Digital Enlightenment Yearbook 2014
K. O'Hara et al. (Eds.)
IOS Press, 2014

© 2014 The authors. All rights reserved.
doi: 10.3233/978-1-61499-450-3-3

Introduction

Kieron O'HARA [a,1], M-H. Carolyn NGUYEN [b] and Peter HAYNES [c]
[a] *Electronics and Computer Science, University of Southampton*
[b] *Microsoft*
[c] *Atlantic Council; contributor, The Economist*

Introduction

This is the third *Digital Enlightenment Yearbook*, part of an annual series which began in 2012 under the auspices of the Digital Enlightenment Forum (http://www.digitalenlightenment.org/). This aims to shed light on today's rapid technological changes and their impact on society and its governance, taking inspiration from Enlightenment thought as well as from the many transformations and evolutions that have taken place since. It examines digital technologies and their application openly with essential societal values in mind. Such values might assume novel forms, taking advantage of both the knowledge and unprecedented access to information which exist today.

The aim of the Yearbook is to track the evolution of digital technology – which sometimes happens so fast that even an annual publication occasionally seems inadequate.The 2012 yearbook focused on trust, privacy and the defence of the values of the World Wide Web. In 2013, the main topic was the value of personal data, to ourselves and to the commercial world, with contributions from academics, technologists and entrepreneurs. This year, the focus has shifted from individuals to their relationships with their networks, as we explore "Social networks and social machines, surveillance and empowerment." In what is now the well-established tradition of the Yearbooks, different stakeholders in society and various disciplinary communities (technology, law, philosophy, sociology, economics, policymaking) bring their very different opinions and perspectives to bear on this topic, forming a basis for inspiring and constructive cross-disciplinary discussions.

The digital world is enabled by technology, but co-constituted by the use of technology by users large and small, corporate and private, motivated by profit and by personal interest. Given that the Internet is a congress of all these entities, using a vast range of protocols, software, apps and devices, it is very hard, not to say pernicious, to generalise. In this introduction, we will highlight some of the themes or interesting phenomena informing the chapters that follow.

1. The Digital World in 2014

The growth of online social networking is a key part of the recent history of the Web, from the launch of invitation-only SixDegrees.com in 1997, to the position of Facebook on its

[1] Corresponding Author: Kieron O'Hara, E-mail: kmo@ecs.soton.ac.uk.

10th anniversary (2014), with over a billion active users and a market capitalisation of $134 billion. On these mediated interaction platforms, users enjoy the benefits of community networking supported by an information infrastructure, and willingly accept these benefits without being fully aware of the risks of surveillance, invasion of privacy, unconstrained data mining, data and identity breaches, and being haunted by the permanence of the various records maintained and increasingly shared by data intermediaries.

However, as these debates unfold, with stronger claims being made about the monetary and non-monetary benefits and costs of data volunteering and sharing, users' calculations of utility are becoming less tractable, and the struggle to maintain democratic protections and other stakeholder expectations more complex and pressing. Sensitivity is increasing with awareness. Events such as the delayed roll-out of the care.data healthcare data-sharing initiative in the United Kingdom, where ideas about privacy have traditionally been relatively relaxed (see Jacqui Taylor in this volume), indicate how attitudes appear to be changing.

Research published in June 2014 by academics from Cornell University, with the cooperation of Facebook, studied induced changes to the emotions of 689,003 experimental subjects, carried out without consent using the fine print of Facebook's data use policy as an ethical fig leaf [1]. Despite the obvious echoes of Stanley Milgram's notorious obedience experiments of the 1960s and others [2], nobody seemed very ashamed. Shortly afterwards, the dating site OKCupid admitted that it had also experimented on its customers without their consent, matching up apparently ill-suited pairs to see how they got on [3]. As they got on reasonably well, customers might consider questioning the company's compatibility algorithms, even if they are not already appalled at being lied to by their matchmaker.

Such events seem to demonstrate the powerlessness of individuals – mere datapoints to be pushed about by academics in search of statistical significance, and companies wishing to hone their products and to remove the last vestiges of unpredictability from the world. Indeed, one could be forgiven for wondering whether the real value of the Cornell researchers' experiment was to boost Facebook's share price by feeding the myth of its omnipotence (after an initial fall over the 4$^{\text{th}}$ July holiday weekend, the price actually rose 8% in the month of July). Certainly much of the experiment seemed to conflate the people and the data – the researchers tracked the vocabulary of users' posts, which seems a relatively remote proxy for emotional state. Had the researchers proved more than the fact that the vocabulary we use is conditioned by the vocabulary others use in similar contexts? That is hardly news. Does anyone care about such fine distinctions now? Perhaps our emotions simply *are* our vocabulary now – or whatever can easily be quantified and mined from social networking data. That will do, until CCTV cameras learn to distinguish smiles from scowls.

However, despite these dispiriting developments, emergent and collective problem-solving at scale on the Web, driven by increasing availability of data and powerful data-handling tools (exemplified by Wikipedia, Ushahidi and Galaxy Zoo), is enabling many individuals and communities to identify and solve their own problems, harnessing collective commitment, local knowledge and embedded skills. They are able to leverage their social networks (often with the help of advanced social networking tools) without having to rely on remote experts or governments. This promises to bring about Tim Berners-Lee's vision of using computers 'to create abstract social machines on the Web: processes in which the people do the creative work and the machine does the administration' [4].

Such social machines have a great deal of potential, but crucially they will depend on the willingness of participants to trust their peers, continue providing data to the systems, and trust that the systems will act to provide services that are deemed fair and appropriate [5].

Meanwhile, Edward Snowden's revelations (and other instances of promiscuous data-sharing practices) have shown how intimate the relationships can be between governments and other large institutions, and how flimsy the technical and legal protections of individual autonomy now seem [6–8]. The level of sharing, and the historically large information differentials created between individuals and the institutions that they rely upon, threaten trust. The free flow of data at scale is jeopardised by citizens' mistrust of data ecosystems in reaction to industrial-scale surveillance, and by the potential for an over-vigorous regulatory response (for instance, with regard to the ongoing negotiations surrounding the EU's revised Data Protection Directive, and any national legislation that may follow). A general reaction against surveillance threatens the services such networking sites provide (and other data-based services in general). Although these services will continue to be in demand, the business models of the companies that run them may be affected by changes in the law, for instance, restricting their ability to move data across borders, amalgamate datasets, preserve data for long periods of time, or craft permissive privacy policies that might benefit their customers. Many socio-economic gains that could be derived from the use of aggregated data sets may also be curtailed.

2. Data and its Social Impact

Data and the technologies that enable its production and use are having a fundamental impact on ordinary life, on our roles as citizens, consumers and individuals, and on our relationships with each other and our communities, as well as with public and private entities. However, much of the discussion on data has focused primarily on the more technical aspects of the data deluge, e.g. the potential impact of data analytics and the need for regulations to minimise the risks resulting from the information differential between individuals and institutions. This is reminiscent of the late 19th and early 20th century focus on science, and Frederick Taylor's *Principles of Scientific Management*, maybe because these are the most easily understood aspects of the complexities associated with data. As Meg Ambrose points out in the opening chapter of this volume, science is seductive in its promise to provide answers to all that is unknown and make everything better.

Throughout history, technology has always evolved at a much faster pace than society and politics are able to assimilate, and technologists are notoriously bad at understanding the unintended human and societal consequences of their innovations (sometimes failing even to entertain the possibility of unintended consequences). Currently, policy makers, with their imperative to create order and predictability out of a hyper-connected, data-heavy world drastically different from the well-ordered hierarchies of living memory, are rushing to rein in the capabilities of the new data technologies and recreate the known and comforting world, instead of taking the time required to understand the role of new technologies and their larger societal impact.

However, in this past year, amidst the hype on big data, there has been an increasing interest in sociological research on the impact and ethics of data use on individuals and society. Perhaps this has arisen out of the recognition of the importance of trust in rebalancing the aforementioned data asymmetry; of building sustainable data ecosystems where all participants must somehow employ mechanisms that respect individual preferences and empower them. Establishing trust requires an understanding of the value of data to the people who are most impacted by its use, not at the abstract big data level, but increasingly at the individual, small data level, where issues are more nuanced and complex.

The existing political dialogues about data have been about simplistic binary approaches: is data personal? is use allowed? is anonymisation sufficient? One approach posits the concept that perhaps the rules and norms governing data use should be more nuanced and context-aware [9–11]. Although there has been growing recognition of the importance of context in data usage, there is little evidence on how individuals define it across cultures, and on factors that would determine the acceptability of a given scenario of data use. Two recent World Economic Forum reports try to shed light on some of these issues through large global surveys of Internet users: one discusses the relative importance of factors that define data use context in different countries around the world; the other addresses individuals' perception of global values, beliefs, and uses of the Internet [12, 13]. Both attempt to provide evidence that a more nuanced discussion on the use of data is needed, and, more practically, some suggestions on how these ideas can be implemented and incorporated into policy frameworks.

Projects such as the Tenison Road Project on 'Data and Its Street Life' take the evidence to a different level [14]. This is an ethnographic study of the role of data in addressing the matters that are most important to the residents and community of a single urban street, in the hope of developing insights into what data means in ordinary life, and how people might meaningfully use it. The researchers' hope is to inform both new innovations in data technologies and policy making through bottom-up and behavioural evidence. Only a few months after its launch, the project identified some fundamental issues in data use that have remained elusive despite years of discussion on privacy and data, including the definition of 'common good', who should determine it, relevant legal frameworks, mechanisms for conflict resolution or for citizens who do not want to participate, the definition of 'privacy of the community' and how this relates to 'privacy of the individual'. Luciano Floridi, among others, has begun to reflect on this notion of group privacy [15]. Nicolas de Cordes (this volume) outlines the potential benefits to the community of making mobile records available – but it is clear that the same data sets can be used to track ethnic group movements in a country at war. What are the guidelines for how these data may be used, and who should determine them? What are the guidelines by which social machines should operate? These questions amplify the need to have a much broader discussion on data and its impact than we are currently seeing.

For instance, the use of machine-learning techniques and algorithms applied more broadly to develop insights and predictions about a complex and imperfectly modelled world can have many unintended consequences, including inadvertent discrimination. Although such algorithms can provide great improvements in efficiency, speed, and response times, an emphasis on solely technical objectives without consideration of the social consequences will not only lead to unsustainable systems [16], but may also further diminish trust. Just because a machine can, does not mean a machine should. Just because researchers can conduct social experiments on a vast population does not mean that they should. There are other considerations, both social and ethical.

The recent White House report on *Big Data: Seizing Opportunities, Preserving Values*, which includes a discussion on big data and discrimination [17], cites the mobile app StreetBump, developed for the US city of Boston, to collect data about potholes and report them to the city's Public Works Department [18]. Although this appeared at first sight to be an innovative use of technologies combined with crowd sourcing, areas primarily inhabited by the poor and elderly risked receivinginadequate services, as these groups are less likely to have smart phones or the skills to use mobile apps. Fortunately, the developers spotted these potential problems, and were able to do a workaround by first deploying the apps to the city's road inspectors.

So how is a 'code of ethics' to be developed which will foster trustworthiness of the tools of big data, and citizens' trust in their use? By what guiding principles can fairness and respect be established? In rethinking policy and regulatory frameworks, there are shorter term pressures to rebalance the asymmetry in power that exists today between people and large institutions, and longer term questions about the balance to be struck between personal agency and communal good. In a digital society, a person can have multiple roles: as consumer, individual, and citizen, with varying levels of civic responsibilities and agencies. The questions are difficult, intertwined, and raise complex issues. A search for definitive answers at this point seems premature if not Quixotic. A diversity of perspectives is what is needed. This serves as one of the many motivations for this volume.

3. Social Machines

This does not mean that we should remain in thrall to computers. As the amount of data that it is feasible to process has grown, so has the number of people that it is feasible to connect within a network, which means we can complement efficient machine-to-machine (M2M) communications with ever more interactions involving people or networked groups. David De Roure (this volume), gives a sense of different interaction modes of computing. Wherever there are more machines to produce the big data paradigm, or more people – as in the social networking paradigm – distribution is inevitable, and hence Web or Web-like technologies are necessary to handle interaction at scale. The technological affordances have, over time, increased in both paradigms. Ultimately we will reach a zone where big data and social networking collide, or collude. At this point, we will need the tools and the paradigms to enable smaller groups to define and solve their own problems.

In this zone we find the networks which have been termed 'social machines' ([4, 19, 20, 21]) – a nascent focus of computing research [22] outlined by De Roure. 'Programming the global computer' or 'global ubiquitous computing' has been recognised as a grand challenge for computing [23], while peer-to-peer technologies flexibly link people and computers, as explored in projects such as SOCIAM (http://sociam.org/), OpenKnowledge (http://www.openk.org/) and the Social Computer community (http://www.socialcomputer.eu/). As we unravel the mysteries of scale and control, we will need not only to understand the emergent phenomena, but to develop means, methods and tools for enabling and managing large-scale phenomena, at least partially [21]. The problem is sharpened by the desideratum that 'programming the social computer' must be achievable from within the social computer – research here should democratise control by enabling people to develop social machines to achieve their own smaller-scale, local, idiosyncratic purposes.

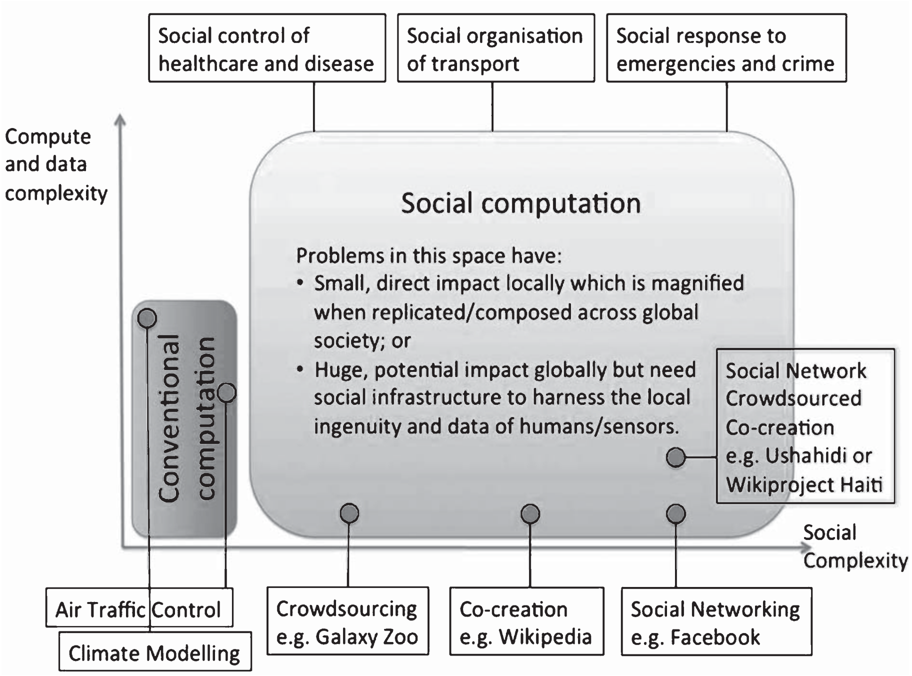
Social control of healthcare and disease
Social organisation of transport
Social response to emergencies and crime
Compute and data complexity
Social computation
Problems in this space have:
• Small, direct impact locally which is magnified when replicated/composed across global society; or
• Huge, potential impact globally but need social infrastructure to harness the local ingenuity and data of humans/sensors.
Conventional computation
Social Network Crowdsourced Co-creation e.g. Ushahidi or Wikiproject Haiti
Social Complexity
Air Traffic Control
Climate Modelling
Crowdsourcing e.g. Galaxy Zoo
Co-creation e.g. Wikipedia
Social Networking e.g. Facebook

Figure 1. The space of social machines [21].

If we unpack that image as Figure 1, we see the potential space for advancement in more detail. We see conventional computation, even highly complex domains such as air traffic control and climate modelling, on the left-hand side, where social complexity is low even where computational complexity is high. Current systems with high social complexity still involve relatively low computational complexity. Crowd sourcing systems, such as the citizen science initiative Galaxy Zoo [24], have a relatively low level of social complexity as well. More complex social arrangements can be found in the co-creation of content, e.g. Wikipedia, and social networking. However, greater complexity can be found, for instance, when social networks act as platforms for crowd sourced co-creation of content, as recently happened with the Ushahidi map of election violence in Kenya in 2007 [25], or the reuse of Ushahidi software to create a post-earthquake map of Port-au-Prince in Haiti in 2010 [26]. As we explore this space of social computation to address perceived issues where there are collective action problems, as with public health, transport or crime, we would expect to find solutions with small impacts locally, which will be magnified at scale as long as the requisite social infrastructure (including Web technologies) is in place.

The idea of a social machine has been implicit throughout the history of the Web. Let us quote Berners-Lee once more at somewhat greater length:

> Real life is and must be full of all kinds of social constraint – the very processes from which society arises. Computers can help if we use them to create abstract social machines on the Web: processes in which people do the creative work and the machine does the administration ([4], 172, Berners-Lee's emphasis).

Many social machines are built on sites such as Facebook, in which human interactions – from organising a birthday party to interacting with a Member of Parliament – are underpinned by the engineered environment. Another type of example is a multiplayer online game, where a persistent online environment facilitates interactions concerning virtual resources between real people. A third type is an online poker game, where the resources being played for are real-world, where the players may be humans or bots, and where the environment in which the game takes place is engineered around a relatively simple computational model. In such systems, (some of) the social constraints that Berners-Lee talks about, currently norm-driven, are administered by the architecture of the programmed environment.

A generalised definition of a social computation is provided by Robertson and Giunchiglia [27]:

> A computation for which an executable specification exists but the successful implementation of this specification depends upon computer mediated social interaction between the human actors in its implementation.

In such an environment, self-organisation (partial or full) becomes viable and scalable, while physical objects, agents, contracts, agreements, incentives and other objects can be referred to using URIs (Universal Resource Identifiers: strings of characters used to identify resources in a context-independent way, which include but are not limited to Web addresses like http://www.digitalenlightenment.org/). 'Programming' the social computer (as opposed to simply supporting and directing interactions in an engineered environment) and integrating larger numbers of people and machines will become increasingly feasible.

As a small example of a social machine, consider reCAPTCHA [28]. A CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart), invented by Louis Von Ahn, is the distorted sequence of letters that someone has to type in a box to verify themselves as a human (e.g. to buy a concert ticket online, or to comment on a blog). This is something that computers cannot do, and so the system stops bots from buying thousands of tickets for later resale, or spam bots from leaving spam messages as comments on blogs [29].

This involves no social computing, but Von Ahn extended the idea of the CAPTCHA to create the social machine reCAPTCHA, which socialises the same principle to solve another problem. Google (which acquired reCAPTCHA in 2009) uses it to scan older books automatically. reCAPTCHA presents the user who wishes to verify themselves as a human with two words, not one. The first is a normal CAPTCHA, and the second is a word from an old book that automatic Optical Character Recognition has failed to identify. If the person succeeds with the first CAPTCHA, then he or she is known to be a human. As humans are reliable at word recognition, the response to the second word will be a plausible suggestion as to what it might be. Presenting the same word to multiple users allows a consensus to emerge. The goal of the social machine is to digitise books – the necessity for users to prove themselves to be human provides the mechanism [28].

However, reCAPTCHA is purely exploitative, as the goal of the machine is independent of the requirements of its human 'components'. As another example, [27] cites the DARPA balloon challenge of 2009, in which the aim was to find ten weather balloons placed randomly around the US (in nine different states from California to Delaware). The rules of the challenge were intended to support the growth of a network of people taking part in the search, enabling a crowd sourced solution. The means of doing this (in the winning solution from Sandy Pentland at the Massachusetts Institute of Technology) was to set out financial incentives, according to a Query Incentive Network Model [30], in which people were incentivised both to look for the balloons and to add more people to the network. Pentland's team began with four people, and using social media had recruited over 5,000 at the point of completion, which took under ten hours [31].

reCAPTCHA and the DARPA challenge were each designed to solve a particular problem, but social machines can, and indeed should [5], solve the problems of the people who constitute them. One could imagine, for instance, a set of computer-mediated interactions enabling a community to provide a social response to problems of crime (such as BlueServo, http://www.blueservo.net/, which uses crowd sourcing to help with policing the Texas-Mexico border), or enabling those suffering from a particular healthcare problem to pool resources and offer support and advice to fellow sufferers (such as http://curetogether.com/). There is a growing number of health-related social machines, as surveyed in detail in [32]. Such efforts will not always be uncontroversial (consider BlueServo, for example). Attempts to crowd source the identities of the 2013 Boston Marathon bombers bordered on farce, and, although the countercultural website 4chan was prominent in the home-made policing efforts with its so-called '4chan Think Tank', its lamentable efforts were soon parodied elsewhere on the same site [33].

The notion of social machines promises to allow an abstract specification of networked interactions at scale between people and technologies. Trust will be a major factor in the success of such machines [5], but it seems possible that they will enable an understanding of such interactions to facilitate the empowerment of communities to define and solve their own problems, rather than having structures imposed upon them.

4. Overview of the Chapters

The aim of the 2014 Yearbook is to give some sense of the position of the individual who may finds themselves in a number of overlapping networks, many of which are valuable, supportive or entertaining, but all of which are now facilitated by instantaneous communications, boundless data storage and incredible machine learning power. As algorithms of increasing sophistication and speed crunch the data, every last drop of significance (for us and for our contacts) is being wrung out of our interactions. The result is a data bonanza, a digital gold rush – but we must remember that in the great gold rushes of the past, the beneficiaries weren't the people who did the mining, but the entrepreneurs who sold the picks and shovels.

The chapters in the Yearbook, by and large, are concerned less with the high politics of the cyber-Klondike – the dangers of totalitarian oppression by an Orwellian overseer, which are nevertheless real – than with the benefits, both social and personal, for citizens and consumers, and with the threats to their identity, autonomy and human dignity. To that extent, the issues up for debate perhaps concern those in wealthy democracies rather than, say, the citizens of Russia (where the main social networking site VK was co-opted by associates of President Putin in 2014), China (where this year saw a crackdown on those spreading 'Internet rumours'), or Turkey (where Twitter and YouTube have been collateral victims in the bitter struggle between newly elected President Erdogan and his Gülenist enemies). The thread running through the book is the question of empowerment versus surveillance; autonomy versus exploitation – the classic issues explored in the political philosophy of the Enlightenment period.

The Yearbook is divided into four parts, bookended by a Prologue and an Epilogue, which provide illuminating perspectives on the discussions in between. The division of the book into sections is, as with all such exercises, not definitive. We have suggested one narrative, but it is clear from the wealth of issues addressed by the 16 chapters that many others are possible.

4.1. Prologue: Lessons From the Past

It is easy to assume that the digital era is a new departure – like the industrial revolution, the telecommunications revolution has given us a year zero and a whole new set of rules. If that assumption is true, then can the past tell us anything? Is there any better place to begin our exploration of the digital world of 2014 than the information world of 1814?

In her opening paper, "From the avalanche of numbers to big data", Meg Ambrose gives the lie to that facile conceit. She explores the increasing facility with, and reliance on, statistics that characterised the 19th century, particularly its second quarter. The dilemmas that we face now also appeared then – the relation of data to actionable knowledge, the apparent devaluing of human dignity and agency, classification effects, the displacement of theory by analysis – and the world changed in complex ways that demanded responses from policymakers. The ramifications of the new world were the subject of great novels: Charles Dickens' *Bleak House* (1852–3) considered the effects on individuals of information embedded in a process-driven legal system, while Franz Kafka's *The Trial* (1914–15) saw such a system as characteristic of modernity. Figures such as William Blake and Edmund Burke railed at the new world, and perhaps their apocalyptic predictions were to some extent borne out in the events of the mid-20th century, as encapsulated by Hannah Arendt's concept of the banality of evil [34].

Today, the literature of big data is beginning to appear – Dave Eggers' *The Circle* (2013) may not be great literature, but its teasing out of some of the trends in the current use of technology means that it will be a cultural reference point for some time to come. The modern classic about data and systems doesn't mention computers at all, but José Saramago's *All the Names* (1997) reminds us of the way in which data sets its own agenda, as the hapless Senhor José finds himself on a search for a woman he has never known, following a data trail in the central registry of a fictional city. Ambrose's historical comparison between the 19th and 21st centuries gives a similarly insightful perspective on today's problems, reminding us – very much in the spirit of the Digital Enlightenment Forum – that there is very little that is new under the sun. We can surely learn from the mistakes of the past.

4.2. Part I: The Individual as Data Manager

A number of chapters in the 2013 Yearbook focused on the issues surrounding the individual's management of his or her own data (however defined), bringing perspectives to bear from law and policy, as well as technical architectures and papers from entrepreneurs developing the market. The idea of personal data stores (PDSs) is becoming more relevant, especially as people are generating more data of their own, sometimes unintentionally (e.g. from their smartphones), but increasingly often via special-purpose devices (e.g.

medical gadgets such as activity trackers, or wifi-enabled scales). The individual will then find themselves controlling quantities of data, which might, if misused, compromise privacy. There will also be issues to do with the use of the data, such as how to integrate one's own PDS with institutional data stores (e.g. of a hospital during a medical consultation), and what rights should exist over the data once such an integration has taken place.

Our first group of chapters in 2014 all focus on the question of how to ensure privacy in a world in which large numbers of inexperienced people are curating their own data. Cathal Gurrin et al.'s "A privacy-by-design approach to life logging" considers it as a basic design issue. Life loggers record as many aspects of their lives as possible, retaining the data for later use, on the principle that you never know what might be useful in future, and data storage is cheap enough. Life logging research has focused on computational issues such as search, and on its potential medical use as prosthetic memory technologies. Privacy is discussed less than one might think in this context ([35] is a significant exception), but the increasing prevalence of data sharing cannot be ignored. In consequence, Gurrin et al. have produced a framework for understanding the information relations in life logging, and consider what an architecture would need to include to preserve a life logger's privacy.

Dave Murray-Rust et al.'s "Social palimpsests" considers essentially the same problem, but in a wider set of contexts – the PDS owners they think about could be anyone with data they wish to share, for example by placing it on a social network. The authors are concerned that this places the data subject at the mercy of large corporations whose legal frameworks are designed to drive a wedge between data subjects and their data. Is there a way for the individual to retain the benefits of being visible to their networks without losing control of their data? Murray-Rust et al. explore the possibility of adjusting the data to create a false account of events – a type of anonymisation or data perturbation. Could such economy with the truth inject the right level of noise into the data – not enough to stop it being useful, but just enough to deter data mining? Can such technologically-enabled mendacity be considered ethical?

Marion Oswald's "Seek and ye shall not necessarily find" reverses the question. Let's not only worry about the privacy of the people generating the data, she argues – there are also the people who are on the receiving end. People with Google Glass, wearable cameras and other such devices collect data about the rest of us, not just themselves. But as she examines the Google Spain decision – perhaps the most important legal judgment of 2014 in this field – she finds precious little support for those of us who are the victims of what is often called 'sousveillance', a democratised form of surveillance conducted by people in a private capacity. Perhaps most disturbing, she finds that the proactive methods of resisting sousveillance that might help to solve this problem may actually be illegal.

4.3. Part II: The Individual, Society and the Market

Whether or not they are enthusiastic curators of their data, all individuals are affected by the world of big data. They receive the benefits of their interactions becoming open, and they become increasingly visible. Furthermore, as charted by recent work in Web science, many macro-scale phenomena across the Web emerge from the micro-scale actions of individuals [21]. The chapters in this section consider how the attitudes, actions and decisions of individuals impact across the Web and wider society, and then in turn how policies and ideology at the macro-level feed back into individual attitudes, actions and decisions. Will such feedback loops produce virtuous or vicious circles? And is it possible to steer these processes, which happen at immense scales, in beneficial directions (assuming we can identify or agree on what is beneficial anyway)? Why are organisations such as Google and Facebook apparently so much more effective in this than policymakers? The chapters in this section consider how markets, regulation and system design can help or hinder the peaceful coexistence of technology and people.

Reuben Binns' "Personal data empowerment and the ideal observer" looks at whether a critical mass of PDS owners will produce genuinely empowering data services, informed by detailed empirical analysis of the current state of the personal data management market. He brings the discussion back to Enlightenment concepts via the work of Adam Smith, and considers whether the information asymmetries and baffling complexities of today's data world can be sidestepped by the use of Smith's concept of the ideal observer. This sees the PDS not as a data holder for a boundedly rational individual, but rather as a model for a rational decision maker to serve as advisor or avatar. In this way, perhaps personal data management functionality can be defined so that it is genuinely empowering for the individual.

Julie Brill's "The Internet of Things" looks at how Internet of Things' (IoT) evolution threatens to outpace the development of security technology and privacy protection, leaving us potentially at the mercy of a set of devices that gather together minute but collectively disclosive quantities of information, and are also used to shape our environment without our overall control – an example of how machine-to-machine communication is especially risk-prone (particularly given its generally weak security at present). As a US Federal Trade Commissioner, she is alive to the immense commercial potential and consumer benefit of this kind of networked technology, while concerned with mitigating the risks. The architectures and human processes – the social machines – behind the IoT are analysed, leading to consideration of how effective legislation and regulation might be created.

The UK government is moving towards a strategy of 'digital by default' in the delivery of its welfare services. In "Watching you watching me", Lizzie Coles-Kemp et al. present a detailed and somewhat dystopian study of how one deprived community interacts with the all-powerful, secure fraud-prevention system protecting welfare payments. Coles-Kemp and her colleagues unpick the attitudes of the recipients of welfare via a focus group, and uncover a sophisticated understanding of the vulnerabilities of the system: the focus-group members had very clear attitudes about how the system could be subverted, and also how the new digital system put them at risk by allowing identity theft or diversion of funds. By listening to the voices of these often disregarded participants in the system, Coles-Kemp et al. sound a warning to the security industry that it is surely better and certainly safer to work with, rather than against, your primary users.

Colin Strong's "The human side of big data" makes a similar point about the commercial world, but from the viewpoint of the system developers. In the commercial world, the use of big data is a brand-management risk. In a famous case that emerged in 2012, the US retailer Target identified that a young girl was pregnant before her father did, on the basis of her purchases at one of its stores. Customers can be 'creeped out' by such apparent omniscience, which may offset the advantages of accurate marketing. Strong writes of the 'uncanny valley', and how consideration of this problem might even bring it about. When Google executive chairman Eric Schmidt argued that he wanted his company's knowledge of its users to 'go up to the creepy line', he in effect dissolved the difference between creepy and non-creepy. If you are trying to be as near-creepy as possible without actually being creepy, then that is a paradigm case of being creepy. This is not an area where self-awareness or reflexivity on the part of the surveillance system is necessarily helpful. Concepts from psychological therapy and existentialist philosophy are pressed into service by Strong to try to steer brand managers away from this paradoxically self-defeating position.

4.4. Part III: Big Data and Open Data

Of course, the most remarked-upon aspect of the digital world of 2014 is big data. Awareness is rapidly spreading about the value of large quantities of data; the business models of numerous companies depend upon it, and many governments are focusing their industrial policy on it. Scepticism is beginning to develop about some of the hyperbole, but there is no doubt that there is promise here.

At the same time, the economic and social impact of data has been given a powerful boost by the movement for open data. Many institutions, particularly governments and government-sponsored ones, have felt the force of moral and political arguments to hand over their data (where it is not personal data, or a state or trade secret) to the public in machine-readable form, via the Web, under an open license. Governments are enabled to collect data and generate information because they are (a) legitimised to do so by citizens voting for them, and (b) resourced to do so by taxpayers funding them. It therefore seems hard to argue against the proposition that if citizens/taxpayers can get value from non-sensitive data, then they should be able to [36]. The result is to open up the Web to much greater quantities of data, including linked data [37]. However, even if transparency advocates can overcome the non-trivial internal difficulties of, for example, ensuring that open data meets the demand for data rather than the state's idea of what should be supplied, ensuring that citizens' privacy is protected during data releases, and making sure governments still collect valuable data once they have transferred some of their services to app developers, there will be many unintended consequences of openness and transparency. The chapters gathered in this section engage in this debate.

Nicolas de Cordes' "The use of big data for development goals" is a practical demonstration of the power of data. He describes a project for using mobile phone data in sub-Saharan Africa – a rich source, as mobile telephony is relatively prevalent in Africa, where it has leapfrogged poorly administered landline networks. De Cordes argues that in this area, many of the first-world concerns of privacy and autonomy discussed in this Introduction might need to be offset against the transformative power of big data for development purposes. He explores the key datasets that might become the pillars of development, and sketches a framework for assessing risks (and, *en passant*, we can see emerging from that framework a method for understanding in advance which data will be socially valuable in emerging nations).

The other two papers in this section reflect on the open data revolution from opposite sides of the barricades. Jacqui Taylor's "Citizen enablement from open data to open policy: a personal view" sets out her vision of open data as a game-changer, allowing us to move from a world of relevance, where data consumers are fed a diet of relevant information, to what she terms a world of resonance, where the user joins in the data-creation effort. She describes her own experience from early efforts in data journalism to a demand-led open data industry. Citizen awareness, participation and inclusion are all increasing, according to her account, although there are still risks from poor communication, as exemplified by the UK's 2014 care.data fiasco.

A more cautious outlook is expressed in Alison Powell's "'Datafication', transparency and the good governance of the data city". Powell highlights the interaction of open data with parallel developments of urbanisation and globalisation which is altering conceptions of citizenship, a corollary of the theory of "communicative capitalism" as an anti-democratic political development. She argues that the ideology of openness is self-defeating, as the promise of transparency and accountability are only realisable given a hard-to-foresee growth in data literacy and the technical expertise required to put open data to use. She warns that the movement of data is as important to monitor as its creation, a thought which brings her ideas a little closer to the open data movement – for example, to thinkers concerned with data accountability, such as Daniel Weitzner [38] and Viktor Mayer-Schönberger [39].

4.5. Part IV: New Approaches

Ambrose's opening chapter tells us that we are not in new territory. History, however, never repeats itself exactly. Our new accommodation to data involves learning to live with unprecedented computing, storage and inferential power – dreaming up new benefits and creating new risks. Our final group of chapters extends our set of analytic tools for coming to understand and critique the digital world we create. Each of them, in their different ways, considers the context of technological deployment, describing the actions and social practices that affect and are affected by the technology. Each is interdisciplinary, though the approaches vary as to how far they involve the conceptual tools of formal computer science itself, and how far to bring in techniques from social science disciplines.

David De Roure's "The emerging paradigm of social machines" motivates the use of social machines as a means of understanding a new set of phenomena: how people interact to solve problems using networked technology. As with our own discussion above, De Roure uses a quote from Berners-Lee – but looks at the context of Berners-Lee's comment. He argues that social machines are with us now, and provides examples from the headlines that cannot easily be described or explained using current analytic tools. Social machines can create social change and disruption extremely rapidly. In this new context, science and social science are at a crisis point – to use a term from Thomas Kuhn's *Structure of Scientific Revolutions* – and according to De Roure's account, social machines could be the new paradigm for understanding this world, embodying values such as empowerment and democratisation. Yet other paradigms are developing in parallel in response to this crisis (which is no bad thing), as reflected in the three remaining chapters of this section.

David Robertson et al.'s "An open system for social computation" considers the problem of 'programming' a social machine. Integrating human and machine problem-solving power so that they complement each other requires abstracting away from particular methods and algorithms. If human abilities for, say, pattern-matching or optical recognition are supported by machine power so that, for example, management of personal data is enhanced by wider inferential analysis, then one can imagine social machines becoming programmable, rather than merely emerging from uncoordinated interaction, or designed from the top down (like reCAPTCHA). Robertson et al. present a core cycle of computation described at a high level of abstraction, and show how useful information – in this case, information about the provenance of the data within the social machine – can be derived at that level.

Mark Elliot and Elaine Mackey's "The social data environment" tackles a recurring problem of data protection. Data sharing is part of the ethos of big data – the value of data is enhanced by its use in new contexts. Yet sharing personal data without consent is an invasion of privacy, and the scale of big data precludes asking for consent for each use. The data protection principle of use limitation is alien to the big data world. The technical solution to this is anonymisation – removing the identifiers from data to render it non-personal. However, we know that with effort and extra knowledge this process is reversible, so there is always a non-zero risk that anonymised data could be shared, and people re-identified [40]. Elliot and Mackey point out that the properties of the anonymised data are less important in working out what that risk would be, than an understanding of the new context for the shared data and the agency deployed within that context. They describe aspects of the social data environment into which anonymised data is placed, to provide the beginnings of a framework to quantify the risk of re-identification of anonymised data. The ultimate aim is to increase the amount of data shared while decreasing uncertainty about the likelihood of invasions of privacy.

Jo Pierson's "Interdisciplinary perspective on social media, privacy and empowerment" considers how the field of media and communications studies could help our understanding and negotiation of (and with) the digital world. Teasing out the various artefacts, practices and social arrangements that make up a type of online interaction, he suggests, can enable us to see its wider links with other social institutions and understand its broader context, breaking the hold of technological determinism. This is especially important for complex issues such as privacy, which cannot be seen as purely a technical or security problem.

4.6. Epilogue: Lessons From the Present

The title of the final chapter, Mark Notturno's "Ten governance concerns about the nature and use of data", is self-explanatory. Notturno reports on the outcome of a focus group of data experts, and summarises their worries as the first nine of his ten concerns. The reader may be reassured that there is little evidence here of technological determinism or unsophisticated positivism; Notturno believes that our big data appears (for the most part) to be in safe hands. Using chairman's privilege, he also adds a tenth concern of his own, and pleads against data-driven policymaking. While such a self-denying ordinance is perhaps unlikely in today's digital politics, the suggestion has much to recommend it (after all, the ludicrous polarisation of US politics is accompanied by the eerie and unlikely ability of both sides to find data to support their positions, so the idea that better use of data might depoliticise some issues seems improbable). Indeed, it is striking that Notturno's suggestion brings us round from our *envoi* back to the conundrums described by Ambrose in her opening historical chapter. If her *dramatis personae* had heeded Notturno's warning from the future, many of the problems of the 20th century might have been nipped in the bud.

5. Glaucon's Dilemma: Sustainable Data-Driven Societies

The increasing datafication of our society is leading to a world where data equates to power, and the lack of transparency and accountability regarding its use raises questions of trust. Without appropriate data stewardship practices and policy frameworks that can balance the potential fear of surveillance with the benefit of empowerment, technologies such as the Internet of Things (IoT) and broad applications of machine learning will only exacerbate the current situation, and undermine the immense promise of emerging data-driven ecosystems. These are socio-technical systems that can only be sustainable if they are perceived to be trustworthy, guided by a discourse and policy frameworks that incorporate a multidisciplinary approach which integrates the humanity, technology, economic, and policy perspectives (see the chapters in Part IV of this volume).

The scenarios of individual/community empowerment by trustworthy systems, and effective surveillance of an alienated citizenry, are not mutually exclusive. It may be, as Socrates reminded Glaucon in Plato's *Republic*, that if the people want 'not only … a state … but … a luxurious state' with all the free services that the Web can provide, then data miners, analysts and marketers 'who were not needed and therefore had no place in the former edition of our State, but are needed now' will be in demand. If the full value of the Web (both to individuals *and* society) is to be realised, perhaps we do need to open ourselves up to use of our data by governments or the corporate behemoths of e-commerce. This is the fundamental dilemma of the digital world in 2014.

Examples of Glaucon's dilemma abound. The IoT will transform the world in which we live, and how we interact with our physical environment. Already the number of smart, connected devices exceeds the estimated world population, and its rate of increase is accelerating [41, 42]. We are clearly entering an era where almost every manufactured object will be able to exchange data with other objects, online services, and individuals, as Julie Brill (this volume) describes. Most of these communications will be by objects measuring or recording aspects of their environment and transmitting that data autonomously – for further analysis or to be commingled with other data. Such detailed information about our environment – and us – has never before been available, and will provide a far more granular understanding of the interplay between many different factors, as well as adaptive and dynamic responses to both micro and macro events, blurring the line between the physical and digital worlds.

Although the IoT is sometimes defined simply as the physical connection of objects, discussions of the IoT should consider the end-to-end systems that also include the software, systems and services that analyse the data these objects have collected; and the behaviours/actions taken by those objects as a result. Almost all these communications, insights and actions will take place without human involvement.

The IoT brings to the physical world the same level of personalisation that is possible in the virtual world. This was eloquently described by Rafi Haladjian, founder of sen.se (http://sen.se/), as the transition from *machina habilis* to *machina sapiens* – from a world where machines respond only to command and control and where individuals make all decisions, to one where machines can be enabled with complex algorithms and adaptive behaviours, and act as intelligent agents on behalf of individuals [43, 44].

That the IoT can communicate in an ad hoc, asynchronous, adaptive, self-configuring and dynamic way – and be equipped with the (artificial) intelligence to operate without human involvement – poses an unprecedented challenge to how technology policy is formulated (and also scares privacy advocates to death [45]). The data-driven economy that was sparked by big data will move into overdrive, and the policy issues encountered in enabling big data will seem simple when compared with the complexity of a world transformed by the IoT. Instead of relatively predictable static systems, policy makers will be faced with technologies whose behaviours are not only unpredictable but essentially *unknowable*. Because establishing appropriate, interoperable, system-level policies for such technologies is so challenging, regulators may default to narrow technology-specific regulations, or they will demand a level of security that undermines the essential business model of networking many simple, cheap devices. In the process they will be in danger of putting a damper on technological innovation, thus undermining the immense socio-economic benefits that the IoT (and other M2M interactions) have the potential to unleash.

The IoT is, in many ways, unknowable. However, not everything unknown is unknowable. Another less welcome trend in the digital economy, which is increasingly grounded in the exchange of data, is to render the ways in which data is collected and analysed even more opaque to the consumer, making the value exchange they are engaged in harder to discern. An individual may have only a vague idea of what data exists about them and what is being done with it – some of it will have been actively volunteered; some will have been obtained passively, with or without their knowledge; and yet more may have been inferred by commingling a range of data sets in ways that expose new information about their lifestyle. Although the individual may receive something in return for this data, the true values of both the data provided and the service returned (e.g. the underlying exchange of value) may be almost impossible to determine.

Worse, for many people any potential value of their data is already largely moot, because they have given away their digital crown jewels for free. Individuals continue to pass vast amounts of personal and other data to large corporations (such as Facebook) with little or no thought as to its true monetary value. And those corporations are making hefty profits as a result, because their 'cost of materials' is essentially zero.

The result is that the greater the role that data play in the global economy, the less the majority of individuals will be worth, and the smaller their role in that economy. As the computer scientist Jaron Lanier has observed, '[T]he dominant principle of the new economy, the information economy, has lately been to conceal the value of information… We've decided not to pay most people for performing the new roles that are valuable in relation to the latest technologies. Ordinary people "share", while elite network presences generate unprecedented fortunes' [46]. If an individual's data doesn't receive a transparent statement of value, Lanier adds, 'a massive disenfranchisement will take place.'

All previous economic revolutions – and we clearly are in the midst of a technology-driven economic revolution – have been based on the idea of an explicit (i.e. transparent) fair value exchange. For instance, in return for $850, early 20th century consumers could obtain a Ford Model T; today a paperback edition of Adam Smith's *The Wealth of Nations* can be secured in exchange for about $15. The costs and benefits to those on both sides of these value equations (usually an individual and a corporation) are both clear and easily discoverable. And the process by which the transaction is executed is well-established: rational, self-interested economic actors determine the price they are willing to pay for goods or services based on their subjective perception of their utility – something that is usually quite simple to determine.

Assessing the fair value of transactions involving data – and in particular personal data – is somewhat harder. Retailer loyalty cards are illustrative. Most consumers know that the discounts they receive via a card are provided in exchange for data they supply to the retailer. But few realise that the *primary* value to the retailer is the ability to track and analyse the spending patterns of both the individual and aggregated data sets of groups of consumers. There are, then, significant information asymmetries in the transaction, and the average consumer lacks the information required to make a rational decision about whether she or he should participate in it.

If a truly sustainable data-driven economy is to be established, the way in which data are traded between individuals and corporations will require a major reset. For such an economy to succeed, individuals will need to receive fair monetary compensation for each specific datum they provide, perhaps with additional payments whenever that datum produces incremental profits for the entity to which it has been passed. Such systems would be technically complex to implement and administer, but will be essential to (re)establish the concept of fair value exchange in a world increasingly dominated by data. It may also go some way to assuaging privacy concerns [47].

Such is the dilemma of our time; no different from that of Glaucon, who wanted the benefits of a sophisticated society without the associated infrastructure of armies, a justice system, a guardian and so on. Yet it remains a research question as to how far it is possible to have the empowerment benefit of datafication without the surveillance cost, and how this dynamic balance is affected by context. The 2014 Digital Enlightenment Yearbook gathers together the science, social science, law and politics of this new environment in order to help us reformulate and address these questions – which are extremely timely and pressing, as Neelie Kroes, Vice-President of the European Commission with responsibility for the Digital Agenda, reminds us in her Foreword.

With this collection of chapters, we hope to facilitate this dialogue, putting forth different perspectives on how individuals can gain control of their personal data, mechanisms (legal, economic, and technical) to enable them to do this, uses of data to empower communities at large, and new approaches to consider for understanding and enabling the social machine. Data and social machines are the themes *du jour* -shiny new techno-toys that promise to solve the problems of the world. With this in mind, Ambrose's prologue, which puts the current age into perspective, reminding us that we have been here before, is essential reading.

Acknowledgements

KOH was supported by SOCIAM: The Theory and Practice of Social Machines, funded by the UK Engineering and Physical Sciences Research Council (EPSRC), under grant number EP/J017728/1. Thanks to Scott L. David and David De Roure for their input into this project and the ideas in this introduction.

Our thanks also to Jacques Bus and the Digital Enlightenment Forum for their confidence in us as editors, to Maarten Fröhlich of IOS Press, and to the legion of reviewers who helped us in our editing tasks.

References

[1] A.D.I. Kramer, J.E. Guillory and J.T. Hancock, Experimental evidence of massive-scale emotional contagion through social networks, *Proceedings of the National Academy of Science* **111** (2014), 8788–8790.

[2] R. Lemov, *World As Laboratory: Experiments With Mice, Mazes and Men*, Hill & Wang, New York, 2005.

[3] C. Rudder, We experiment on human beings! *OKTrends*, 28th Jul, 2014, http://blog.okcupid.com/index.php/we-experiment-on-human-beings/

[4] T. Berners-Lee, *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web*, HarperCollins, New York, 1999.

[5] K. O'Hara, Trust in social machines: The challenges, in *Proceedings of the AISB/IACAP World Congress 2012: Social Computing, Social Cognition, Social Networks and Multiagent Systems (SOCIAL TURN/SNAMAS)*, http://eprints.soton.ac.uk/339703/, 2012.

[6] R.A. Clarke, M.J. Morell, G.R. Stone, C.R. Sunstein and P. Swire, *The NSA Report: Liberty and Security in a Changing World*, Princeton University Press, Princeton, 2014.

[7] G. Greenwald, *No Place to Hide: Edward Snowden, the NSA and the Surveillance State*, Hamish Hamilton, London, 2014.

[8] K. O'Hara and N. Shadbolt, Who guards the guardians? *Science* **345**(6195) (2014), 387.

[9] H. Nissenbaum, Privacy as contextual integrity, *Washington Law Review* **79** (2004), 119–157.

[10] The White House, Consumer Data Privacy ina Networked World: A Framework for Protecting Privacy and Promoting Innovation in the Global Digital Economy, U.S. Government, Washington, D.C., 2012.

[11] Federal Trade Commission, *Protecting Consumer Privacy in an Era of Rapid Change*, U.S. Government, Washington, D.C., 2012.

[12] World Economic Forum, Rethinking Personal Data: Trust and Context in User-Centered Data Ecosystems, Geneva, 2014.

[13] World Economic Forum, The Internet Trust Bubble: Global Values, Beliefs, and Practices, Geneva, 2014.

[14] http://www.tenisonroad.com/project

[15] L. Floridi, Open Data, Data Protection, and Group Privacy, *Philosophy & Technology* **27** (2014), 1–3.

[16] J.C. Scott, *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*, Yale University Press, New Haven, 1998.

[17] The White House, Big Data: Seizing Opportunities, Preserving Values, Washington, D.C., 2014.

[18] http://www.newurbanmechanics.org

[19] J. Hendler and T. Berners-Lee, From Semantic Web to social machines: A research challenge for AI on the World Wide Web, *Artificial Intelligence* **174** (2013), 156–161.

[20] N. Shadbolt, D. Smith, E. Simperl, M. Van Kleek, Y. Yang and W. Hall, Towards a classification framework for social machines, in *Proceedings of SOCM2013: The Theory and Practice of Social Machines*, Rio, http://eprints.soton.ac.uk/350513/, 2013.

[21] K. O'Hara, N.S. Contractor, W. Hall, J.A. Hendler and N. Shadbolt, Web Science: understanding the emergence of macro-level features on the World Wide Web, *Foundations and Trends in Web Science* **4** (2013), 103–267.

[22] A. Bernstein, M. Klein and T.W. Malone, Programming the global brain, *Communications of the ACM* **55**(5) (2012), 41–43.

[23] M. Kwiatkowska, R. Milner and V. Sassone, Science for global ubiquitous computing, *Bulletin of the European Association of Theoretical Computer Science* **82** (2004), 325–333, http://eatcs.org/images/bulletin/beatcs82.pdf, 2004.

[24] C.J. Lintott, K. Schawinski, A. Slosar, K. Land, S. Bamford, D. Thomas, M.J. Raddick, R.C. Nichol, A. Szalay, D. Andreescu, P. Murray and J. Vandenberg, Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey, *Monthly Notices of the Royal Astronomical Society* **389** (2008), 1179–1189.

[25] O. Okolloh, Ushahidi, or "testimony": Web 2.0 tools for crowdsourcing crisis information, *Participatory Learning and Action* **59** (2009), 65–70.

[26] N. Morrow, N. Mock, A. Papendieck and N. Kocmich, *Independent Evaluation of the Ushahidi Haiti Project*, Development Information Systems International, http://ggs684.pbworks.com/w/file/fetch/60819963/1282.pdf, 2011.

[27] D. Robertson and F. Giunchiglia, Programming the social computer, *Philosophical Transactions of the Royal Society A: Mathematical Physical and Engineering Sciences* **371**(1987) (2013).

[28] L. Von Ahn, B. Maurer, C. McMillen, D. Abraham and M. Blum, reCAPTCHA: human-based character recognition via Web security measures, *Science* **321** (2008), 1465–1468.

[29] L. Von Ahn, M. Blum, N.J. Hopper and J. Langford, CAPTCHA: using hard AI problems for security, in E. Biham (ed.), *Advances in Cryptology: EUROCRYPT 2003*, Springer-Verlag, Berlin, 2003, 294–311.

[30] J. Kleinberg and P. Raghavan, Query incentive networks, in *Proceedings of the 46th Annual IEEE Symposium of Foundations of Computer Science* (FOCS'05), Pittsburgh, 2005, 132–141.

[31] G. Pickard, I. Rahwan, W. Pan, M. Cebrian, R. Crane, A. Madan and A. Pentland, *Time Critical Social Mobilization: The DARPA Network Challenge Winning Strategy*, arXiv.org 1008.3172v1, http://hd.media.mit.edu/tech-reports/TR-660.pdf, 2010.

[32] M. Van Kleek, D. Smith, W. Hall and N. Shadbolt, "The crowd keeps me in shape": social psychology and the present and future of health social machines, in *Proceedings of SOCM2013: The Theory and Practice of Social Machines*, Rio, http://eprints.soton.ac.uk/350511/, 2013.

[33] P. Walker, Boston bombing identification attempts on social media end in farce, *The Guardian*, 19th April, 2013, http://www.guardian.co.uk/world/2013/apr/19/boston-bombing-suspects-reddit-social-media

[34] H. Arendt, *Eichmann in Jerusalem: A Report on the Banality of Evil*, Penguin, Harmondsworth, 1977.

[35] G. Bell and J. Gemmell, *Total Recall: How the E-Memory Revolution Will Change Everything*, Dutton, New York, 2009.

[36] K. O'Hara, The information spring, *IEEE Internet Computing* **18**(2) (2014), 79–83.

[37] N. Shadbolt and K. O'Hara, Linked data in government, *IEEE Internet Computing* **17**(4) (2013), 72–77.

[38] D.J. Weitzner, H. Abelson, T. Berners-Lee, J. Feigenbaum, J. Hendler and G.J. Sussman, Information accountability, *Communications of the ACM* **51**(6) (2008), 82–87.

[39] V. Mayer-Schönberger and K. Cukier, *Big Data: A Revolution That Will Transform How We Live, Work and Think*, Houghton Mifflin, New York, 2013.

[40] A. Narayanan and V. Shmatikov, Privacy and security: myths and fallacies of "personally identifiable information", *Communications of the ACM* **53**(6) (2010), 24–26.

[41] T. Ray, Internet of Things: Mammoth Morgan Stanley Note Tries to Explain It, *Tech Trader Daily*, 26th September, 2013.

[42] United Nations, *World Population Prospects: The 2012 Revision*, New York, 2013.

[43] http://open.sen.se

[44] R. Haladjian, presentation at LeWeb Paris 2012, http://jumahe.tumblr.com/post/37325037882/rafi-haladjian-founder-of-violet-nabaztag-and

[45] K. O'Hara, The fridge's brain sure ain't the icebox, *IEEE Internet Computing* **18**(6) (2014).

[46] J. Lanier, *Who Owns the Future*, Simon & Schuster, New York, 2013.

[47] K. O'Hara and N. Shadbolt, Privacy on the data web, *Communications of the ACM* **53**(3) (2010), 39–41.