of the existing scholarly discourse landscape. Thus the use of qualitative research methods gives participants the freedom to provide detailed feedback, in turn producing a rich dataset of perspectives that can be used to answer the research questions.

When considering evaluating an idea via feedback from experts, it is also important to consider how expert participants will be sampled. The 251st American Chemical Society National Meeting & Exposition Computers in Chemistry conference (Americal Chemical Society, 2016) provided such a venue for the study, with many relevant experts in attendance, including academics, those who work in academic publishing, and software developers all of whom are aware of many of the issues that pertain to academic publishing. However, in sourcing study participants from a conference, the sample is biased towards those people that are sufficiently interested in the use of information technologies and their application within chemistry and thus are likely to be enthusiastic about

its use, which may not truly represent the larger populations of the types of stakeholders being queried.

Taking the study sample into consideration, two approaches were therefore used to gather data: focus groups and semi-structured interviews. Firstly focus groups were designed to gather feedback on the wide range of areas that demand evaluation. Running focus groups to gather feedback presents a number of advantages; whilst they can be used as "a quick and convenient way to collect data from several people simultaneously", the group dynamics of a focus group presents an opportunity to "examine not only what people think but how they think and why they think that way" (Kitzinger, 1995) and provides "insights into the sources of complex behaviours and motivations" (Morgan, 1996), a relevant approach when challenging a deeply established system such as the one that has formed around scholarly discourse. Kitzinger (1995) goes on to not only sum up the key advantages of focus groups, but to also describe how "group dynamics can allow for a shift from personal, self blaming psychological explanations to the exploration of structural solutions" - a relevant facet when considering the complex relationship between the individual stakeholders in scholarly discourse, what each of the stakeholders are trying to achieve by participating, and how the structure of the system can be seen to have inhibited innovation in this area.

Another key aspect in the evaluation of disaggregation and the Disaggregator prototype lies in the understanding of the language and vocabulary used to discuss these ideas. The term disaggregation has been derived to place the focus on the notion that the traditional scholarly units, of which journal articles are a typical example, can be broken down into constituent parts each of which can exist independently of their aggregated whole. However, the different stakeholders may see the different interpretations of scholarly discourse in different terms which informs not just both future work concerning the Disaggregator idea, but also how the stakeholders of scholarly discourse relate to one another. Focus groups provide a method for teasing out the norms and language used by different parties: "group work ensures that priority is given to the respondents' hierarchy of importance, their language and concepts, their frameworks for understanding the world", which in turn allows the group to "highlight (sub)cultural values or group norms" (Kitzinger, 1995).

When running focus groups decisions need to be made as to how to segment participants – who should be in which group. As discussed by Morgan (1996), "the most obvious kinds of segmentation capture something about the research topic itself", with the natural boundaries for segmentation in scholarly discourse existing between researchers, those who work for institutions and facilitate research, such as subject librarians, and publishers who redistribute the researchers' work. Segmentation in this fashion creates a number of advantages, both in how it "facilitates discussion by making the participants more similar to each other" through to adding a "comparative dimension" to the data analysis stage (Morgan, 1996).

Semi-structured interviews also offer an opportunity to gather detailed feedback from individuals. Conducting the research at a conference venue to take advantage of the density of experts available means a flexible approach to data gathering is required: it may not always be possible to organise focus groups with a number of individuals at once and as such semi-structured interviews may be used to gather data from participants where running a focus group would be otherwise infeasible. Not only does running interviews provide the flexibility required to gather data from a busy conference venue, they are also a suitable method of research. As described by Gill et al. (2008), "the purpose of the research interview is to explore the views, experiences, beliefs and/or motivations of individuals on specific matters", presenting a method which allows for disaggregation to be probed with respect to its different facets and its potential affect on different stakeholders. Semi-structured interviews provide a further advantage, helping to "define the areas to be explored, but also allow the interviewer or interviewee to diverge in order to pursue an idea or response in more detail" (Gill et al., 2008); a useful trait when considering the expert nature of study participants and the broad topic under discussion.

6.1.1 Focus Group Procedure

As touched upon previously, the focus group participants are experts from the domain of academic publishing and thus already very familiar with its existing mode of operation. Therefore to consider new ideas may prove challenging when the ideas are disruptive in nature, requiring significant shifts in behaviour from a range of stakeholders. Thus to encourage an open minded perspective, participants are explicitly asked at the beginning of the focus group to consciously discard their existing preconceptions concerning academic publishing and instead asked to imagine a scholarly ecosystem that sounds not unlike the disintermediation approach proposed and evaluated previously: a system without conventional publications, a system where researchers could release any research output on the Web at any point and retain control over who has access to them. Participants, having had the scenario proposed to them may then be asked to consider its impact on their role, the impact on other stakeholders in the system and what the advantages and disadvantages of such an approach would be.

Once this idea has been explored with the focus group, the discussion can move on to the more moderate proposal that disaggregation poses. Having asked participants to abandon their preconceptions previously, it is hoped that at this stage participants will retain an open mind when considering disaggregation and the paradigm shift it represents, rather than disregard it (which is not to say that negative findings are not to be expected or desired, but the study has been designed to discourage them from arising too rapidly). Depending upon time available, at this stage the idea of disaggregation can be introduced to the focus group either via a brief description presented by the researcher

running the focus group, or via a brief demonstration of the Disaggregator prototype. Participants are then asked to consider how disaggregation may be incorporated into their workflows and how they imagine it might (or might not) operate within the confines of real world academic publishing systems. Thus the two stages of the focus group are designed to elicit thoughts and opinions on how academic publishing's stakeholders may be affected by such changes, what drives them and what they think motivates others, what potential exists for changes to the system in an ideal scenario and how we might reach this scenario when considering the context in which academic publishing is bounded.

6.1.2 Semi-Structured Interview Procedure

The semi-structured interviews follow a similar agenda to that of the focus groups and are designed to be used for when a potential study participant is unable to take part in a focus group, but may be available to discuss the ideas of disaggregation at another time. Interviewees are asked to briefly explain their role within academic publishing, before being asked to consider the hypothetical world of scholarly discourse that features elements of disintermediation. Then interviewees are asked to evaluate how disaggregation may exist alongside the conventional approaches to scholarly discourse that are commonly practised today. The interviews adopt a semi-structured position to let the expert interviewees elaborate on their specialist expertise where appropriate.

For a full interview guide, please see Appendix D.

6.2 Analysis

To allow for a broad and reliable analysis, audio recordings of the focus groups and interviews and subsequent textual transcriptions are used. Methods for analysing an interview or focus group transcript are numerous ranging from approaches that focus on the *meaning* of what is said, including meaning coding, content analysis, meaning condensation, meaning interpretation; and those which focus on *language* such as linguistic analysis, conversation analysis, narrative analysis, discourse analysis and deconstruction (Kvale, 2007), or are simply more free form such as thematic analysis (Bryman, 2012). Many of these approaches to analysis have been designed to elicit information from study participants concerning their relationships with social systems which are under investigation and thus approaches such as linguistic analysis aim to look beyond the explicit statements made by participants in an attempt to uncover latent insight that is not immediately obvious through a simpler reading of the text (Kvale, 2007), whereas the purpose of the focus groups and interviews to be conducted as part of this study is to gain an insight into a theory that has already been proposed. Therefore, as suggested by

Kvale (2007) a "theoretical reading" may be the best approach to adopt whilst analysing the transcripts, where "recourse to specific analytical tools becomes less paramount with an extensive theoretical knowledge of the subject matter of an investigation". However this approach lacks the nuance demanded by the other consideration that will be taken into account when conducting the focus groups and interviews, and for which the semi-structured approach has been chosen; the need to retain an exploratory perspective that ideas of disaggregation and disintermediation may stimulate further conversation concerning scholarly discourse that may introduce new concepts, or frames previously discussed theories within new contexts. Kvale (2007)'s notion of "interview analysis as bricolage" addresses this point, permitting a mixed methods approach "where the interpreter moves freely between different analytic techniques" and which is ultimately often a "common mode of interview analysis".

Therefore to facilitate analysis of the interview and focus group transcripts, a coding approach is adopted, which as described by Bryman (2012) involves "reviewing transcripts [...] and giving them labels (names) to component parts that seem to be of potential theoretical significance and/or that appear to be particularly salient within the social worlds of those being studied". Coding offers the flexibility to both continue with the "analytic induction" thread that weaves through this thesis so far (whereby hypotheses are redefined or reformulated to conform with the data) and will continue as a means for the evaluation of disaggregation; but it also permits the introduction of grounded theory as a means to formulate results, which itself "relies on coding in an emergent approach, not preconceived" (Bryman, 2012). Coding in this manner also has much in common with thematic analysis, which requires interpreters of transcripts to identify patterns "found in the information that at minimum describes and organizes the possible observations and at maximum interprets aspects of the phenomenon", with the themes themselves either "generated inductively from the raw information or generated deductively from theory and prior research" (Boyatzis, 1998); a trait that ideally suits the combined evaluative and exploratory nature of this research. The codebook guiding the analysis is therefore comprised of codes and categories that are derived from the research questions presented in the previous chapter (and which are reflected in the focus group and interview schedules), and those codes which emerge over the course of conducting the analysis. For the full codebook, please see Appendix E.

Once the coding is complete, the resultant codes may be analysed and overall themes elicited. These themes are then outlined, using supportive quotes, to describe their context and how they have emerged. Having established the themes, each theme is discussed further, elaborating upon the findings to place them within the context of the literature previously discussed, and ideas previously proposed within in this thesis. Through doing so new insights may be gleaned with regards to the role of the Web in scholarly discourse, the parts and practises of the various stakeholders and the feasibility and implementation of disaggregation.

A final thing to consider for the analysis of the focus group and interview transcripts is the differing nature of these two means of data collection. As has already been touched upon, the two techniques have their advantages and disadvantages, with the extra dimension of having numerous participants in a focus group who will react and respond to one another being the key difference to consider when conducting an analysis. The focus group schedule has been designed to elicit participant's thoughts on academic discourse and therefore it is hoped that the relationship between participants is unlikely to affect their response.

6.3 Disaggregation Evaluation Results

6.3.1 The 251st American Chemical Society National Meeting & Exposition Computers in Chemistry Conference

Whilst the American Chemical Society (ACS) National Meeting & Exposition Computers in Chemistry conference (Americal Chemical Society, 2016) provided a venue from which to source participants for the study, the presentations given as part of the Chemical Information (CINF) conference track also served as a valuable source of information. Many of the talks presented covered issues pertinent to this thesis, with themes covered including the issues that currently afflict academic publishing contrasted with potential that the Web affords to improve scholarly discourse, the roles of different stakeholders who participate, the possibilities that various technologies enable and how behaviours and practises of those who operate in academia may be encouraged to change.

Many of the presentations made reference to the problems of scholarly discourse which it should now be possible to solve through effective use of the Web. Carsten Kettner of the Beilstein Institut, Frankfurt spoke of "publications' focus on positive and novel findings" for example and a "lack of comprehensiveness, comparability and reproducibility" that arises when "negative results are left out". Mark Parsons of the Research Data Alliance also commented upon the need for research results to be reusable, elaborating further by highlighting the demands of "grand challenges", imposing research problems that "require diverse data to be shared and integrated across cultures, scales and technologies". The issue that much chemistry data is "locked up in text", a founding notion behind the hypotheses which have inspired the Disaggregator was also put forward by a number of presenters including Evan Bolton from the National Center for Biotechnology Information, along with numerous comments concerning the lack of suitability in using PDFs for sharing data from speakers such as Stuart Chalk from the University of North Florida and Santiago Dominguez from Mestrelab Research, also furthering the general use case proposed for the Disaggregator. Many of the goals that scholarly discourse should strive for were summarised by Imperial College's Henry Rzepa, who outlined the "importance of FAIR" (Rzepa, 2016), that is to say that research should be "Findable", "Accessible",

"Interoperable" and "Re-usable", which emphasises the importance of these traits and the manner in which they should be considered when implementing new approaches and technologies for scholarly discourse (Wilkinson et al., 2016); and Maryann Martone from the University of California elaborates on FAIR defining a Web enabled "scholarly commons" that takes a holistic perspective to the problem of scholarly discourse (much like the earlier chapters of this thesis), stressing the need to consider both "Non-Digital" entities such as the stakeholders involved, how data and knowledge is communicated via "protocols" and DOIs and the array of networks and platforms that are required to support such an approach. However, the economic idea of the commons invites the problem of the tragedy of the commons and how that issue may be addressed; Martone recognises that "scholarship rests on formalised citations" stating that it is required for scholarly outputs beyond traditional articles, a solution which also lies behind the reward and recognition hypothesis that has influenced the Disaggregator.

The role of the citation and its ability to act as an incentive was not just discussed within the context of fostering a scholarly commons; the use of citations as a tool for both rewarding researchers who deposit data and as a means for accessing and describing data was one widely discussed, with Ann Gabriel from Elsevier stating that the "biggest incentive of all is citation" (perhaps in keeping with the role a publisher plays in scholarly discourse). Along with a discussion of citations and the practise of citing comes the topic of identifiers. Sara Bowman, Center for Open Science, Virginia highlighted the need for "persistent, citable identifiers", along with Henry Rzepa emphasising the value of DOIs, the Royal Society of Chemistry's Valery Tkachenko recognising that "identifiers are vital for systems to work and Patricia Cruse from DataCite, Berkeley who described the "second class" treatment of data as scholarly artefacts resulting in the publication of data not having the same overall impact that the publication of a journal article might have. The need for change in the research community's view of data was further elaborated upon by Stuart Chalk, invoking the "Joint Declaration of Data Citation Principles" derived by the FORCE11 group, which endeavour to "cover purpose, function and attributes of citations" and include such notions of ensuring data is "accorded the same importance in the scholarly record as citations of other research objects", the roles of "credit and attribution" and the need for "persistence", "specificity and verifiability" and "interoperability and flexibility" (Martone, 2014). However, Cruse also discusses the need for citations to have context asking if the manner in which the citation is being used is either for or against the research outcome being cited. This limitation of citations exposes the lack of information that is contained within the meta-data of scholarly discourse and yet upon which a degree of significance and influence is placed. How such a problem should be addressed however is left unclear by the speakers at the ACS National Meeting; is the solution to devise new approaches to citations that describe a context or to simply not count them when considering a researcher's impact? Or perhaps, whilst the context of a citation is not represented, a citation nevertheless remains as a symbol of transferring knowledge and informing the thought processes of

peers, leaving the nature of the knowledge transfer a, potentially useful, but ultimately unrecorded and unimportant extra item of metadata.

Whilst the role of citations and expanding upon the ways in which citations can be deployed was cited as a means for encouraging a more open and data-centric approach to the dissemination of research outputs, conversely a number barriers to change were also identified by the various speakers. Mestrelab's Dominguez highlighted that whilst all stakeholders would benefit from greater "preservation and sharing of research data", it is likely to be hindered by the barriers to change that are the "labour intensive" approach that is often required, the "lack of common formats" with which to share and reuse data, and highlighted problems concerning the overall "accessibility to data". And as such, when conversation turns to tackling the obstacles in the way of change, so it turns to the roles and responsibilities of the various stakeholders in academic discourse. Serin Dabb from the Royal Society of Chemistry places responsibility on stakeholders other than the researcher, with institutional repositories becoming responsible for "data management and open access", whilst also commenting on the extra difficulties added by funding agencies with their differing requirements and noting publishers' potential for being able to alter scholarly discourse through imposing higher standards and necessitating data is deposited in repositories. The role of publishers is one also identified by Ian Bruno, Cambridge Crystallographic Data Centre (CCDC), who states that "many journals require authors to provide derived data" and in the sub discipline of crystallography, "most publishers require data to be deposited with CCDC pre-publication", thus evidencing the influence a publisher can wield. The roles and responsibilities of researchers themselves did not go unmentioned however, Linda Galloway from Syracuse University Libraries noted the rise of academic social networks, their increasing levels of influence and the need for academics "to keep some sort of profile on academic networking tools and demonstrate their impact", highlighting not only how researchers need to take an active role in disseminating research outputs, but also how they should embrace new methods and platforms for doing so (and in how doing so they may further their own careers).

Underlying all of the above discussions however is the role of technology in encouraging and facilitating the open sharing of research data and thus technology also acts as another stakeholder capable of eliciting change from those who engage in the academic discourse system. The technologies chosen have implications for how data is stored (which in turn affects its ability to retain its meaning, be persistently identified and remain accessible), how research data is managed both past and present, how research data can be used by humans and machines alike, and how new practises can be integrated with researcher's day-to-day workflows. The potential of electronic lab notebooks (ELNs) to facilitate such processes was highlighted, with Leah McEwen, librarian at Clark Library, Cornell University indicating how appropriate deployment of ELNs may result in

"improved accuracy of data collection", "easier processing of outputs", and "more consistent record keeping" amongst other benefits with an overall effect of helping to "improve intra-lab communication" and an aim to "instil better habits in students". However as Chalk points out, implementing ELNs that are effective requires software that can handle a broad range of types of data beyond research results, which include "workflows", the "research story", "hypotheses" and information about the people involved with the work including "expertise, provenance, integrity" and "eminence".

Thus from a brief analysis of the presentations given at the CINF division of the ACS National Meeting, an overall context of scholarly discourse in the chemistry community can be derived, indicating the driving forces of change and the approaches that are considered most important. Overall themes that emerged include data as second class objects and what is required for research data to be treated with the same degree as traditional journal publications (from a range of perspectives from how it is cited, what it can tell us about researchers and institutions and how it can be re-used and integrated with systems to help enact better science), the roles and responsibilities of different stakeholders in encouraging and actually implementing changes, and the potential of technology, including best practises for how it may be deployed and what needs to be further developed. It should be noted however that the conference had multiple speakers talking in parallel sessions and as such the sample of opinions and approaches of the scholars and academics described above can only represent a cross-section of the community.

The presentations given at the conference have some initial implications for disaggregation. The need for a process like disaggregation is made evident by comments given on the need to make more research data available to encourage principles such as FAIR. However the talks also give rise to some concerns about the way in which disaggregation has been proposed, including the difficulty of extracting data from PDFs, and if data is to be FAIR and citable (the importance of which was not understated by the speakers) then the Disaggregator would need to provide both a robust infrastructure for making disaggregated outputs citable and guidance on how this should be achieved. A theme of all stakeholders needing to take a role in encouraging researchers to be more open with their research data also emerged, an approach the disaggregator avoids by appealing directly to the researchers themselves. The presentations given imply it is pressure from stakeholders such as institutions and publishers that are responsible for changing researcher behaviour, rather than incentives born out of new initiatives.

6.3.2 Overview of the Implemented Study

A number of focus groups and interviews were held both before, during and after the aforementioned American Chemical Society (ACS) National Meeting & Exposition Computers in Chemistry conference (Americal Chemical Society, 2016), with different stakeholders from academic discourse. For all bar one of the individual instances of the study carried out an audio recording and transcript were produced, with an audio recording of one focus group unable to be taken due to the impromptu nature around which the focus group formed at the ACS conference; notes of all the points deemed salient to the researcher were taken however for future analysis and the roles the participants have in respect to scholarly discourse were also recorded to aid this analysis.

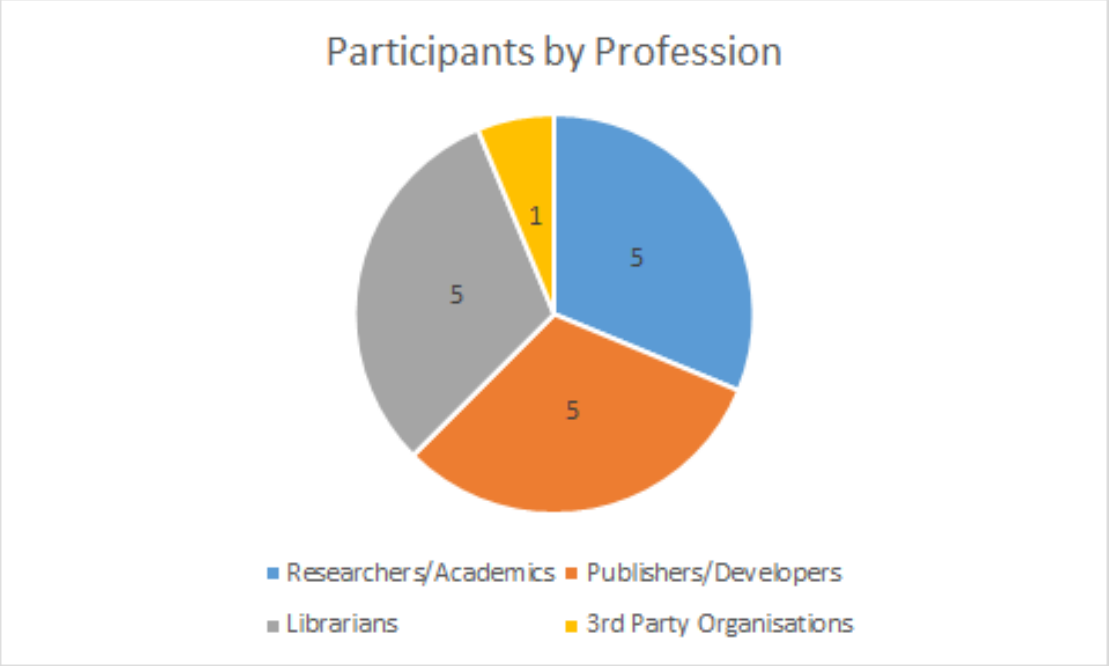

FIGURE 6.1: The number of participants and their stakeholder roles.

Figure 6.1 shows the number of participants categorised by their roles as different stakeholders. The category "3rd Party Organisations" refers to non-profit organisations that provide scholarly services but could not be categorised as either publishers or academic institutions. It should also be noted that whilst the participants who took part in the study could typically be ascribed a particular role in scholarly discourse, be it a publisher, an academic or a librarian for example, many of the participants have fulfilled alternative roles in the past, and, during the course of the focus groups and interviews, reflected back on their experiences from past roles. The very nature of this area of study and the experiences necessary to fulfil the scholarly discourse stakeholder roles, means that nearly all of the participants do at least have a background as a postgraduate researcher and so many of the comments are made with an awareness of this perspective in mind. Another aspect of the participants to consider is that many of the participants

were attending the Chemical Information (CINF) track of the ACS conference and as such are (by design of the study) already interested in matters of scholarly discourse and thus any findings reported or new theories generated may not reflect the opinions of chemists whose interests can be found further afield from this particular topic.

Having coded the transcripts of the interviews and focus groups using the codebook outlined in Appendix E, the coded extracts were then studied to elicit what themes may be present within the texts. Appendix E also details how often each of the codes where referenced in the transcripts and Figure 6.2 shows the top five most frequently referenced codes. This gives an initial insight, into what areas the participants focused on, which will be explored in more detail as this chapter goes on to elaborate on the themes, with the aid of quotes taken from the transcripts.

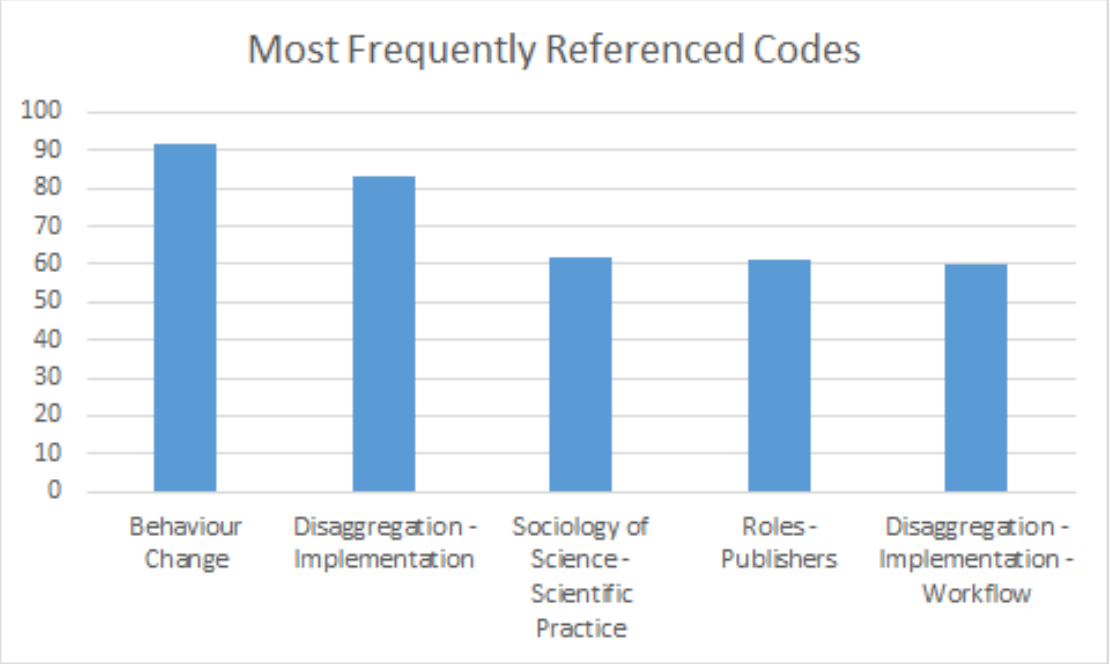

FIGURE 6.2: The most frequently referenced codes, highlighting how the topics of changing researcher behaviour, the implementation of disaggregation and the roles of publishers were all frequently referred to.

6.3.3 Norms and Behaviour

6.3.3.1 Results

Cultural Norms

A section of the interview and focus group agendas was set aside to discuss how researchers may be encouraged to adopt new approaches to disseminating their research outputs. However, ideas concerning why researchers (and the other stakeholders of scholarly discourse) behave in the manner in which they do, and why they make particular

decisions was a theme that peppered the participants' responses. This theme first begins to become apparent when participants talk about norms:

"one way or the other you're going to have to either honour it or shift that cultural norm and that's huge"

The comment above was given in response to being asked about disintermediation, yet the message that it conveys nevertheless remains relevant when considering any type of change to scholarly discourse, placing emphasise on the influence that cultural norms play within scholarly discourse. To what extent cultural norms are influential becomes clearer still when one participant presents a potted history of scholarly discourse and the ways in which we choose to exploit the medium of our time:

"Now we have digital technologies, it is very clear that digital technologies, those representations that were best suited to the printing press need not be best suited, but one of the problems we have is [...] what they [ethnographers] would call stabilised habitual behaviours"

Thus an emphasis is placed on the established protocols of science and the manner in which they can shape the practises of researchers. One area where the apparent influence of cultural norms is particularly prevalent was exposed in one of the focus group sessions, where the participants began to discuss the quirks of peer review:

Participant 1: "I'm trying to think what are the incentives to do anonymous peer review for someone else?"
Participant 2: "Yeah it's because it's expected as something that you do."
Participant 1: "Yeah I suppose it would be [...] you feel like you're important if someone asks for your opinion..."
Participant 2: "Sometimes you get to see interesting research in a similar area [...] it acts as part of doing your [...] professional duty."

With certain cultural norms identified, contradictions that exist within scholarly discourse also began to emerge:

"that's the odd thing, that on one hand you have this very competitive careerist thing, and on the other hand, people still have this culture you must do the peer review"

Need for Change

Having established that elements of scholarly discourse are bound by conventions, study participants also recognised that change is required:

"there's a huge amount of value there, just in getting the community to start talking about this in more intelligent ways and to really start seriously thinking about, do we want continue to communicate in the 21st century in the way we were communicating in the 18th century?"

"it's a very interesting time because we need to change the model of science in a fundamental way I think and that's going to take a long time to do."

"I think what you need to do is [. . .] take these things in small steps."

"it's one of those things, where if you can get a consensus that something is the new way to do things then everyone will decamp there en masse and it will be great, but finding that magic formula is going to be very difficult"

For the norms themselves can be inhibitive, not suited for all researchers or institutions:

"not that the current model fails but that maybe it's a higher barrier to being able to highlight your work, [. . .] you can publish in a regular journal and that's from an undergraduate institution perspective that's . . . very important thing to do obviously for tenure, but . . . in of itself, it's not as easy to do obviously as it would be at a major research institution because my undergraduate students that work with me in the lab, they nominally will stick around for 9 months and then they move on and then so from a practicality standpoint its somewhat difficult sometimes to actually get a project to conclusion so you can publish it"

"for someone like myself who is tenured, who just needs to get to full professor but you know, that's a motivation but it's not something that's absolutely critical at this moment in my career"

Similarly not all research suits being disseminated in the conventional manner, as described by one participant recounting an interview they had conducted with a researcher as part of their own study:

"he said from his perspective if you know, you can still enable the scientist to have, you know, to be able to trace their provenance so that they can, you know, be given credit for it and have the conversations with their peers if someone wants to use it – but why put it behind a firewall? Then he said "I wish we had an arXiv for pedestrian chemistry," is the phrase he used"

Which in turn leads to new advantages:

"that would facilitate a tonne of new research opportunity because it would be open and that's actually where most of the chemistry happens and probably where most of those. . . the little problems still need to be solved, optimisation stuff and all that, needs to be worked out here, it's just not the cutting edge stuff"

Role of Education

With a requirement for a change in the culture of scholarly discourse identified, study participants began to discuss how such a change might begin to take place, with many focusing on the role of education to provide a sustainable and meaningful change.

"we're in the third digital revolution and . . . and so my personal opinion is that we need to educate the youth, the youth are actually going to move this forward not the elders"

"I think it has to be part of their . . . of their sub. . . almost their subconscious practical thinking actually"

The comments on education also provided an indication as to what researchers need to be made aware of, from helping to allay fears, to providing scientists with the skills to take advantage of the affordances new approaches to dissemination have to offer:

"I think we have to educate them and say, you know what, as much as traditional publishing is important. . . you can make contributions in chemistry in other ways than publishing research papers"

"changing the culture more might be a matter of educating people and saying actually this is not a massive risk and actually its all linked to your name so its cited, can be cited to you and nobody can scoop you on it because you've already published"

"There are a lot of these fears about moving to a different system that people just. . . they're gut reactions and they're not actually based on any evidence or research about how the system works and how this might affect it"

"this is one of the problems we have, to really do e-science you have to know some programming"

"I'm trying to do now is figure out how to get more cheminformatics into the undergraduate curriculum"

Role of Peers

Some of the comments focused on the nature of learning from one's supervisors and peers, which not only begin to highlight the importance of the overall scientific workflow in changing scholarly discourse, but also highlight how local champions may be effective as discussed in the earlier chapters of this thesis:

"I mean the, the sort of vision I have is of the young, enthusiastic PhD student wanting to publish anything they can get their hands anywhere and the supervisor being largely absent and the post doc in charge of them saying "no that's not the way you do things, you do things the way I was taught to do them when I was a PhD student""

"there's a lot of managing your personal data information stuff is. . . idiosyncratic and ad hoc and even when you've got some big machine that reduces data, the data management from that is a lot less strict in academia"

"so you really need exemplars of people doing this approach and that is. . . that will transmit so much to the community, people will really take notice and start thinking of doing it that way themselves"

Role of Mandates

When discussing how to encourage researchers to adopt new behaviours whilst disseminating research outputs, the use of mandates was queried, but invariably received similar responses:

"if you try and mandate it then there is always going to be a backlash"

"you tend to find that if somebody has an inherent value in doing it then they will just do it, whereas if somebody is mandated to do it they will give you the absolute the bare minimum required to fulfil their mandate"

Interviewer: "And so how does this all contrast with things being mandated, is that not . . . that not essentially as effective do you think?"
Participant: "Oh not at all"

Thus the initial theme focusing on cultural norms and the behaviours of researchers once again reiterates how difficult to initiate, yet necessary, change is within the sector of academic publishing. The comments given highlight how scholarly discourse is a system that has developed over hundreds of years and as such as developed quirks and attributes that protect it from the disruptive influence of the Web. Yet they also provide an insight into how it is possible to change, with many of the norms and behaviours embedded within the scientist's day-to-day work practises.

6.3.3.2 Discussion

The theme of norms is a pertinent one to begin with because it helps to reiterate many of the problems that this thesis has concerned itself with and exposes the system of scholarly discourse as a curious network of stakeholders that has developed certain characteristics over many years. The nature of scholarly discourse is summed up in one participant's comments concerning the ongoing practises of the 18th century found in the 21st, reiterating the discussion in the earlier chapters of this thesis and in previous studies (Fyson et al., 2013), that the Web largely facilitates more conventional forms of scholarly discourse; it simply allows journal articles (and similar artefacts) to be disseminated more quickly and easily, rather than revolutionising the market. Following on from this conclusion, this thesis explored the potential for the Web to disrupt scholarly communication, through the means of disintermediation, and then sought to understand why such disruption had not occurred using ideas such as game theory and knowledge management, when the affordances of the Web otherwise indicate it should be possible. The participants' comments further reveal why such a disruption is so difficult to instigate, bringing attention to the cultural norms of scholarly discourse that have resulted in many of the exchanges that take place to become ossified. Thus whilst earlier chapters have examined why the researcher acts in the way that they do and learn from

these ideas to propose new methods of encouraging open discourse, through the medium of disaggregation, it becomes clear that the analysis and ambition of change requires a grander scope that cultural norms suggest. Burns et al. in their report *Advocacy to Benefit from Changes*, also identify the change that is required as being a "cultural change", noting that "embedding new behaviours takes time and needs a detailed knowledge of not only the discipline but also the local infrastructure and culture. Advocacy plans need to take a long term approach, to gain buy-in at an institutional level and be tailored to the local culture and infrastructure".

Framed in this manner, the participants reveal a theme of conflict in the relationship between a long term shift in cultural norms and a short term shift in changing behaviours and decision making. As such a number of questions arise: to what extent can appealing to stakeholder's motivations be used to encourage behaviour change; to what extent do cultural norms inhibit the effects of new approaches enabled by specific technologies or policy changes; and how may cultural norms themselves be adapted to pursue the changes in scholarly discourse that would help to solve some of the previously identified problems?

To understand how it may be possible to shift the cultural norms, it is necessary to understand what they are founded in. An example that may give some insight, as indicated by the participants, is the seemingly contradictory nature of peer review, an important part of the scientific process, and a role of the researcher, which itself was another ongoing theme that needs to be explored in more detail separately. Being such a fundamental facet of scholarly discourse, it is one that is ultimately achieved through fulfilment of "professional duty", displaying researchers' passion and adherence to the scientific method and the workflows that produce their research outputs. However it is not the requirements of the scientific method alone that ensure peer review continues; the vanity and curiosity of researchers also plays a part, which itself could be prescribed as different manifestations of the the role of recognition in science, the only "property right" of scientists (Merton, 1957), a notion identified by (Ben-David and Sullivan, 1975). Nevertheless peer review provides evidence that the scholarly community does not need explicit incentives to carry out essential elements of their work; it can simply be achieved by developing the appropriate culture.

Thus the question inexorably alters from how do we encourage researchers to disseminate their research outputs in more appropriate formats and more openly, to how might we shift the cultural norms of scholarly discourse such that more open practises that facilitate reproducibility and greater access to knowledge become the researchers' professional duty, a requirement recognised by many of the participants. That is not to say however that the Web has not had an impact on the culture surrounding academia, as demonstrated by the likes of the Budapest Open Access Initiative (Chan et al., 2002) and the many organisations and technologies that have formed to facilitate the dissemination of research outputs such as FORCE11 and the Research Data Alliance. Yet engendering

a cultural shift is evidently a process that takes time and careful planning, an argument that is backed up by both the participants' responses and the Web's history with respect to scholarly discourse.

It is not simply the passage of time that delivers progress however, with study participants rejecting the idea that old practises are replaced by new ones as one generation of researchers is succeeded by the next. A shift of culture is a time consuming process that requires constant pressure to gradually alter the status quo and thus it is perhaps therefore expected that many of the participants acknowledged the importance of education, which offers the opportunity to raise awareness of the issues and present researchers with the tools and skills required to engage with more open, digitally enhanced methods of scholarly discourse. As seen in the quotations presented above, education is not just concerned with teaching researchers new practises, but may also play a role in helping to allay fears or prejudices that researchers may have about new approaches to scholarly discourse – a point that corresponds with similar findings when examining knowledge management with respect to disintermediation (Cabrera and Cabrera, 2002). The indications of the quotations above is that for a shift in the culture to take place, it is ultimately the policy decisions of institutions that educate early career researchers that will hold significant influence, an influence that may prove to be much more effective than introducing new technologies such as the Disaggregator tool. However, not only does education place a focus on the role of the institution, it also places an emphasis on the workflow of the researcher, which is informed by researcher's peers and supervisors. With knowledge and practises being passed on from one generation of researchers to another, it becomes apparent that a cultural shift can become sustainable if new techniques are embedded within existing workflows and seen to benefit the researcher.

However, whilst education and academic institutions may have a key role to play in helping to influence a paradigm shift in the manner in which researchers disseminate and use research outputs, as this thesis has previously demonstrated, they are far from being the only stakeholder with influence in the scholarly knowledge cycle.

6.3.4 Stakeholders

6.3.4.1 Results

Throughout this thesis, the varying roles of the stakeholders in scholarly discourse and the influence each possess have frequently informed ideas and discussions, having arrived early on at the conclusion that only by understanding the interplay between the stakeholders is it possible to properly understand the Web's unique role. Thus the parts of stakeholders in scholarly discourse and how they would be affected by the changes implied by the implementation of disintermediation and disaggregation were key topics in the interviews and discussions held.

Researchers

Upon discussing the role of the researcher one academic, contributed the same premise that underpins disintermediation:

"We are the ones who create the knowledge and... and write the papers and we're the readers of the papers and we're the ones who are hiring our fellow professors and we're the ones making the decision on who gets tenured and who gets promoted. And we're the ones writing the reviews of the grant proposals, that are used to fund the research"

With the interview participant continuing further:

"if the community decided we're going to make some dramatic changes the... the publishers will have to roll with whatever we do because we're both the supplier of their content and the buyer of their content [laughter]. But I... you know, this is a huge change."

The many roles of the researcher was a theme amongst participants' responses:

"I interact with publishers both as a reviewer and of course as somebody who tries to publish and read papers"

"The peer review process is largely governed by the community as it stands anyway, it is directed by... publishers, ostensibly... in truth we pick our own reviewers to some level or other"

The fact that researchers are burdened with many tasks also featured in some responses when discussing how new systems or approaches, such as disaggregation, may be deployed in a bid to change the scholarly knowledge cycle, with an emphasis placed on the researcher's main passion of conducting their research:

"I'm still very interested in, [an] integrated system that can really allow the scientist to do the science they want to do but manage the data capture and dissemination. I mean [...] that's really what we need. You know scientists really just want to do science, they really don't like writing it up."

"I feel like if you were doing something like that and uploading single reactions and they had to be sort of checked and commented it would almost feel like you would have to go through peer review twice for the same thing where you do it once for the molecule, once for the reaction, once for the paper and that's a lot of extra burden on researchers"

The passion for doing science was also cited as a typical reason for why researchers enter into the profession, but an individual researcher's motivations were also noted to shift during their career:

"I think the altruistic bit is a part of the jigsaw and can get people, sets them off and can still be a part of the types of questions they're trying to answer, that are... because part of that is all these sort of global challenge type things of curing lots of diseases and

feeding the world and... yeah powering energy, yeah, good clean energy for everyone, so yeah that's still an important part of a motivator"

"I mean the trouble is you have these two views, you have this idealistic view of doing all this great research and then you have the fact that everyone else has this idealistic view as well and there are far more people wanting this glorious scientific career than there are places available, so if you don't do the careerists stuff and don't get into the careerist mindset then, then you end up working somewhere else."

"maybe at a PhD level you still got a lot of idealistic people but the higher up you go on the ladder, the more you find people who are motivated by their reputation"

"the reality is... that you need to put food on your table, you need to put clothing on your back and a roof over your head and if you've got family, you've got to support them too and so you've got to play in the real world and the real world right now... counts papers, published in peer review journals and until that culture changes, nothing else changes"

Nevertheless as one participant pointed out:

"people want to talk about the things they're working on because they're excited about them"

Not only does the role and the motivation of the researcher become more established, but through interviewing participants with different backgrounds it is possible to glean an insight into how the different stakeholders view each other. A common thread in researcher participant responses was the ability of the publishers to be agents of change:

"my first step would be to now go and... evangelise journal editors and work exclusively with editors and editorial boards and say look, we need, you need to recognise the data is the coin of the realm, that it's how science gets done and we now have tools to enable sharing of data that we didn't have before and you need to make it mandatory to archive your data if you want to publish and I think that's the first step for moving in this direction."

"If the journal editors at the very best journals said, you want to publish here, you have to deposit your data, everybody bitch and moan because its extra work but they would do it. Compliance would be 100% because everybody wants to publish in those best journals."

" the number of A+ journals out there is small and if you want to publish there, you'll do whatever they say you have to do"

"a publisher can help a bit to formalise some of those expectations, put some things in place but a lot, a lot of it has to come from... you know from the expectations of that particular group unless you do this this way or this that way"

Publishers

However the researchers' views on publishers differs from the publishers' and learned societies' own perspective:

"[learned societies have] less concrete influence in terms of, terms of the society, in terms of what is best practice for chemical scientists to, so that tends to be very general rather than more specifics"

"there's a sort of element of contact there with our community and bringing in people together around specific areas of the chemical sciences and defining out where problems, although that's probably in terms of research areas, things we need to encourage, funding of one sort or another, but that's more general in terms of trying to influence the amounts of money and different bits of chemistry."

"even with mandating to provide enough supplementary information the publishers are struggling for years, my question, is why don't we request or require authors to submit enough supplementary information means "oh oh our authors will leave us, they will go to another publisher" [. . .] so if you will try to introduce a rigid processes in what seem to be pretty relaxed life for a researcher before you will find a huge push back."

It should also be noted that participants commented upon the changing nature of the academic publishing industry. For example, a developer working for one publisher said:

"the primary thing that we're trying to do, is to manage and electronic supplementary information"

Whereas one academic reflected on the changing business model for publishers as the Web continues to be used for scholarly discourse and the importance of data is emphasised:

"I think the publishers are just going to migrate their models to be hybrid, you know it will be regular publishing, open, or open access publishing and it will be data publishing, slash database generation. . . and access for scientists because data is where the money is."

Funders

One of the focus groups held also elicited comments on the role of funders, with the focus group participants remarking on how the "comfort level" of what research data can and cannot be shared depends on how the research is funded. It therefore follows that funders have an element of responsibility to define how research outputs they fund should be disseminated and thus they are also influential stakeholders in the scholarly knowledge cycle – a theme that many of the interview participants commented upon:

"Here in the United States, the national science foundation requires as part of a grant proposal a document indicating how you plan to distribute the data that's created in this work."

"if that document is not part of the proposal when you submit it, it gets rejected right out of hand. Right, the rest of it could be nobel prize stuff but if you don't have this data management document in there, its just, its just tossed aside. So mandates... mandates will work."

Interviewer: "And there's not the fear that the researchers will just go elsewhere then if... if one funder is mandating it the researcher will say "oh I'll go to this other funder" or vice versa [...]?"
Participant: "If the funding situation changes dramatically and there's a lot more funders out there."

Although the extent to which they exert their influence is questioned:

"if there's more follow up from funders for that sort of thing that's what slowly will cause cultural change"

"if funders saying, look we want to see evidence of impact of your work and blog postings and twitter and... other social media outlets and how much stuff you've put on Figshare, we're going to value that, that will encourage people to do it."

Other Stakeholders

Whilst most comments centred around the more significant stakeholders of the researchers, publishers and funders, some participants had reflections on some other stakeholders too, including commenting upon the role of industry and the influence of organisations such as IUPAC (the International Union of Pure and Applied Chemistry):

"I think it's going to take articulation from major chemistry bodies like IUPAC, like the major publishers, like the major societies to... to promote the idea that even though journal articles are still... an important way in which scientists contribute their work to the community [...] that we should and could embrace other... mechanisms of publishing, or other formats of publishing and...] some kind of statements as to... how equivalent they are"

"with chemistry I notice that academic researchers I worked with, they're very sensitive to industry connections that they have, because they're really interested in having their science... implemented and transferable, so you know, they want to know what the industry is doing, but also how the industry... what's important to the industry is important to them"

"I don't know if this statistic is true, but certainly the industry uses journals a lot, we use them too, but I think they really use them a lot and they're not contributing back. They pay the publishers but they're not contributing content back"

"in general the times where you see standards emerge it is because a body has set one. [...] The initial prototypes for those standards certainly emerged from the community but somebody eventually had to sit down and say, hey this is how it is."

With one librarian interviewed elaborating on their role:

"I support research chemists, I support chemical education at the undergraduate and graduate levels and I manage... one and half million dollar budget purchasing... you know research materials, which includes a lot of published journals and databases in chemistry, physics and many other related fields"

"I do talk a lot with publishers in terms of licensing, and pricing and quality"

Adding Value

As elaborated upon earlier in the thesis, the stakeholders can be largely defined by their interactions with one another, which formed another aspect on the theme of stakeholders in the participant's responses. In this instance, these comments typically involved the notion of adding value and the extent to which added value is communicated to the stakeholders at large:

"I think a lot of that, discussion about where the value is in the publication process is not being properly articulated."

"I think that's the real key question to all of this is... is for every step here what is the value add"

"publishers have to be more clear at laying out you know, these are what... services we are providing and... and we need to be paid for those services"

"if you're still valuing peer review, if you're still valuing an editor sifting through this and trying to craft a journal that has a point of view... if you're talking about copy editing, if you're talking about copyright protections... you know... you're worried about plagiarism, you know, all these other things that... that publishers will argue that they do"

With faults being identified of both sides of the conventional researcher–publisher relationship and beyond:

"I think scientists have been bad players because we just haven't thought about this at all"

"You know, we've washed our hands of it and let the publishers and librarians take care of it for us."

"I think publishers have, by and large, also done a bad job in... in articulating the value add that they provide"

"I think its incumbent on publishers to let us know what value they're providing and its incumbent on the scientists to say what values we want."

"I think librarians have done a very poor job of educating people about all the services they provide and why they need money."

"Yeah but I'm not sure we [researchers] thought about that, because I think we still are thinking about the journal as a delivery mechanism. This was... this was you know the [...] 18th century role which was to aggregate, print and bind and then ship. Well now they're aggregating and the printing and binding now means putting it up on a website, but I think most people are still thinking of that as the predominant role of the publisher."

As ideas of value added and issues of clarity in stakeholder exchanges emerges, one comment from a participant draws out what researcher's value the most:

"that's because scientists value those things and scientists have to really start thinking about what is it that we want to have happen in our communication process, what's important to us. And right now, I think its top of the list, reputation is number one."

6.3.4.2 Discussion

The roles and interactions of stakeholders has been a key theme that has influenced many of the ideas, especially with the proposal of disintermediation. The participants' comments back up much of the analysis of the stakeholder interactions, also coming to the conclusion that the researcher fulfils many of the requirements of scholarly discourse and could, if acting as a cohesive community, exert a disruptive influence. However they also make clear that researchers are keen to do science more than any thing else – writing up their results to publish journal articles becomes an inconvenience, necessitated by a requirement to bring a focus to the work and disseminate research outputs. This in turn reveals an inherent value that publishers and journal articles add to the system of scholarly discourse. By providing a venue to disseminate from and prescribing what a final output that the research is aiming towards should look like, the mere presence of the publishers adds value to scholarly discourse, helping to lubricate the smooth running of all the other functions that stem from it.

What also becomes apparent from participants' reflections is the near corrupting influence the scholarly knowledge cycle has on researchers. Study participants indicated that many researchers join the profession for "altruistic" reasons, with a desire to help resolve "global challenge[s]", but the pressures of a competitive research environment causes researchers' attitudes to shift – a long standing issue as documented by Altman and Melcher (1983) when describing the "publish or perish" nature of academia over 30

years ago. Such pressure derives from socio-economic realities – the number of vacancies at academic institutions is limited and researchers like anyone else need to earn money.

Discussions with various stakeholders about other stakeholders' roles also resulted in differing perspectives at odds with one another. The first emerges when considering the role of mandates for encouraging changes to way in which researchers disseminate their findings. Previously participants commented that mandates are ineffective ways of creating change, which researchers will ultimately be reluctant to commit to. Yet both funders and publishers alike seemingly have the ability to impose changes. Funders can exploit the rarity of their position in the market for research knowing that researchers are unlikely unable to seek funding from those with less stringent requirements on dissemination of research outputs – the RCUK's requirement that "authors' outputs must have been deposited in an institutional or subject repository" (Higher Education Funding Council for England et al., 2015) is a good example of this, although the effect of this policy is yet to be determined. The reasoning behind mandates can also be quite simple: where research is publicly funded, the public at large should benefit, a key argument for OA mandates (Department for Business Innovation & Skills, 2012; Stebbins, 2013; Van Noorden, 2013). Another example comes from the Structural Genomics Consortium (SGC), formed of funders, academic institutions and industry partners alike, who working together have mandated to make "all its research output available to the scientific community with no strings attached" (Consortium, 2016), demonstrating how stakeholder collaboration to enact a particular policy can simply change the way in which results are communicated.

Similarly, those publishers with sufficiently strong journal brands are also in short supply; by the very nature that there can only be a limited number of journals considered to be the best in the field, and as such journal editors are in a position to impose publication requirements. Making the case for publishers as enablers of cultural change is harder however. They are under no obligation to change their procedures. Yet as commented upon by one participant, as demand for more access to data increases among researchers, publishers are likely to adapt their business models to better profit from researcher's new desires and funders' new mandates. Publishers may also wish to avoid negative influences in academic publishing such as journal impact factors as more evidence is provided about their unsuitability as demonstrated by Nature (Callaway, 2016) in response to a paper further highlighting impact factor flaws (Lariviere et al., 2016).

Publishers changing the service they provide reflects a retroactive change as opposed to a proactive one; it will depend on researchers making it clear that they are not satisfied with the utility received from their arrangement with publishers. For such an exchange to be meaningful however, requires stakeholders on both sides of the relationship to be clear about the value that is being generated through the exchange, yet communication between these two stakeholders is unclear. This ambiguity begins with librarians acting

as intermediaries, meaning many researchers are unaware of the total cost of journal subscriptions and it continues with researchers' perceptions of value added. As described by one participant, many researchers see the publisher's role as being one of aggregating and shipping research findings, as well as proposing that for researchers' priorities "reputation is number one" and these are services expertly supplied by academic publishers. If researchers continue to see this exchange as being a fair one, then it is unlikely that pressure will be placed upon publishers to help encourage a shift in scholarly discourse that enables a more open and data-centric approach.

Thus a discussion of stakeholders, as it did so earlier in this thesis, reveals a number of reasons why a cultural shift is so hard to effect in scholarly discourse. The conclusion previously was the proposal of disintermediation – researchers are responsible for many of the tasks of the scholarly knowledge cycle and so this role should be made explicit and self sustaining through the use of the Web. Several reasons were discussed as to why such an outcome of the Web is not possible however, with the allocation of reputation deriving from published articles being a key reason. While this still holds true, the participant's perspectives add further information, indicating how overburdened researcher's find the current system to meet their needs and that funders and publishers have a key role to play in instigating a change, which cannot involve them being left out of the loop. Thus if a cultural shift is going to occur it therefore needs engagement from across the different stakeholders introducing a political angle to encouraging change rather than an approach centred around appealing to researcher's motivations. As noted by Collins (1985), "Science policy ought not to be about maintaining efficient institutions, but about maintaining the sort of cognitive community that will produce the desired scientific products. The second, interrelated, criticism is that [...] measures of scientific output are inappropriate, since they reflect only the internal reward system of science; optimum performance according to such measures is not necessarily optimum for science policy."

6.3.5 Reputation

6.3.5.1 Results

The role of reward and recognition is a theme that has been an undercurrent throughout this thesis and thus one that became a focus of the interviews and focus groups.

Assessment via Publication

As observed via an analysis of the cultural norms of science, recognition is the means by which the success of scientists is measured, and participant's comments revealed how this holds true for academic institutions making hiring or tenure decisions:

"it's true in the UK, it's true in Europe, it's true here, it's true in Asia, that people get, in academia people get, get their jobs and get promoted and get grants because the publications"

"you know, we're pushing younger people to publish more and more, the result is likely that they're chopping things up into smaller and smaller pieces so they get more and more papers out of it"

"I suspect though at... certainly at all the top institutions and it's probably filtering further and further down now, that if you're not publishing in those really best journals you're going to have a hard time [...] getting tenure now already"

"I know people who lost tenure because they were not published. [...] I went and published for the sole reason that I needed tenure."

"reputation is the real thing that researchers want to get out of publishable, out of disseminating their work right cause reputation gets them promotions money and tenure and advancement in their career"

Part of the reason behind why academic institutions use a publication record for such decisions is due to the convenience they offer:

"if you take away impact factors entirely people are still going to want a quick and easy way to do that first assessment of 50 CVs and say here are the 10 who are most promising."

And as observed when examining the researcher's role in scholarly discourse, this may not accurately reflect an individual's contribution to their field:

"I'm at a small school where you have to teach 4 classes, ok [...] minimal resources and this guy created an entire course in medicinal chemistry... and our school was like a teaching school... and he was just point blank, they said "you haven't published anything""

Familiar Metrics

However, such contributions are difficult to measure and participants' comments reveal that an advantage for publications is that they are a familiar and understood mechanism for assessing a researcher or an institution's output.

"we will still need a high level, a core of high level journals [...] we know they have scientific editorial control around right, and they have the peer review process and the peers are high level peers and people can really feel comfortable that that's where the best science is going to happen and it's going to be presented in the best way that they could make it happen"

"I suspect that most [...] people would say no, we need, we want you to publish in Nature, we want you to publish in... in these top two journals because we know what that means and we know that our peers around the world knows what that means."

"they [researchers] have been very careful, they feel that preparing that information that went into those publications and how these sort of fuzzy metrics where you really can't see how they count and how things are weighted and... you know, I don't think they are able to respect those metrics very well and if they don't understand what the authority of the metric is then it's, you know, off their radar already"

This familiarity partly hinges on the nature of citations, which make it clear that one piece of work is acknowledging another and an aspect of research dissemination which may need to be developed further if more open, data-centric approaches are to be adopted:

"I think they [researchers] like to be read and I think, [...] there's all kind of problems, with [...] citation and also downloads as a sort of surrogate for counting how much you're read... but it's not unrelated to that and I think that's why they tend to buy into that a little bit more. I mean they can at least understand what that connection is."

"changing the culture more might be a matter of educating people and saying actually this is not a massive risk and actually its all linked to your name so its cited, can be cited to you and nobody can scoop you on it because you've already published"

"there seems to be this dual stage thing where you use the arXiv for actually getting your research your results out, and then once they're out you then get, try to get your paper in a proper journal so that you've got the nice, nice shiny solid, citation to back it all up"

"But it's always useful to... well its essential to have something traceable to something because if you, if you're doing something and you use someone else's synthetic method then then you have to have something to cite and because people go... how do I cite this if its not a proper journal citation then you want a journal to cite."

Publications not only provide a formal, consistent mechanism for recognising contributions, but they can also act in an informal manner too, further establishing their influence:

"if you don't have formal impact factors you still have the formal "I have heard of this journal, I have not heard of this journal, I have heard of this journal but it has a reputation of being terrible""

But informal approaches to reputation may also be damaging traditional scholarly processes such as peer review:

"I think peer review is basically on reputation. You know, "oh that guy's a famous chemist he won't do anything wrong, his lab is a good lab, we'll trust him". "Never heard of that guy, he's from...' I won't say where, [...] the only way you can peer review it is to rerun the experiment, why are you going to rerun the experiment? It ain't your work."

Publish or Perish

Ultimately, one participant summarised many of the points presented above when discussing the topic of the Disaggregator and its potential as a platform from which to disseminate research outputs:

"it's a matter again of you know, how are you going to be recognised within the community you're working in and you know, if you're working at a chemistry department at a university and you're a brand new professor and you need to work for tenure and promotion, is your department or your colleagues going to... recognise these pieces as valuable and will they reward you for it and will they promote you for it?"

Thus it is once again clear that the act of publishing is hugely influential in academia and as mentioned by one participant it drives many of the researcher's actions:

"I think people by and large try to get their papers into the very best journals they think they can, sometimes they even overreach."

"you don't get into the grade A journal then you'll keep moving down the list until it gets published somewhere"

"a total synthesis lab [...] might spend like 4 or 5 years trying to get from here to some sort of important drug or something you know, that's 20 steps of synthesis to make a synthetic version of a bio product or something like that right, [...] they're not going to survive if they only publish once every 5 years"

"people like publishing in places where the people they would like to publish alongside are"

New Routes to Reputation

But not only does the need to publish put pressure on the researcher's workflow, the workflow also represents a new opportunity for researchers to enhance their reputation and a chance to use the incentive that reputation provides to encourage open and data-centric approaches to research:

"if they [researchers] can demonstrate that they're, you know, that they're doing valid science you know, just through their practise I think that would be really powerful, I think that would speak as much to their reputation as how many times their output got cited, actually I think, because I think there is a big scare... about, you know, are people doing real science or are they just you know... doing dirty science!"

"the better they [researchers] can write and the better they can organise their research, the better their reputation and the better example they're setting"

"I would be looking to do that with a mechanism to be able to obviously... get a reputation or you know at least communicate that my... the way in which I do things and my, you know, thought process behind it is done in a scientific and technical way that is appropriate for what's being done so that in... in the end it would promote the kind of things that I'm doing and it would encourage people to... to follow the work that I publish through that mechanism"

Nevertheless if new reputation mechanisms are to be introduced, as one participant commented these will need to be clearly understood as per the recognition that publishing bestows now:

"if somebody puts out this amount of data or you know, puts online these many experiments, [...] if its got community review of X then that means it should be counted equivalent to a regular peer reviewed journal article or something like that, or at least maybe just general guidance that... that departments should consider other mechanisms of publishing as important to the tenure and promotion process"

6.3.5.2 Discussion

The notion of reputation as a driving force behind many of the decisions and actions taken within scholarly discourse is a familiar one at this stage, forming the basis of the disaggregation. The participants' comments reveal further insight into the role of publishing journal articles however. What first becomes apparent is the extent to which reputation via publication is used. As discussed earlier in this thesis, a good publication record was highly influential for previous REF processes (Corbyn, 2008) used to rank and fund academic institutions in the UK, and as participants' answers have shown, the institutions themselves also use a researcher's publication record as a means for making hiring and tenure decisions. The use of a publication record as a convenient measure in a competitive environment is also cited by Lariviere et al. (2016) stating how indicators such as journal impact factors are "often viewed as a convenient proxy for 'quality' by busy academics perennially faced with sifting large numbers of grant applications or CVs"

Thus it becomes clear that the academic institutions and funders, who so far have been identified as being crucial to instigating any cultural shift in scholarly discourse, need a convenient solution for assessing researcher's contributions and any more multi-faceted approach will likely introduce undesirable complications. And it is not just institutions who may benefit from reputation as a shorthand for assessing research outputs: as commented upon by participants, and backed up by previous surveys such as those by Henderson (2010) and Peters and Ceci (1982), reputation can assist time poor researchers

unable to reproduce results in making value judgements on the work they are building upon. Rightly or wrongly reputation is a vital aspect of many scholarly processes and through various social and economic pressures cannot be avoided – a revelation that illustrates furthermore the cultural norms within which scholarly discourse operates.

Thus developing the Disaggregator around the principle that it will provide researchers with a new platform with which they can highlight even more of their research outputs may ultimately prove to be an erroneous proposal. The intention behind the Disaggregator was to deconstruct the articles and research outputs produced as a means of being published and use the disaggregated components to build a portfolio of research contributions, complementing existing publishing procedures so as to minimise the risk to researchers adopting disaggregation. However the portfolio, in attempting to provide a more complete and thorough representation of a researcher's body of work, competes with conventional research outputs such as articles rather than complements them. The nature of publications in their ability to provide a concise representation of research leads them to be used by various stakeholders makes it very difficult for a new means of disseminating research to gain traction, which will often be less effective. It does not matter that the portfolio of disaggregated components may represent a more complete version of a researcher's body of work if the stakeholders who the researcher wishes to present their work to, do not count or recognise work presented in this way. This in turn highlights why some academic social networking sites have seen success, utilising the network effect such that they gain relevance among a significant proportion of the research community (Van Noorden, 2014a). One participant commented upon the need for "general guidance" and a means to calculate equivalence should new presentation formats be accepted. This would be useful, but such guidelines would demand cross-stakeholder input and engagement, which, like the network effect, would require a significant proportion of the community to push for new standards to be set.

Whilst the study participants' responses may have shown how appealing to reputation as the driving force behind the scholarly knowledge cycle may be mistaken, if not implemented carefully, they have also shown alternative approaches as to how reputation can be leveraged to encourage more open, data-centric dissemination. For example, as highlighted by one participant the need to publish frequently does not suit all researchers' workflows: "a total synthesis lab [...] might spend like 4 or 5 years trying to get [...] some sort of important drug [...] they're not going to survive if they only publish once every 5 years". Therefore being able to demonstrate processes was identified as a "powerful" way to add to one's reputation and another participant commented on the desire to "get a reputation or you know at least communicate that [...] the way in which I do things and my, [...] thought process behind it is done in a scientific and technical way that is appropriate for what's being done", with both responses hinting at new ways in which researchers could build their reputation whilst also helping to reduce concerns

about the validity of science being published. The idea of sharing workflows was proposed by the *myExperiment* project (Goble and Roure, 2007), although workflows were "mainly contributed for philanthropic purposes" (Goble et al., 2010). Developing new ways for researchers to enhance their reputation via a demonstration of good process also has the added benefit of helping to shift cultural norms – if reputable scientists are seen to be conducting science in an open and organised fashion then, as per one participant's comments concerning publishing where those you look up to publish (as well as from the comments about cultural norms and developing practises based upon your supervisors and peers) then this may have a beneficial knock-on effect.

Another advantage of publications is that they have an established means for acknowledging the work of others through citations, a mechanism which, as the participants indicated is familiar and well understood. As indicated by one participant, these are an effective form of acknowledgement; the cultural norms section touched upon researchers' vanity and this arises once again, researchers "like to be read" and thus a citation explicitly states that one researcher has read and been influenced by the works of another. However, in one the focus groups held, the limitations of citations were discussed with participants highlighting their ambiguous nature. Firstly citations frequently fail to indicate what particular aspect of the cited paper is being referred to, an issue that arose when discussing the Disaggregator with a focus group, and which was framed as a problem the Disaggregator could fix – by breaking apart documents it becomes easier to cite a particular facet of the document. Secondly citations cannot portray context and as such there is an inherent risk in converting citations into reputation – the citation could be rubbishing the cited text but this would not show in traditional metrics. Nevertheless what citations do achieve is the provision of a "traceable" link and a recognised way of conveying appropriate metadata, fundamental requirements for a tool such as the Disaggregator and for data sharing in general, as discussed previously in the overview of the conference presentations.

Therefore if the use of reputation is an unavoidable convenience for researchers and institutions who need to make quick decisions, then bringing the workflow in to the reputation making mechanisms may be a step forward in resolving some of the issues of scholarly discourse. As per the discussion of cultural norms, the importance of the researcher's workflow is once again emphasised and may be an area where developments may be made to help researchers fully take advantage of the Web's affordances when disseminating their research.

6.3.6 The Researcher's Workflow

6.3.6.1 Results

Comments about researchers' workflows have emerged across all three of the themes so far, most notably concerning the way in which it embodies many of the cultural norms of scholarly discourse, but also how it presents opportunities to feed in to the drive for reputation. Thus it becomes an important theme of its own, as indicated by how many of the study participant's alluded to it within their responses.

Workflow Efficiency

As has already been observed, the way in which a researcher goes about their work is often influenced by their supervisors and their peers and as commented upon by one participant, becomes established over time:

"in a general sense, you know a scientist, certainly one that's further along in their career, kind of has their own workflow, they have their own expectations of what their students will do and what they then need to do in addition to that and not that they're not willing to change that but that they are willing to change it, it has to, it has to be some benefit and in a sense it also probably has to be at least as easy as what they were doing but hopefully easier because then its delta g is negative, you have a spontaneous process"

The indication given in the above response suggests that for a researcher to adopt a change within their workflow requires the change makes their job simpler to perform. Other ways of appealing to the researcher may also be possible however, helping researchers to manage their workflow so that they may be more efficient and effective at communicating:

"if you're able to give an organised presentation of yourself to your peers whether, you're talking or how your papers read, that's really, you know, its like everyone says, "oh we like to listen to a dynamic speaker", that's what I mean, one that's very clearly organised, maybe you know, I think that comes through in that way, and they're probably organised because they're already pretty systematic about their thinking, so I think it comes out in a lot of different ways"

"the better they can write and the better they can organise their research, the better their reputation and the better example they're setting."

Mid-Workflow Outputs

Through talking about the research workflow, it also becomes apparent that research outputs are produced throughout the research cycle:

"you'll end up producing several different sorts of research outputs, [...] as well as producing papers I produced various bits of software and I produced [...] various datasets and I mean one of the datasets was quite nicely tie-able to a specific journal article that people could very much look at the journal article and not the dataset attached to it"

"you'd write these internal report and this was back in the days where there was this great big filing cabinet of things about that thick bound, bound in blue card that sat in the file somewhere and once in a blue moon someone would take them down and say "so and so is working on that, yes that does look good". I'm guessing maybe they just went in a file and never got seen again"

"several of the different labs had intra-lab outputs, just for their lab business or their presentations they were giving"

"there are really critical points now where things are stamped or approved or whatever you know, the patents do it, you register compounds, there are all these points, these control points in the whole process"

Intra-Lab Communication

Developing on from this, there were also a number of responses that highlighted the nature of intra-lab communications:

Participant 1: "often the research group is existing to make a lot of individual projects happen under the same roof where they can sharing in skills and supervision and, and equipment and resources."
Participant 2: "so there's no real incentive to collaborate on a broader scale."
Participant 1: "there's no real incentive to have these gigantic collaborations"

"you're group has the 3 or 4 systems which they can do all sorts of exciting tricks with and some other group has 3 or 4 slightly different systems and there's a lot of cross fertilisation of ideas"

"Do you think that all this work is properly indexed within the laboratory, so people are doing, doing the same work and the laboratory manager is approving or disapproving all the works that have been done before, all the theses? I don't believe there is, there is a perfect system in place which allows to identify the work that's been done redundant every day"

"a lot of the times, one research group does not see the actual numbers [...], that another research group is doing or especially might be a year or so lag until something is published"

Knowledge Management

As well as comments that concern knowledge management within the lab group environment:

"there was something one time in my PhD where I had an interesting reaction result and my supervisor happened to recognise the NMR spectrum and he's like "oh yeah one my students 5 years ago had this" and was like... found synthesise it by completely different route and I'd happened by chance to come across it a much much simpler synthetic route to this molecule"

"I would have had no idea what it was except for the fact that my supervisor [...] would have had to go and figure it out and do all that stuff, but my supervisor was like no, no I know this one and got out the old PhD thesis and sure enough the same NMR spectrum... but like if it hadn't been the same supervisor hadn't been there or if like the guy had been in another subgroup of our research group or you know I might not have known so if there was some way to yeah... look at other stuff on a network drive or find similar things and... that could be really useful. And our lab group was only about 25 people so it wasn't like a huge group"

Interviewer: "So if you did want to collaborate with your colleagues and share particular files, what, where would you upload them to, just a network drive?"
Participant 1: "Yeah we just had network drives... yeah would just be like it's here on the network drive for... "
Participant 2: "You ought to hand it over on a floppy or a USB stick or mail it to them"

"I think that this one lab was definitely buying into that idea, they saw that by breaking down their stuff and, into a package that just covered what was relevant for that piece of data, that they would help them with their planning and in a strategic way"

Workflows and Research Outputs

One final aspect of the comments provided on the nature of research workflows was the manner in which they reflect the final outcome of the research process, an aspect of workflows which has already been touched upon

"I think our workflows today... are dramatically linked to the ultimate result which is we got to write up a paper"

With one response concerning workflows and their relationship with the final product given in answer to a question on disaggregation:

"but again it's a cultural change, if we're living in a disaggregated model then [...] your ultimate product is different to the paper as it is today and so your workflow is going to reflect that"

Thus the responses provided by study participants highlight the importance of the research workflow in the overall system that forms the scholarly knowledge exchange. The researcher's workflow naturally has a strong connection to the final research outcomes that are disseminated – the processes that form the workflow determine what

gets produced, but conversely the final product shapes the workflow itself. Participants' comments reveal however that the workflow is an area where it may be possible to revolutionise the overall processes of scholarly discourse and is confronted by issues of intra-lab communication and knowledge management.

6.3.6.2 Discussion

It is perhaps not surprising that research workflows have emerged as a theme; an emphasis on workflows can be explained due to the rigorous and well defined nature of the scientific method that demands certain structures are in place and a participant's comments on the nature of the workflow where outputs are "stamped or approved" only serves to reinforce this. Thus if the cultural norms around the scholarly knowledge cycle are to be changed, then it is adjustments to the workflow that are required – the holistic approach adopted earlier in this thesis focused on the exchanges between the stakeholders which together form a large part of this workflow, however, what an analysis of the stakeholder interactions fails to recognise is the work that goes on within the lab itself, ultimately resulting in a position where the importance of the researcher's day-to-day workflow was neglected, a result which can be observed in the approach to disaggregation that was proposed that focuses on research outputs and not research processes.

However, moulding the workflow to help the processes of scholarly discourse better take advantage of the Web presents a challenging problem. As noted by one of the participants, "our workflows today... are dramatically linked to the ultimate result" and with publications being the primary source of reputation this is unlikely to change. Thus any new approaches to adapt the workflow will need to accommodate existing workflows rather than expecting them to change – an approach taken by the Disaggregator proposal, but applied to the very end of the researcher's workflow rather than permeating it. The discussion on stakeholder roles has already observed how "scientists really just want to do science, they really don't like writing it up", and as such suggested changes should be aimed at making a researcher's day-to-day work life easier, as opposed to asking them to reconsider their means and methods of dissemination.

Earlier discussions on the evaluation of disintermediation allude to workflows and knowledge management and have both helped to explain why the Web's potential in this area has not been fully maximised, and informed the disaggregation approach to encouraging more open and flexible scholarly discourse. As stated by Cabrera and Cabrera (2002), knowledge management can help organisations increase market effectiveness and improve product innovation and it is clear that academic institutions also endeavour to achieve this through the organisation of lab groups. Yet the participants' comments highlight where these efforts are found to be lacking. The sharing of data among colleagues is often difficult or disorganised, backing up the findings of the *Support the Changing Research Practices of Chemists* report (Long and Schonfeld, 2013), and there are many

outputs that are produced throughout the workflow which researchers share manually with one another. This in turn creates inefficiencies in the research process, possibly leading to duplicate work being conducted or connections between separate items not being identified.

Therefore adapting the workflow in a way that appeals to researchers by helping them get on with their science, involves devising knowledge management solutions and tools for providing oversight. As remarked by one participant, laboratories need systems that index the work being produced, helping laboratory managers identify what work is going on and what is redundant. This comment aligns with the findings of the *Supporting the Changing Research Practices of Chemists* report which found that "chemists' research work is conducted collaboratively at the level of the lab group" and that there is room for innovation with respect to data management practises: "Data are often stored in impermanent formats that are at risk of technological obsolescence" and "Most universities are still in the early stages of developing support services to help scholars deal with data management" (Long and Schonfeld, 2013). Disaggregation as proposed ultimately fails to take advantage of this innovation space.

6.3.7 Disaggregation

6.3.7.1 Results

The main purpose of the interviews and focus groups was to evaluate the idea of the disaggregation and the prototype Disaggregator tool that was demonstrated to the participants where possible. The results and respective discussions shall be separated in to two sections, firstly a theme concerning the concept of disaggregation itself, followed by another for its implementation.

Disaggregation Proposal

What first becomes clear is that many of the participants' comments agree with some of the ideas that lie behind the disaggregation:

"I think it is fair to think of a journal article as an aggregate of a lot of different things"

"I... really like the idea of... dividing the paper into chunks that can be re-utilised and re-purposed"

"I could well see people being very tempted to put preliminary stuff or slide stuff or whatever in to various disaggregatory things"

"there is no reason therefore why one could not... disaggregate say the method and the exacted experiment"

With some participants alluding to previous projects they had worked on that were trying to achieve similar aims:

"I have actually built a program similar to this [...] but it wasn't for dissertations, we were working with published data that was in tabular format – supposedly in tabular format – because these were publications that had been published, some of them in the 60s so they were in a table [...] scanned and sometimes lopsided and so things would be on different lines and stuff and so [...] we wrote a script that would take... – called them rules, which would define a regex to actually go and search for text in a particular area, pull it out and then we would use some semantics to actually say where does this text go and how does it link to all the other data."

"so what we did, we allowed to feed, not PDF data, but we fed Word documents I believe, and some raw text or HTML, and we pass it through OSCAR and OSCAR marked up what it thought our chemical terms and then after that immediate step was the user interface, very primitive user interface but still which allowed you to curate those terms and say yes, yes this is right, this is wrong... "

Advantages of Disaggregation

Some of the comments given, indicated how there were advantages to be gained from adopting disaggregation:

"getting data, these chunks, into each other's hands makes for... reuse, checking that data is correct, finding errors, finding fraud"

"One is good data which you think is good enough to use in a publication, then there's good data that either isn't good enough to use in a publication or is... – doesn't necessarily support your hypothesis, so its data that you're not going to use, and then there's data which is complete crap so... you know we don't want, we don't want the data that's useless but we do want the data that's scientifically... appropriate meaning it was actually gathered with the experiment set up correctly, it just was never published because it didn't provide what the researcher was looking for, that's still useful information to, so putting that up as well would be appropriate"

"I do believe that disaggregation, disaggregated pieces can be reassembled with some new insight by being packaged in this new way. And I do think there is knowledge to be gained in doing that."

"the institution is there to enable knowledge gain and I think this, disaggregation approach enhances knowledge acquisition and so the institution benefits"

"more data being available is always a good thing because you can obviously... if you have more data available to you, for instance when I do informatics work I can then make a more educated guess on what an end result may be and therefore can possibly produce better data myself"

"It's quite easy as a cheminformatician to see the value of doing this with new documents going forward because that's easy and I can just adopt it into my workflow quite readily"

Knowledge Management Potential

Many of the comments tapped in to the as yet unexplored potential that disaggregation may offer. Some of the comments referred to its potential to play a role in knowledge management, a potential route to changing cultural norms as discussed earlier:

"that use case as a discovery tool for your own body of work, is it a way of allowing more granular commenting on the work that's being done"

"how many times do we have to write an introduction to a paper that's just regurgitating essentially what was done in our own papers before in other papers, you know, we can just grab that intro from paper X and use it for paper Y that you're now writing"

"I believe there what your system can achieve is to help index content and then it is up to people or PIs how to use that indexing whether for control or dissemination or for anything else."

"And I think that's another place where you're focusing on that smaller package that then could be used in a lot of different ways – to assemble for publishing or to assemble a knowledge base that you can do inferencing or... whatever you want to do with it right"

Other comments allude to how disaggregation can help a wide range of stakeholders:

"researchers they just have all of their... all of their reaction and... purification in one like text block, and that's good enough for them and its just one paragraph for the most part because its pretty standard methods and they can read it, understand it and they don't feel like they need to break it down into a form, in fact actually that would be longer, irritating to read across different fields so from their perspective they're happiest with the paragraph. My environmental safety colleagues on the other hand, they want to know every task, that was there, if you did something, did you stir this, did you pour, did you... you know, so for them they would just love to break out that level, practically almost every word or the unit with the... you know... measurements with units and temperature conditions, and all that stuff, they want to break it out like that. They don't care actually about your schema so much but they care about your... so you can see how they have different priorities... "

"I can see the way around where you start off with the article and then you spin off various bits you think might be useful for particular audiences"

New Research Potential

Some comments indicated disaggregation may provide a new way to analyse research outputs:

"The components, so you have the component of chemical, the component of person, the component of equation. What you would probably find is that after you've done 20 or 30 dissertations, you've got a list of 40 or 50 components and now every new dissertation, is these ones, or these ones, so in a sense there will be a... you could have a bag of components that eventually could capture all of the dissertations... "

"So then you get 50 different, then you get 50 dissertations and you can start mapping between them."

"much of cheminformatics is computational, but the fourth paradigm of science is e-data, data exploration, okay? And until you decouple the data from the science of the first paradigm, [...] one of my hypotheses is what I think your research is [...] which is that the fourth paradigm is going to change the first paradigm and that by changing how we represent our data will influence how we run our experiments. [...] So being able to deaggregate the data from the script of the experiment allows explorations in new ways, it allows correlations, you may suddenly discover that the barometric pressure was influencing the experiments, because people in different labs had the same pattern that correlated to the atmospheric pressure, [...] to the humidity... and that never would have come up if you weren't able to map one person's lab with another person's lab, with another person's lab with the seemingly unrelated thing."

In the unrecorded focus group session, one participant also mentioned how disaggregation may enable new forms of post-publication peer review, by allowing people to comment or criticise a particular part of the paper.

Selling Disaggregation

With participants identifying potential benefits of the disaggregation approach, many of the comments focused on how it may be possible to sell the idea of disaggregation to researchers. The hypothesis that informed disaggregation initially was to appeal to researchers' desire to better their academic reputation through the ability to publish more of their research outcomes. Participants' comments indicated different approaches may be appropriate however:

"if you're, if you're really going to be able to sell this idea that you should be able to, and should, publish... as much of your work as you feel is of appropriate quality through some kind of website in as quickly as possibly as the timeframe you were comfortable with then you've got to provide them with a tool to do that and I think the best way to sell it to them is to sell it as, if you use this you will improve the quality of the science that you do."

"what would be necessary aspect to persuade most academics and certainly non chem-informatician academics to do this is some use case where by you demonstrate real, practical value in time saved or money saved or increase in research outputs that arise from such a system."

"I don't know about increasing the appeal of it but it might enable you to have a test case by where you can demonstrate – or not – . . . the utility of having disaggregated documentation."

Dissemination via Disaggregation

The unrecorded focus group session also indicated a number of problems with using disaggregation as a means for dissemination. When discussing how disaggregated components could be used the focus group participants first acknowledged that for disaggregation to be useful in this context, opinions need change that allow researchers to gain credit from sharing pieces of a manuscript in this way, a comment that reflects earlier thoughts on how reputation works in scholarly discourse. The focus group participants added to the complexity of this issue by then raising the question as to how one might compare the credit awarded to a researcher for two different types of research outputs, making it clear that measuring the reputation of researchers, who frequently have so many strands to their work, is a challenging problem. However, disaggregation was acknowledged to possibly be beneficial for dissemination of research, when one focus group participant commented on how it would allow a researcher to cite a particular aspect of the work in question rather than the entire artefact, introducing an element of flexibility in an otherwise coarse mechanism.

Terminology of Disaggregation

Finally, one question in the interview and focus group schedule asked about the term disaggregation itself and whether or not it was a useful or suitable name for the process. Whilst there was some agreement that is was fair to identify traditional scholarly outputs as aggregate collections of other research outputs, there was nevertheless some aversion to the name disaggregation:

"it would be maybe more compelling if there was a name which described what it was, less about what the process was and more about what the. . . the end result or potential benefits were"

"it's a word about pulling things apart which maybe doesn't have the greatest positive connotations even if it is going to have positive effects in the long term"

"I think. . . I think. . . I think the word is probably technically appropriate but whether for . . . for scientists it . . . it really works as a word that allows them to really appreciate and understand what you're talking about I'm not sure about that. And I'm also not sure about quote, unquote, just taking a research paper and taking it. . . – putting it into pieces. . . is necessarily equivalent because. . . as the phrase goes, . . . the sum is greater than the parts"

"I think if you could find a different word for disaggregation that would be better . . . I'm not sure what the right word is though."

Thus the concept of disaggregation presented to users garnered a generally positive reaction, and whilst participants elaborated on its potential they were also able to identify a number of limitations with disaggregation in the manner in which it has been proposed.

6.3.7.2 Discussion

Overall the study participants responded positively to the idea of disaggregation, agreeing with the notion that conventional, scholarly publications are aggregate collections of smaller research outputs. This is perhaps unsurprising; the previous theme, discussing the way in which the workflow is shaped by the final product of the published article, already alludes to this. Typical research cycles of conceptualising, analysis, initial results and formalising (Research Information Network, 2009), and the scientific method itself, tend to be quite granular in nature, and so the notion of aggregation and disaggregation are a natural fit. The notion that other aspects of a scientist's work can be disseminated beyond the results has been explored before, for example the notion that "workflows are scientific assets" that should be "pooled to be drawn upon to be reused, re-purposed and recycled by others" (Goble and Roure, 2007).

The importance of the article was something that was still stressed by participants on the subject of disaggregation however – "the sum is greater than the parts" – and so the use of disaggregation as a means for disseminating research was questioned. Disaggregation was designed as a new outlet for researchers to disseminate their findings, in a more open and thorough fashion, helping researchers establish a portfolio of their work that truly represented their experiences and abilities in the field, with the incentive to take part laying with this opportunity to enhance one's reputation through evidencing with research outputs. Earlier discussion of reputation has shown this to be a fallacy however, and comments on disaggregation only served to reinforce this – the disaggregated components would need to be recognised and guidelines in place for researchers to understand how the impact on a researcher's reputation should be interpreted. Both a social framework as well as technological tools need to be developed for disaggregation to successfully operate as a means for dissemination.

Nevertheless, benefits of disaggregation were identified by the study participants, who once again highlighted the importance of the researcher's workflow. The benefits can be broadly categorised as those that enhance existing scholarly processes and those that offer brand new potential to conducting science. Comments that fit into the first category highlighted how a greater level of granularity allows researchers to be more precise and encourages a greater degree of communication. As described by one participant, data "chunks" assist "reuse, checking data is correct, finding errors, [and] finding fraud". It presents researchers with an opportunity to be more specific when reviewing or citing the work of others. However whilst this may be a possible outcome of a disaggregated approach it would seem naive to assume that it would be welcomed by the research

community. Comparing the possibilities offered to facets of the research cycle like peer reviewing with the mechanisms of reputation as just discussed, indicates that methods of citing and reviewing are based around negotiating a balance between what is sufficient for the community and what is convenient for individuals. More thorough and nuanced practises may be made possible, but will they be recognised by time poor academics and resource stretched institutions? Participants' responses would indicate that they would not.

The comments that fall into the second category of offering new approaches to conducting science, suggest that researchers or primary investigators may benefit from a more data driven approach. With data organised in a structured fashion, it presents either researchers or project investigators with new means for analysis, acting as either " a discovery tool for your own body of work" or a mechanism through which it may be possible to identify patterns across a broad range of documents. These advantages of disaggregation may also feed into advocacy efforts; participants mentioned the need to be able to demonstrate that results may be formulated via disaggregation that would otherwise be impossible in order to sell the benefits of it, an approach to advocacy that is recommended by Burns et al. suggesting the need to provide "evidence (for example in the form of case studies) of how new developments in technology or resources could help the researcher [...] in their work". Taking advantage of a data driven approach however requires skills and resources many stakeholders may not be able to leverage, as highlighted by one participant's comment on the "fourth paradigm" referring to the works of Hey et al. (2009). Therefore disaggregation may not offer a means for greater dissemination of research outputs, but it may present a new approach to providing data for e-science techniques.

The final aspect of discussion focused on how the idea of disaggregation should be presented to researchers so as to encourage new forms of dissemination, with many participants recognising the need for disaggregation to be shown to enhance researchers' scientific practice, enabling them to be more efficient or reach results unobtainable via traditional methods. Through discussions about how to present the idea of disaggregation, along with the discussions that have accompanied previous themes, the answer to the research question of "What motivates researchers to engage with new approaches to scholarly discourse?" becomes increasingly clear: improving a researcher's ability to do science and secure good publications is key; appealing directly to reputation is flawed due to the extent to which publishers already furnish researchers with this need. Allowing researchers to be able to better use their own data is the space within which innovation can take place however, and assist researchers in publishing more frequently. Through doing so the benefit to researchers becomes immediately obvious.

Therefore from discussions around disaggregation as a concept, participants are in favour of disaggregation – drawing the conclusion of disaggregation from a study of the theories and literature that were used to critique the idea of disintermediation, resulted in a

viable solution. However what also begins to emerge is that the chosen implementation of disaggregation may be mistaken.

6.3.8 Disaggregation Implementation

6.3.8.1 Results

Disaggregation Proposal Issues

Disaggregation was proposed and presented to the study participants as a means for extracting information from the articles and other scholarly artefacts that researchers produce as part of their conventional dissemination practises. This lead to a number of concerns being raised:

"I mean if you wanted to disaggregate stuff it might have to be in place of a journal otherwise you feel like you're doing a bunch of extra work"

"disaggregation of existing information would probably not fit well with the existing academic workflow"

"It's difficult to extract text from PDF"

"the process of disaggregation shouldn't be built into researcher's workflow because they shouldn't be burying their data into PDFs, they should be dealing with it in a different way in the first place"

The concerns presented above match the manner in which the participants viewed disaggregation as discussed in the previous section. Disaggregation should not work as a tool that exists after the conventional research output has been produced, but as a tool that assists the researcher earlier on in their workflow. Nevertheless some elements of the Disaggregator's prototype implementation were seen positively:

Interviewer: "Yeah, so we can either save them [text-mined components] or discard depending on whether they're interesting or not."
Participant: "That's quite a nice approach actually. I quite like that."

But ultimately the vision of disaggregation presented to participants was seen as problematic. Some comments focused on the practical implementation of such a service:

"How do you get people to know where it is and how do you get to it? And there's also the cost factor, who's going to pay for this?"

"There is some kind of problem in that, sometimes, you know, how do you indicate where it is cause sometimes these publications are moved or else, they're only going to be there for a certain amount of time, so that presents somewhat of a problem"

"unless you said, some kind of web page or something where everyone who's working on a certain problem, deposits all their information there, whereas another group working on something completely different goes some place else, but you know to find what's out there would be the overriding problem."

Disaggregation and the Role of the Researcher

Other comments reflected on stakeholder roles, reiterating the researcher's desire to do science:

"I'm not sure this is up to researchers to do, digest the content of PDFs and and then to create something"

"ultimately I think the only intervention is going to have to be when the researcher hits the button and says ok that's for public release now and... and the message goes out to the repository and ok, you can make this publicly available

"if it's getting chunked and metadata is being assigned and ... data is being filtered and processed and stored ... as the authoring is taking place then... then that gets delivered to your system ready preassembled and then the intervention you need to make can be on a higher level."

Interviewer: "That's what we want to do of course, we want to spend just as much time in the lab doing our chemistry."
Participant: "Correct! That's right, no, no, no, that's absolutely correct and again that's why I've been arguing for whatever tools we need to build, it's got to be invisible to the scientist"

"I really very firmly believe in is that as much of this process has to be made invisible to the scientists"

Disaggregation in the Researcher's Workflow

The final comment above, reiterating that the researcher's role should be minimised, one of simply deciding whether or not an item should be released or not, provides some insight in to an alternative approach for implementing disaggregation, one that involves integrating it with the researcher's workflow:

"the authoring tool that generated that first article that gets deposited to you, could have things built into it to generate metadata to start chunking it up before it gets delivered"

"creation of the little components and where they get stored and how they get distributed that can all happen behind the scenes without much intervention

"if we move in the direction that you're advocating it's going to change the workflow and that's where the tools need to come into place to, enable that workflow to take place so

that you manage where all these pieces are and then the managing the distribution to the world

An approach which matches some of the comments provided on the nature of researcher workflows:

"it's already kind of disaggregated in that they have very clear sections and then we could build a template around it and then a template in the back end is each of those as a piece in it, it's got pretty granular data structure and metadata around it so you can mix and match and that was part of the whole idea, you could, you can put it in and it facilitates input and then you can mix and match the bits as you want depending on what you want the output to say"

"even the analogue labs, are already doing that. I've looked through a lot of print notebooks and they're pretty consistent about you know, within the one lab, depending on what their research methodology is, they're pretty consistent about their notebooks in terms of format, so it doesn't, it's not hard for me to imagine how you would be able to pull it apart pretty consistently"

"if the mechanisms were in place to allow it then yes that would be the place, I think it would, . . . be easiest to implement. If your lab culture was such that you were recording all of this information anyway. . . then that's ideal because then you're only doing everything once which minimises the impact on the workflow"

Through doing so an opportunity is presented to help enable some of the workflow knowledge management changes suggested previously:

"that's actually the PI's way of managing her students, is this is the kind of information I want you to be collecting, this is where I want you to put it . . . and if we all do it like this then its much easier for us to communicate and I know at least you didn't miss something and stuff like that. So in a sense that's already disaggregated thinking, so they're already doing it and that's the really important thing."

"you can publish all that data and then make it reusable to the end user within the tools that they're using for their own research"

Implementing Disaggregation

With participants focusing on the challenging nature of extracting text from PDFs and the problematic role of the researcher in post-publication disaggregation, some comments alluded to a different implementation approach:

"would favour arguably just a disaggregated approach from the get go, whereby you just input your data, marked up or in an otherwise disaggregated way which designates that separation."

“ in the sense of a research group it could be that be defined by the faculty member at the top or a postdoc that’s there or it could be left open to the individual students, how they want to set it up”

“But over time I think what you could do is you say ok, if an individual research group said “hey, we really need a module to do this”, and the system where set up in a way in which they could nominally build any module they wanted then they could potentially publish that module and it be available for others to download and use in their particular version of the same system.”

“So that it’s very easy for a researcher to be able to say “ok, let me, I think I want to publish this data online”, ok, and then you could have a special screen that says ok what data do you want to publish and what annotation do you want to write with this, how do you want to describe it, you know, there’s checkmarks that indicate that it was verified and that it was authentic and all that kind of stuff and then you just hit a button and publish it.

Incorporating the disaggregation process earlier on in the workflow may also introduce new benefits:

“if we’re talking about researchers, this is probably the place for your system to catch errors in the early stages.”

“being able to do some of those substructure searches right from in one experiment, ok we did this in this experiment, we would like to play with this and do some, you know, move some bonds around right in there and populate a whole new template out that would respond, you know, respond to what they did based on, you know, that would be ideal for them, it had some experimental planning support based on new variables that they put in, whether the variables are different substructure or different conditions and you know, they could be drawing at least on their own experimental knowledge base in the system to start with, or if we could pull the literature in that would be great”

Disaggregation and ELNs

To implement the disaggregation process in to the workflow, an ELN solution will be required, with one participant detailing the complexities involved with running an ELN, first from the perspective of industry:

“industry has had something equivalent to... well has had ELNs for quite a long time [...] the reason why industry has ELNs now is because it’s a huge market... obviously there’s a lot of software vendors that are able to build software for that particular application because there’s huge market of customers that will pay for it and those customers have a lot of money, you know, industrial ELNs typically run $100,000 at least and that’s maybe initial start up price and then of course there’s a maintenance fee every year and stuff like that dependent on how many users. So there’s a lot of... it’s a big

business, it's a lot of money to get into it and it also has a lot of capabilities that rarely an academic scientist doesn't need"

And then comparing this to the requirements of academia:

"there's a whole bunch of features that are really geared very much towards the industrial section but that don't work properly for the academic setting and that's why in most cases academics don't buy existing ELNs they... they're just too much money"

6.3.8.2 Discussion

Whilst it is clear that the implementation of disaggregation in the form of the Disaggregator prototype is misguided, some elements of its design were well received. The semi-automatic nature of the text-mining, with users able to accept or reject suggestions was identified as a useful feature given the technical limitations of extracting data from documents, and the ability to disaggregate components of the research process such as individuals involved as well as data from documents was also an aspect the participants found useful. One participant also reflected on the need for such a system to be modular, allowing different research groups to implement their own approaches to disaggregation, something which the Disaggregator prototype achieves through its components system. Similarly many of the talks at the ACS conference reflected on the use of semantic web technologies, something which the Disaggregator takes advantage of to provide meaning to the relationships that form the components. However, these elements all need to be expanded upon if disaggregation as a concept is to be useful to researchers, especially given disaggregation's new context as a means for assisting a researcher in the lab as opposed to a means for disseminating research outputs.

It is this new context that the flaws of the Disaggregator prototype implementation are most easily demonstrated. The Disaggregator works in a retrospective fashion and asks researchers to extract information from publications under the assumption that it is the researchers who stand the most to gain from doing so, both from enhancing their reputation and from having access to more data. This implementation exposes the stakeholder roles further still however: it is the publisher's role to analyse documents and the researcher's to conduct science, reiterating the publisher's role as the sole provider of reputation in academia and demonstrating that they are in a position to add value too. The retrospective approach also limits the data that may be extracted; disaggregation of a thesis may provide a wealth of data, but disaggregation of articles will be limited and perhaps most importantly of all, disaggregation of a PDF is a difficult process, hindered by the complications that arise when trying to text-mine the PDF format.

Discussion of the implementation also exposed a number of practicalities concerning the disaggregation process; namely how would such a system be funded and how would the data hosted be discoverable. This was an aspect of the Disaggregator that was left quite

vague, with disaggregation initially being explored as an idea for opening up scholarly discourse, with the intention placed on examining the validity of the idea, as opposed to the practical implementation. The implication for disaggregation is that it is a service for researchers to take advantage of, and as such it may be the role of institutions to provide the service to researchers, with the component interface designed to allow different items to be disaggregated dependent on the given context. However as with any form of research dissemination, results need to be citable. As noted by Smith et al. (2003), "it is considered crucial that citations to archived material, whether found in printed articles or online, remain valid for long periods", a practice which is unlikely to be achieved if instances of the Disaggregator are left to be maintained by individual research groups.

Therefore what might a more agreeable implementation of disaggregation look like? One participant's response may be key in beginning to frame this: "if it's getting chunked and metadata is being assigned and ... data is being filtered and processed and stored ... as the authoring is taking place then... then that gets delivered to your system ready pre-assembled and then the intervention you need to make can be on a higher level". In the context of a chemistry lab, an electronic lab notebook (ELN) is therefore the likely candidate for assuming the role of a disaggregator. Another participant commented upon the role of ELNs in academia however, describing how they are ideal in an industrial context, where organisations have the resources to develop them and where ELNs assist chemists in conforming to strict regulations, whereas in academia, ELNs are deemed too expensive to maintain and posses unnecessary features – a comment verified by Butler (2005) and Schraefel et al. (2004) writing on the topic of ELNs in academia. Butler (2005) describes how "industry's massive investment in e-notebooks has helped the technologies to mature", whereas "academia has been slower to respond", citing the fact that "most software needs to be customized to the working habits and research needs of individual groups which means an investment of time and energy" as an issue for ELN uptake in academia: "Academics simply don't have the funds for software that needs customizing to their situation, and there are almost as many different situations as there are labs". The introduction to ELNs described by Butler (2005) also describes a number of benefits of ELN adoption that mirror comments made by participants in previous themes such as the ability to make it easier to find and manage one's own work, but that also "scientists won't adopt e-notebooks, if it means taking on additional computer tasks, unless it also makes their job markedly easier".

Implementing disaggregation as part of the workflow therefore answers many of the participants' concerns about disaggregation – one no longer needs to worry about extracting data from PDFs and it can appeal to researchers' desire to make their day-to-day work easier and enhance knowledge management. But this suggested implementation makes it unclear how the original aim of encouraging openness and modern approaches to disseminating in scholarly discourse is achieved. A mechanism will need to be in place to

ensure that the disaggregated data stored in ELNs can be published to a platform hosted by stakeholders such as publishers or funders. Whilst deploying disaggregation within an ELN context may help a researcher conduct better science and publish more frequently, it would also help other researchers and thus ensuring the data is disseminated is a key consideration for a new approach to disaggregation.

6.3.9 Social or Technical Determinism

6.3.9.1 Results

Throughout the evaluation of disaggregation, another theme began to emerge, concerning the interplay between the role of technology and the role of individuals and social structures in the development of scholarly discourse.

Role of Technology

Some of the participants' responses indicated that advances in technology are key to developing a more open scholarly knowledge exchange:

"the instrument should create a dataset that is ready for redistribution automatically and in the ideal world it would just pipe into whatever archive there is for that kind of data and just shoot it out automatically"

"I'm still very interested in, is a really, truly... integrated system that can really allow the scientist to do the science they want to do but manage the data capture and dissemination. I mean that's that's really what we need. You know scientists really just want to do science, they really don't like writing it up."

One participant gave an example of how changes in technology can enable new services to take place, with the example given being a key component of what makes the Disaggregator prototype work:

"one of the real powers of Microsoft Word for example is that in this last generation it was creating XML without us knowing its creating XML ok. That's what I'm talking about behind the scenes."

Amongst other examples of how technology has influenced scholarly discourse:

"So there's been a lot of great development over the last 15 years creating technologies to support a lot of interesting ideas for new publication. Data sharing, open access, ... preprint servers, institutional repositories. Lots and lots of it and a lot of technologies to enable that today."

Limitations of Technology

However there were also comments that indicated the technology that is really required to shift dissemination practises is not yet sufficiently advanced:

"the instrument spit out the data and send it to a repository instantaneously that's you know, that's decades down the road before we're ready to do that, but that's kind of the ideal goal"

And if technology were to be influential in a researcher's approach to disseminating their research outputs it would have to work seamlessly:

"that's what academics would want, academics want something that's free in most cases and that works mostly."

"From the rest of information, the chemical names what you are already doing, ... it would be useful probably, but the extraction [of] the record should be better, so while you're still offering the manual correction phase in your work, its ..., people are expecting that it works like magic so when something you are using, like the OSCAR version is not proper, doesn't work, they'll just abandon your system."

"When you want to sell some new product to your potential customers, what do you do? First you [indistinguishable] their pains, that is what you're doing right now. And then you suggest some solution which at least on the surface can ease that pain, so if you show a system which contributes more problems than solutions, then it unlikely will grasp the attention, but if you can show that something you are doing can save their time, or may save their reputations that would really be appreciated.

Roles of Stakeholders

Conversely, as per the discussions of the various themes thus far, many comments refer to the responsibility of various stakeholders to encourage change:

"my first step would be to now go and ... evangelise journal editors and work exclusively with editors and editorial boards and say look, we need, you need to recognise the data is the coin of the realm, that it's how science gets done and we now have tools to enable sharing of data that we didn't have before and you need to make it mandatory to archive your data if you want to publish and I think that's the first step for moving in this direction"

"I think always the big challenges is not the technology, the big challenge is always the community and I mean all the community, stakeholders on that research life cycle"

Technological and Social Alignment

Thus ultimately there appears to be a conflict between the roles of individuals and the facilitating role of technology, a conflict which needs to be overcome:

"we want more metadata to go with the data that we generate no matter whether that means it's going to be published electronically or via a website or if it's going to be published in a more traditional mechanism but . . . but the scientists don't want to put that extra metadata, [. . .], they're not focused on data, on the annotation of data. Now yeah sure they do this and they don't realise they're doing it when they do their lab notebook"

"so I would be all over that [disintermediated approach], especially if there were two things: 1) a standard set of guidelines and expectations relative to doing things in that way, kind of like we have already for doing things in chemistry the traditional way, [. . .] and secondly . . . that there were tools to be able to capture the research that I was doing in a way that maybe automated the process of collecting the metadata and adding the provenance and the authenticity and you know. . . checking for . . . checking that data was not corrupted and all that kind of stuff"

An example where social and technological forces have worked in conjunction can be seen in other disciplines, as recounted by one participant on the history of the arXiv repository and its use in physics:

"where it [arXiv] started from was some physicists, I believe it was at Harvard, would act as a clearing house of preprints. Physical preprints, you would, kind of send them there and . . . and they would kind of collate and copy them and send them out. [. . .] it was the high energy physics community so there were only a handful of people across the US and probably a few in Europe that were involved in this as well. So, maybe it was 20-30 groups, so you were physically mailing out to, you know, a small number of . . . of collections of preprints. But that. . . that was happening from the [19]40s and so people were used to . . . shooting off their preprints to their friends through this community and then submitting it to the journals. So, [. . .] the best physics journals did not come in and say, you cannot . . . you can't disseminate through this mailing system prior to publication . . . if you do that we're going to consider that prepublication already. That was just accepted as part of how their culture is, so when, the Internet came around and this arXiv distribution method became so much simpler, well other physics community near the high energy community kind of bought in to that culture as well."

With chemistry in comparison having a different history:

"we have a really long heritage of a commercial aspect"

"depositing your stuff in a preprint server, for chemists, has already been shown not to work. People have tried preprint servers and they have failed"

6.3.9.2 Discussion

The both complementary and contrasting roles of technology and society are a theme not just emergent in this analysis of scholarly discourse, but also in general for Web Science, as introduced via the "concept of co-constitution" in Halford et al. (2010)'s "Manifesto for Web Science": "Technology shapes society. Society shapes technology". The manifesto is particularly pertinent in this instance for it goes on to illustrate this concept in the scientific domain, describing the work of Collins (1985) and Latour (1984) on the influence of "social processes shaping scientific practice in the laboratory".

What became apparent at the ACS conference was that there is a whole suite of technologies available to researchers, publishers, librarians and the many other stakeholders to facilitate the open exchange of research outputs in a useful and reusable manner. The Disaggregator prototype is just one beginning in demonstrating this and many others have done work in this area taking advantage of the semantic web and ontologies to formally describe research outcomes; standards and identifiers such as ORCIDs and InChIs; and formats for conveying detailed information succinctly in a machine readable way such as JSON-LD. However whilst the technology is present it is not being properly utilised other than in a few, niche use cases. The technologies that are used frequently require significant investment from those wishing to benefit from them and for many researchers this will ultimately prove too distracting from the need to conduct science and publish their results.

Consequently the social pressure – the demand from researchers or other stakeholders – is also not sufficient to push technological developments further. What benefits there are to be gained from adopting new technologies and methods are not seen to provide sufficient economic utility; as described by one study participant, "people are expecting that it works like magic". Analysis of the stakeholder roles and functions, show that whilst the scholarly knowledge cycle in its current form has flaws, it does perform the intended consequences as far as many of the involved stakeholders are concerned, and as such there is insufficient drive to develop the technologies to solve the problems. This is an issue compounded further still by the unique socio-economic structuring of academia, where no one organisation stands to benefit from developing technologies which will inevitably help all, a public good market failure. Many of the participants commented that if an approach such as disaggregation were to be adopted by the community at large, then a powerful use case would be required, demonstrating the ability to do science that would previously be impossible, thus in turn creating the social force that will drive the technology to mainstream sensibilities.

As a result the shaping forces of technology and society reach an impasse. There is not enough influence or sufficient resources from stakeholders to develop the technology further, nor is the technology sufficiently advanced for researchers to observe the benefits of its use. Consequently progress in the use of the Web for disseminating research

outputs moves along at a slow pace and only affects the small numbers of individuals who actively engage with it. This will likely only be overcome with the arrival of a significant disruption on either side. A social force to disrupt the current mechanisms may require an overwhelming influence exerted from one particular set of stakeholders. Sci-Hub, "the world's largest pirate site for academic papers" (Bohannon, 2016), may be an example of such an influence, but its illegal nature makes it difficult to predict the final outcome of such a disruptive influence. For an innovative technology to arrive and change the way in which researchers disseminate their work, the technology will have to meet academics' desires of being "something that's free in most cases and that works mostly", an improbable investment for any technology company. That is not to say that the right combinations of social and technological determinism cannot form; as noted by one participant, the physics community makes use of a very successful preprints repository in the form of arXiv, which developed as a result of social changes, which laid the ground work for future technological ones. For chemistry however the situation is quite different, the community never conspired to create a preprints server and its "commercial" heritage, discourages open sharing of research data. Being a discipline that is focused on the individaul as opposed to the wider community, gives no larger stakeholder an incentive to provide a sufficiently advanced technological solution.

Thus unblocking the path to co-constitutional change in the chemistry community may ultimately require a greater degree of engagement amongst stakeholders, to better establish, what it is they value from scholarly discourse, what they expect to receive and provide to the overall system, and how they may build on this to better serve everyone's needs. Disaggregation's focus on the researcher means that the approach is flawed without sufficiently advanced technology to make the burden on the researcher non-existent and disaggregation as proposed presents no other stakeholder with an incentive to develop such technology.

6.4 Conclusion

To sum up the findings from the disaggregation evaluation study, each of the research questions defined before the study will be revisited.

1. How would the process of disaggregation operate alongside conventional processes and workflows already established for the various stakeholders of academic discourse?

The participants' responses make it clear that disaggregation as proposed does not sit well alongside conventional scholarly workflows, where it is too heavily based towards the latter, dissemination focused, portion of the scholarly knowledge cycle. One of the most immediate issues that arises, evident both from talks given at the ACS conference as well as responses from participants is that data is too difficult to extract reliably

at this stage, with most research being distributed in the form of a PDF. The second major issue for disaggregation in the manner in which it has been proposed results from the extra burden it places upon the researcher. Not only do they have to conduct the disaggregation process, but if the extracted data is to be considered reliable it may need to undergo another peer review process, ultimately creating more stages for a researcher and their community in the overall process of disseminating research; a significant issue when researchers are already expected to record their research outputs in a number of different locations.

However, it is not impractical for disaggregation to operate alongside conventional processes. Disaggregation as proposed, could resolve many of the issues by becoming an operation performed by publishers, who could use it as a value adding service and have the resources to host the results of disaggregation on a discoverable and citable platform. But implementing disaggregation in this manner may prove to be costly to researchers or institutions if publishers choose to charge for access, and may limit the source material for disaggregation to just articles that publishers publish. Retrospective disaggregation may ultimately only be possible with sufficiently sophisticated text and data-mining algorithms, such that the entire process may be automated, but as observed by study participants this is not realistic at this moment in time.

Conversely, the participants' responses made it clear that disaggregation has a place in the day-to-day workflows of the researcher, acting as a knowledge management tool, that will help principal investigators manage their lab and assist researchers with being organised so that they can ultimately better present their results.

2. How might disaggregation be implemented with respect to current services, standards and protocols?

Whilst it may be clear that the proposed method for implementing disaggregation may be flawed, disaggregation may nevertheless take advantage of existing services and standards. Although the study participants had little to say with regards to existing services and standards, a number of organisations and standards were mentioned such as IUPAC and InChIs, as well as the notion that chemistry has a variety of standards to match the variety of forms data may take.

However, with regards to the implementation, participants did refer to ELNs as a means for enabling disaggregation to take place within the laboratory context and provided mechanism for allowing the data contained within to be discoverable and accessible. Existing technologies such as OAI-PMH, as used by the repository community may help enable access to data (Sompel et al., 2004), but mechanism for pushing data to other platforms may also be essential to make results more discoverable when the user wishes to make the data openly available to their research community. A key theme that emerged from discussing disaggregation is the need for disaggregated data to be discoverable and

citable and pushing to existing services would be necessary in making this happen if disaggregation were implemented in a distributed fashion across individual labs.

3. How might disaggregated items be used? How should they be stored and presented to maximise their utility?

Disaggregation is in a position to enable knowledge management within labs and within an individual's own research output. Not only may this introduce efficiencies within the lab, but through structuring data in machine readable formats e-science techniques for analysing the data become available. Once again however the efficacy of disaggregation is dependent on its implementation, providing researchers with an incentive to use it – now thought to be through lab management – but also a mechanism for enabling data to be made discoverable and reusable. Therefore to enable knowledge management and indexing of a lab's research outputs, disaggregated data should still be stored at a local level, using local infrastructure so that the disaggregation process may be configured to suit the particular needs of the lab. For the disaggregated data that researchers want to be made available to the wider community however, a mechanism needs to be established for hosting that data on a wider platform, known to the appropriate scholarly community and managed by a stakeholder with the required resources.

4. What effect would disaggregation have on the functions of scholarly discourse, including the dissemination and assessment of research outputs, and assignation of reputation?

Participants' responses indicate that disaggregation would ultimately have little effect on the dissemination of research outputs, the focus would still lie on existing forms of communication such as articles and theses. Where disaggregation may be able to assist dissemination of research however is through the provision of supplementary material, presenting users with structured, organised data derived from their research and through allowing researchers to organise their research outputs and better plan the writing of publications.

With disaggregation unlikely to assist the dissemination of research outputs, it is consequently unlikely to help a researcher boost their reputation either. Publishing disaggregated items to a researcher's portfolio page will not help enhance their reputation if the platform they are being hosted on is not recognised by the various stakeholders of the scholarly knowledge cycle, or if it is not discoverable. It also becomes difficult to compare and contrast researcher outputs when disaggregation may take different forms in different contexts. With academia already not fairly measuring the impact of researchers who make their contributions to their community or society in ways that extend beyond publishing articles in journals with a high impact factor, introducing yet further nuance through disaggregation is unlikely to help resolve this issue. Reputation mechanisms work at present because they are simple and convenient, offering quick insight when it comes to deciding which papers to read and which researchers to hire.

5. What motivates researchers to engage with new approaches to scholarly discourse?

The publication of articles is a major driving force for researchers. Not only do they provide reputation for publishing researchers, but they also provide a form through which researchers can structure their work. However, appealing directly to researchers' pursuit of publishing, as the proposed method of disaggregation does, is ineffective. Researchers know that it is publications which are ultimately counted by their peers and other stakeholders.

As such, it is making the process of getting published easier that is the most effective way of getting researchers to engage with new approaches to scholarly discourse. Providing use cases of new technologies and practices in action and showing the beneficial results may be crucial to convincing researchers that they may enhance one's publication record. It is also important to ensure that adopting these processes is as convenient as possible. For such changes to be feasible however would require technological advances that make disaggregation processes seamless for researchers, requiring investment from the stakeholders of scholarly discourse. Education, advocacy and mandates may also provide routes to shifting cultural norms. Funder mandates have demonstrated how open access uptake may be encouraged, but for a process such as disaggregation placed within a lab context, institutional policies and education may be better placed to encourage researchers to adopt new processes for aiding knowledge management, and consequently scholarly discourse.

Chapter 7

Disaggregation: A New Approach

7.1 Proposals

Through initially proposing disaggregation and evaluating it, a number of conclusions and proposals for those seeking to encourage more open and thorough scholarly discourse in the future, not just for disaggregation, are established.

1. Current methods of dissemination do not suit all research, researchers or institutions. Scholarly discourse needs to be more agile.

Academic inquiry is an international endeavour, comprising of researchers working across a range of disciplines and sub-disciplines, and in a range of different types of institutions. Research is also a fundamentally collaborative practise, with knowledge built on past discoveries, and yet it is also driven by competition, in a race to be the first to contribute new findings and to secure access to funding and resources. As such it becomes vital to disseminate, review and cite research findings in a consistent manner that has a recognised meaning around the world, a function accomplished by academic publishing.

Yet, this one size fits all model for disseminating research presents an unfair system for many. Previous research has already indicated how notions such as impact factors may fail to reflect social or economic impact (Smith, 2001; Seglen, 1997) and this was an idea reinforced by the disaggregation study. Participants described how for example their institution was not organised around the notion of publishing high impact academic journal articles due to a high turn around of students who were unable to commit to long term projects. One participant described how an academic was penalised for working on teaching materials, rather than on publishing articles; and another described how valuable, but ultimately "pedestrian" results, may be best served by maximising access to them, rather than putting such results behind a paywall. This problem becomes greater still when considering it is not just a problem of contributing to research, but gaining access to it as well, as indicated by the serials crisis.

Therefore scholarly discourse needs to adopt a more agile, flexible approach to disseminating research findings and yet maintain a consistent, internationally recognised approach to citing works. The FORCE11 community has already outlined how this may work (Martone, 2014), but as commented upon by Neylon (2016), "the current limited scale of data citation means that it will be some time before a critical mass of high quality information is available". As such, an element of responsibility falls upon academic institutions, who are in a position to educate and support researchers about modern approaches to citing and to being cited, with the latter being something an institution can additionally support by providing platforms where valuable research results, which have nevertheless been rejected by conventional publishers, may be published and cited. Such tools, like the Disaggregator prototype that aims to facilitate a similar service, should make clear how work presented there is available to be cited and how this should be done.

2. Only certain venues for research are recognised. New services for scholarly discourse should respect these.

Disaggregation was designed with the intent to build new approaches to dissemination on top of the existing system in the hope of complementing current approaches rather than compete with them. However through doing so, it simply reveals the insignificance of this new approach; if a new method is to be counted, all stakeholders have to recognise it, requiring stakeholder collaboration or a critical mass of stakeholder engagement. The reputation that a researcher develops through publishing, concerns the place of publication as much as the content – publishing in renowned journals is an internationally recognised currency for reputation in academia.

This was ultimately the approach adopted by the Open Access (OA) movement, which did not seek to disrupt scholarly discourse by changing the manner in which it operates, but simply aimed to make it more open. Gold OA through pay-to-publish journals, or green OA chiefly enabled by institutional repositories, built on top of the existing publishing system and required very little extra effort required from the researchers themselves to achieve the benefits of OA. Through proposing OA in this manner, a feasible solution with obvious benefits is presented, making it easier to gather the support of other vital stakeholders such as institutions and funders.

However, whilst only certain venues for presenting research are recognised and will contribute to a researcher's reputation, this does not mean that arguments focused on helping a researcher better their reputation are irrelevant. Publishing remains important to researchers and any mechanism that can be put in place that may help them become published should be considered.

3. Build into the researcher's workflow.

Any new system that aims to make scholarly discourse more open, needs to be built on the idea that it is making the researcher's life easier. When discussing disintermediation it became clear that the researcher lies at the centre of the various interactions that make up scholarly discourse. However the researcher is not primarily concerned with their role as a small cog in a machine that produces research and effects academic, economic and societal impact. Instead they are focused on the day-to-day work of their research and any changes which aim to enhance a societal good, should appeal to making researchers' work in the lab easier.

The process of building openness and dissemination into the researcher's workflow has added benefits. As identified during the interviews and focus groups, a researcher's processes may speak as much about their reputation as do their outputs. This provides a whole new way of helping researchers to enhance their reputation that complements traditional means not by adding more of the same types of research outputs at a different location, but by providing new metrics. However when building new processes into the researcher's workflow it is important to consider how data may be extracted from the workflow and disseminated to the wider community. This may be achieved simply through helping the researcher to publish more frequently and thus more of their research is made open, but this may result simply in more journal articles of a lower quality. Alternatively knowledge management solutions that operate for intra-lab communications could nudge researchers to communicate their data more openly, or the benefits of disseminating data at an earlier stage in the scholarly knowledge cycle, namely being able to receive feedback on your work practises from the community and act on it before the final work is complete, should be communicated to researchers.

4. An ELN solution needs to be open source.

Within the chemistry context, building new approaches into the researcher's day-to-day workflows involves an ELN solution. However, as previously discussed, ELNs can be expensive to develop and maintain and often lack flexibility. Therefore an ELN for academia, that may adopt disaggregation principles should be open source and should endeavour to foster a community of developers and practitioners around it who will be able to sustain development, whilst minimising cost and risk for any individual institution, an approach that has proven successful for the OA repository community (Smith et al., 2003).

5. Disaggregation needs to be flexible.

Finally, the idea of disaggregation has the potential to permit a number of benefits, facilitating e-science approaches and knowledge management, by clearly defining data through the process of curating disaggregated components and therefore allowing greater indexing and control of the data. However this process needs to be flexible. Different

labs have different specialities and operate within different contexts – for example some may be more teaching oriented whereas others may be more research focused – and as such disaggregation should aim to be agile, helping research groups best reflect their work.

However, a flexible approach to disaggregation is also required if the process is to be a success. OA works because it requires little extra effort from researcher's to achieve the benefits (and even the little effort that is required is often deemed burdensome (Björk et al., 2014; Rowlands and Nicholas, 2005)). What disaggregation is trying to achieve, the transformation of how research is communicated, extending beyond the conventional artefacts of discourse such as journal articles, demands further input from researchers, who need to engage with a new approach. As such the system needs to be flexible enough to work even if not everyone is participating entirely, or only has the time or the will to participate partially. In this manner, the approach to disaggregation is not unlike Berners-Lee's for the World Wide Web, sacrificing "control over the hypertext structure and data" to allow a "massively decentralized" system (Whitehead, 1999). Through doing so the Web saw wide-scale adoption, ultimately benefiting from the network effect to become the dominant hypertext system (Whitehead, 1999) – with recognition from stakeholders being a crucial factor of uptake in scholarly discourse, a flexible system which may encourage network effects may be a route for success.

7.2 A New Proposal for Disaggregation

Building upon the evaluation of disaggregation as it was initially presented, it should be possible to devise a new approach for disaggregation that aims to take the lessons learnt and produce something that may be more amiable to the scholarly knowledge cycle stakeholders.

The vision of disaggregation presented in Figure 7.1 is one where the research group's ELN is a key component. However the world outside the lab group, here consisting of researchers, their support staff and PIs, also has a role to play, via a new facet introduced to disaggregation, the 'Component Store'.

7.2.1 Component Store

The component store is an important new feature to the new proposal for disaggregation, representing a Web service whereby any of the stakeholders of scholarly discourse, from funders through to publishers, can define their components to be used as part of disaggregation. The component descriptors, collect together a group of fields, which together represent some concept or artefact that may be disseminated; working in the

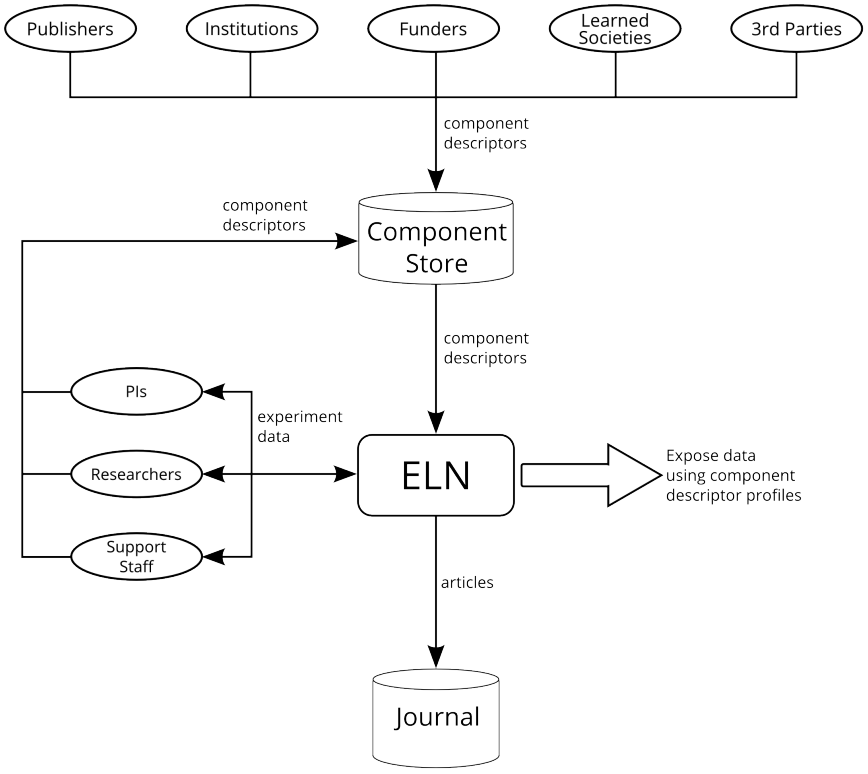

FIGURE 7.1: A new approach to disaggregation, with the process of disaggregation built into an ELN, but driven by components specified by a wide range of stakeholders.

same manner as the Disaggregator prototype. Table 7.1 lists the the fields needed to define a component.

However whereas with the Disaggregator prototype the components were locally defined to that instance of the Disaggregator, here they are available to all with anyone able to define new ones and users choosing which may be relevant for them. Through doing so, funders, publishers and institutions are able to define the data they are interested in, making it clear to researchers that the data they are gathering together is recognised. "Deposit Location" is a key field in enabling this, allowing those who seek to define disaggregation components to specify where they would like to see extracted data deposited, presenting stakeholders with an opportunity to build platforms, from which disaggregated data may be discovered, and where reviewing and reputation mechanisms may be established.

Table 7.2 shows an example of a component descriptor along with example values. A component descriptor's 'Name' and 'Description' describe the type of object being described, in this instance a chemical structure for the Royal Society of Chemistry's ChemSpider database. The 'Publisher' field is filled with the value of the individual or organisation making the deposit (in this case the Chemistry department at the University of Southampton) and the 'Deposit Location' field describes where the disaggregated component can be stored and accessible to others where appropriate (in this case it

Name*	A name for the disaggregator component that concisely indicates what types of artefacts it may be used to describe.
Description*	A more detailed description of the disaggregator component to assist users in deciding whether or not the component is relevant within their laboratory context.
Publisher*	The individual or organisation who has published the disaggregator component.
Deposit Location	A platform where an ELN may push extracted content to, so that it is available to the wider community. Component publishers can build and then specify platforms. The platforms they build will help make disaggregated components discoverable and, where such platforms have been built by recognised organisations or communities, will lend legitimacy to the items published there.
Profile*	A list of fields, their types and whether or not they are mandatory or may store multiple values. The fields listed within the profile can be comprised of the different field types as used by the Disaggregator prototype, but more are likely to be added as the various community's needs become more complex.

TABLE 7.1: The fields used to create a disaggregator component descriptor. Fields marked with a * are mandatory.

refers to the ChemSpider website, but in an implementation of this multi-stakeholder encompassing version of disaggregation this may refer to an API endpoint that ELNs may be able to deposit records to). This is then followed by the descriptor's profile, listing the fields required for this component along with some example values (in this example case the chemical structure details for aspirin).

It is not just the major stakeholders of scholarly discourse who can define components however. To assist with knowledge management within the research group, project principal investigators (PIs) and support staff, such as the "environmental safety" officers mentioned by one disaggregation interview participant, may also be in a position to prescribe disaggregation components that will be of use in their specific lab context. However, in the same way that Apple's AppStore allows them to crowdsource app development "to a global base of software developers" (Bergvall-Kåreborn and Howcroft, 2013), so too does the disaggregation component store, allowing anyone to contribute and publish components and enabling the possibility of common standards to emerge through reuse and refinement.

A key concern however, should a research group's ELN be taking advantage of numerous component descriptors from the component store, is the issue of researchers being obliged to repeatedly scan their work to extract the disaggregated components. Such an issue could be resolved however by allowing for mappings to be defined that transcribe data held in one component to that defined by another. Ultimately, however the day-to-day implementation of this new approach to disaggreagtion could be left to research group or institution policy, ensuring that researchers only have to define data in one particular way

Name	ChemSpider Chemical Structure
Description	A chemical structure for the Royal Society of Chemistry's ChemSpider database.
Publisher	Chemistry, University of Southampton
Deposit Location	http://www.chemspider.com/
Profile	• Name: Aspirin • Molecular Formula: C9H8O4 • Average Mass: 180.157 Da • Monoisotopic Mass: 180.042252 Da • Mol File: '/home/documents/molfiles/2157.mol'

TABLE 7.2: An example of a disaggregated component. The type of component is a ChemSpider database chemical structure which has fields for its 'Name', 'Molecular Formula', 'Average Mass', 'Monoisotopic Mass' and 'Mol File'. The publisher indicates where the component comes from, and the deposit location indicates the site the component can be deposited at (in a real-life implementation this may be a link to an API endpoint allowing the ELN to automatically deposit the record).

and some other solution may be in place for fulfilling any other component requirements that may arise.

Another concern with the component store is which organisation should be responsible for hosting such a service. This may be an opportunity for a learned society to adopt a value adding position for example, but with the component store aiming to be a flexible platform, open for all stakeholders to take advantage of, this may create a biased platform. With the component descriptor comprising of a range of different fields, the responsibility may be better placed with a non-profit organisation such as FORCE11, who are already proposing how data should be cited.

7.2.2 The ELN

Whilst ELNs have been used with success in industry, they remain unpopular in academia, as suggested both by interview participants, "in most cases academics don't buy existing ELNs they... they're just too much money", and by previous research: "We wouldn't touch those [commercial "eLab book systems"] with a barge pole. They're horrible... Not for us" said one of the participants from Schraefel et al. (2004)'s *Making Tea* study. Schraefel et al. (2004) provides an overview of different approaches to implementing an ELN, proposing all ELN solutions "can be charted between two axes: the degree to which paper is kept/replicated or entirely replaced on one axis, and the degree to which the system for the device is personal (like a lab book) or distributed (like the web)". Schraefel et al. (2004), through their original ethnographic methodology, ultimately opt

to design "a digital lab book that affords the functional and experiential qualities of a paper-based one, while transparently introducing additional benefits for the lab practice available in a digital system". Alongside this approach to building an ELN, an RDF (Resource Description Framework) ontology was devised to provide a flexible, extensible and machine readable structure of experimental data, allowing for "automated logical inferences" and thus potential e-Science benefits (Hughes et al., 2004).

More recently, the ELN LabTrove adopted a blogging platform for bringing a linked data approach to ELNs and found that whilst this did pose problems – lack of metadata support, poor representation of non-Web objects – it also offered a number of advantages such as supporting linked data, as well as a simple "single post for each physical sample and each data file" set up (Milsted et al., 2013). LabTrove has also enjoyed success amongst its user base, despite admitting the "user interface has not been a priority for development and remains rather crude" and that "aspects could be enhanced by including more visual editing tools, improving the drag and drop functionality and offering a range of accessible visualisation tools" (Milsted et al., 2013). It ultimately takes a different approach to the *myTea* project, with LabTrove being "semantically unaware; the system itself has no understanding of the content of the posts, the meaning behind their structure and connections or the logic of the metadata", citing that "freedom, and the ability arbitrarily to add or modify metadata keys and values is critical to the recording of what is planned and what actually occurs in the small scale experimental laboratory". Nevertheless, Milsted et al. (2013) go on to demonstrate how "it is feasible to map the local, free-form metadata onto external controlled vocabularies".

The approach adopted by LabTrove, of a web-based platform that operates in the a web browser, presents a promising starting position for building disaggregation in to the researcher's day-to-day workflow, with the Disaggregator prototype already making use of a browser-based interface. It is also recommended that an open source approach to developing the ELN is adopted. This places the other stakeholders, key to making this approach to disaggregation workable and valuable to other stakeholders, in a position to contribute towards it and modify it, ultimately improving the sustainability and effectiveness of the project (Smith et al., 2003).

The approach taken by a disaggregation-based ELN to impose semantics on its contents is different to those developed by Schraefel et al. (2004); Hughes et al. (2004) and Milsted et al. (2013). It presents a very flexible approach to semantically enriching data. By not trying to formally describe everything the ELN documents, and instead only describe and disseminate data by selecting bits that can complete the fields defined by a component store descriptor, researchers can engage with the process as much or as little as they like. Disaggregation becomes a layer on top of their ELN's contents, a layer researchers may be encouraged to engage with as lab group policy, or their sense of knowledge management, dictates. Disaggregation simply becomes a tool to organise and categorise, and to segregate data, with the user choosing which components they wish

to disseminate further and which they wish to remain private. Taking this approach also allows metadata to be retroactively applied – as new components are introduced a researcher or automated disaggregating agent may pick and choose data to fit a new component descriptor – creating a system that may respond to new institutional or funder policy requirements for example. Similarly with data, organised and structured using disaggregator components for day-to-day knowledge management use, these already categorised and described components may be disseminated at a later date when any reasons for keeping them confidential (such as researcher anxiety over scooping, or industry partner confidentiality clauses) may have expired.

7.3 Conclusion

An evaluation of disaggregation suggested that the idea that research outputs could be broken down into constituent parts was a feasible one. By describing data in structured ways it would be possible to enable a number of benefits for different stakeholders and would help to facilitate e-science techniques. However, the way in which disaggregation was proposed was flawed. Applying disaggregation to the final stages of the scholarly knowledge cycle means trying to extract data once it has been packaged in PDFs. It also places an extra burden on researchers without providing sufficient incentives for their engagement, introducing complexities for comparing and measuring researchers' contributions and created issues concerning the hosting of the data, namely its discoverability, permanence and financial costs.

The new proposal for disaggregation aims to fix the above issues. Instead of applying disaggregation to articles and theses that have already been written, it is instead used as a knowledge management service, built into ELNs. In doing so, laboratories can define their own ways of structuring data, relevant to the stakeholders within their local lab context, assisting with the indexing of data and helping create a more efficient laboratory, in turn helping boost the researchers' publication prospects. However, the disaggregation evaluation also highlighted the importance of engaging all stakeholders within the scholarly knowledge cycle to ensure that new technologies or services are recognised and therefore relevant. Therefore any stakeholder may contribute components to be used as part of the disaggregation process and may host disaggregated data on their own platforms or repositories, suiting individual funder, institution or publisher needs, whilst also creating an environment where popular standards may emerge.

It is therefore possible to describe disaggregation in two different contexts: 'top-down' as epitomised by the proposal in Chapter 5, and 'bottom-up' as described above. 'Top-down' disaggregation is hindered by a number of issues, mostly caused by its position in the overarching scholarly knowledge cycle, taking place when the research is complete and bound within a publication. It fails to engage many or academic publishing's

stakeholders and through doing so fails to present an enticing prospect for any of them, with researchers expecting to take on the burden of disaggregation ultimately with little reward. 'Bottom-up' disaggregation solves this, encouraging participation from a range of stakeholders and presenting a flexible approach that is built in to the researcher's day to day workflows. However it has problems of its own; principally the difficulty with disseminating and validating content which is being produced at early stages of the research process and that much of the technology required for such a process to work seamlessly – wide uptake of ELNs within the academic community, and automatic text-mining techniques which are not yet sufficiently sophisticated to not require researcher intervention – does not yet reflect the situation in labs.

Thus the 'top-down' approach to disaggregation epitomised by the proposal in Chapter 5 is ultimately just a stop-hap solution until the 'bottom-up' approach suggested above is feasible, through both increased uptake of ELNs within an academic context and through greater involvement across the stakeholders.

The new approach also raises a number of new questions concerning disaggregation however, which will fall outside of the scope of this thesis. Firstly, by incorporating a centralised aspect of disaggregation in the component store, it becomes the responsibility of a stakeholder to host such a service, which may be difficult to resource given academia's unique structure. Similarly, encouraging stakeholders to create platforms from which users may access disaggregated resources, presents a new approach to data repositories, which may present stakeholders with problems such as how to curate and review data they receive in this manner. Secondly, the effectiveness of a flexible approach to semantically enriching data is currently unknown – disaggregation built into an ELN in this way may result in useless data for many potential users, or allowing a flexible approach may ultimately result in more data being published as there are fewer requirements placed upon the depositing researcher. Developing an ELN that supports disaggregation will also require further research into ELN solutions both with regards to user interfaces and underlying data storage.

Chapter 8

Conclusions

8.1 Overall Discussion

Throughout the course of this thesis, the topic of the Web and its relation to the scholarly knowledge cycle – how researchers formulate ideas, assess the background literature, perform their research and disseminate the results – has been analysed from a number of different perspectives and with varying evaluative agendas. Having now proposed and evaluated the relationship between the Web and scholarly discourse through a number of different mechanisms, an opportunity presents itself to revisit some of these ideas to establish what has been learnt.

8.1.1 The Web – A Scholarly Utopia?

As an open, free-to-use platform, designed in CERN to facilitate the sharing of results, the Web should be an ideal medium for scholars, whose sole aim is to contribute to the ever expanding body of knowledge. The digital medium allows research results to be presented in context specific ways. Data storage is so cheap that not only may vast amounts of data be recorded and made available, but so too can the provenance that lies behind it, enhancing the scientific method's tenets of reproducibility and verifiability. Yet despite a vast array of technologies available to the research community, the realm of scholarly discourse is swathed in debates and confronted by a number of issues, for the potential of the Web is largely used to distribute conventional scholarly articles, published in academic journals. And whilst the Web eliminates the traditional marginal cost of distributing research – the printing, binding and shipping of journals – much research lies behind pay-walls, inaccessible to those who cannot afford to pay, even when that research has been publicly funded.

The most significant attempt to help alleviate the problem of inaccessible research, in an era where the Web makes communication of information so cheap and easy, is the open

access (OA) movement, designed to remove "access barriers" to "peer-reviewed journal literature" (Chan et al., 2002), predicated on researchers' desire to push the boundaries of knowledge. OA has changed the scholarly landscape; recent developments including a statement from the EU Ministers of Research and Innovation "to support a transition to immediate open access as the default by 2020"(Council of the European Union, 2016) and the post-2014 Research Excellence Framework specifying submissions will only be eligible if they have been deposited in an open access repository (Higher Education Funding Council for England et al., 2015), mean that publishers are having to adapt their business models and that more and more people have access to research. However progress towards OA has been slow moving (initially proposed in 2002, with the benefits of the aforementioned policies not arriving until over a decade later) and OA does not utilise the full potential of the Web. By focusing solely on the access issue, OA fails to address the many other facets of academic publishing: researchers developing their reputations and careers, the allocation of funding to institutions, and the reviewing of research outputs to assess their validity and importance.

What has also been revealed through an evaluation of the ideas of disintermediation and disaggregation however is that the existing methods of scholarly discourse fail many researchers. This is not a new discovery, but it highlights why the Web and associated movements such as OA, may have failed to enable the scholarly utopia they have the affordances to permit. Researchers make their contributions to their field in a myriad of different ways, but the extent to which so called high impact publications are valued over other contributions to the scholarly domain, creates a bias in favour of publications, even where high impact publications may not be a main priority of an institution. The overall effect is to create a research environment in which researchers and research institutes lack mobility – a result which is not an indictment on the efforts of the OA community or any others who have endeavoured to change the way in which scholarly communication in the age of Web operates, but instead on the innate conservatism of the research community. Where disciplinary differences allow for new approaches to sharing knowledge to emerge this has taken place. For example, within physics, where research problems require vast amounts of data from large collaborations of researchers around the world, OA repositories such as arXiv prove very successful. It is the established community norms, rather than the Web that present problems when trying to fully take advantage of the knowledge sharing potential the Web has to offer.

The scholarly community, comprised of so many different stakeholders needs to begin to not only more explicitly acknowledge that all research outputs are important – which it is beginning to do as funding councils begin to change the way in which they assess their projects (Research Councils UK, 2014b) – but also to understand that the medium of dissemination that works for high impact academic work, does not necessarily suit the less "high impact", but nevertheless still important academic work, or researchers contributions to higher education, industry collaboration or social outreach programmes

(to name just a few ways in which researchers may choose to expend their skills and knowledge). If it is to do so, the community needs to devise new ways of recognising such contributions, identify there are different contexts within which a researcher may work, often governed by the type of institution they choose to work at, and provision the community with the appropriate tools and techniques for context-sensitive dissemination.

8.1.2 Disintermediation Critique

The Web may be the ideal platform for academic knowledge exchange and yet it has failed to produce the scholarly utopia that would seem inevitable. To try and understand why this might be, this thesis chose not to look at how the Web is being used by researchers to communicate their research findings, but instead look at the overall system, adopting a holistic perspective that recognises that academia is formed from a collection of stakeholders each with their own roles and agendas. Such an analysis reveals the researcher's central role in academia; they are the consumers and producers of the knowledge that is exchanged and largely responsible for quality assurances processes too.

As individuals, researchers are burdened with the primary tasks of scholarly discourse, the nature of their position as those who are exploring and expanding the boundaries of knowledge inevitably means that their actions are largely of interest and are largely only assessable by their peers, before the knowledge is ready to be processed and applied. The other stakeholders such as the publishers, academic institutions and librarians are simply responsible for facilitating these processes, either by collecting research together, providing facilities and resources, or by managing access to research. The Web therefore is in a position to preside over a phase of disintermediation, the removal of intermediary stakeholders to allow the primary stakeholder that is the researcher to carry on engaging with their conventional processes by interacting with one another directly, changing the academic publishing landscape as it does so.

However the assessment of stakeholder interactions that reveals this potential reconfiguration of the processes and roles played fails to acknowledge a number of implicit traits and behaviours that the stakeholders exhibit. There are various reasons for researchers not wanting to disseminate their research outside of the formal, acknowledged format of the journal article (or equivalent scholarly artefacts such as conference proceedings) chief among these being the effectiveness at which a journal's brand can bestow reputation, a facility the research focus groups and interviews showed to be too inconvenient, when assessing research, not to use. Journal publications provide a concise means through which to communicate complex discoveries and manage the externalities that arise: they provide a simple (if perhaps flawed) way of measuring the contribution, a method of collecting research outputs so that they are easy to find, have permanence

and can be cited in a recognisable fashion. They also provide a goal for researchers to aim towards and a structure around which their research can take place, ultimately helping researchers to define the scope of their research and draw it to a conclusion. Without this structure research may be difficult to both discover and to understand.

Thus, the slow progress of open access and the general lack of taking advantage of the affordances of the Web becomes clear. Why risk publishing in new open access journals which have yet to establish a reputation? Gold OA publishing presents an opportunity cost that needs to be evaluated by the publishing researcher who need to consider if the open nature of their publication will draw in more citations and improve their impact, or is there the chance that they are paying to publish in a predatory journal. Similarly journals cannot be bypassed completely for how will others find the work and know at a glance if the work can be trusted, for the scientific method's principal of reproducing the work of others is one that is fundamental, if not always invoked in practise as revealed by interview participants. If a shift in the practises of disseminating research outputs is going to occur, it will likely therefore need to take into account both the central nature of the researcher's role but also the immovable status of journal publications in academia.

8.1.3 Disaggregation Critique

The process proposed for encouraging a more open approach to scholarly discourse that aimed to take advantage of the potential the Web has to offer was disaggregation. Disaggregation combined the lessons learnt previously, hypothesising that researcher's strive for journal publications in order to enhance their reputations and that the articles and other scholarly artefacts they produce through doing so contain a wealth of data that could be disseminated in more widely accessible and appropriate means. Through taking this approach the researchers need to produce publications from their work is not hindered and they are presented with an opportunity to complement their publications with a more thorough and detailed portfolio of their work. To illustrate, inform and evaluate this hypothesis, the disaggregator prototype was developed, which took the ideas behind disaggregation and implemented them as part of a Web based service. In doing so a number of technologies were incorporated into disaggregation, including the use of text-mining algorithms, identifiers such as ORCIDs, and semantic web technologies.

When evaluating disaggregation and the disaggregator prototype however it became clear that a number of elements of the implementation were misguided. Expecting researchers to disaggregate their work at such a late stage in the workflow imposes not only an extra task on a stakeholder who is already overburdened with their various roles, but one which is difficult to accomplish when the data is often locked in the closed PDF format. Endeavouring to provide a complementary source of information by appealing to researchers' desire to enhance their reputation was also mistaken, underestimating

the monopolistic nature of the convenient and entrenched journal publications, which makes them vital to the many stakeholders of scholarly discourse.

What the evaluation of disaggregation also revealed was the nature of the change that is required within scholarly discourse if the affordances of the Web are to be properly utilised. Building new tools or services is a technically deterministic approach which will not be sufficiently disruptive to cause a change in the culture of scholarly discourse that is required. Innovative technology is capable of disrupting markets, as observed numerous times with the Web, but in this instance where the social constructs that it is attempting to dislodge are so pervasive, technological determinism may be hindered. However social determinism is also inhibited; the unique funding situation of academia makes it difficult for any one organisation to implement or develop technologies that can improve scholarly discourse. There is also a lack of pressure to change emerging from the community who are effected, the cost of journal access is often abstracted away from the researchers themselves by librarians, and the reputation mechanism provided by journals drives researchers who are otherwise only interested in conducting science.

8.2 Disaggregation – A New Approach

With an evaluation of the proposed approach to disaggregation highlighting the important role of the researchers' workflow, but generally recognising that disaggregation provides a valid means for disseminating data with the aim of enabling a number of advantages including e-science research techniques, attention turns to how else disaggregation may be implemented. Within the chemistry context, building disaggregation into an electronic lab notebook (ELN) allows the process to be embedded within a researcher's day-to-day processes and at a stage prior to formalising the knowledge through the authoring and publishing of an article – an approach to disaggregation that emerged through several study participants' comments.

Building ELNs within academia is problematic however as revealed by the study participants' comments. Whilst ELNs are successful within an industrial context, they are often too expensive and contain unnecessary features for academia which deter usage. Academia also presents a different approach to structuring research activities, with a single institution comprising of many small units each specialising in completely different areas of research, whereas within an industrial context, research may typically be focused on a particular product area, presenting more challenging (and ultimately more expensive) development and support demands upon ELN suppliers. However the OA movement saw success by adopting an open source approach to building the infrastructure it required (institutional repositories) and this may be a solution for ELN development too. An open source approach to developing an ELN, has the potential to deliver an ELN which can be configured at low cost to suit any laboratory context and

thus support a new approach to disaggregation. It also has the benefits of allowing the various types of stakeholders to contribute their own ideas and resources as to how a practice such as disaggregation should take place.

There are still a number of problems yet to be addressed by the the new approach to disaggregation however. For example, researchers' desire to have automatic extraction of results and seamless interoperability with lab equipment, so that ultimately researcher intervention is minimal presents a problem for which the technology is still "decades down the road" as one interview participant described. As such, changing the culture of scholarly discourse, even at the day-to-day workflow level in the ELN may still present a challenge for the co-constitutional nature of the Web, with the social pressure to develop technology not forming due to a lack of potential demonstrated by technology that is immediately beneficial to the research community as a whole.

8.3 Limitations

The scope of the work presented in this thesis, has two significant limitations. Firstly, the ideas proposed are very difficult to evaluate. They involve large numbers of stakeholders, long term changes and substantial risks to be undertaken by the stakeholders if they are to be evaluated through real-world experiments. The idea of disintermediation for example cannot be implemented in reality; researchers need to be able to conduct their science and disseminate their results via the means that are known to work and are expected to use. Similarly whilst disaggregation can be implemented in a basic form, as per the Disaggregator prototype, a sophisticated implementation that would give a more accurate impression of the concept would require resources outside of the scope of a PhD project and would need to be embedded within laboratory contexts.

The second limitation comes from the range of views collected on the idea of disaggregation. This has necessarily had to be restricted; scholarly discourse is a vast topic and this thesis has chosen to look at it primarily through the lens of one discipline, chemistry. Thus disaggregation has been proposed as a concept that applies to chemistry and as such the participants for the interviews and focus groups all had to be individuals who are familiar with the chemistry domain. The nature of expert interviews may also lead to potentially biased results, failing to truly reflect the views of the community at large. However, the disruptive nature of the ideas in question benefited from an evaluative perspective where those critiquing the ideas are already familiar with the topic being studied. The extent to which disaggregation is applicable to other disciplines also remains unknown, although the results form the study suggest it should work in other scientific disciplines.

8.4 Future Work

Future work will inevitably focus on the new proposal for disaggregation, building disaggregation into an ELN to act as a tool for knowledge management within the lab, but also as a platform from which other stakeholders, such as funders or publishers, are able to contribute. Therefore to develop this further a first step would be to start analysing the researcher's day-to-day workflows from the perspective of disaggregation to build upon existing findings in the field of ELN development, and through doing so construct and evaluate a new disaggregation-based ELN prototype. An important aspect of developing the ELN would be to ensure that all code is released under an open source compatible licence and advocacy efforts should also be considered to encourage development support from other organisations and institutions where possible.

The new proposal for disaggregation is also very firmly entrenched within a chemistry context. It suggests that disaggregation should take place in the day-to-day work of the researcher which within chemistry is can be centred around the ELN. The extent to which this may apply to other disciplines should be investigated however if a similar principle is to be applied. The approach may be easily adapted to suit other scientific disciplines such as biology or physics, but may struggle to operate within more contrasting disciplines of social sciences and the humanities. The new approach to disaggregation has been established through interviews with stakeholders who are based within chemistry and as such interviews with equivalent stakeholders in other subject areas may be a useful place to start when considering how disaggregation may be applied elsewhere.

A second aspect to building and evaluating the new approach to disaggregation would be to consider what component descriptors it would be useful to define and use to help prepopulate a 'Component Store' service as proposed. To do so would require talking to a number of key stakeholders and compiling reports of the platforms and services they already operate, which a disaggregation-based ELN could interoperate with.

8.5 In Conclusion

Throughout this thesis, both a social and technological perspective on the Web and how it relates to the scholarly knowledge cycle have been taken, encompassing socio-economic theories and proposals for new means of disseminating research results. Disaggregation was a result of much research into different approaches to publishing and insight into the behaviour of researchers and what drives them. Yet the implementation proposed for disaggregation was not quite compatible with the stakeholders and the system that is scholarly discourse. Trying to gain access to data and enable e-science techniques, after conventional scholarly articles or theses have already been constructed, results in an approach that is too late in the scholarly knowledge cycle and as such has to

overcome more barriers, namely the difficulties in extracting relevant data from PDFs and incentivising researchers to participate.

Instead, as the results of the disaggregation evaluation study made clear, new approaches to scholarly discourse need to be built into the researcher's day-to-day workflows, making them a part of scientific practise. Doing so presents new challenges in designing ELNs that are sufficiently complex that they have the flexibility of their paper-based counterparts along with added semantic sophistication to leverage the benefits of disaggregation. However to ensure that the connection to the Web is not lost and that a change to the researcher's scientific practises does result in more open and accessible dissemination of research, the principles of the Web need to be embedded within the ELN, allowing users to easily link to the works of others, create and publish content, be flexible, simple and open to use. Such a service could then be further enhanced by encouraging collaboration and reuse of resources, allowing users to publish and share their own approaches to disaggregation.

The Web does not just present an open and effective platform for achieving these aims however. It also represents a platform where many organisations and individuals are already present and trying to enhance scholarly discourse. Through the Web, an ELN would be able to link concepts, results, people and organisations together. Taking a holistic approach to revolutionising scholarly discourse is the only valid approach, when considering the mesh of the stakeholders, motivations and agendas, and a Web based ELN has the potential to assist all. However, the introduction of new technologies to bring the stakeholders together is by itself insufficient to produce lasting change. A cultural shift needs to take place, and this too demands a holistic approach, where stakeholders collaborate with one another. The open access community has already shown how this may be achieved, with pressure from OA advocates leading to lasting change on the behalf of research funders and publishers alike. A similar change should be possible in scientific practice, but first the technology needs to be developed in an open, sustainable way, such that the beneficial implications are evident to all.

Appendix A

Preliminary Questionnaire and Responses

The questionnaire and responses below were carried out as part of a preliminary study for the work in this thesis.

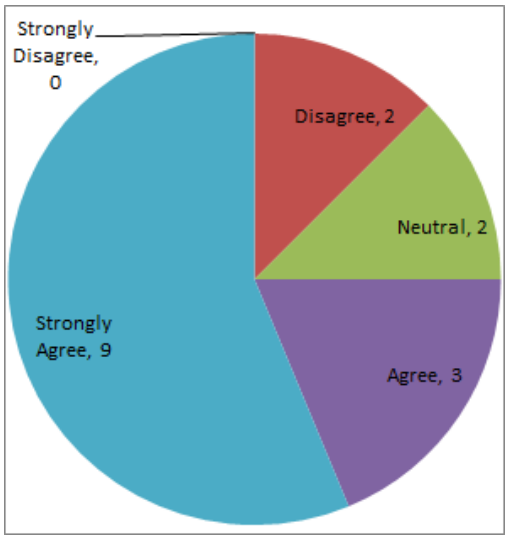

FIGURE A.1: Q1. One should record all laboratory work, including both planning and execution

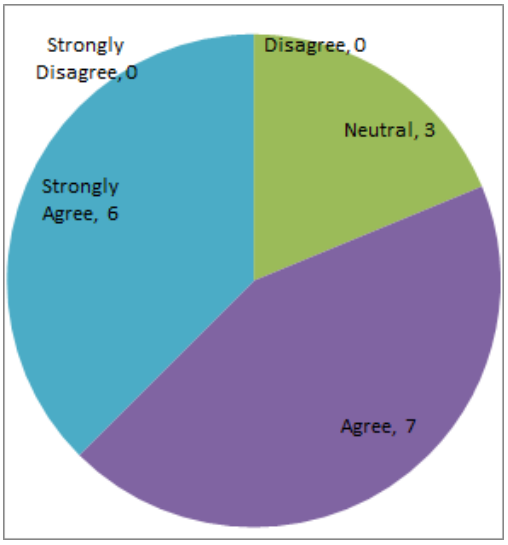

FIGURE A.2: Q2. I should make everything I record available to my supervisor

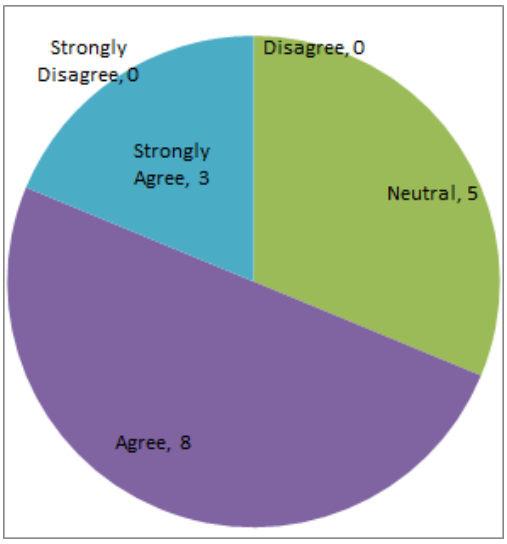

FIGURE A.3: Q3. I should make everything I record available to my research group

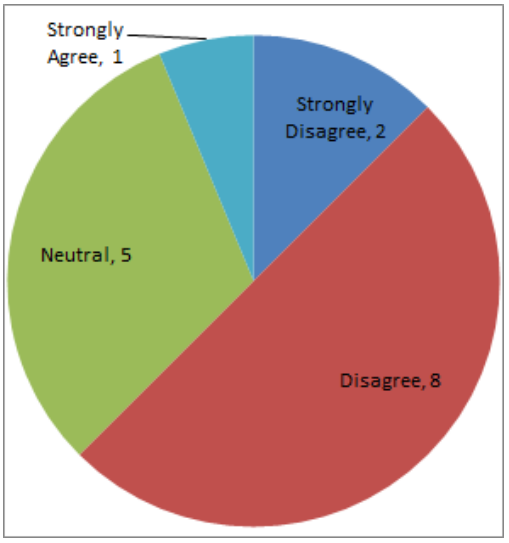

FIGURE A.4: Q4. I should male everything I record availably publicly

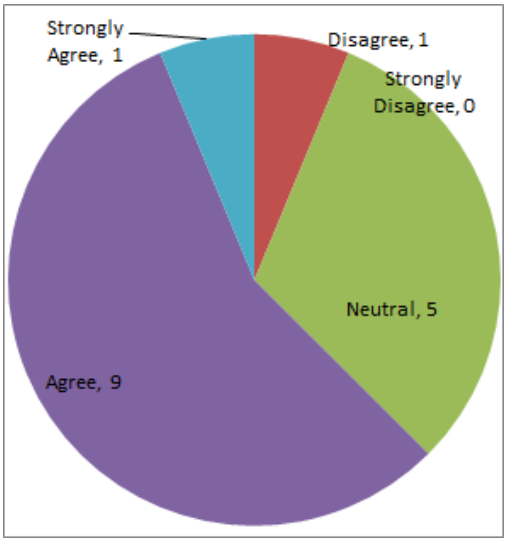

FIGURE A.5: Q5. Access to other researchers' experiment records would be useful

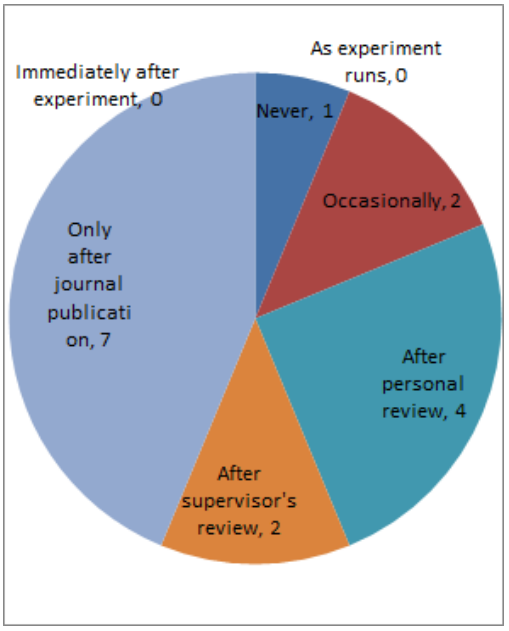

FIGURE A.6: Q6. When would you be willing to make your recorded work available?

Q7. What impacts (good or bad) do you perceive to informally publishing research findings (i.e. not publishing in a journal, but online)?

Old ideas are lost into the public domain, so cannot be published at a later date if revisited.

lack of recognition for work. difficulty of publishing 'already published' work

Very useful to find out information on techniques and other processes that may not be elaborated on in publications. However, do not want to give away information that may compromise your own research.

can be good for collaborations but my get shafted if competitors get info

Bad for career, publication in a good journal is still regarded as important.

Other researchers might steal your ideas before you have chance to publish them.

It should not enforced, as some of the work is novel and commercial. There is a risk of poor quality work and in accurate records, must as least be a review process.

A negative factor to publishing findings informally could be that there is just too much data available making it quite difficult to find what you're looking for. On the other hand, you occasionally look at a journal paper and wonder how they actually got those results (sometimes results/experimental details are not well explained) or if they did other related experiments that the reader might be interested in - informal publishing would be good in this respect.

The lack of a recognition structure, plus the lack of feeling that you have published into something with a true impact factor [and in the annals of science]

It would be difficult to publish it in a Journal since it would be alreaqdy in the public domain which means no recognition

Publishing informally online does not provide the same degree of validation or accredidation. I would expect results are highly more likely to be questioned from within an informal publications in comparison to a published journal.

It would be vastly beneficial. Sometimes someone else sees something in a set of results that it might have taken another researcher months to spot, so the more expert opinions, the better.

Publishing information would be quicker, and collaboration between unforeseen groups would become more feasible; as has been seen in the open source software communities, sharing of information leads to success. However, in the current paradigm where private corporations hold the keys to high-impact journals, even where research is paid for by the taxpayer- and publications to those journals are the criteria used for career judgement, researchers have little option but to wait until after the journal publication- or risk being denied since the information has then already been published. Of course, once it has been published, the journal then holds the copyrights to that publication, not the researcher, and this could lead to other IP problems if the lab records were then self-published elsewhere.

Q8. Do you have any other comments on informal, online publishing or on making research data available?

Lots of inaccurate, non-robust data will be published

patenting

no

There must be a balance between research data published informally and formally, one which is decided on by supervisor and research group and finally sanctioned/authenticated by supervisor.

no

I would prefer not too

I've always though that having an electronic lab book would save so much time and would be a much more organised way to record experimental methods and data. Even if the information is only available to the supervisor or research group and parts of it are available publically, it would make sharing information a lot easier. It would also save the need to store hundreds of thick paper lab books and then having to flick through them to find something you did 2 years ago!

It'd be a case of it doesn't matter what I think, as my supervisor has the final say in what does and doesn't get published!

If the purpose of the studies is to make them available online then yes, otherwise it does make sense for the same reasons I mentioned above.

I do agree that making research data is beneficial to many researchers but there should still be a degree of security to protect the researcher who gathered the information so that there would be no question that they would be accredited for the research that they had carried out. There should be the system that basically ensures that the person who wants to access the information must credit the appropriate people should their data be used.

I work jointly between chemistry and the High Voltage Laboratory (part of ECS). There are many more opportunities in Electrical engineering to publish conference papers rather than full journal papers. I have published a number of conference papers in the last two years and as it is generally expected that they are less "polished" results compared to journals, I have found the input of other lecturers and researchers as my research project develops extremely helpful. It seems a shame that very few chemistry conferences publish full conference papers rather than just books of abstracts. I think research can progress much faster with the input of other researchers both in the group and externally.

It is desirable, but not practicable, until there is a paradigm shift in the way that scientific careers are judged.

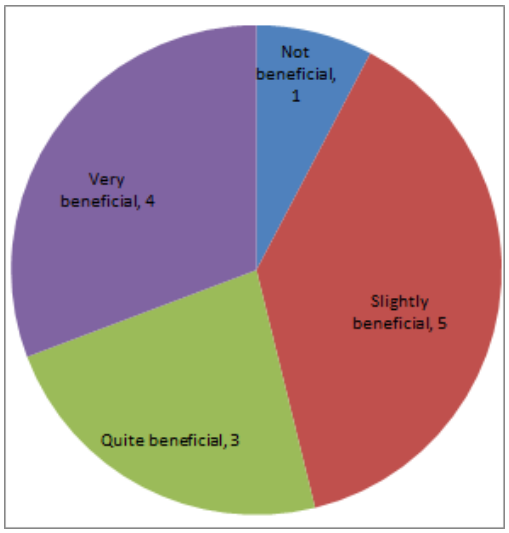

FIGURE A.7: Q9. Increased preparation and thought during the planning stages of an experiment

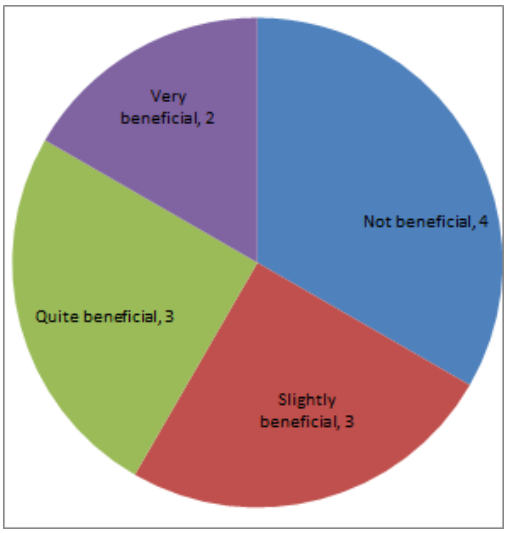

FIGURE A.8: Q10. Online "research passport", a portfolio of all your research activities

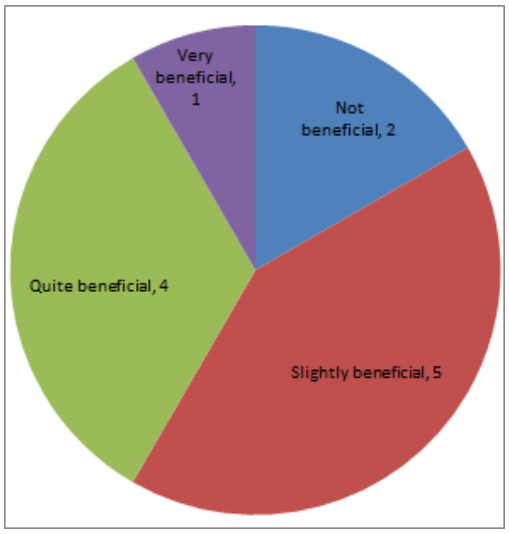

FIGURE A.9: Q11. Easy remote access for your supervisor to monitor your research

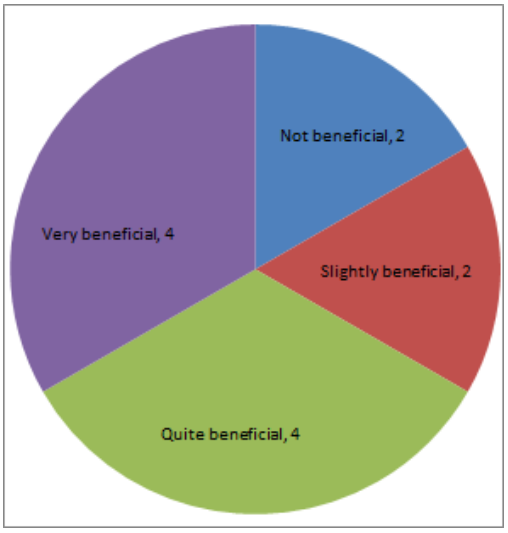

FIGURE A.10: Q12. Automated generation of "experimental" sections of a paper or thesis

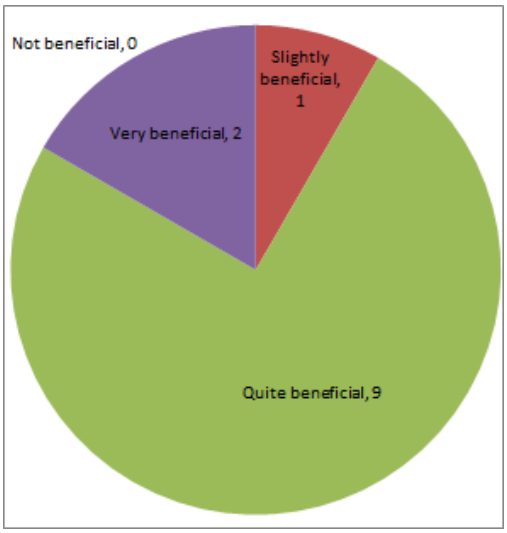

FIGURE A.11: Q13. Reduced possibility of introducing error when recording or transcribing results

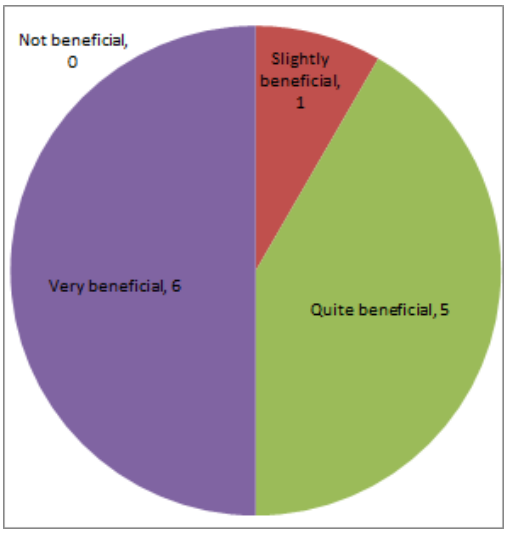

FIGURE A.12: Q14. Greater repeatability of both your experiments and those conducted by others

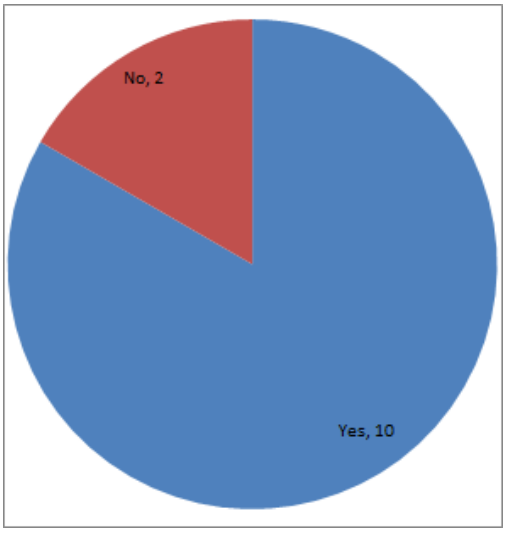

FIGURE A.13: Q15. Would you be willing to post all of your research online (this could be to an audience of your choosing!)?

Q16. If you answered "no" to question 2.2 [Q15], what are your concerns with recording and posting research online?

Ideas which you may decide to return to could have been performed by a 3rd party

someone could copy my results and publish as there own/intellectual thief. This will viod any patent as the information will be in the public domain.

Q17. Do you have any other comments on motivations to recording and publishing research activities?

have to be very careful about what you make available to certain people

Cynically, if there was some benefit as a result of doing so, then it would be something that the scientist and their supervisor would be happy to do - as a means of furthering their image [or looking more hi-tech and in with the times]

I would need to know that my research would be safe and secure online. Only those who I have selected to view my research should be able to see it, at least until my research was complete.

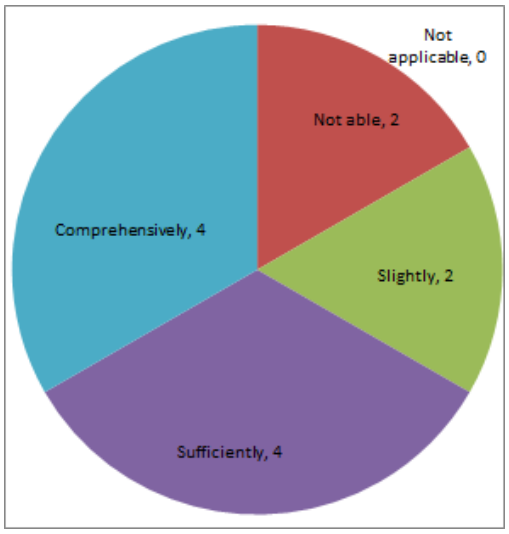

FIGURE A.14: Q18. Ability to record planning

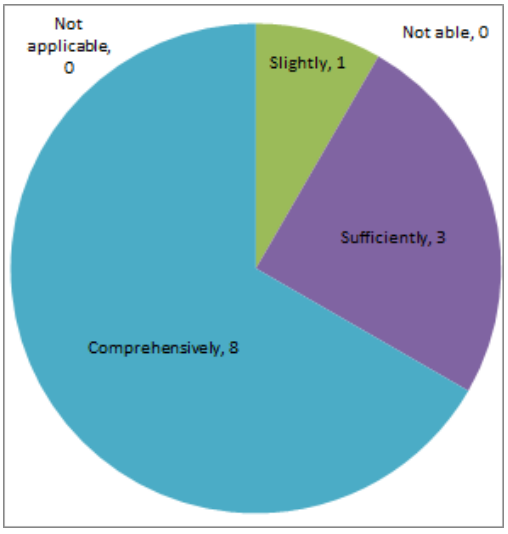

FIGURE A.15: Q19. Ability to record enactment/execution

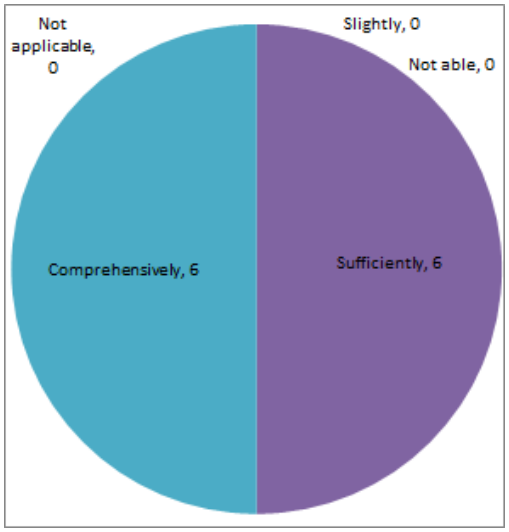

FIGURE A.16: Q20. Ability to record oberservations

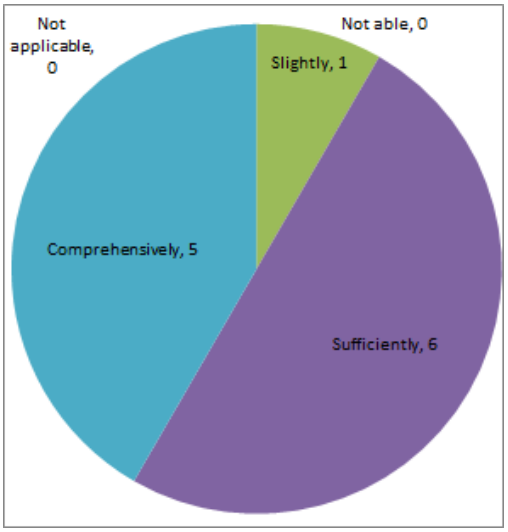

FIGURE A.17: Q21. Ability to record outcomes

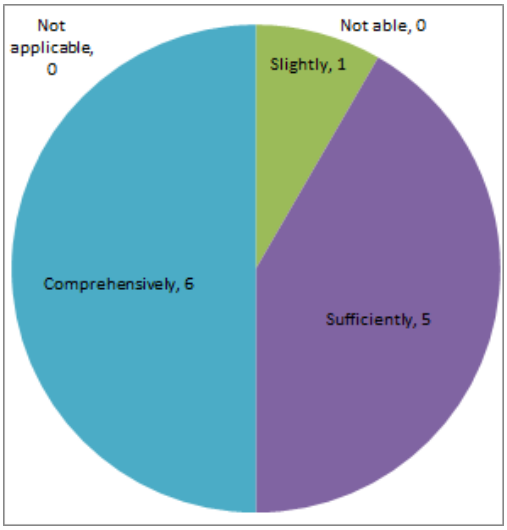

FIGURE A.18: Q22. Ability to record measurements and/or characterisation

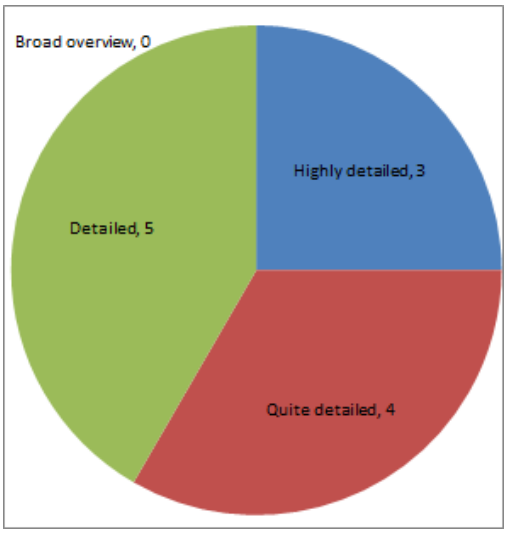

FIGURE A.19: Q23. If you could record everything about an experiment, where would you stop?

Q24. Please explain the reasons for your answer to question 3.2 [Q23]

All what is needed is for repeatability. There is no point working to conditions such as ISO9001 when there are links in the chain beyond our control which are less stringent.

There comes a point of diminishing returns where so much time is spent recording ambient data etc that the point of the experimenmt is missed

There is no harm in having too much information

to be useful, the record needs to be comprehensive but not too extensive, otherwise it becomes hard to look back on.

It is ridiculous to web cam every experiment and as for the rest, experiments would take far too long. it would be very time consuming and produce a pointless amount of information.

If I could record every single detail of an experiment without wasting too much time (if it was mostly automated), then I would. It is often the case a while after the experiment that you want to know something that you didn't record in the experiment.

In computational chemistry, there is little detail that you can describe that isn't already logged in the output files.

technical problems of recording notes in digital format working at the same time in the laboratory under unsafe conditions (toxic chemicals, solvents, etc)

I believe recording everything to the point of it being quite detailed is sufficient. I did not stretch to highly detailed as using bar-codes for test tubes seems a tad extreme however I could imagine that it would be very useful for some industries (e.g forensics). I do like

the idea of using a webcam to record the experiment as this would allow others to see exactly how an experiment was carried out for the sake mimicking or possible seeing not how to do it in some cases, should things go wrong.

I think that bar coded test tubes may be necessary for super-sensitive reactions, but generally I find that with comprehensive cleaning of all equipment and recording procedures and ambient conditions, good repeatable data can be obtained for my research.

The devil is in the detail, the more information you have to analyse, especially with the modern analyses that can be performed with computers (admittedly, however, this is not open to everyone, requires certain skills and cannot be done by hand), unexpected factors are more likely to be caught if more detail is recorded.

Q25. What level of detail do you think is necessary to record an experiment procedure, such that it could be recreated by somebody else in the future?

Basics of procedure, quantities.

The level I answered in Q.2

from an organic perspective quantities, temperatures, time intervals and addition rates

whatever information is required to repeat the work

moderate amount detail - procedure

It does not have to be highly detailed. The simplest case would be just recording the concentrations of chemicals, quantities and other parameters like temperature, pressure, stirring rate, duration etc.

Enough that it is possible for someone to realise where they can alter the procedure to fit their own budget or equipment [so using twice as much half-strength reagent, etc.]

It has to be comprehensive to the other person who will continue or repeat experiment, so all procedures, steps, quantities, etc need to be specified correctly

Either by writing or typing up a step-by-step instruction of how I carried out the experiment. Including details of temperatures, pressures and substances used. Also have an advisory section of methods which potentially did not give the best outcome had the method needed changed at any point and if there were any problems which resulted while carrying out the experiment (as a guideline of what not to do).

I think "quite detailed" would be important, it is better to have too many details about a procedure. A broad procedure seems fine, until a researcher finds they can't reproduce the results, and it isn't possible to work out why.

As much as is possible.

Q26. Do you feel there is a divide between the ease with which digital processes can be recorded as compared to physical experimental processes?

yes

Yes, unless completely automated, digital processes tend to be harder to record at present.

Yes

no

no

Not sure what you're asking here.

It will not be as easy to describe where an experiment goes wrong, without a lot more examination. This is compared to something digital, where all you need to do is use breakpoints and debug tools.

technical problems of recording notes in digital format working at the same time in the laboratory under unsafe conditions (toxic chemicals, solvents, etc)

I think it lies with people's preferences. Everyone has their own way of doing things.

Not particularly, on my placement we all had to record experiments electronically with photos and detailed experimental procedures, this made report writing and results sharing much easier.

Yes, but that is a matter of time invested in arranging such recording of information more than an inherent gap.

Q27. What tools and resources would make recording experiments easier, for both digital and physical processes?

PC workstations in labs.

seamless integration of digital note-taking, this will demand hardware and software advances a few years in the future yet!

ELN

dont think any tools are really required, but certain processes could use very specific tools... eg digital slr cameras recording camera settings with picture is very useful. however in a laboratory setting you could generally note down the relevant info fairly readily.

a net-book perhaps, this would cost too much.

electronic lab book

Probably something to remind you to do such a thing. Also, a way of exacting a file/directory structure that makes it easier to recall past information/output files at a moment's notice. My supervisor has had many a tantrum, because he wanted output data from an obscure experiment I have done a year before, and I spent too long finding it!

tablets with stylus (but this is fragile to chemicals), remote cameras, voice recording, etc

Tools such as live script (pen which records everything you write on the allocated notebook) provide a good bridge between digital and physical processes. A camera to record different parts of experiments would also be useful.

Perhaps a template with each experiment/ page having a unique number or identifier, which could be cross referenced with a paper note book for rough calculations etc when its not possible to type on a laptop. A LaTeX template which is easy to insert photos into would be good, and would produce very professional looking experimental procedures.

Electronic lab notebooks, and recording hardware and software (depending upon the information trying to be recorded.)

Appendix B

Preliminary Interview Schedule

These interviews were ultimately only used on a small sample of participants sourced from the University of Southampton's chemistry department and as such any results are only indicative as opposed to significant.

Early Career Researcher Interview

The purpose of the interview questions is to gain an understanding as to what extent early career researchers in organic chemistry, feel able to present their work and the impact they have made to the field. This is with the aim of investigating possible sources of motivation that will encourage researchers to make contributions to online repository sites, such as ChemSpider Synthetic Pages (http://cssp.chemspider.com/), with the ultimate hope of making more data available to more people, exploiting the long tail nature of the Web. If researchers are to make use of such repositories however, it can be assumed that there will have to be an intrinsic benefit to do so, adding value to the original contribution made. One such way in which this might be achieved is to help the author more easily take advantage of the information they have submitted, allowing it to be presented in more appropriate ways, and be recognised and rewarded for their contribution, not just to the repository, but to the field in general.

In summary, we want to find out:

- *What do researchers choose to present, when trying to provide a representation of themselves as a scholar for a particular discipline?*
- *What research output do researchers find it difficult to represent?*
- *Do researchers feel able to accurately reflect their work and impact in the field?*
- *How do researchers manage their research outputs?*
- *What do researchers think of some of the reasons provided by sites like ChemSpider*

- *Synthetic Pages, to encourage users to deposit their data?*
- *What would incentivise researchers to deposit their data?*

1. Tell me about the academic achievements you are most proud of...
   (a) What achievements are most related to the actual practice of chemistry (e.g. datasets, reactions)?
   (b) What achievements are most related to academia (e.g. posters, papers).
2. Tell me about how you are required to present your achievements...
   (a) For example, would you present your work when applying for a new position?
   (b) Does the work presented change to reflect the position you are applying for?
   (c) Are there any aspects to your work that you find it difficult to present?
3. Tell me about the process of collecting together and compiling your record of achievements...
   (a) Do you have to go to lots of different resources?
   (b) Is it time consuming and do certain items get missed out?
4. Tell me about any steps you may have taken to try and increase your status or visibility in the field...
   (a) Do you feel you could further exploit certain achievements to promote yourself as a competent chemist?
5. Tell me what you think about online repositories such as ChemSpider?
   (a) Are these a useful outlet for expressing one's work?

Support Staff Interview

The purpose of this interview is to establish what motivates researchers to publish their work beyond publishing in high impact journals. Are there other approaches to be recognised for one's contributions and are their new ways to communicate these? What is required for new methods to be accepted?

The first part of the interview will focus on research outputs and notions of reward and recognition.

1. What areas of your work do you not get much recognition for?
2. How well does a journal article represent the content of your thesis?
3. How well does a thesis represent your larger body of work (e.g. your lab notebook)?

4. How do you get the items of work you think are important to work their way up to the level of a journal article?
   (a) Do you choose bits that are relevant and discard other bits?
   (b) Does everything get included at some point?
   (c) When a project comes to an end, are certain elements never seen again?
5. How many of these smaller contributions are there? How difficult would it be to collate them all and make accessible in some form?
   (a) Do you feel there is a value in making small contributions available to the wider community?
6. Imagine you had a portfolio of your work, what would be in there?
7. What has been valuable to your career so far?
   (a) a. Is there anything that could have made your career progress faster or in a more desirable direction?
8. Imagine a world where journal articles do not exist. What would you try to make more publicly accessible and how would you go about it?

The following questions will focus on new means of academic publishing and open access

1. What do you think is more important professionally, being a member of a learned society such as the RSC or cultivating and maintaining an online presence?
2. What do you think of online repository sites such as ChemSpider or Reaxys? Are they a useful outlet for expressing work?
3. How restrictive is the need for confidentiality? Are copyright and IP preventing sites such as ChemSpider Synthetic Pages from being successful?
4. Would the recent developments in OA publishing (RCUK mandates publishing in a GOLD OA journal) make you more inclined to self-publish?
5. Do you feel that journal publishers add sufficient value to journal articles for the price charged?

The final part of the interview focus on the relationship between students and their supervisors, and general knowledge management within the lab

1. Is it generally the case that a supervisor calls the shots? Are students present to do the lab work, whilst supervisors choose what bits they want to publish?

2. How good are electronic lab notebooks at getting out information and how powerful are the tools that are available?

3. How smooth is the paper authoring process? What do the students do? What do supervisors do?

4. Do the approaches to publishing vary among the sub-disciplines of chemistry? (i.e. organic, inorganic, physical, etc.)

Appendix C

Disaggregator Screenshots

The following screenshots demonstrate the various interfaces of the Disaggregator prototype, developed to further explore and evaluate the disaggregation proposal.

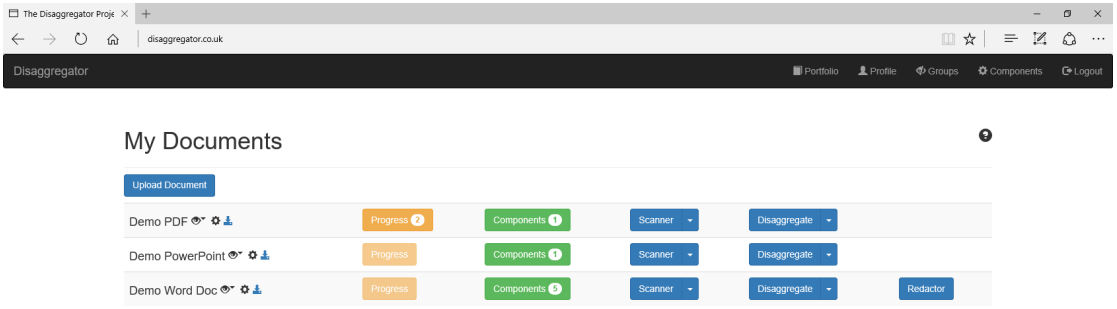

FIGURE C.1: The Disaggregator homepage, showing the documents a user has uploaded and the various actions that can be taken with those documents. Along the top of the screen lies a navigation bar with links to "Portfolio", "Profile", "Groups", "Components" (and "Logout").

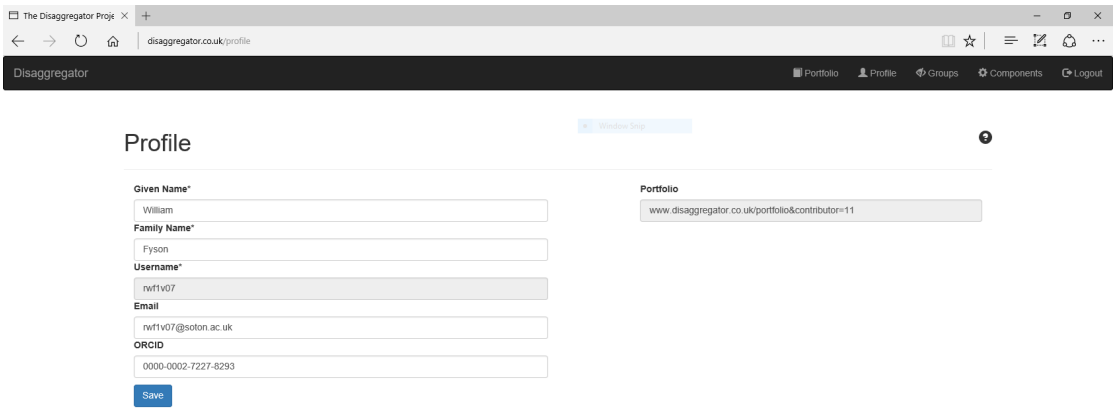

FIGURE C.2: A user's profile page, where they can specify their name and email as well as an ORCID, to allow for a user's contributions to be more easily recorded and queried.

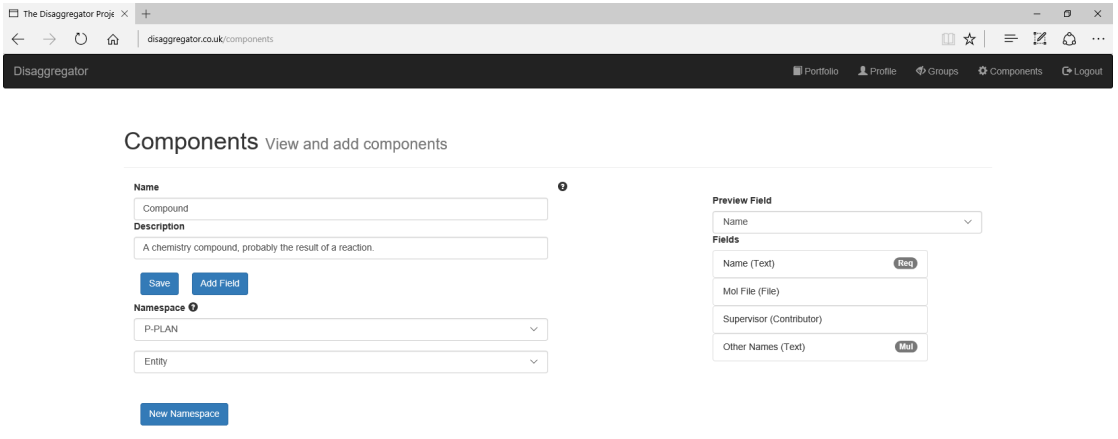

FIGURE C.3: The components interface of the Disaggregator, showing a "compound" component that comprises of the fields "Name", "Mol File", "Supervisor" and "Other Names", allowing a range of data and metadata to be captured about the compound. The user may also add new references to new semantic web namespaces and select a class based from the selected namespace.

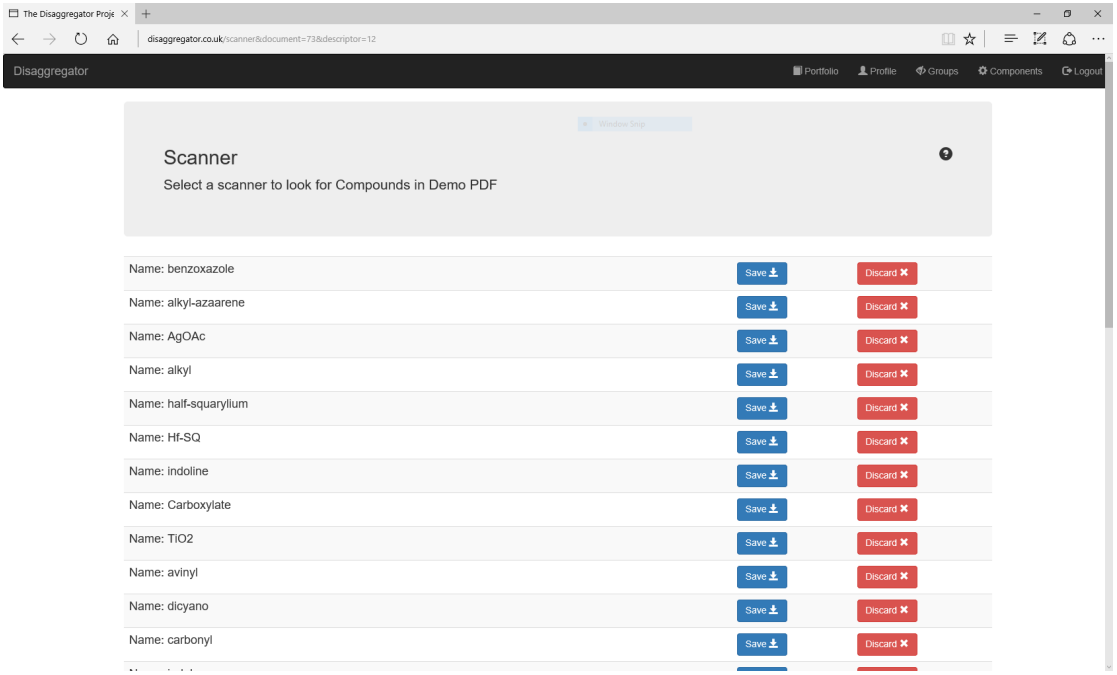

FIGURE C.4: The results screen that is shown after the user scans a document for a particular component. For each result that is returned to the user, they may either save the result as a component that has been extracted from the document (so that they may later add more details or add it to their portfolio, or the user may discard the suggestion if it is irrelevant.

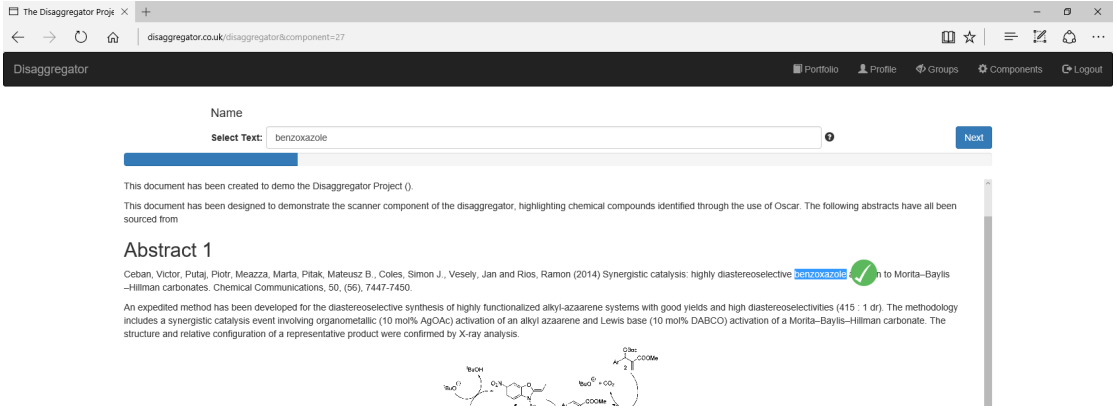

FIGURE C.5: Users may either enter text into the text field at the top of the page to complete the component's field, or they may select text from the document, which is rendered in the browser, by highlighting it.

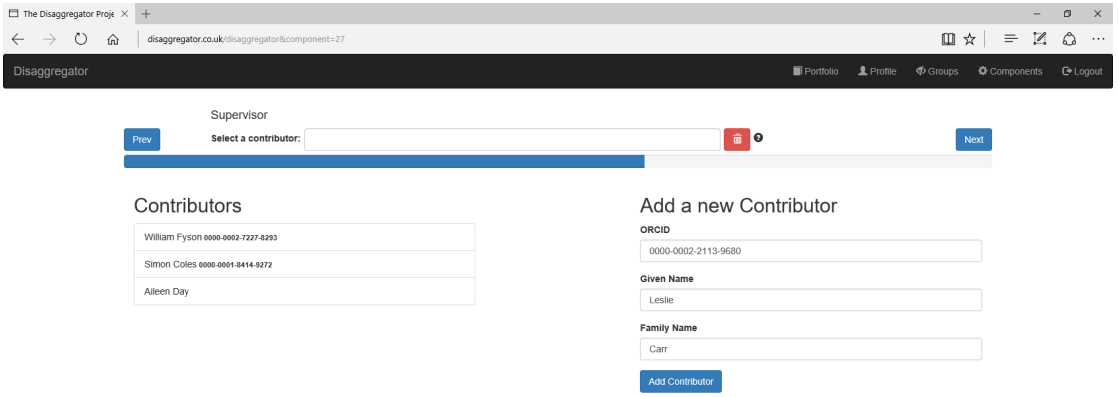

FIGURE C.6: Contributors may also be added to disaggregated components, allowing for metadata to be stored from beyond the contents of the uploaded document. The name of the field used when defining the component may give an indication as to the role of the contributor, with people being added as values for the a contributor field either by selecting an existing member registered in the system (as shown on the left hand side of the screenshot, or by entering their ORCID, allowing their details to be fetched from http://www.orcid.org.

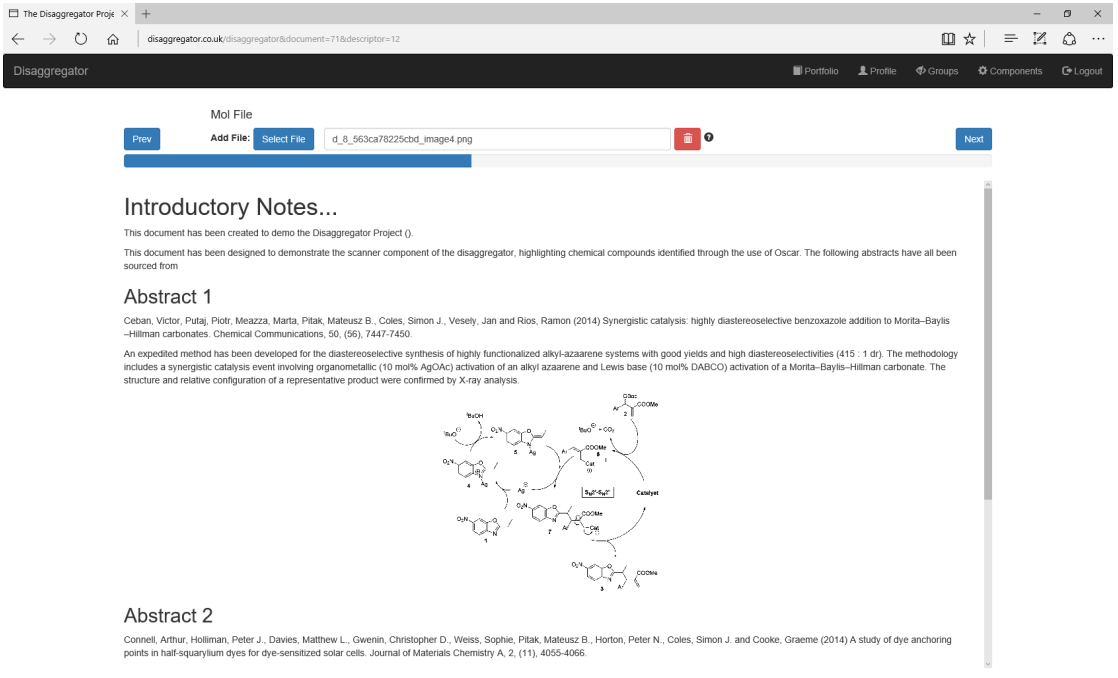

FIGURE C.7: Files may be disintegrated either be selecting an image from the document's contents or by clicking the "Select File" button and uploading a file from the user's device.

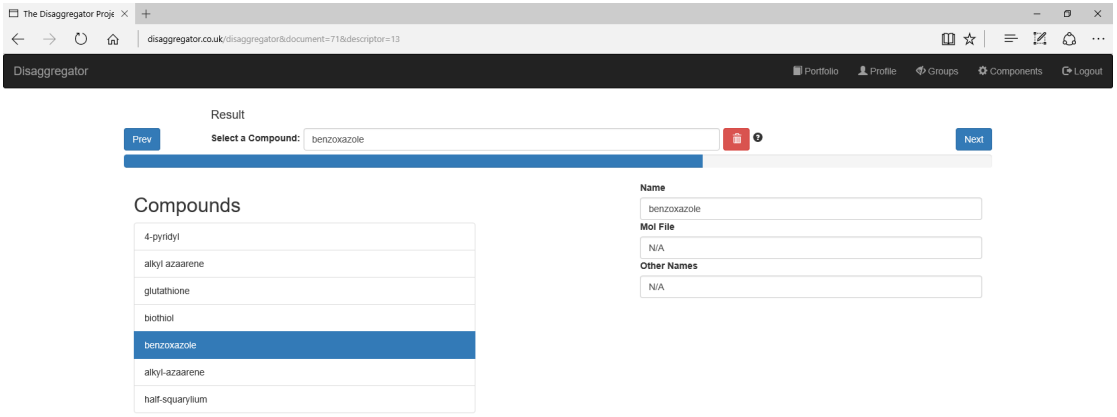

FIGURE C.8: Instances of components that have already been disaggregated from documents may also be added as values for fields defined by components. In this example, when adding a "Reaction" the user is asked to select a compound for the "Result" field, with options being populated from those compounds which already exist within the Disaggregator.

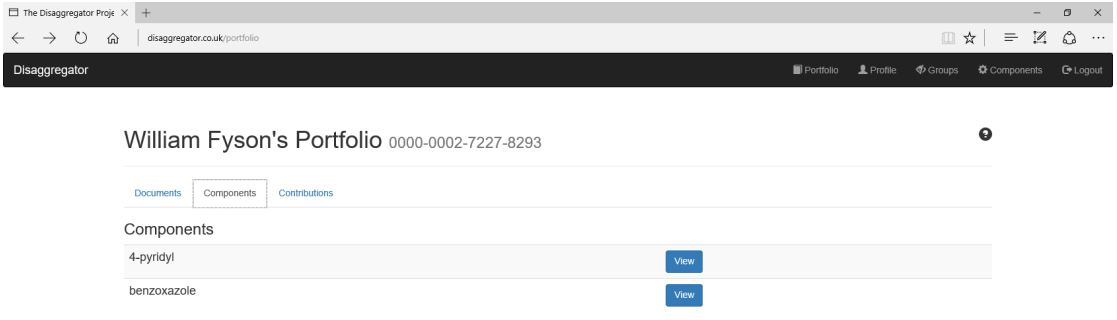

FIGURE C.9: The portfolio presents users with an opportunity to display their research outputs that they have recorded in the Disaggregator. Outputs are categorised into one of three tabs to show either documents, components or contributions. From this page, documents may either be downloaded, or components and contributions may be viewed.

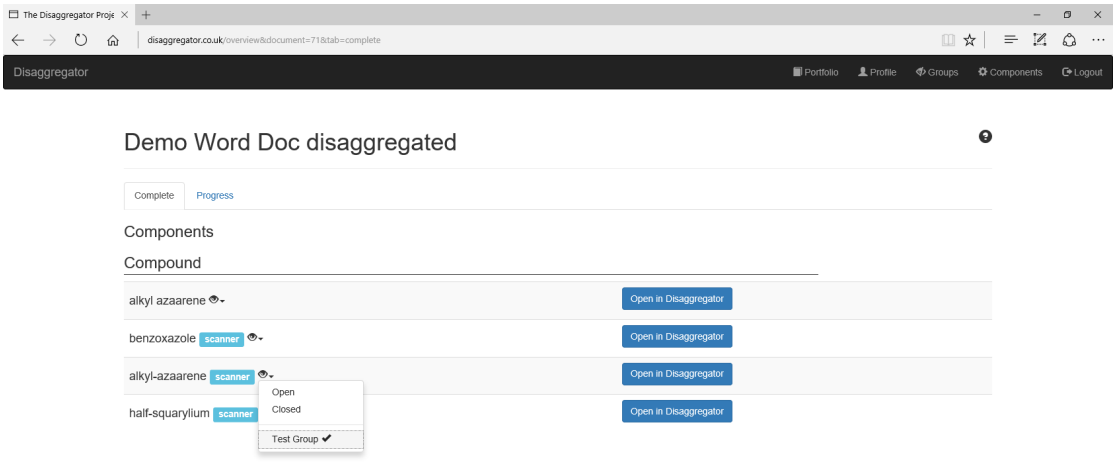

FIGURE C.10: The above interface shows an overview of the components that have been extracted from a given document. By clicking the eye symbol next to each component listed, the user may select who is able to view the component in their portfolio, with the options for permissions including "Open", "Closed" or the name of a specific group (which represents a subset of register users).

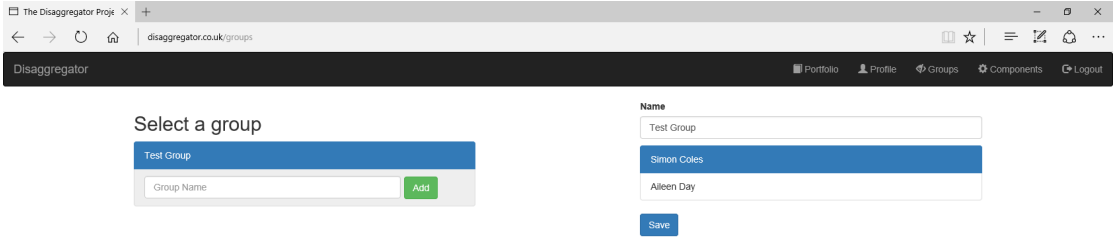

FIGURE C.11: The groups interface allows the user to create or edit groups of users. When a group is selected (on the left hand side of the screen, the system's registered users are listed on the right hand side of the screen and can be selected or deselected based to determine if they should be part of the group and thus able to view the documents and components that have been assigned this particular group.

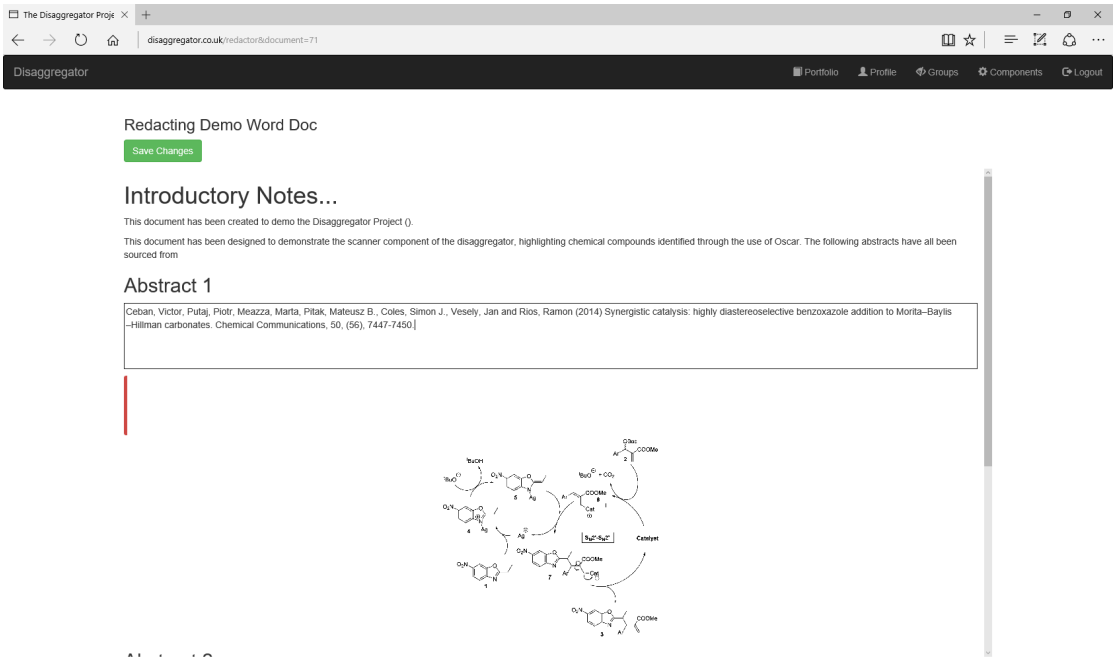

FIGURE C.12: Uploaded Word documents may also be redacted. The above screenshot shows that when a paragraph from a document is clicked a text box appears allowing the paragraph to be edited. Removed paragraphs are highlighted via the red bar on the left hand of the screen which showed where a paragraph was present previously. Once redaction is complete a new version of the document appears in the user's library of documents, which they could then change the visibility permissions for, making the redacted version visible whilst keeping retaining the confidentiality of the original.

Appendix D

Interview Guide

Part 1

1. What is your role within academic publishing?
2. What are the other roles in your immediate network?
   (a) And what are the common interactions between the stakeholders?

   Forgetting everything you understand about conventional academic publishing, imahgine researchd ata could be released at any time, and in any format via the Web.
3. For each of the following functions of academic discourse, consider the following questions:

   Functions:

   - Dissemination of knowledge
   - Assessing quality of work
   - Assessing contributions of researchers
   - Ranking of institutions
   - Allocation of funding
   - Achieving non-academic impact
   - Any other functions?

   Questions:

   (a) How might interactions between stakeholders be adapted to better suit the above functions? How might the interactions described previously be short circuited?
      i. How would this affect your role?

ii. How would this affect researchers/institutions/publishers?

iii. How would if affect the way in which research data is released?

4. To sum up this part of the interview, thinking about the different functions of acadmemic discourse and the interactions between stakeholders, what are the services that need to be provided?

Part 2

One approach to scholarly discourse is the idea of "Disaggreagtion", taking existing publishable items (such as journal artices or conference posters or theses) and breaking them down into smaller chunks to be distributed separately.

1. How would "Disaggregation" affect your role?
2. How would "Disaggregation" affect researchers/institutions/publishers?
3. How would "Disaggregation" affect the different functions of scholarly discourse discussed earlier?
4. What are the different levels of granularity that "Disaggregation" may operate with?

Part 3

A brief demo of the disaggregator tool follows.

1. *Main view shows documents – everything is dervied from documents and from here we can upload new documents and control document visibility. Visibility can be controlled by selecting documents to be visible to all who visit the site, to a particular, pre-defined group of users, or by being closed to all but the depositing user.*
2. *Components: Here we can define the types of things we want to extract from a document. E.g. Spectrum (Type, Image, Compound, Supervisor). Components can have URIs assigned to them to describe their type and fields can be related to components using properties to formally describe the relationship between components and fields.*
3. *Scanner: Use the OSCAR and ChemicalTagger based text-mining mechanism to identify components that can be extracted from the demo documents.*
4. *Manual Disaggregation: Open up a document and fill in the stages. We can disaggreagte from either Word documents, PowerPoint presentations or PDFs. When adding a contributor, we can select from existing contributors recorded in the system, or we can add a contributor using their ORCID.*

5. *Portoflio: This shows the user's visible documents, components extracted from documents and contributions (those components that others have extracted but have listed the user as a contributor). The portoflio page may be accessed via an API, using an ORCID to get the user's details and returning JSON, listing all the components and contributions.*

Part 4

For this next part of the interview I would like to discuss how might be possible to implement some of the features of "Disaggreagtion" in the real world of scholarly discourse. Please note however that the software demonstrated previously is a prototype designed to facilitate discussion and is not intended to be integrated with any real life systems.

1. How would new practises of dissemination sit alongside conventional practises such as publishing journal articles? How might existing workflows need to adapt?
2. What services already exist that the process of "Disaggregation" could be built into?
3. What standards and protocols does the scholarly community require or already possess that would faciltate such an approach?
4. How should "Disaggregation" be presented to researchers for it to be adopted into their workflows?
   (a) What is required to draw researchers to it?
   (b) And how does this contrast with it being imposed upon them?
5. What do you perceive to be the benefits of "Disaggregation"?
6. And what problems may arise from "Disaggregation?"

Part 5

1. I've introduced the idea of breaking down traditional, publishable items into smaller ones for dissemination as "Disaggregation", but are there other terms that may be more appropriate or meaningful?
2. Is disaggregation a one-way process? Is it possible for individual items to be aggregated together to form a bigger item such as a journal article?
3. Is there any that we should have talked about, but have so far failed to mention?

Appendix E

Code Book

No.	Name	Description	Emergent	References
1	Scholarly Discourse	Comments relating to general ideas around scholarly discourse, including notions of disseminating research outputs and the other functions that are byproducts of this process (e.g. allocation of recognition).		21
1.1	Disciplinary Differences	Comments upon how the roles of scholarly discourse and the manner in which it is implemented differ between academic disciplines.	Yes	22
1.2	Economics	The economics of scholarly discourse, how do different stakeholders perceive value in this area.	Yes	11
1.3	Functions	The functions of scholarly discourse.		1

Continued on next page

Table E.1 – continued from previous page

No	Name	Description	Emergent	References
1.3.1	Dissemination	The dissemination of research outputs allowing researchers to build upon the ideas of their peers and forebears.		24
1.3.2	Impact	Comments on ideas concerning the socio-economic impact of research outcomes.	Yes	15
1.3.3	Reputation	The act of recognising the contributions of researchers or their institutions through the research outputs they have produced.		49
1.3.4	Review	The process of reviewing research outputs to assess their validity and originality, typically carried out via peer review. Can also include comments on the need for trust in data that is disseminated and ideas around provenance.		28
1.4	Negative Results	Comments on the idea of negative results - research outputs that didn't go as expected and are often left unpublished.	Yes	3
1.5	Open Access	Cooments on Open Access.	Yes	11

Continued on next page

Table E.1 – continued from previous page

No	Name	Description	Emergent	References
1.6	Problem Areas	Any comments that highlight particular problems that arise in the scholarly discourse and areas that any of the stakeholders may be trying to resolve.	Yes	11
1.7	Requirements	The tools, services and procedures required for scholarly discourse to take place and fulful the various roles required of it.		32
1.8	Reward and Recognition	The notion that researchers are driven to contribute to the overall body of work in a bid to boost their recognition, which in turn will lead to career progression. Typically measured by published articles via the means of impact factors and citations.		16
1.8.1	Citations	The role of citations in scholarly discourse (and beyond).	Yes	7
1.9	Roles	Different roles undertaken by those who partake in scholarly discourse.		0

Continued on next page

Table E.1 – continued from previous page

No	Name	Description	Emergent	References
1.9.1	Funder	The funder, who is responisble for funding researchers and expects to see impactful results.		9
1.9.2	Interactions	Interactions between different stakeholders of scholarly discourse.	Yes	19
1.9.3	Librarian	The role of the librarian, who manages subject resources and journal subscriptions.		10
1.9.4	Publishers	The role of the publisher as stakeholder chiefly responsible for the dissemination of result through the publication of journal articles.		61
1.9.5	Researcher	The role of the researcher in scholarly discourse as someone who produces and consumes research outputs.		24
1.10	The Web	Comments on the role of the Web in scholarly dicourse, both how it is used and what it enables.	Yes	17
2	Sociology of Science	Comments on how science is structured from a social perspective.	Yes	58

Continued on next page

Table E.1 – continued from previous page

No	Name	Description	Emergent	References
2.1	Knowledge Management	Comments that arise concerning issues surrounding knowledge management, both where it is done well and where it is done poorly.	Yes	6
2.2	Motivational Factors	Comments on what drives people to act in the way in which they do.	Yes	36
2.3	Scientific Practice	Comments and observations upon the way in which science is conducted - both in theory and in practice.	Yes	62
3	Behaviour Change	Comments upon how to encourage behaviour change among the research community.		92
3.1	Negative Influences	Negative approaches to encouraging behaviour change (e.g. by mandating new approaches).		18
3.2	Positive Influences	Positive approaches to encouraging behavour change (e.g. helping researchers improve their reputation).		18

Continued on next page

Table E.1 – continued from previous page

No	Name	Description	Emergent	References
4	Chem Informatics	Comments on the field of chem informatics and how it is perceived by those in the wider chemistry community.	Yes	5
5	Disintermediation	Exploring the idea that researchers are responsible for many of the functions of scholarly discourse and that other stakeholders are intermediaries whose role could be shifted through the disruptive influence of the Web.		14
5.1	Conventional Practises	The effect of disintermediation upon conventional scholarly discourse practises.		0
5.2	Potential Benefits	The benefits afforded by disinteremediation.		4
5.3	Potential Problems	The problems that may arise through disintermediation.		5

Continued on next page

Table E.1 – continued from previous page

No	Name	Description	Emergent	References
6	Disaggregation	Exploring the idea of disaggregation, that suggest a journal article or some other conventional scholarly artefact may be broken down into smaller items, each of which could be distributed in their own right, presenting an opportunity to more openly disseminate research outputs in a manner that takes advantage of the affordances of the Web.		26
6.1	Concept	Comments upon the general concept of disaggregation.		0
6.1.1	Granularity	Comments upon what can be disaggregated and the level of granularity.		46
6.1.2	Potential Benefits	Benefits to arise from disaggregation.	Yes	39
6.1.3	Potential Problems	Problems that may arise as a result of a disaggregated approach to scholarly discourse.	Yes	13
6.1.4	Reaggregation	The reverse of disaggregation, whether or not individual items can be put together to form a single, publishable item.		5

Continued on next page

Table E.1 – continued from previous page

No	Name	Description	Emergent	References
6.1.5	Scholarly Artefacts	What are research outcomes, how do they develop through time and what can be disseminated?		16
6.1.6	Terminology	Comments upon the term disaggregation and the insightthis provides with regards to which conventional scholarly artefacts can be considered as aggregated collections of items.		20
6.2	Effects	The effects of disaggregation upon scholarly discourse, its stakeholders and its functions.		0
6.2.1	Dissemination	The effects of disaggregation on the dissemination of research.		18
6.2.2	Reputation	The effect that disaggregation mayhave on a researcher trying to build up their reputation.		16
6.2.3	Review	The effects of disaggregation on the research community's ability to assess the validity of the work of their peers.		14
6.3	Implementation	How should the process of disaggregation be implemented.		83

Continued on next page

Table E.1 – continued from previous page

No	Name	Description	Emergent	References
6.3.1	ELNs	Comments on Electronic Lab Notebooks	Yes	1
6.3.2	Technologies, Services and Standards	Disaggregation's relationship to existing technologies, services and standards.		38
6.3.3	Workflow	The nature in which disaggregation should work alongside existing scholarly workflows.		60
6.4	Previous Work	Examples of where disaggregation, or processes similar to it, already take place or have been experimented with.	Yes	12

Bibliography

Jeffrey C Alexander. Commentary: Structure, Value, Action. *American Sociological Review*, 55(3):339–345, 1990.

David B Allison, Andrew W Brown, Brandon J George, and Katherine A Kaiser. A Tragedy of Errors. *Nature*, 530(7588):27–29, 2016. ISSN 1098-6596.

Larry Altman and Laurie Melcher. Fraud in science. *Environmental science & technology*, 286(6383):2003–2006, March 1983. ISSN 0013-936X.

Altmetric. Altmetric - We Make Article Level Metrics Easy, 2015. http://www.altmetric.com/ [Accessed: 16 April 2015].

Americal Chemical Society. San Diego - Spring 2016 - American Chemical Society, 2016. http://www.acs.org/content/acs/en/meetings/spring-2016.html [Accessed: 18 January 2016].

Alexander Ardichvili, Vaughn Page, and Tim Wentling. Motivation and barriers to participation in virtual knowledge-sharing communities of practice. *Journal of Knowledge Management*, 7(1):64–77, 2003. ISSN 1367-3270.

ArXiv. arXiv monthly submission rates, 2016. https://arxiv.org/stats/monthly_submissions [Accessed: 17 October 2016].

T. K. Attwood, D. B. Kell, P. McDermott, J. Marsh, S. R. Pettifer, and D. Thorne. Utopia documents: Linking scholarly literature with research data. *Bioinformatics*, 26(18):568–574, 2010. ISSN 13674803.

Robert M Axelrod. *The evolution of cooperation*. Basic Books, New York, 1984.

Monya Baker. 1,500 scientists lift the lid on reproducibility. *Nature*, 533(7604):452–454, May 2016. ISSN 0028-0836.

K. M. Bartol and a. Srivastava. Encouraging Knowledge Sharing: The Role of Organizational Reward Systems. *Journal of Leadership & Organizational Studies*, 9(1):64–76, January 2002. ISSN 1071-7919.

Jeffrey Beall. Predatory publishing is just one of the consequences of gold open access. *Learned Publishing*, 26(2):79–84, 2013. ISSN 09531513.

Sean Bechhofer, Iain Buchan, David De Roure, Paolo Missier, John Ainsworth, Jiten Bhagat, Philip Couch, Don Cruickshank, Mark Delderfield, Ian Dunlop, Matthew Gamble, Danius Michaelides, Stuart Owen, David Newman, Shoaib Sufi, and Carole Goble. Why linked data is not enough for scientists. *Future Generation Computer Systems*, 29(2):599–611, February 2011. ISSN 0167739X.

Sean Bechhofer, David De Roure, Matthew Gamble, Carole Goble, and Iain Buchan. Research Objects: Towards Exchange and Reuse of Digital Knowledge. In *The Futue of the Web for Collaborative Science*, Raleigh, USA, February 2010.

Joseph Ben-David and Teresa A. Sullivan. Sociology of Science. *Annual Review of Sociology*, 1:203–222, 1975.

Monica Berger and Jill Cirasella. Beyond Beall's List: Better understanding predatory publishers. *College & Research Libraries News*, 76(3):132–135, 2015. ISSN 00990086.

Theodore C Bergstrom, Paul N Courant, R Preston McAfee, and Michael A Williams. Evaluating big deal journal bundles. *Proceedings of the National Academy of Sciences*, 111(26):9425–9430, July 2014. ISSN 0027-8424.

Birgitta Bergvall-Kåreborn and Debra Howcroft. The Apple business model: Crowdsourcing mobile applications. *Accounting Forum*, 37(4):280–289, 2013. ISSN 01559982.

Andrés Bernal and Edgar E. Daza. On the epistemological and ontological status of chemical relations. *HYLE, International Journal for Philosophy of Chemistry*, 16(2): 80–103, 2010. ISSN 14335158.

Tim Berners-Lee. Information Management: A Proposal, 1989.

Tim Berners Lee, Dave De Roure, Stevan Harnad, Nigel Shadbolt, Derek Law, Peter Murray-Rust, Charles Oppenheim, and Yorick Wilks. Journal publishing and author self-archiving: Peaceful Co-Existence and Fruitful Collaboration, 2005. http://eprints.soton.ac.uk/261160/.

Tim Berners-Lee and Nigel Shadbolt. There's gold to be mined from all our data, December . *The Times*, http://www.thetimes.co.uk/tto/opinion/columnists/article3272618.ece.

Rachel Bernstein. Updated: Sexist peer review elicits furious Twitter response, PLOS apology, 2015. *Science Insider*, http://news.sciencemag.org/scientific-community/2015/04/sexist-peer-review-elicits-furious-twitter-response [Accessed: 30 April 2015].

Holly M. Bik and Miriam C. Goldstein. An Introduction to Social Media for Scientists. *PLoS Biology*, 11(4):1–8, 2013. ISSN 15449173.

Bo-christer Björk, Mikael Laakso, Patrik Welling, and Patrik Paetau. Anatomy of Green Open Access. *Journal of the Association for Information Science and Technology*, 65 (2):237–250, 2014.

Bo-Christer Björk and David Solomon. Open access versus subscription journals: a comparison of scientific impact. *BMC Medicine*, 10(73):1–10, 2012. ISSN 1741-7015.

Bo-Christer Björk, Patrik Welling, Mikael Laakso, Peter Majlender, Turid Hedlund, and Guðni Guðnason. Open Access to the Scientific Journal Literature: Situation 2009. *PLoS ONE*, 5(6), June 2010. ISSN 1932-6203.

John Bohannon. Who's Afraid of Peer Review? *Science*, 342(6154):60–65, October 2013. ISSN 0036-8075.

John Bohannon. The frustrated science student behind Sci-Hub. *Science*, 352(6285): 511, 2016. ISSN 0036-8075.

Rick Bonney, Caren B. Cooper, Janis Dickinson, Steve Kelling, Tina Phillips, Kenneth V. Rosenberg, and Jennifer Shirk. Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy. *BioScience*, 59(11):977–984, 2009. ISSN 0006-3568.

Mark Borkum, Carl Lagoze, Jeremy Frey, and Simon Coles. A Semantic eScience Platform for Chemistry. In *2010 IEEE Sixth International Conference on e-Science*, pages 316–323, Brisbane, AU, 2010. IEEE.

Nathan Bos, Ann Zimmerman, Judith Olson, Jude Yew, Jason Yerkie, Erik Dahl, and Gary Olson. From Shared Databases to Communities of Practice: A Taxonomy of Collaboratories. *Journal of Computer-Mediated Communication*, 12(2):652–672, January 2007. ISSN 10836101.

Raymond Boudon. What Middle-Range Theories Are. *Contemporary Sociology*, 20(4): 519–522, 1991.

Richard E Boyatzis. *Thematic Anaysis and Code Development Transforming Qualitative Information.* Sage Publications, Inc, London, 1998.

Robert Boyle. A Further Account of Mr. Boyle 's Experimental History of Cold. *Philosophical Transactions*, 1:46–52, 1665.

Axel Bruns. *Blogs, Wikipedia, Second Life, and Beyond: From Production to Produsage.* Peter Lang Publishing Inc, New York, 2008.

Alan Bryman. *Social Research Methods.* Oxford University Press, Oxford, fourth edition, 2012. ISBN 978-0-19-958805-3.

Peter J Buckley and Mark Casson. *The future of the multinational enterprise.* Macmillan London, London, 1976.

Robert Burgess. Review of the implementation of the RCUK Policy on Open Access. Technical report, Research Councils UK, Swindon, UK, 2015. http://www.rcuk.ac.uk/RCUK-prod/assets/documents/documents/Openaccessreport.pdf.

John C Burnham. The Evolution of Editorial Peer Review. *JAMA: The Journal of the American Medical Association*, 263(10):1323–11329, March 1990. ISSN 0098-7484.

Lara Burns, Nicki Dennis, Deborah Kahn, and Bill Town. Advocacy To Benefit From Changes: Scholarly Communications Discipline-Based Advocacy. Technical report, JISC Scholarly Communications Group. http://www.jisc.ac.uk/media/documents/aboutus/workinggroups/scadvocacyfinal%20report.pdf.

Business Innovation and Skills Committee. Government mistaken in focusing on Gold as route to full open access, 2013. http://www.parliament.uk/business/committees/committees-a-z/commons-select/business-innovation-and-skills/news/on-publ-open-access/ [Accessed: 10 September 2013].

D Butler. The writing is on the web for science journals in print. *Nature*, 397(6716): 195–200, January 1999. ISSN 0028-0836.

Declan Butler. Electronic notebooks: a new leaf. *Nature*, 436(7047):20–21, 2005. ISSN 1476-4687.

A. Cabrera and Elizabeth F Cabrera. Knowledge-Sharing Dilemmas. *Organization Studies*, 23(5):687–710, September 2002. ISSN 0170-8406.

Ewen Callaway. Beat it, impact factor! Publishing elite turns against controversial metric. *Nature*, 535(7611):210–211, July 2016. ISSN 0028-0836.

J M Carroll and P A Swatman. Structured-case : a methodological framework for building theory in information systems research. *European Journal of Information Systems*, 9:235–242, 2000.

Arturo Casadevall and Ferric C. Fang. Causes for the Persistence of Impact Factor Mania. *mBio*, 5(2):e00064–14, March 2014. ISSN 2150-7511.

Centre pour la Communication Scientifique Directe. Episciences.org, 2013. http://episciences.org/ [Accessed: 19 March 2015].

Leslie Chan. Supporting and Enhancing Scholarship in the Digital Age: The Role of Open Access Institutional Repository. *Canadian Journal of Communication*, 29(3): 277–300, 2004. ISSN 1499-6642.

Leslie Chan, Darius Cuplinskas, Fred Eisen, Yana Genova, Jean-Claude Guédon, Melissa Hagemann, Stevan Harnad, Rick Johnson, Rima Kupryte, Manfredi La Manna, István Rév, Monika Segbert, Sidnei de Souza, Peter Suber, and Jan Velterop. Budapest Open Access Initiative, 2002. http://www.opensocietyfoundations.org/openaccess/read.

Simon J. Coles, Jeremy G. Frey, Michel B. Hursthouse, Mark E. Light, Andrew J. Milsted, Leslie A. Carr, David DeRoure, Christopher J. Gutteridge, Hugo R. Mills, Ken E. Meacham, Michael Surridge, Elizabeth Lyon, Rachel Heery, Monica Duke, and Michael Day. An E-Science Environment for Service Crystallography - from Submission to Dissemination. *Journal of Chemical Information and Modeling*, 46(3): 1006–1016, 2006. ISSN 15499596.

Simon J. Coles and Philip A. Gale. Changing and challenging times for service crystallography. *Chem. Sci.*, 3(3):683–689, 2012. ISSN 2041-6520.

H. M. Collins. The Possibilities of Science Policy. *Social Studies of Science*, 15(3): 554–558, 1985.

European Commission. Horizon 2020 sections, 2016. https://ec.europa.eu/programmes/horizon2020/en/h2020-sections [Accessed: 11 October 2016].

Structural Genomics Consortium. Pioneering Science to Inspire Pioneering Medicines, 2016. http://www.thesgc.org/about/what_is_the_sgc [Accessed: 6 July 2016].

Zoe Corbyn. RAE table will be shaken by use of journal rankings, 2008. *Times Higher Education*, http://www.timeshighereducation.co.uk/news/rae-table-will-be-shaken-by-use-of-journal-rankings/403502.article [Accessed: 23 April 2015].

Cornell University Library. Nondisclosure Clauses, 2015. https://www.library.cornell.edu/about/inside/policies/nondisclosure [Accessed: 29 May 2015].

Council of the European Union. Council Conclusions on the Transition Towards an Open Science System. Technical report, Council of the European Union, Brussels, BE, 2016. http://data.consilium.europa.eu/doc/document/ST-9526-2016-INIT/en/pdf.

Iain D. Craig, Andrew M. Plume, Marie E. McVeigh, James Pringle, and Mayur Amin. Do open access articles have greater citation impact? A critical review of the literature. *Journal of Informetrics*, 1(3):239–248, 2007. ISSN 17511577.

Ryam Crow. The Case for Institutional Repositories: A SPARC Position Paper. *Association of Research Libraries Bimonthly Report*, 223, 2002.

Suenje Dallmeier-Tiessen, Robert Darby, Bettina Goerner, Jenni Hyppoelae, Peter Igo-Kemenes, Deborah Kahn, Simon Lambert, Anja Lengenfelder, Chris Leonard, Salvatore Mele, Malgorzata Nowicka, Panayiota Polydoratou, David Ross, Sergio Ruiz-Perez, Ralf Schimmer, Mark Swaisland, and Wim van der Stelt. Highlights from the SOAP project survey. What Scientists Think about Open Access Publishing, 2011. http://arxiv.org/abs/1101.5260 [Accessed: 16 May 2012].

Paul J. Darwen and Xin Yao. Co-evolution in iterated prisoner's dilemma with intermediate levels of cooperation: Application to missile defense. *International Journal*

of Computational Intelligence and Applications, 2(1):83–107, March 2002. ISSN 1469-0268.

Department for Business Innovation & Skills. Allocation of science and research funding 2011/12 to 2014/15. Technical report, 2010. https://www.gov.uk/government/publications/allocation-of-science-and-research-funding-2011-12-to-2014-15.

Department for Business Innovation & Skills. Government to open up publicly funded research, 2012. https://www.gov.uk/government/news/government-to-open-up-publicly-funded-research.

Peter Doshi, Kay Dickersin, David Healy, S Swaroop Vedula, and Tom Jefferson. Restoring invisible and abandoned trials: a call for people to publish the findings. *BMJ*, 346, June 2013. ISSN 1756-1833.

M du Plessis. Drivers of knowledge management in the corporate environment. *International Journal of Information Management*, 25(3):193–202, June 2005. ISSN 02684012.

Patrick Dunleavy. The Future of Joined-up Public Services. Technical report, 2020 Public Services Trust, London, 2010. http://eprints.lse.ac.uk/28373/1/The_Future_of_Joined_Up_Public_Services.pdf.

Economic and Social Research Council. ESRC Research Data Policy. Technical report, ESRC, 2015.

Leo Egghe. The Hirsch Index and Related Impact Measures. *Annual Review of Information Science and Technology*, 44(1):65–114, 2010.

Elizabeth Eisenstein. *The Printing Press as an Agent of Change*. Cambridge University Press, Cambridge, 1980. ISBN 9780521299558.

Elsevier. Your Guide to Publishing Open Access with Elsevier. Technical report, Elsevier. https://www.elsevier.com/__data/assets/pdf_file/0003/78492/openaccessbooklet.pdf, year = 2015.

EPSRC. UK University Facilities and Equipment Open Data, 2016. http://equipment.data.ac.uk/ [Accessed: 30 June 2016].

Gunther Eysenbach. From intermediation to disintermediation and apomediation: new models for conusmers to access and assess the credibility of health information in the age of Web2.0. *Studies in Health Technology and Informatics*, 129(1):162–166, 2007.

F1000Research. F1000Research - An innovative open access publishing platform offering immediate publication and open peer review., 2016. https://f1000research.com/ [Accessed: 24 October 2016].

Daniele Fanelli. Negative results are disappearing from most disciplines and countries. *Scientometrics*, 90(3):891–904, 2011.

Janet Finch. Accessibility, sustainability, excellence: how to expand access to research publications. Technical report, Research Information Network, 2012. https://www.acu.ac.uk/research-information-network/finch-report-final.

Annie Franco, Neil Malhotra, and Gabor Simonovits. Publication bias in the social sciences: Unlocking the file drawer. *Science*, 345(6203):1502–1505, 2014.

Richard Fyson. *Uncensoring Open Science? Addressing the negative psychology of publishing "undesired" results.* Masters dissertation, University of Southampton, 2011.

Richard Fyson. What is ORCID?, 2015. *Software Sustainability Institute Blog*, http://www.software.ac.uk/blog/2015-08-24-what-orcid [Accessed: 22 October 2015].

Richard Fyson. GitHub - Disaggregator2, 2016. https://github.com/wfyson/Disaggregator2.

Richard Fyson, Simon J Coles, and Leslie Carr. Dissemination through disintermediation. In *Digital Futures 2012 The Third Annual Digital Economy All Hands Conference*, pages 1–2, Aberdeen, UK, 2012.

Richard William Fyson, Simon Coles, and Les Carr. AltOA: a framework for dissemination through disintermediation. In *Proceedings of the 5th Annual ACM Web Science Conference on - WebSci '13*, pages 79–88, Paris, France, 2013. ACM Press. ISBN 9781450318891.

Rita Gardner. Open access and learned societies. In Nigel Vincent and Chris Wickham, editors, *Debating Open Access*, pages 13–29. British Academy for the humanities and social sciences, 2013.

Yassine Gargouri, Vincent Larivière, Yves Gingras, and Stevan Harnad. Green and Gold Open Access Percentages and Growth, by Discipline. *Proceedings of 17th International Conference on Science and Technology Indicators*, 1:285–292, 2012.

Susan Gibbons. Benefits of an institutional repository. *Library Technology Reports*, 40 (4):11–16, 2009. ISSN 00242586.

P Gill, K Stewart, E Treasure, and B Chadwick. Methods of data collection in qualitative research: interviews and focus groups. *British Dental Journal*, 204(6):291–295, 2008. ISSN 1476-5373.

Carole Goble and David De Roure. myExperiment: Social Networking for Workflow-using e-Scientists. In *Proceedings of the 2nd workshop on Workflows in support of large-scale science*, pages 1–2. ACM, 2007. ISBN 9781595937155.

Carole Goble, Chris Wroe, and Robert Stevens. The myGrid project: services, architecture and demonstrator. In *Proceedings of UK e-Science All Hands Meeting 2003*, pages 595–602, 2003.

Carole a Goble, Jiten Bhagat, Sergejs Aleksejevs, Don Cruickshank, Danius Michaelides, David Newman, Mark Borkum, Sean Bechhofer, Marco Roos, Peter Li, and David De Roure. myExperiment: a repository and social network for the sharing of bioinformatics workflows. *Nucleic Acids Research*, 38(2):W677–W682, July 2010. ISSN 0305-1048.

Tim Gowers. The Cost of Knowledge, 2012. http://thecostofknowledge.com/ [Accessed: 9 April 2012].

Tim Gowers. Why I've also joined the good guys, 2013. *Gowers's Weblog*, http://thecostofknowledge.com/ [Accessed: 1 March 2013].

Tim Gowers. Elsevier journals - some facts, 2014. *Gowers's Weblog*, http://thecostofknowledge.com/ [Accessed: 19 June 2016].

Stephen Grace, Michael Whitton, Sara Gould, and Rachael Kotarski. Mapping the UK thesis landscape: Phase 1 project report for Unlocking Thesis Data. Technical report, University of East London, London, 2015. http://roar.uel.ac.uk/4307/.

Mark Granovetter. Threshold Models of Collective Behavior. *American Journal of Sociology*, 83(6):1420–1443, 1978.

Colin R. Groom and Frank H. Allen. The Cambridge Structural Database in Retrospect and Prospect. *Angewandte Chemie International Edition*, 53(3):662–671, January 2014. ISSN 14337851.

Paul Groth, Andrew Gibson, and Jan Velterop. The anatomy of a nanopublication. *Information Services and Use*, 30(1-2):51–56, 2010. ISSN 01675265.

Feng Gu and Gunilla Widén-Wulff. Scholarly communication and possible changes in the context of social media. *The Electronic Library*, 29(6):762–776, 2011. ISSN 0264-0473.

Michael B Gurstein. Open data: Empowering the empowered or effective data use for everyone? *First Monday*, 16(2), January 2011. ISSN 13960466.

Bernard H. Gustin. Charisma, Recognition, and the Motivation of Scientists. *American Journal of Sociology*, 78(5):1119–1134, 1973.

Laurel L. Haak, Martin Fenner, Laura Paglione, Ed Pentz, and Howard Ratner. ORCID: A system to uniquely identify researchers. *Learned Publishing*, 25(4):259–264, 2012. ISSN 09531513.

Susan Halford, Catherine Pope, and Leslie Carr. A Manifesto for Web Science. In *WebSci10: Extending the Frontiers of Society On-Line*, pages 1–6, Raleigh, US, 2010.

Carol Hansen Montgomery and Donald King. Comparing Library and User Related Costs of Print and Electronic Journal Collections. *D-Lib Magazine*, 8(10), 2002.

Stevan Harnad. Scholarly skywriting and the prepublication continuum of scientific inquiry. *Psychological Science*, 1(6):342–344, 1990.

Stevan Harnad. The Subversive Proposal. In Ann Okerson and O'Donnell James, editors, *Scholarly Journals at the Crossroads: A Subversive Proposal for Electronic Publishing*, volume 08540, pages 13–14. Office of Scientific & Academic Publishing, Washington, DC, 1995. ISBN 0918006260.

Stevan Harnad. On "Overlay Journals", "Epijournals" and "Diamond OA", 2013. `http://openaccess.eprints.org/index.php?/archives/974-On-Overlay-Journals,-Epijournals-and-Diamond-OA.html` [Accessed: 1 March 2013].

Stevan Harnad and Tim Brody. Comparing the Impact of Open Access (OA) vs. Non-OA Articles in the Same Journals. *D-Lib Magazine*, 10(6), 2004.

Stevan Harnad, Tim Brody, François Vallières, Les Carr, Steve Hitchcock, Yves Gingras, Charles Oppenheim, Chawki Hajjem, and Eberhard R. Hilf. The Access/Impact Problem and the Green and Gold Roads to Open Access. *Serials Review*, 30(4):310–314, 2004. ISSN 00987913.

Stevan Harnad, Tim Brody, François Vallières, Les Carr, Steve Hitchcock, Yves Gingras, Charles Oppenheim, Chawki Hajjem, and Eberhard R. Hilf. The Access/Impact Problem and the Green and Gold Roads to Open Access: An Update. *Serials Review*, 34(1):36–40, 2008. ISSN 00987913.

Steven Harnad. Post-Gutenberg Galaxy: The Fourth Revolution in the Means of Production of Knowledge. *Public-Access Computer Systems Review*, 2(1):39–53, 1991.

Harvard University. Faculty Advisory Council Memorandum on Journal Pricing, 2012. `http://isites.harvard.edu/icb/icb.do?keyword=k77982&tabgroupid=icb.tabgroup143448` [Accessed: 7 July 2012].

Susanne Hauschild, Thomas Licht, and Wolfram Stein. Creating a Knowledge Culture. *The McKinsey Quarterly*, 74, 2001.

Lezan Hawizy, David M. Jessop, Nico Adams, and Peter Murray-Rust. ChemicalTagger: A tool for semantic text-mining in chemistry. *Journal of Cheminformatics*, 3(1), 2011. ISSN 1758-2946.

Mark Henderson. Problems with peer review. *BMJ*, 340(c1409):1–4, March 2010. ISSN 0959-8138.

Alfred Hermida and Neil Thurman. A Clash of Cultures. *Journalism Practice*, 2(3): 343–356, 2008. ISSN 1751-2786.

A J G Hey and A E Trefethen. The Data Deluge: An e-Science Perspective. In F Berman, G C Fox, and A J G Hey, editors, *Grid Computing - Making the Global Infrastructure a Reality*, chapter 36, pages 809–824. Wiley and Sons, 2003. ISBN 0470853190.

Tony Hey. Cyberinfrastructure for e-Science. *Science*, 308(5723):817–821, May 2005. ISSN 0036-8075.

Tony Hey, Stewart Tansley, and Kristin Tolle, editors. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research, Redmond, Washington, 2009. ISBN 9780982544204.

Higher Education Funding Council for England. How we fund research, 2016. http://www.hefce.ac.uk/rsrch/funding/ [Accessed: 11 October 2016].

Higher Education Funding Council for England, Scottish Funding Council, Higher Education Funding Council for Wales, and Department for Employment and Learning. Policy for open access in the post-2014 Research Excellence Framework. Technical report, Higher Education Funding Council for England, London, 2015. http://www.hefce.ac.uk/pubs/year/2014/201407/#d.en.86771nhttp://www.hefce.ac.uk/media/hefce/content/pubs/2014/201407/HEFCE2014_07.pdf.

J E Hirsch. An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102(46):16569–16572, November 2005. ISSN 0027-8424.

Brant Houston. The future of investigative journalism. *Daedalus*, 139(2):45–56, April 2010. ISSN 0011-5266.

Gareth Hughes, Hugo Mills, David De Roure, Jeremy G Frey, Luc Moreau, m. c. schraefel, Graham Smith, and Ed Zaluska. The semantic smart laboratory: a system for supporting the chemical eScientist. *Organic & Biomolecular Chemistry*, 2(22):3284, 2004. ISSN 1477-0520.

Jane Hunter. Post-publication peer review: opening up scientific conversation. *Frontiers in computational neuroscience*, 6(63), 2012. ISSN 1662-5188.

David Hutchens. Trust and the Prisoners Dilemma, 2007. http://www.davidhutchens.com/BizWriting/articles/trustandthepriso.html [Accessed: 20 February 2012].

Impactstory. Impactstory: Share the full story of your research impact., 2015. https://impactstory.org/ [Accessed: 16 April 2015].

David M. Jessop, Sam E. Adams, Egon L. Willighagen, Lezan Hawizy, and Peter Murray-Rust. OSCAR4: A flexible architecture for chemical textmining. *Journal of Cheminformatics*, 3(41):1–12, 2011. ISSN 17582946.

Alok Jha. Academic spring: how an angry maths blog sparked a scientific revolution, 2012. http://www.theguardian.com/science/2012/apr/09/frustrated-blogpost-boycott-scientific-journals [Accessed: 9 April 2012].

Paul Jump. Open access will cause problems for learned societies' journals, accepts Finch, 2013. http://www.timeshighereducation.co.uk/422395.article [Accessed: 23 August 2013].

Ralph Katz and Thomas J. Allen. Investigating the Not Invented Here (NIH) syndrome: A look at the performance, tenure, and communication patterns of 50 R & D Project Groups. *R&D Management*, 12(1):7–20, January 1982. ISSN 0033-6807.

Inge Kaul, Isabelle Grunberg, and Marc a. Stern. Knowledge as a global public good. In Inge Kaul, Isabelle Grunberg, and Marc A. Stern, editors, *Global Public Goods: International Cooperation in the 21st Century*, pages 308–325. Oxford University Press, New York, 1999. ISBN 0195130510.

Clint D Kelly and Michael D Jennions. The h index and career assessment by numbers. *Trends in Ecology & Evolution*, 21(4):167–170, April 2006. ISSN 01695347.

Key Perspective Ltd. Data dimensions: disciplinary differences in research data sharing, reuse and long term viability. Technical report, Digital Curation Centre, 2010.

Jihyun Kim. Finding documents in a digital institutional repository: DSpace and Eprints. *Proceedings of the American Society for Information Science and Technology*, 42(1), October 2006. ISSN 00447870.

King's College London and Digital Science. The nature, scale and beneficiaries of research impact. Technical report, Higher Education Funding Council of England, London, 2015. http://www.hefce.ac.uk/media/HEFCE,2014/Content/Pubs/Independentresearch/2015/Analysis,of,REF,impact/Analysis_of_REF_impact.pdf.

Nancy Kingsbury and John Scanzoni. Structural-Functionalism. In Pauline Boss, William J Doherty, Ralph LaRossa, Walter R Schumm, and Suzanne K Steinmetz, editors, *Sourcebook of Family Theories and Methods*, pages 195–221. Springer US, New York, 1st edition, 1993.

Gill Kirkup. Academic blogging: academic practice and academic identity. *London Review of Education*, 8(1):75–84, 2010. ISSN 1474-8460.

J Kitzinger. Qualitative research. Introducing focus groups. *BMJ: British Medical Journal*, 311(7000):299–302, 1995. ISSN 0959-8138.

Jens Klump, Roland Bertelmann, Jan Brase, Michael Diepenbroek, Hannes Grobe, Heinke Höck, Michael Lautenschlager, Uwe Schindler, Irina Sens, and Joachim

Wächter. Data publication in the open access initiative. *Data Science Journal*, 5 (15):79–83, 2006. ISSN 1683-1470.

Paul Knoepfler. Revewing post-publication peer review. *Trends in Genetics*, 31(5): 221–223, April 2015.

Karin Knorr-Cetina and Aaron Cicourel, editors. *Advances in Social Theory and Methodology Toward an Integration of Micro- and Macro- Sociologies*. Routledge & Kegan Paul, Boston, 1981.

John Kratz and Carly Strasser. Data publication consensus and controversies. *F1000Research*, 3(94):1–21, April 2014. ISSN 2046-1402.

Tim Kreider. Slaves of the Internet, Unite!, 2013. http://www.nytimes.com/2013/10/27/opinion/sunday/slaves-of-the-internet-unite.html [Accessed: 13 November 2015].

David Abraham Kronick. *Scientific and technical periodicals of the seventeenth and eighteenth centuries*. Scarecrow Press, 1991.

Steinar Kvale. *Doing Interviews*. Sage Publications Ltd, London, 2007. ISBN 978-0-7619-4977-0.

Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. What is Twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web - WWW '10*, pages 591–600, New York, New York, USA, 2010. ACM Press. ISBN 9781605587998.

Mikael Laakso and Bo-Christer Björk. Delayed Open Access: An Overlooked High-Impact Category of Openly Available Scientific Literature. *Journal of American Society for Information Science and Technology*, 64(7):1323–1329, 2013. ISSN 14923831.

Carl Lagoze and Herbert Van De Sompel. The Open Archives Initiative: Building a Low-Barrier Interoperability Framework. In *First ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'01)*, pages 54–62, Roanoke Virginia, 2001. ISBN 1607255308.

Carl Lagoze and Herbert Van de Sompel. The making of the Open Archives Initiative Protocol for Metadata Harvesting. *Library Hi Tech*, 21(2):118–128, June 2003. ISSN 0737-8831.

Charles Landesman. *An Introduction to Epistemology*. Blackwell Publishers Inc, Cambridge, 1997.

Vincent Lariviere, Veronique Kiermer, Catriona J MacCallum, Marcia McNutt, Mark Patterson, Bernd Pulverer, Sowmya Swaminathan, Stuart Taylor, and Stephen Curry. A simple proposal for the publication of journal citation distributions. *bioRxiv*, July 2016. http://biorxiv.org/content/early/2016/07/05/062109.abstract.

Bruno Latour. *The Pasteurization of France*. Harvard University Press, Cambridge, MA, 1984.

Bruno Latour. *Science in action: How to follow scientists and engineers through society*. Harvard University Press, 1987.

Bryan Lawrence, C. Jones, Brian Matthews, S. Pepler, and S Callaghan. Citation and peer review of data: moving towards formal data publication. *International Journal of Digital Curation*, 6(2):4–37, 2011. ISSN 1746-8256.

Stuart Lawson and Ben Meghreblian. Journal subscription expenditure of UK higher education institutions. *F1000Research*, 3(274), July 2015. ISSN 2046-1402.

Carole J Lee, Cassidy R Sugimoto, Guo Zhang, and Blaise Cronin. Bias in peer review. *Journal of the American Society for Information Science and Technology*, 64(1):2–17, January 2013. ISSN 15322882.

Na Li, Leilei Zhu, Prasenjit Mitra, Karl Mueller, Eric Poweleit, and C. Lee Giles. oreChem ChemXSeer : A Semantic Digital Library for Chemistry. In *Proceedings of the 10th annual joint conference on Digital libraries*, pages 245–254, New York, 2010. ACM. ISBN 9781450300858.

Maurice B. Line. The Publication and Availability of Scientific and Technical Papers: an Analysis of Requirements and the Suitability of Different Means of Meeting Them. *Journal of Documentation*, 48(2):201–219, 1992. ISSN 0022-0418.

Claudia Loebecke, Paul C. Van Fenema, and Philip Powell. Co-opetition and knowledge transfer. *ACM SIGMIS Database*, 30(2):14–25, March 1999. ISSN 00950033.

Matthew P Long and Roger C Schonfeld. Supporting the Changing Research Practices of Chemists. Technical report, Ithaka S+R, New York, 2013. http://www.sr.ithaka.org/wp-content/uploads/2015/08/Supporting-the-Changing-Research-Practices-of-Chemists-FINAL.pdf.

Clifford A Lynch and Joan K Lippincott. Institutional Repository Deployment in the United States as of Early 2005. *D-Lib Magazine*, 11(9):1–11, 2005.

Dena P. MacMynowski. Pausing at the brink of interdisciplinarity: Power and knowledge at the meeting of social and biophysical science. *Ecology and Society*, 12(1), 2007. ISSN 17083087.

Nicola Maffulli. More on citation analysis. *Nature*, 378(6559):760, December 1995. ISSN 0028-0836.

Michael J. Mahoney. Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research*, 1(2):161–175, June 1977. ISSN 0147-5916.

Apoorva Mandavilli. Trial By Twitter. *Nature*, 469(7330):286–287, July 2011. ISSN 1095-9203.

Ben R Martin. The Research Excellence Framework and the 'impact agenda': are we creating a Frankenstein monster? *Research Evaluation*, 20(3):247–254, 2011. ISSN 0958-2029.

M Martone. Data Citation Synthesis Group: Joint Declaration of Data Citation Principles, 2014. https://www.force11.org/group/joint-declaration-data-citation-principles-final [Accessed: 5 May 2016].

Natalie Matosin, Elisabeth Frank, Martin Engel, Jeremy S Lum, and Kelly A Newell. Negativity towards negative results: a discussion of the disconnect between scientific worth and scientific culture. *Disease Models & Mechanisms*, 7(2):171–173, 2014. ISSN 1754-8411.

Marshall McLuhan. *Understanding media: The extensions of man*. MIT press, 1994.

Christian Meier zu Verl and Wolfram Horstmann, editors. *Studies on Subject-Specific Requirements for Open Access Infrastructure*. Bielefeld, 2011. ISBN 9783943363005.

Robert K Merton. Priorities in Scientific Discovery: A Chapter in the Sociology of Science. *American Sociological Review*, 22(6):635–659, 1957.

Robert K Merton. *Social Theory and Social Structure*. The Free Press, New York, 1968.

Robert K Merton. Behavior Patterns of Scientists. *The American Scholar*, 38(2):197–225, 1969.

David Michaels. Politicizing Peer Review: The Scientific Perspective. In Wendy Elizabeth and Rena Steinzor, editors, *Rescuing Science from Politics: Regulation and the Distortion of Scientific Research*, pages 220–221. Cambridge University Press, Cambridge, 2006.

Thaddeus R Miller, Timothy D Baird, Caitlin M Littlefield, Gary Kofinas, III Chapin, F.Stuart, and Charles L Redman. Epistemological Pluralism : Reorganizing Interdisciplinary Research. *Ecology And Society*, 13(2):46, 2008.

Andrew J. Milsted, Jennifer R. Hale, Jeremy G. Frey, and Cameron Neylon. LabTrove: A Lightweight, Web Based, Laboratory "Blog" as a Route towards a Marked Up Record of Work in a Bioscience Research Laboratory. *PLoS ONE*, 8(7):1–18, 2013. ISSN 19326203.

Henk F. Moed. *Citation Analysis in Research Evaluation*. Springer, Dordrecht, NL, 2005.

David L Morgan. Focus groups. *Annual Review of Sociology*, 22:129–152, 1996. ISSN 1744-6155.

Sally Morris. Getting Started in Electronic Journal Publishing. Technical report, 2006.

Noah Moxham. Innovative Oldenburg, 2015. *History Today*, http://www.historytoday.com/noah-moxham/innovative-oldenburg [Accessed: 29 May 2015].

Peter Murray-Rust and Henry Rzepa. The Next Big Thing: From Hypermedia to Datuments. *Journal of Digital Information*, 5(1), 2006.

San Murugesan. Understanding Web 2.0. *IT Professional*, 9(4):34–41, July 2007. ISSN 1520-9202.

National Science Foundation. About Funding.

National Science Foundation. Data Management & Sharing Frequently Asked Questions (FAQs), 2010. https://www.nsf.gov/bfa/dias/policy/dmpfaqs.jsp [Accessed 21 October 2016].

Nature Publishing Group. Policies (journals), 2016. http://www.nature.com/openresearch/about-open-access/policies-journals/ [Accessed: 19 June 2016].

Isaac Newton. A Letter of Mr. Isaac Newton, Professor of the Mathematicks in the University of Cambridge; Containing His New Theory about Light and Colors: Sent by the Author to the Publisher from Cambridge, Febr. 6. 1671/72; In Order to be Communicated to the R. Socie. *Philosophical Transactions*, 6(69-80):3075–3087, 1671.

Cameron Neylon. Data Citation - what is the state of the art in 2016?, 2016. *Research Data and related topics*, https://researchdata.jiscinvolve.org/wp/2016/04/14/data-citation-what-is-the-state-of-the-art-in-2016/ [Accessed: 3 July 2016].

Ikujiro Nonaka. Dynamic Theory Knowledge of Organizational Creation. *Organization Science*, 5(1):14–37, 1994.

B. A. Nosek, G. Alter, G. C. Banks, D. Borsboom, S. D. Bowman, S. J. Breckler, S. Buck, C. D. Chambers, G. Chin, G. Christensen, M. Contestabile, A. Dafoe, E. Eich, J. Freese, R. Glennerster, D. Goroff, D. P. Green, B. Hesse, M. Humphreys, J. Ishiyama, D. Karlan, A. Kraut, A. Lupia, P. Mabry, T. Madon, N. Malhotra, E. Mayo-Wilson, M. McNutt, E. Miguel, E. Levy Paluck, U. Simonsohn, C. Soderberg, B. A. Spellman, J. Turitto, G. VandenBos, S. Vazire, E. J. Wagenmakers, R. Wilson, and T. Yarkoni. Promoting an open research culture. *Science*, 348(6242):1422–1425, 2015.

Dan O'Connor. Apomediation and the Significance of Online Social Networking. *The American Journal of Bioethics*, 9(6-7):25–27, June 2009. ISSN 1526-5161.

Mancur Olson. *The Logic of Collective Action*. Harvard University Press, Cambridge, MA, 1965.

Open Scholar. Developing the first Open Peer Review Module for Institutional Repositories, 2015. http://www.openscholar.org.uk/developing-the-first-open-peer-review-module-for-institutional-repositories/ [Accessed: 22 October 2015].

OpenDOAR. Growth of the OpenDOAR Database - Worldwide, 2016a. http://www.opendoar.org/onechart.php?cID=&ctID=&rtID=&clID=&lID=&potID=&rSoftWareName=&search=&groupby=r.rDateAdded&orderby=&charttype=growth&width=600&height=350&caption=GrowthoftheOpenDOARDatabase-Worldwide [Accessed: 24 June 2016].

OpenDOAR. Open Access Repository Types - Worldwide, 2016b. http://www.opendoar.org/onechart.php?cID=&ctID=&rtID=&clID=&lID=&potID=&rSoftWareName=&search=&groupby=rt.rtHeading&orderby=TallyDESC&charttype=pie&width=600&height=300&caption=OpenAccessRepositoryTypes-Worldwide [Accessed: 23 June 2016].

ORCID. What is ORCID?, 2015. http://orcid.org/about/what-is-orcid [Accessed: 22 October 2015].

Tim O'Reilly. What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. *Communications & Strategies*, 1:17, 2007.

Susan R Owens. Revolution or evolution? *EMBO reports*, 4(8):741–743, 2003.

Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The PageRank Citation Ranking: Bringing Order to the Web. Technical report, Stanford InfoLab, November 1999.

Perakakis Pandelis. New forms of open peer review will allow academics to separate scholarly evaluation from academic journals, 2013. http://blogs.lse.ac.uk/impactofsocialsciences/2013/08/20/libre-project-open-peer-review-perakakis/ [Accessed: 20 August 2013].

M. A. Parsons, O. Godoy, E. LeDrew, T. F. de Bruin, B. Danis, S. Tomlinson, and D. Carlson. A conceptual framework for managing very diverse data for complex, interdisciplinary science. *Journal of Information Science*, 37(6):555–569, 2011. ISSN 0165-5515.

Harry E. Pence and Antony Williams. Chemspider: An online chemical information resource. *Journal of Chemical Education*, 87(11):1123–1124, 2010. ISSN 00219584.

Douglas P. Peters and Stephen J. Ceci. Peer-review practices of psychological journals: The fate of published articles, submitted again. *Behavioral and Brain Sciences*, 5(2): 187–195, February 1982. ISSN 0140-525X.

Wendy Piatt. Government response to the Finch Report, 2012. *Russell Group*, http://www.russellgroup.ac.uk/russell-group-latest-news/151-2012/5324-government-response-to-the-finch-report/ [Accessed: 3 December 2012].

Gabriele Piccoli, Rami Ahmad, and Blake Ives. Knowledge management in academia: A proposed framework. *Information Technology & Management*, 1(4):229–245, 2000. ISSN 1385951X.

Stephen Pinfield, Jennifer Salter, and Peter Bath. The "total cost of publication" in a hybrid open-access environment: Institutional approaches to funding journal article-processing charges in combination with subscriptions. *Journal of the Association for Information Science and Technology*, 2015. ISSN 23301635.

Heather Piwowar. Altmetrics: Value all research products. *Nature*, 493(7431):159, 2013. ISSN 0028-0836.

Jason Priem and Bradley M Hemminger. Scientometrics 2.0: Toward new metrics of scholarly impact on the social Web. *First Monday*, 15(7), 2010.

Jason Priem, Dario Taraborelli, Paul Groth, and Cameron Neylon. altmetrics: a manifesto, 2011. http://altmetrics.org/manifesto/ [Accessed: 5 November 2012].

David Prosser. Too soon to tell: Review of the RCUK OA policy, 2015. *Research Libraries UK*, http://www.rluk.ac.uk/about-us/blog/rcuk-oa-policy/ [Accessed: 17 April 2015].

Danny Quah. Digital Goods and the New Economy, 2003. *CEPR Discussion Paper*, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=410604 [Accessed: 22 September 2015].

Jerome Ravetz. Sociology of science: Keep standards high. *Nature*, 481(7379):25, 2012.

RCUK. Cross-Council Research, 2014. http://www.rcuk.ac.uk/research/xrcprogrammes/ [Accessed: 6 November 2015].

RCUK. Concordat On Open Research Data. Technical report, RCUK, Swindon, UK, 2015. http://www.rcuk.ac.uk/documents/documents/concordatopenresearchdata-pdf/.

Redecentralize.org. Redecentralize Conference - taking back the net, 2015. http://redecentralize.org/conference/ [Accessed: 26 November 2015].

RELX Group. Annual Reports and Financial Statements 2015. Technical report, RELX Group, London, 2015. http://www.relx.com/investorcentre/reports2007/Documents/2015/relxgroup_ar_2015.pdf.

Allen H Renear and Carole L Palmer. Strategic reading, ontologies, and the future of scientific publishing. *Science*, 325(5942):828–832, August 2009. ISSN 1095-9203.

Allen F. Repko. *Interdisciplinary Research*. Sage Publications Inc, Los Angeles, 1st edition, 2008.

Research Councils UK. RCUK Policy on Open Access and Supporting Guidance. Technical report, 2013. http://www.rcuk.ac.uk/RCUK-prod/assets/documents/documents/RCUKOpenAccessPolicy.pdf.

Research Councils UK. Pathways to Impact - Research Councils UK, 2014a. http://www.rcuk.ac.uk/innovation/impacts/ [Accessed: 23 April 2015].

Research Councils UK. Research Outcomes Common Question Set, 2014b. http://www.rcuk.ac.uk/RCUK-prod/assets/documents/documents/ResearchOutcomesCommonQuestionSet.pdf.

Research Councils UK. Research Outcomes Overview, 2014c. http://www.rcuk.ac.uk/research/researchoutcomes/ [Accessed: 16 April 2015].

Research Councils UK. What is Researchfish?, 2014d. http://www.rcuk.ac.uk/research/researchoutcomes/researchfish/ [Accessed: 1 January 2015].

Research Information Network. Patterns of information use and exchange: case studies of researchers in the life sciences. Technical Report November, 2009. http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Patterns+of+information+use+and+exchange+:+case+studies+of+researchers+in+the+life+sciences#0 [Accessed: 27 November 2012].

Research Libraries UK. Library Spend on Journal Big Deals, 2014. http://www.rluk.ac.uk/news/university-spend-big-deals/ [Accessed: 19 June 2016].

Ian Rowlands and David Nicholas. Scholarly communication in the digital environment. *Aslib Proceedings*, 57(6):481–497, 2005.

Royal Society of Chemistry. RSC Roadshow 2014 – Southampton, 2014. http://www.rsc.org/images/SouthamptonRoadshowagenda2014_tcm18-236478.pdf.

Royal Society of Chemistry. National Compound Collection, 2016. http://www.rsc.org/blogs/escience/national-compound-collection [Accessed: 27 October 2016].

Russell Group. Russell Group evidence: Independent review of the implementation of RCUK Policy on Open Access. Technical report, 2014. http://russellgroup.ac.uk/media/5106/54-russell-group-response-to-review-of-the-implementation-of-rcuk-oa-policy.pdf.

Henry Rzepa. Discoverability and reusability of FAIR chemistry research data as a key outcome of registering persistent identifiers and standardised metadata with DataCite., 2016. http://www.ch.ic.ac.uk/rzepa/talks/acs16-cinf/index.html [Accessed: 5 May 2016].

Arthur Sale, Marc Couture, Eloy Rodrigues, Les Carr, and Stevan Harnad. Open Access Mandates and the "Fair Dealing" Button. In Rosemary J. Coombe and Darren Wershler, editors, *Dynamic Fair Dealing: Creating Canadian Culture Online*. University of Toronto Press, 2014.

Miklos Sarvary. Knowledge Management and Competition in the Consulting Industry. *California Management Review*, 41(2):95–107, 1999. ISSN 0008-1256.

Leo Sauermann, Richard Cyganiak, and Max Völkel. Cool URIs for the semantic web. Technical report, DFKI Deutsches Forschungszentrum für Künstliche Intelligenz, Kaiserslautern, 2007. http://www.dfki.uni-kl.de/~sauermann/2007/01/semweburisdraft/uricrisis.pdf.

Robert W Scherer, Cesar Ugarte-Gil, Christine Schmucker, and Joerg J. Meerpohl. Authors report lack of time as main reason for unpublished research presented at biomedical conferences: a systematic review. *Journal of Clinical Epidemiology*, 68(7): 803–810, 2015.

m. c. Schraefel, Gareth V Hughes, Hugo R Mills, Graham Smith, Terry R Payne, and Jeremy Frey. Breaking the book. In *Proceedings of the 2004 conference on Human factors in computing systems - CHI '04*, volume 6, pages 25–32, New York, New York, USA, 2004. ACM Press. ISBN 1581137028.

Thomas A Schwandt. Constructivist, interpretivist approaches to human inquiry. In N K Denzin Y S Lincoln, editor, *Handbook of qualitative research*, pages 118–137. Sage Publications, Inc, Thousand Oaks, CA, US, 1994. ISBN 0-8039-4679-1 (Hardcover).

P O Seglen. Why the impact factor of journals should not be used for evaluating research. *BMJ: British Medical Journal*, 314(7079):498–502, 1997. ISSN 0959-8138.

Steven Shapiro. JSTOR, University Presses, and the Serials Crisis. *Journal of Electronic Resources Librarianship*, 25(3):240–242, 2013. ISSN 1941-126X.

Clay Shirky. *Here comes everybody: How change happens when people come together*. Penguin, London, 2009a.

Clay Shirky. Newspapers and Thinking the Unthinkable, 2009b. http://www.shirky.com/weblog/2009/03/newspapers-and-thinking-the-unthinkable/ [Accessed: 1 June 2012].

David Shotton and Paolo Ciccarese. CiTO, the Citation Typing Ontology, 2013. http://vocab.ox.ac.uk/cito [Accessed: 2 September 2013].

Jonathan Silvertown. A new dawn for citizen science. *Trends in ecology & evolution*, 24 (9):467–71, 2009. ISSN 0169-5347.

John W T Smith. The Deconstructed Journal. In *Proceedings of the ICCC/IFIP Conference on Electronic Publishing*, pages 73–84, 1997.

MacKenzie Smith, Mary Barton, Mick Bass, Greg Branschofsky, Margret McClellan, Dave Stuve, Robert Tansley, and Julie Harford Walker. DSpace: An Open Source Dynamic Digital Repository. *D-Lib Magazine*, 9(1), 2003.

Michael D. Smith, Joseph Bailey, and Erik Brynjolfsson. Understanding Digital Markets: Review and Assesment. In *Understanding the Digital Economy*. MIT Press, Cambridge, MA, 2001. ISBN 0-262-02474-8.

R Smith. Measuring the social impact of research: Difficult but necessary. *BMJ: British Medical Journal*, 323(7312):528, September 2001. ISSN 0959-8138.

Richard Smith. Classical peer review: an empty gun. *Breast Cancer Research*, 12(4): 1–4, January 2010. ISSN 1465-542X.

David J Solomon and Bo-christer Björk. A Study of Open Access Journals Using Article Processing Charges. *Journal of American Society for Information Sciences and Technology*, 63(8):1485–1495, 2012. ISSN 14923831.

Herbert Van De Sompel, Michael Nelson, Carl Lagoze, and Simeon Warner. Resource Harvesting within the OAI-PMH Framework. *D-Lib Magazine*, 10(12), 2004.

Springer Nature. Recommended Data Repositories, 2016. `http://www.nature.com/sdata/policies/repositories` [Accessed: 21 October 2016].

Michael Stebbins. Expanding Public Access to the Results of Federally Funded Research, 2013. *The White House Blog*, `https://www.whitehouse.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research` [Accessed: 19 June 2016].

Peter Suber. Open Access Overview, 2004. `http://legacy.earlham.edu/{~}peters/fos/overview.htm` [Accessed: 13 November 2015].

Alma Swan. The Open Access citation advantage. Technical report, University of Southampton, 2010. `http://eprints.soton.ac.uk/268516/`.

Kin Bee Tay and John Chelliah. Disintermediation of traditional chemical intermediary roles in the Electronic Business-to-Business (e-B2B) exchange world. *The Journal of Strategic Information Systems*, 20(3):217–231, September 2011. ISSN 09638687.

The Cambridge Crystallographic Data Centre. The Cambridge Structural Database, 2016. `http://www.ccdc.cam.ac.uk/solutions/csd-system/components/csd/` [Accessed: 18 July 2016].

The Economist. No peeking... A publishing giant goes after the authors of its journals' papers, 2014. *The Economist*, http://www.economist.com/news/science-and-technology/21593408-publishing-giant-goes-after-authors-its-journals-papers-no-peeking [Accessed: 17 February 2014].

The Royal Society. Science as an open enterprise. Technical report, The Royal Society, London, 2012. https://royalsociety.org/topics-policy/projects/science-public-enterprise/report/.

The Royal Society. The Future of Scholarly Scientific Communication Part 1, 2015. https://royalsociety.org/{~}/media/events/2015/04/FSSC1/FSSC1_schedule.pdf [Accessed: 24 April 2015].

The Wellcome Trust. Costs and business models in scientific research publishing. Technical report, 2004. https://wellcome.ac.uk/sites/default/files/wtd003184_0.pdf.

Thomson Reuters. The Thomson Reuters Impact Factor, 2012. http://wokinfo.com/essays/impact-factor/.

Ramine Tinati, Olivier Phillipe, Catherine Pope, Les Carr, and Susan Halford. Challenging Social Media Analytics: Web Science Perspectives. In *Proceedings of the 2014 ACM Conference on Web Science*, WebSci '14, pages 177–181, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2622-3.

Paul Trevorrow and Dietrich a Volmer. Dispelling the myths surrounding the Research Excellence Framework. *Rapid communications in mass spectrometry*, 26(4):399–402, February 2012. ISSN 1097-0231.

Rob Tucker. Disintermediation: The disruption to come for Education 2.0, 2010. *Radar*, http://radar.oreilly.com/2010/05/disintermediation-risks-trends.html [Accessed: 11 March 2011].

Bryan S Turner. Preface to the New Edition. In *The Social System*, pages xiii–xxx. Routledge, London, 1991.

University of Southampton. Registry of Open Access Repositories, 2015. http://roar.eprints.org/ [Accessed: 17 April 2015].

University of Southampton. Welcome to eCrystals - University of Southampton, 2016. http://ecrystals.chem.soton.ac.uk/ [Accessed: 17 July 2016].

P J Van Diest, H Holzel, D Burnett, and J Crocker. Impactitis: new cures for an old disease. *Journal of Clinical Pathology*, 54(11):817–819, 2001. ISSN 0021-9746.

John P. van Gigch. Comparing the epistemologies of scientific disciplines in two distinct domains: modern physics versus social sciences. I: The epistemology and knowledge characteristics of the physical sciences. *Systems Research and Behavioral Science*, 19 (3):199–209, April 2002a. ISSN 10927026.

John P. van Gigch. Comparing the epistemologies of scientific disciplines in two distinct domains: modern physics versus social sciences. II: Epistemology and knowledge characteristics of the âĂŸnew' social sciences. *Systems Research and Behavioral Science*, 19(6):551–562, August 2002b. ISSN 10927026.

Richard Van Noorden. Open access: The true cost of science publishing. *Nature*, 495 (7442):426–9, 2013. ISSN 1476-4687.

Richard Van Noorden. Online collaboration: Scientists and the social network. *Nature*, 512(7513):126–129, 2014a. ISSN 0028-0836.

Richard Van Noorden. The arXiv preprint server hits 1 million articles, December 2014b. ISSN 1476-4687. *Nature*, http://www.nature.com/doifinder/10.1038/nature.2014.16643 [Accessed: 17 October 2016].

G. Veletsianos. Higher education scholars' participation and practices on Twitter. *Journal of Computer Assisted Learning*, 28(4):336–349, 2012. ISSN 02664909.

P Vinkler. Evaluation of some methods for the relative assessment of scientific publications. *Scientometrics*, 10(3-4):157–177, 2005.

G Walsham. Actor-Network Theory and IS Research: Current Status and Future Prospects. In AllenS. Lee, Jonathan Liebenau, and JaniceI. DeGross, editors, *Information Systems and Qualitative Research*, IFIP – The International Federation for Information Processing, pages 466–480. Springer US, 1997. ISBN 978-1-4757-5487-2.

Geoff Walsham. The Emergence of Interpretivism in IS Research. 6(4):376–394, 1995. ISSN 10477047.

Mark Ware. Peer review: benefits, perceptions and alternatives. Technical report, Publishing Research Consortium, London, 2008. http://publishingresearchconsortium.com/index.php/prc-documents/prc-research-projects/35-prc-summary-4-ware-final-1/file.

Martin Weller. *The Digital Scholar*. Bloomsbury Academic, London, 2011.

Christine Wennerás and Agnes Wold. Nepotism and sexism in peer-review. *Nature*, 387 (6631):341–343, 1997. ISSN 0028-0836.

E. James Whitehead. Control choices and network effects in hypertext systems. In *Proceedings of the tenth ACM Conference on Hypertext and hypermedia : returning to our diverse roots returning to our diverse roots - HYPERTEXT '99*, pages 75–82, New York, USA, 1999. ACM Press. ISBN 1581130643.

Rolf T Wigand. Electronic Commerce: Definition, Theory, and Context. *The Information Society: An International Journal*, 13(1):1–16, 1997.

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(160018), March 2016. ISSN 2052-4463.

Mark Williamson. ChemicalTagger Overview, 2016. `https://bitbucket.org/wwmm/chemicaltagger` [Accessed: 8 January 2016].

Alicia Wise. Elsevier welcomes the Finch report, 2012. **Research Blogs*, `http://exquisitelife.researchresearch.com/exquisite_life/2012/06/elsevier-welcomes-the-finch-report.html` [Accessed: 20 March 2015].

Alice Witt, Nicolas Suzor, and Patrik Wikström. Regulating ride-sharing in the peer economy. *Communication Research and Practice*, 1(2):174–190, 2015. ISSN 2204-1451.

Michael Witt. Co-designing, Co-developing, and Co-implementing an Institutional Data Repository Service. *Journal of Library Administration*, 52(2):172–188, 2012. ISSN 0193-0826.

Arch G. Woodside. Journal and author impact metrics: An editorial. *Journal of Business Research*, 62(1):1–4, January 2009. ISSN 01482963.

Paul Wouters. What is the matter with e-Science?-thinking aloud about informatisation in knowledge creation. 23, 2006. `http://www.pantaneto.co.uk/issue23/wouters.htm`.

Xin Yao and Paul J. Darwen. An Experimental Study of N-Person Iterated Prisoner's Dilemma Games. *Informatica*, 18(4):435–450, 1994.

Jun Zhao, Chris Wroe, Carole Goble, Robert Stevens, Dennis Quan, and Mark Greenwood. *Using Semantic Web Technologies for Representing E-science Provenance*, pages 92–106. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004. ISBN 978-3-540-30475-3.