nominettrust

Open data and charities

A STATE OF THE ART REVIEW written for Nominet Trust by Wendy Hall, Nigel Shadbolt, Thanassis Tiropanis, Kieron O'Hara and Tim Davies

This document is designed to be read with **Adobe Acrobat**

Foreword

Open data provides opportunities to develop new ways of addressing persistent social challenges. By utilising networked and computation technology, we can begin to understand the challenges in more detail and find new ways to work together to develop more effective interventions.

That said, the understanding, use and publishing of open data within the third sector is still very much in its infancy. Yet the promise of these emerging technologies and practices demand further investigation and effort. As a starting point, we're grateful to the range of authors for putting together this review, and we look forward to working with you to build on the recommendations and challenges it sets out.

Dan Sutch
Head of Development Research
Nominet Trust - July 2012

About the series

Nominet Trust State of the Art Reviews are undertaken by leading academics to collate and analyse the latest research at the intersection of the internet and society. Drawing on national and international work, these reviews aim to share the latest research to inform the work of the Trust, those applying to the Trust and our wider partner organisations.

We value your comments and suggestions for how to act on the recommendations in these Reviews, and how we can build the series, so that it is more useful to us all as we explore how digital technology can be used to design radically new solutions to address specific social problems.

We look forward to your comments and suggestions at:

developmentresearch@nominettrust.org.uk

Contents

Executive summary

Open data involves a paradigm shift in the way organisations manage their information and data: moving from a default of charities keeping data resources locked up in under-used internal systems, to building a shared 'Web of Data'. The emergence of the open data movement has supported powerful new models of creativity, innovation and public engagement.

Although most of the recent stories of progress and success in the open data field come from government and research where open data is more established, this report sets out to explain the ways in which open data is increasingly coming to play a role in the charitable sector. Existing open government data can be used by charities to add value to their work, to target services better, to improve advocacy and fundraising, and to support knowledge sharing and collaboration between different charities and agencies. Crowdsourcing of open data also offers a new way for charities to gather intelligence, and a wide range of freely available online tools can support charities to analyse open data resources. Realising the potential of open data will require charities to meet a number of technical and organisational challenges. Indeed many charities will need to address key issues relating to open data, whether they choose to pursue benefits from open data or not (as regulatory, funding and performance indication is published online by researchers, by government and by others in the sector).

This report reviews open data as it relates to the charitable sector, drawing on long experience of developing open data in research and government, as well as early work exploring open data with charities and third-sector organisations. It defines open data, describes the background context of a knowledge economy, and outlines key opportunities and challenges of open data in the charity sector.

Existing open government data can be used by charities to add value to their work, to target services better, to improve advocacy and fundraising

On the basis of this overview and analysis of the field, the report sets out in more detail a number of options for the further development of open data practices in the charitable sector, via five recommendations derived from the analysis, as follows:

— Open data will be an increasingly important part of the information landscape. **Charities will need to understand this new landscape** and be able to operate in it. They can benefit from open data in a wide range of ways, and should look to develop their capacity to be both creators and users of open data. Organisations of any size can take the first steps to produce and consume open data.

— There are many opportunities to increase the flow and sharing of relevant datasets within the charity sector, to support engagement between charities and the public, government, funding bodies and other charities. The deployment of open data portals for the UK third sector could help **bridge the information sharing gap** on a national or regional level and it could provide for more intense public engagement, innovation, and impact. Infrastructure organisations have a key role to play in bridging the data gap, though much will also need to take place through gradual, bottom-up work on task-specific open data sharing in particular contexts.

— Charities may need to i**nvest in capacity building** in the open data world. Apart from open data portals, it is likely that they will have to invest in the development of light-weight standards to ensure harmonisation of their data infrastructures and better information sharing. This should involve a focus on web standards and open standards. Funders have an important role to play in supporting the development of standards and promoting their adoption.

— Charities need to be aware of, and to have implemented, **good information processing practices** (even in the absence of open data). At the same time, they should be aware of safeguarding issues when publishing open data and have strategies for managing related risks. They should be focused on retaining trust and maintaining the privacy of data subjects, as well as complying with their responsibilities under the Data Protection Act.

— Publishing open data will produce gains for charities whose data will help inform the decisions of others about their areas of expertise, and about the issues for which they take responsibility. As with most public goods[1], the tangible gain from a particular publication event will be small for a charity – the gain is less easily measurable, and over a longer term, and consists of the network effects of a growing quantity of linked data with a growing online community. Network effects are not inevitable, and **continued attention and action to support their emergence will be important** to make sure the charity sector benefits to the maximum possible extent from open data.

1
In economics, a 'public good' is a good that is both non-excludable and non-rivalrous in that individuals cannot be effectively excluded from use and where use by one individual does not reduce availability to others. **http://en.wikipedia.org/wiki/Public_good**

Introducing open data

Network effects: Web 2.0 and the Web of Data

The web initially made it possible for people to publish and access documents online. The higher the number of users publishing documents on the web, the higher the value of the web to its users. In turn, the higher the value of the web to its users, the higher the number of users who joined in bringing additional value (this is an instance of Metcalfe's Law, that the benefits to users of a network are proportional to the square of the number of connected users, so that each new addition to the network has a disproportionately positive effect on the value of the network). This process of increased adoption leading to increased value and vice-versa is described as a 'network effect' and has been instrumental to the growth of the web right from the beginning. The conditions that can instigate such network effects have been the subject of analysis since the inception of the web, but evidence suggests that the adoption of standards, the ease of navigation with simple tools such as web browsers, and the availability of content have all been significant components.

Early web users were primarily consumers of online resources or services. However, the emergence of 'Web 2.0' online tools and services such as blogs, wikis and social networks led to many users adding content to the web – creating value on Web 2.0 social platforms. For example, YouTube allows users to publish videos online without having to worry about running and maintaining their own web servers. The more videos uploaded by users, the higher the value of YouTube as a service to existing and new users. The main characteristic of Web 2.0 has been the emergence of services that derive value from user generated content and that leverage network effects. With Web 2.0, a number of new business models have emerged that employ a number of revenue models with free services, premium services or a mix of both ("freemium" services)[2].

The abundance of Web 2.0 services, the growth of content available on the web and the openness of web standards has also enabled a new generation of services that combine data from different sources to provide added-value services. For example, services that combine mapping data with geo-tagged photos and videos to display content related to a specific geographic area.

The main characteristic of Web 2.0 has been the emergence of services that derive value from user generated content and that leverage network effects

[2]
It is worth noting that often Web 2.0 firms offer considerable discounts to charities using their platforms.

These services (commonly called 'mashups') demonstrate the value of publishing data on the web in formats that are easy to combine with other data. In addition, the emergence of mashups has shown that an increasing number of web users are able to programme novel applications leveraging published datasets.

We are now witnessing a new stage of web evolution that is leading to a Web of Data. An increasing volume of data is becoming available on the web as open data in standard formats whilst an increasing number of tools enable the average web user to contribute and to process data online. In addition, large content repositories are making metadata available about the content they hold, supporting not only discovery of abstract facts in datasets, but also supporting more effective discovery of content and information (Powell et. al., 2012). For example, the hierarchical structure of topics available on the user-generated encyclopaedia Wikipedia is available as structured and linked data through the DBpedia service, supporting a wide range of platforms to re-use Wikipedia information in novel ways. The growth of available data on the web, their potential for value-added services and the availability of standards seem to set the conditions for new network effects. In the Web of Data the user is not just a producer of content but also a producer of data and a producer of applications that combine available data sources to provide innovative applications. It is therefore possible to 'crowdsource' not only contributions to open datasets but also the development of visualisations, applications and analysis that will add value to them.

DEFINING OPEN DATA

All modern organisations rely on data: from lists of members or service users, to funding and finance records; from geographic information and maps of where services are delivered, to performance statistics, feedback and research data. Often this data is held in disparate databases and spreadsheets stored on servers, laptops or USB sticks.

An increasing volume of data is becoming available on the web as open data in standard formats whilst an increasing number of tools enable the average web user to contribute and to process data online

The World Wide Web initially provided people a way to publish documents online, helping to take documents from disparate organisational silos and to make them available online, connected up through hyperlinks. Over recent years technologies and practices have developed that seek to do the same for datasets: unlocking the value of data. Just as it is possible to share documents on the web that are only accessible to a few people, these technologies can often support private data sharing. However, as experience and research on the web shows, it is when content is openly accessible and linked that network-effects are generated which add value to both the content and to the network as a whole.

Generally, open data is defined by three criteria. A dataset is open data if:

1. it is made accessible online
2. it is published in an open machine readable format
3. it is licensed to allow others to re-use it.

(2) and (3) are important because they reduce the friction of working with and combining different datasets, increasing the chance that a dataset can support innovation, collaboration and added-value services. Publishing in an open, machine readable format (2) means that someone wishing to use a dataset shouldn't need expensive proprietary software to access and use it. Often this will be in CSV (comma separated values) format instead of Excel[3], although the linked data principles outlined in Section 3.2.1 offers a 'gold standard' approach to choosing an open format.

A range of licences exists for open data. A licence sets out what someone can and can't do with the data without additional consent. When data does not have an explicit open licence, people who want to use it cannot be sure what they are allowed to do with the data or not, and the data cannot be re-used in other openly licensed tools and resources. For example, Wikipedia uses an open licence. If you happened to publish a research dataset, and someone wanted to draw upon it to create a page on Wikipedia about the research topic, the lack of a clear licence could prevent them from drawing on your research.

It is when content is openly accessible and linked that network-effects are generated which add value to both the content and to the network as a whole

3

Although most desktop computers have a copy of Excel on, the Excel file format is controlled by Microsoft. That means that third-party tools that want to work with spreadsheet data have an extra hurdle to overcome if it is in Excel format. CSV is an open standard for spreadsheet-like data that any third-party tool can freely use to exchange data.

Vignette 1
Licences

Open licences work by the owner of the copyright or intellectual property rights over a digital asset (e.g. a database) publicly stating the permissions they grant to others over that asset. Without a licence, each re-user of a dataset would have to seek individual permission from its owner. There are a range of different licences that can be used for open data.

The Open Database Licence (ODbL) grants permission to third-parties to re-use a dataset, subject to the requirement that the re-user attributes the source of the data, shares any derivative works under the same licence (i.e. the re-user can't, without separate permission, restrict re-use of value-added products created on the back of the dataset), and maintains the openness of any redistributions of the data.

The UK Government has development an 'Open Government Licence' that could be applied to all government data and content. The OGL licence includes additional terms requiring that re-users do not use the data in any way that suggests governmental endorsement, or that would "mislead others or misrepresent the Information or its source".

The Creative Commons Zero (CC0) licence, or Public Domain Dedication allow dataset owners to waive their intellectual property rights over content or a dataset in their entirety, and to permit any future distribution, re-use or remixing of the data.

For more on open data licences see **http://opendatacommons.org/licenses**

ODC Open Database License (ODbL) Summary

This is a human-readable summary of the ODbL 1.0 license. Please see the disclaimer below.

You are free:

To Share: To copy, distribute and use the database.

To Create: To produce works from the database.

To Adapt: To modify, transform and build upon the database.

Many kinds of (open) data

Data can have many different properties. Datasets might be a single spreadsheet updated once a year, or might be held in large-scale, constantly updated database systems. The table below[4] sets out some different terms used to describe different types of data. Most of these can be combined, so that you could have, for example, open real-time big data; or linked open data.

	Definitions	Potential implications
Open data	Datasets that are made accessible in non-proprietary formats under licences that permit unrestricted re-use (OKF – Open Knowledge Foundation, 2006). Open government data involves governments providing many of their datasets online in this way.	Third-parties can innovate with open data, generating social and economic benefits. Citizens and advocacy groups can use open government data to hold state institutions to account. Data can be shared between institutions with less friction.
Big data	Data that requires 'massive' computing power to process (Crawford and Boyd, 2011). Massive computing power, originally only available on supercomputers, is increasingly available on desktop computers or via low cost cloud computing.	Companies and researchers can 'data mine' vast data resources, to identify trends and patterns. Big data is often generated by combining different datasets. Digital traces from individuals and companies are increasingly captured and stored for their potential value as 'big data'.

	Definitions	Potential implications
Raw data	Primary data, as collected or measured direct from the source. Or Data in a form that allows it to be easily manipulated, sorted, filtered and remixed.	Access to raw data can allow journalists, researchers and citizens to 'fact check' official analysis. Programmers are interested in building innovative services with raw data.
Real-time data	Data measured and made accessible with minimal delay. Often accessed over the web as a stream of data through APIs (Application Programming Interfaces)	Real-time data supports rapid identifications trends. Data can support the development of 'early warning systems' (e.g. Google Flu Trends; Ushahidi). 'Smart systems' and 'smart cities' can be configured to respond to real-time data and adapt to changing circumstances.
Linked data	Datasets are published in a format (for instance RDF) facilitating the use of URLs (web addresses) to identify the elements they contain, with links made between datasets (Berners-Lee, 2009; Shadbolt, Hall and Berners-Lee, 2006).	A 'web of linked data' emerges, supporting 'smart applications' (Allemang and Hendler, 2008) that can follow the links between datasets. This provides the foundations for the Semantic Web.

	Definitions	Potential implications
Personal /private data	Data about an individual that they have a right to control access to. Such data might be gathered by companies, governments or other third-parties in order to provide a service to someone, or as part of regulatory and law-enforcement activities.	Many big and raw datasets are based on aggregating personal data, and combining them with other data. Effective anonymisation of personal data is difficult, particularly when open data provides the pieces for 'jigsaw identification' of private facts about people (Ohm, 2010).

The only combination that requires particular caution is 'raw personal data'. Whilst there are discussions about opening this data up to the individuals who the data is about (for example, in the Government's MiData programme of personal data access), this sort of data should never be directly published as openly licensed and accessible data without explicit consent of the individuals covered in the data. Section 2.4.3 highlights some of the privacy issues involved in opening up datasets based on personal data.

OPEN DATA IN THE KNOWLEDGE ECONOMY

The successful deployment of open data builds upon the policies and practices that have developed alongside the World Wide Web. The web has ushered in a different model of information economics. In the traditional economy, the value of information depended on scarcity, and our legal framework largely reflects that.

Copyright, trade secrets and patents all allow the creators of data and information to establish information asymmetries, and to develop income streams by exploiting temporary or partial monopolies, or by licensing information to others for a fee. In the knowledge economy, the value of information is based not on scarcity but on abundance.

The web in particular, a technology based on copying and information sharing via the Internet, has broken down the boundaries that formerly preserved information asymmetries. A new economy is developing in, in which (a) information can be shared far more easily, (b) preventing the spread of information is more difficult, and (c) it is easier to amalgamate information from heterogeneous sources to create larger, more complete bodies of information. The underlying principle of the knowledge economy is that such bodies of information, more comprehensive and developed from a more diverse range of view points, can support realisation of much greater economic and social value, as well as offering broader and less siloed views of particular domains.

Making data accessible online, in common formats, and licensed for re-use supports serendipitous discovery and reuse of information. This can dramatically increase the range of users of certain information, and the uses that information is put to: realising value that is locked up when proprietary scarcity-based strategies to manage data are adopted. In the knowledge economy, value does not come from hoarding or selling raw data, but from transforming it into actionable information and knowledge: putting it into use.

The emerging knowledge economy has to coexist with a legal framework based on enabling the artificial creation of information asymmetries: copyright is not going away, but with the use of careful licensing it can be used to promote rather than to restrict sharing. Management methods and information storage policies need to be reconfigured from a presumption of proprietary restrictions to a presumption of openness. Organisations (and individuals) need to explore where they can add value to data through innovation, synergies and co-operative ventures: identifying the areas where they have a competitive advantage in creating (and possibly monetising) knowledge, innovations and services based on large quantities of data and information.

In the knowledge economy, value does not come from hoarding or selling raw data, but from transforming it into actionable information and knowledge: putting it into use

The 'Open Data Movement' (a broad based coalition bringing together private sector data-driven businesses, and civic-technologists) has developed in recent years as a force pushing for a switch to a genuine knowledge economy. Open data, taking inspiration from other ideologies of openness such as open source and open access publishing, articulates the idea that data should be usable by anyone, not just the data owner (or 'data controller' in the language of the Data Protection Act). There are many alternative stakeholders with respect to particular datasets, including data subjects whose details may be in a dataset, the funders of projects that generate data, and those who could benefit from innovative services developed from data. Open data advocates argue that making datasets accessible on the web can provide the foundations for a wide range of innovations and value-generating activities in the knowledge economy.

Over recent years, key foci of open data activity have included e-science data and government data, seeking to unlock these datasets and to make them available as part of a global infrastructure of datasets on the web. Both research and government generate very large quantities of data as a result of their operation, and each is publicly funded (or at least mostly so in the case of science). This introduces an additional argument for why they should publish their data openly (aside from the extrinsic value added by sharing data). Is it legitimate to try to restrict access to information that has been collected and stored using taxpayers' money, or should this data not be contributed to supporting creation of shared value? This argument has been particularly focused on datasets such as the Ordnance Survey national mapping data, although it has been applied across government datasets.

Pressure for open government data has also emerged independently from another source: the transparency agenda. Transparency of government is a key foundation of modern democracy, with citizens given access to information needed to hold governments to account. If citizens are to perform this function easily, then they need access to the relevant information as data that they can freely reuse. Hence the data that is needed to evaluate government function is also being posted online in greater quantities.

Yu and Robinson (2012) have highlighted the important distinction between interpreting 'open government data' as data that comes from government (e.g. to support a knowledge economy), and 'open government data' as data that is used to promote more open and accountable government. Only some of the data governments publish is likely to be used in this second way.

OPEN DATA IN ACTION: EXAMPLES FROM GOVERNMENT, RESEARCH AND BUSINESS

Open data practices are now well established in many governments. The US and UK led the field with the respective development of data.gov and data.gov.uk data portals in 2009 and 2010, proving access to large quantities of public data, and there are now over 100 open government data initiatives worldwide[5]. Government datasets cover a wide range of areas of activity including transport, health, agriculture, business, law and education. In the UK, the Prime Minister has also asked local government to publish substantial quantities of open data, including details of all local public spending over £500, and details of senior staff salaries. Many local authorities are also exploring proactive publication of other datasets, including, amongst other things, the location of public services, local grant allocations and recycling statistics. Much of this data can be useful to the charity sector.

5
http://datacatalogs.org

Vignette 2
Open Charities, Open Corporates and Openly Local -
making the connections

Chris Taggart is an open data innovator, and a member of the UK Government's Local Public Data Panel. He has developed a number of open datasets including:
- **OpenlyLocal.com** – contains details of local authorities, councillors and public spending (disclosed under the governments requirement that local authorities publish as open data details of all spending over £500)
- **OpenCharities.com** – contains a copy of the charity register
- **OpenCorporates.com** – collates details on company registration from over 50 registration agencies.

Making the connections between these open datasets, and a charity's own data, offers powerful opportunities. For example, an infrastructure organisation could match details of its member charities to records in OpenCharities (using automatic matching tools), and could then draw on additional, automatically updated information about them - from their annual accounts, to their Twitter and YouTube addresses. Searching for the same charities in OpenlyLocal's public spending database can locate details of funding they have recieved from different local authorities.

In a demonstrator built with Esmee Fairbairn grants data, data from OpenCharities could be used to visualise the rough geographic spread of grants over time (as OpenCharities gives a postcode for the Charities registered location), and to compare the current size of organisations based on their most recent accounts with the size of grant they received. This mapping also draws upon open postcode-to-location datasets released as part of the UK Government's open data drive. Previously, this data was only available to buy or through subscription services at high annual cost.

Many of the government data portals that have been launched also provide access to applications that have been built on top of 'raw' open data from governments. For example, data.gov.uk and the data.london.gov.uk portals link to applications that use real-time Transport for London open data on the city's cycle-hire scheme to help citizens locate a station of the London cycle hire network and to check availability of bikes there. Similar applications exist for the London Tube, and many regional bus networks. These applications, created by independent third parties using government data, may be freely accessible online, or might be for sale as mobile applications in 'app stores'. Independent third parties are often able to develop tools using government data far more quickly and cheaply than if government developed them direct. Projects like RewiredState (**www.rewiredstate.org**) have been running events since 2009 that bring developers together to create prototype applications with public data, and to show how open government data can support innovation in public service delivery. One of the first large-scale application competitions in the United States, Apps for Democracy, in 2009, was estimated to have generated $2,600,000 in value to the City of Washington DC for an investment of $50,000[6].

In the process of opening up data the UK Government has also been able to identify problems and errors with key datasets. When it released the NAPTAN database of bus stops, it turned out that the dataset contained over 18,000 inaccurate stop locations. When the data is public, these errors can the spotted, and corrected, more easily. The NAPTAN data has been made available on the community-provided Open Street Map platform, where anyone can update the data to make corrections, creating the possibility of crowdsourcing more accurate data (Shadbolt, 2011).

In the academic field, data collection is often one of the most expensive aspects of research. Peter Murray-Rust has outlined the importance of open science data, highlighting that data availability allows findings to be checked, can support communal activities and can accelerate the pace of research (Murray-Rust, 2008). Many research questions, particularly in physics, biology and medical sciences can only be answered through large-scale collaborations involving data sharing, and opening science data can support that collaboration.

Independent third parties are often able to develop tools using government data far more quickly and cheaply than if government developed them direct

6
See **www.appsfordemocracy.org/about** (accessed 23 April 2012)

In the case of the Human Genome Project, the release of the sequenced genome into the public domain at the end of the project supported a wide range of different institutions, public and private, to collaborate in the sequencing research. Secondary access to research data and secondary access by researchers to datasets created in the processes of government can also support new research discoveries. Many publically funded social science datasets are now deposited in the UK Data Archive where they are available for re-use. Where these datasets contain information that might be disclosive they may not be fully 'open'. However, where possible, governments and funders are increasingly asking that data from public-funded research should be available in open formats and under open licences.

Businesses are also realising what the academic community has realised regarding open access to scientific work and data. Companies are already providing access to data related to their offerings, enabling third parties to use these data and offer value-added services to users (e.g. price comparison websites that draw on 'Good Relations' open data). One benefit for companies that allow third parties to use their data is the potential to reach additional customers (users of the third-party services); they may also benefit from savings by providing data about open supply chains, and enhance their brand reputation through making open data available (Shadbolt, 2011). A recent McKinsey report outlined how firms can gain an edge over their competitors by sharing information with customers and suppliers strategically over the web (Bughin and Chui, 2010). In addition, by providing open data companies can crowdsource added-value applications that further enhance their offerings. To some extent, firms like Twitter are built on opening access to their core data (tweets) so that third parties can build applications – making Twitter accessible on a wide range of platforms, leading to the development of a wide range of value-added applications: creating an ecosystem of applications and services.

Open data provides an opportunity to produce social value. We must now go on to ask what specific opportunities and challenges are emerging for charities to benefit from open data, as both consumers and producers.

A recent McKinsey report outlined how firms can gain an edge over their competitors by sharing information with customers and suppliers strategically over the web

Open data for individual charities: opportunities and challenges

Whilst open data has been theorised as an issue for the public sector (legitimacy and transparency) and the private sector (improving efficiency while preserving competitive advantage; supporting new economic models), not much has been written about how open data, and the emerging knowledge economy more generally, will impact, positively or negatively, upon the charitable sector.

The charity context

Charities have a specific legal structure, operating subject to a non-distribution constraint (Hansmann, 1980) that means any surpluses generated by the charity cannot be appropriated by employees or shareholders, but must instead be used for the stated mission and purpose of the organisation. The rationale for non-profit provision of services typically revolves around the need to address asymmetric relationships between producers and consumers. Relationships of trust are vital to many services, and are more likely, it is said, when a service user (e.g. a patient) knows that the judgement of a professional (e.g. a doctor) is not going to be clouded by financial gain. Charities also operate where it is difficult to specify prices, products and outcomes through market mechanisms, providing social value that would otherwise be undersupplied by market mechanisms. Furthermore non-profit legal forms can facilitate the involvement of a range of stakeholders in organisational governance, including workers, volunteers, and service users. Voluntary organisations have also, historically, been a means whereby the needs of minority, vulnerable, unpopular or hard-to-reach groups have been met, in circumstances where there is no democratic mandate for public provision of assistance.

Identifying ways in which open data will impact charities is complicated by the extremely heterogeneous nature of charitable bodies. Alongside a small number of very large national and international organisations that have an annual turnover in excess of £100 million each, there are thousands of smaller charities working across the country.

From clubs and community groups, through to grant making trusts, research and campaigning organisations, and to service providers and the infrastructure organisations that support them, the charity sector is made up of diverse organisations operating many levels of scale. By number though, we must keep in mind that the vast majority of registered charities in England and Wales have an income of £10,000 or less, most organisations are small and volunteer-run, and involve individuals working on a one-to-one basis with users of services. The extent to which different charities are impacted by the developing knowledge economy, and the extent of their capacity to engage with it, will depend both on their thematic areas of work, and upon their size and scale, as well as on the support available to them. In this report we take a necessarily broad approach to identify potential impacts of open data on charities, focusing on general learning that is likely to apply in some form across charitable contexts, even though it may need to be additionally contextualised for particular groups.

The extent to which different charities are impacted by the developing knowledge economy, will depend both on their thematic areas of work, and upon their size and scale, as well as on the support available to them
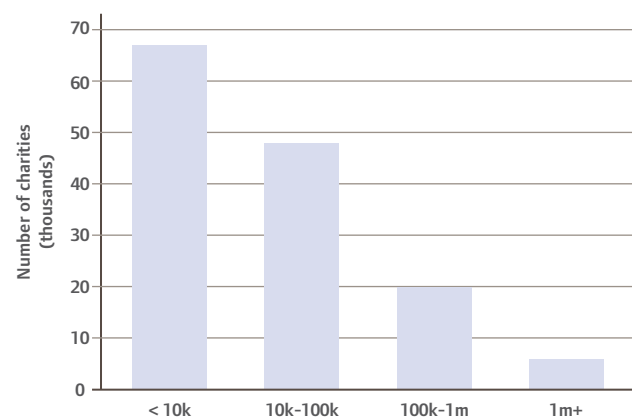

Figure A: Number of charities (in thousands) by 2009 income (£)

Note: Vertical axis shows number of charities, in thousands. 2009 income categories: <10k: less than £10,000; 10k-100k: between £10,000 and £100,000; 100k-1m: between £100,000 and £1,000,000; 1m+: more than £1,000,000. Where an organisation performs more than one role, it appears in more than one panel of the figure.
Source: analysis of Charity Commission Register. c.142,000 charities with a reported income figure considered, from a total charitable population of c.162,000 main charities.

Figure B: Number of charities (in thousands) by 2009 income (£), for charities performing different roles

Note: Vertical axis shows number of charities, in thousands. 2009 income categories: <10k: less than £10,000; 10k–100k: between £10,000 and £100,000; 100k–1m: between £100,000 and £1,000,000; 1m+: more than £1,000,000. Where an organisation performs more than one role, it appears in more than one panel of the figure.
Source: analysis of Charity Commission Register. c.142,000 charities with a reported income figure considered, from a total charitable population of c.162,000 main charities.
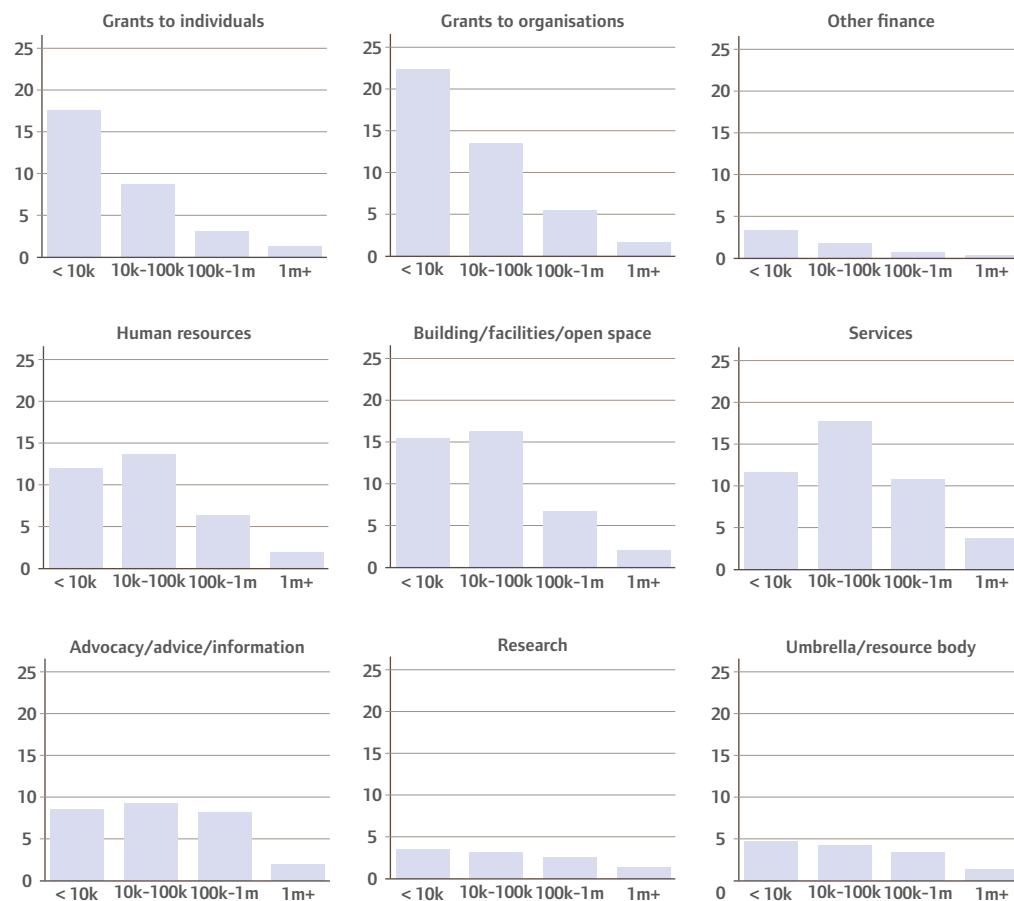
A number of charities are already exploring the potential of open data: for research and intelligence; to facilitate collaboration; to support campaigning and advocacy work; to increase accountability and streamline reporting; to improve fundraising; and to innovate in more effective and targeted service delivery. In early 2012 Nominet Trust funded ten 'Open Data Day' pilots to explore how open data might support small and medium sized charities, looking both at how charities can consume and produce open data, and identifying key challenges for charities to address. This section steps through the potential benefits of open data for charities, before looking at the practical issues of consuming and producing open data at the level of an individual charity. In the following sections we look at how open data may impact on a whole sector, and wider issues of open data ecosystems.

Vignette 3
Open Data Days

In early 2012 Nominet Trust sponsored a series of 'Open Data Days' with ten charities. Each charity received a day of consultancy from an open data expert, who met with them to identify opportunities to create or use open data, and sought to explore the potential of open data by creating rapid prototypes using open data over the course of the working day. Prototypes created included a new 'members map' for Urban Forum, drawing on their existing organisational membership database and making it available as open data which could be visualised on an interactive map, and a tool to help a local Council of Voluntary Organisations to clean up their database of community services.

Many of the open data days also included discussions about open data strategies, looking at ways in which charities could publish raw data alongside their annual reports to allow the creation of interactive graphs, and exploring opportunities for open data to support more effective commissioning of services and collaboration between local councils, the NHS and charities.

You can read more about the open data days at **http://opening-doors.posterous. com** and **www.timdavies.org.uk/2012/01/10/exploring-open-charity-data-with-nominet-trust**

Potential benefits of open data for individual charities

In general terms, the authors of this report believe there is little doubt that the charitable sector can benefit from open data. Open data has potential advantages for the charity sector in a number of ways including:

1. **Driving policy formation and planning.** Access to a range of data, from government, business and other charities, can allow for longitudinal analysis of trends and comparative analysis of local areas. This supports the identification of emerging needs or underperformance in particular settings. Access to open data can help organisations to set and monitor strategic priorities.

2. **Joining up responses and supporting collaboration.** In-depth data from heterogeneous sources about a particular area or issue can support coordination between different organisations and service providers. Regular reporting of open data on how charities are responding to particular needs can help identify overlaps or gaps in provision, and can support more agile service delivery.

3. **Providing more integrated services.** Similarly, particular services do not have to be delivered to clients in isolation, but can be tailored for the particular circumstances of clients. For example, someone's healthcare issues are likely to connect with child-care and housing issues. When key data from organisations are openly available, citizens can be empowered and supported to access more joined up services. Technical intermediaries can develop services which provide joined-up interfaces to a mixed ecology of public, private and charitable services.

4. **Supporting advocacy and campaigning.** Governments and companies often have an informational advantage over communities, individuals and charities that can make it hard to challenge decisions. Open data can help remove the information asymmetries, and can make it easier for an opposing case to be assembled. Charities can also produce open data that challenge an official interpretation of events, or adds to the shared evidence base supporting particular policy change.

5. **Enhancing impact evaluation.** Where it has been collected, opening good data can be extremely useful in showing the positive (and negative) effects of even small-scale experiments in the charitable sector, and enabling third-parties to get involved in providing data analysis or carrying out secondary data analysis across a range of evaluations. Open data can drive standardisation of metrics, and could become extremely important, for instance, in establishing the bona fides of social impact bonds, where transparency of evaluation will also be important. Publishing open research data can contribute to the shared knowledge base on particular forms of intervention.

6. **Improving fundraising and communication.** Open data can be used to demonstrate the effects of particular programmes in particular regions, or across the national scene: supporting the creation of accessible and interactive information on charitable activities. A number of organisations have explored how open data can play a part in novel fundraising approaches[7]. Open data also plays a role in increasing the accountability of charities to the public and their funders, and can streamline reporting.

7. **Mediating between local, regional and national.** Data analysis can reveal trends at varying levels, from the very local to the wider national scale. Trends may be discernible at one level without necessarily being discernible at others. The charitable sector is dispersed across all levels, and open data is potentially very useful for determining exactly how the phenomena at these different levels relate to each other.

8. **Making commissioning more efficient.** Access to open data from government can help charities bidding to provide services to make better decisions in their proposals. Charities commissioning or procuring services based on open data and open standards are likely to benefit from a more competitive market, where proprietary access to data or proprietary standards are not able to inflate prices.

7
For example, the Akvo crowdfunding platform in the Netherlands, which makes open data feeds of project information available to syndicate funding opportunities across the web, and which provides a reporting tool to funded projects to help them provide feedback to the individuals who have funded them. The microfinance grants platform Kiva.org also provides detailed open data to support third parties and researchers to analyse the data and innovate in ways that could increase engagement and income for their platform.

9. **Harnessing volunteer and private sector innovation and effort.** Open datasets can be co-created and crowdsourced from supporters. Volunteers may be more likely to contribute to a project that takes place 'in the open', and open data projects can bring new volunteers to a charity. Open data can also support collaboration with the private sector and social enterprises, allowing third-parties to provide value-added services based on charitable data, without restricting socially useful access to that data.

Of course, the charitable sector has always used data in many of the ways outlined above, but with limited free access to data, or handling data in ways that accommodate only restricted re-use of data (Cole, 2012). Open data shifts the default restricted and closed model of data management, and offers an established framework for collaboratively working with data, maximising opportunities for re-use of data in large and small quantities, and supporting charities to link their own data into deeper and richer contexts.

Open data is often discussed alongside ideas of crowdsourcing: the idea that data will not only be contributed to, and maintained, by established organisations – but that 'the crowd' may be involved in creating, curating, updating and working with data. By 'the crowd', we mean large numbers of heterogeneous people – heterogeneous in interests, knowledge and context. No single member of the crowd can provide a global overview of a particular area or problem. Similarly, no single member will be able to challenge the expertise of someone who has studied that problem for many years. However, given a sufficiently sensitive way of collating data, the local knowledge of context of each individual in the crowd can together produce a more accurate view of the wider domain. The wider knowledge is distributed across the crowd, and the aggregation mechanism renders the knowledge usable. This has been referred to as the wisdom of crowds (Surowiecki, 2004) and indicates particularly powerful opportunities for collaboration across the sector.

Open data offers an established framework for collaboratively working with data, maximising opportunities for re-use of data in large and small quantities, and supporting charities to link their own data into deeper and richer contexts

Dutton (2011) has pointed out that many initiatives that draw on ideas of crowdsourcing are more generally seeking to make use of 'distributed public expertise' – searching for input across common organisational or disciplinary boundaries. In these cases, which Dutton terms 'citizen sourcing' in the context of input to government, it is not so much the aggregation functions of the wisdom of the crowds that matters, but the creation of networks that engage distributed expertise to input into particular problem solving.

We might then draw a number of connections between crowdsourcing and citizen sourcing and the development of open data. Firstly, a number of datasets are the product of crowd contributions, or of distributed sourcing from a wide range of actors who would not otherwise feed into closed organisational dataset construction. Secondly, aggregation of content from crowds, or the analysis of existing content by crowds, can support the identification of patterns and trends within data. There is an emerging industry developing applications and tools that can perform rapid and in-depth analysis of data aggregated from 'crowds' and aggregated from across the many publishers of open data – treated as a crowd. Important to note here is that, with the right aggregation processes, even if not all data is 'correct', it is possible to derive actionable meaning from large quantities of data. This is related to ideas of big data processing. Thirdly, just as financial markets are designed to express and circulate information for the efficient allocation of resources, crowd-sourced data may play a role in supporting non-market, and non-planned creation and allocation of value (Benkler, 2006). Lastly, crowd-input might improve the quality of published data by generating feedback loops that address concerns about releasing low quality data. This would show that the maxim from open source development that 'many eyes make bugs shallow' (Raymond, 2000) can also be applied to open datasets. That is, a crowd of users should be able to improve the quality of datasets, even without being domain experts with relation to the dataset in question.

It is not so much the aggregation functions of the wisdom of the crowds that matters, but the creation of networks that engage distributed expertise to input into particular problem solving

The practical implications of open data for charities

Consuming open data

Many organisations are unaware of the open government data resources that are now available to them online: from local domestic information held by local authorities, to regularly updated national data from central government, and detailed global data gathered by multilateral institutions. Add to the mix data from researchers, private firms and NGOs, and just about any charity will find open data that could improve its research, intelligence and planning capacity. This data is often (but not always) organised into data portals, such as:

— The London Data Store (**http://data.london.gov.uk**) which provides access to data from the City of London.

— data.gov.uk (**http://data.gov.uk**) which provides access to both local and national government datasets.

— World Bank Data Portal (**http://data.worldbank.org**) which provides access to a wide range of global development indicators.

— IATI Registry (**www.iatiregistry.org**) which is a sector specific portal providing data on aid projects published through the International Aid Transparency Initiative (IATI) to a common IATI Standard.

— The Data Hub (**www.thedatahub.org**) which is a community-led project cataloguing a wide range of data and which has 'groups' collecting 'civil society' and 'international development' datasets together.

All these (and hundreds of other data catalogues) are listed at **http://datacatalogs. org** and provide access to a wealth of datasets. Even when a dataset is not catalogued in one of these sites, it might be accessible on the web.

The Open Data Cook Book provides a step-by-step guide to sourcing open data at **http://opendatacookbook.net/wiki/recipe/sourcing_ingredients**, highlighting additional routes to access data, such as requesting it from governments through the Freedom of Information Act, or scraping data using tools like ScraperWiki that make it possible to extract open data from structured web pages.

Small open datasets can be easily analysed using familiar desktop tools such as spreadsheet software. However, much of the value in consuming open data comes from the ability to use it with a wide range of new data processing tools. When datasets are openly licensed and in standard formats it becomes possible to use a wide range of off-the-shelf tools to combine different datasets, to analyse and to visualise data, including creative interactive interfaces to explore data. As the Nominet Trust Open Data Days demonstrated, with open data, initial exploration of the possibilities can often take place at low cost in a matter of hours, rather than requiring large-scale or expensive projects (cf. also Alani et al., 2008).

> When datasets are openly licensed and in standard formats it becomes possible to use a wide range of off-the-shelf tools to combine different datasets, to analyse and to visualise data

Vignette 4
A selected open data toolbox

During the Nominet Trust Open Data Days a wide range of open data tools were used, including:

– **Google Spreadsheets.** Google spreadsheets makes it easy to upload a tabular datasets and to visualise it using in-built charting tools. The charts can be embedded in a blog as interactive content. Google spreadsheets can also pull in related data from open data sources across the web using =ImportData formula.

– **Google Fusion Tables.** Fusion Tables makes it possible to merge two different datasets on a common column, and provides easy geocoding (looking up addresses or postcodes and plotting the location on a map) and visualisation tools.

One example of Fusion tables use is to take a list of member organisations with their charity name, and to merge this with the 'Open Charities' register of charities to get detailed addresses and descriptions of the organisations.

- **Google Refine.** Run as a desktop program, Refine is a power tool for working with data. It allows you to automate common operations on data, including fetching data from the web. For example, Refine could be used to tidy up common errors in address details and then to look up the parliamentary constituency an address is in using an open data source.

- **IBM Many Eyes.** Many Eyes makes it easy to perform multiple-variable analysis of tabular datasets. Datasets can be uploaded to Many Eyes and then visualised with a range of interactive bubble charts and graphs.

- **R.** R is an open source and free statistical programme with power to match many of the large-scale commercial statistical packages that can cost upwards of £1000. Although R does have a fairly steep learning curve, accessible interfaces like R Studio are making it more user friendly, and there are many free online tutorials.

At the higher end, large scale open source tools for working with open data are also available, such as Hadoop for processing very big datasets to perform advanced analysis.

As all these tools are freely available online, when using them with data that is already published as open data there are fewer data security issues to consider.

Charities can also take advantage of applications that have already been built to make open data accessible. For example, the Seme4 SeeUK application[8] takes a wide range of local ward and local-authority-area crime, transport and education statistics and makes them accessible on an interactive map, making it possible to drill down to find out about the particular issues facing a local area, and to compare it with neighbouring areas. Because the application is built on open data, if a charity uses this application to spot overall trends, but then wants to perform a deeper analysis, the raw data should be available to support this. Charities can also come together to pool resources to commission tools that will help analyse and make use of data, or could develop innovative value-added and revenue-generating services based on available data. For example, charities publishing data using the IATI Standard (see Section 4.1.1) are currently exploring ways to pool resources to commission a range of widgets that will visualise the location of the projects they fund to supporters.
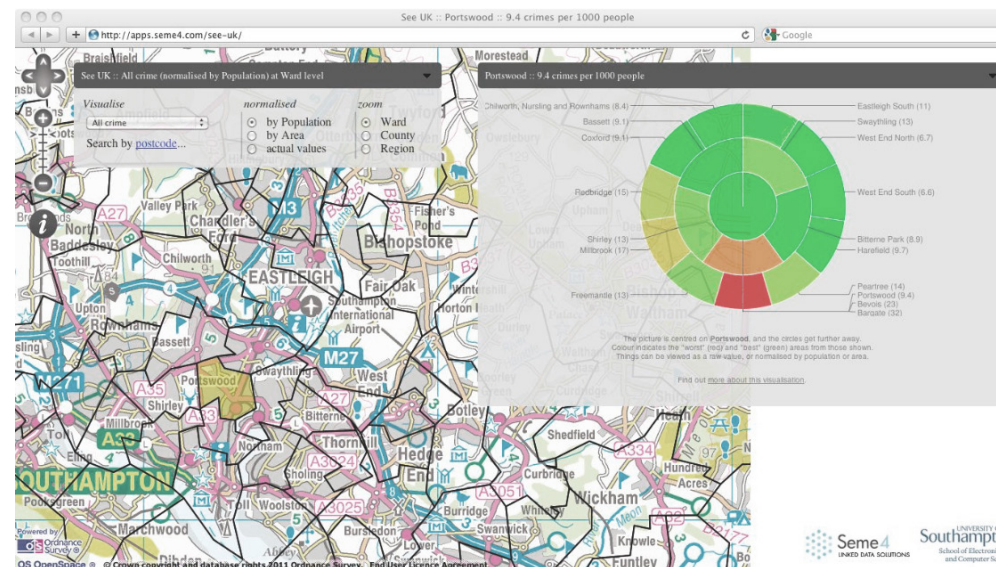

Figure 1: See-UK representing crime in the Southampton area

8
Available at **http://apps.seme4.com/see-uk** (accessed 23 April 2012)

Much of the day-to-day operational value of open data comes when an organisation can link third-party datasets with data it already holds. Where charities regularly connect their data to third-party datasets, then as the reservoir of available open data improves, it will be increasingly important to organise internal information storage and representation in ways that take advantage of possible linkages. Even if the charity was not planning to publish its own open data, it is important to identify common 'keys' and identifiers that can be used to connect datasets together and to manage data in structured and standardised ways.

This means as a first step beginning to understand the power of data as expressed in formats that make the smallest number of assumptions. Data becomes more valuable when it is possible to reuse it in unanticipated contexts – and hence it must be represented in such as a way as to facilitate this. Linking to related data online, and using other online resources is one way to provide important context for data, and to help other users understand the meaning and reference of the terms with which it is expressed. For example, instead of managing the 'location' column in a spreadsheet of projects using whatever the person entering the data types in (so that the 'location column' might contain a mix of 'London; london; Westminster; and Greater London;' to all refer to roughly the same location), using a controlled list of places, and ideally one that maps to a widely used code-list, such as ONS local authority codes (00BK in the case of Westminster), allows data to be linked to a wide range of third-party datasets that also use these codes[9].

The move toward flexible non-proprietary standards of data representation and markup is an important first step toward an organisation's exploitation of open data. As we see in figure 4 and figure 5, there are many sources of open data all of which can enhance a charity's understanding of the contexts in which it operates, and the environment of its clients. Even if a charity, in the first instance, is aiming to exploit these data sources without making a contribution to the public good, it needs to represent data in such a way as to facilitate linking into open data.

The move toward flexible non-proprietary standards of data representation and markup is an important first step toward an organisation's exploitation of open data

[9]
And even if a third-party dataset does not use exactly these codes, many standard sets of codes are mapped together and charities can leverage this open mapping. For example, the MapIT service provided by MySociety makes it possible to access detailed geographical codes for wards, local authorities and other areas. See for example
http://mapit.mysociety.org/area/2504.html

As with most technological developments, there can be adjustment costs that can look threatening, especially for organisations working with limited resources. The opportunities of open data and linking data across the web bring with them a new type of interaction with information, requiring a fresh approach to information management, information technology and data storage. Exploiting open data could require mastering techniques of linking datasets, moving from reasonably well-known proprietary data formats (e.g. Excel spreadsheets) to an understanding of open standards and formats (such as CSV format and the Resource Description Framework [RDF]). It may also require capacity building around issues of coordinating and mapping different data vocabularies together, so that an organisation is able to work with heterogeneous data sources consistently and effectively.

The pragmatics of consuming open data are not overwhelmingly difficult (Alani et al 2008), but for small organisations with limited resources (which makes it harder to devote time to back office issues of all kinds, including ICT) the task of integrating new information management paradigms can seem daunting. It is estimated that although around 750,000-1,000,000 people work in the voluntary sector, most voluntary organisations actually have no employees: at most, perhaps one third do, and where voluntary organisations have employees at all, typically they will only have one or two staff. If the charity sector is to gain the most from the opportunities of consuming open data, support is needed to progressively up-skill in data management, developing basic data management skills in small organisations, and building capacity for advanced data use in larger organisations, networks and infrastructure organisations. Supporting the emergence of open data intermediaries who will provide the tools and support to make open data accessible to small and large organisations, and brokering the development of markets that will provide data-driven analysis and services to charities may also be important to explore.

If the charity sector is to gain the most from the opportunities of consuming open data, support is needed to progressively up-skill in data management

Vignette 5
DataBridge

The DataBridge project in Brighton and Hove explored how communities, voluntary organisations and social enterprises could make use of open data, to "explore it, combine it with their own information, to help build a bigger, truer picture, to help with funding and finance, influencing [and] proving need.[10]"

Working with six different small and medium sized local organisations, DataBridge identified a number of opportunities for organisations to more proactively make use of open data sources to create analysis for internal management and business development, as well as for funders and commissioners. Many organisations were unaware of the data available from the local authority and government, even though sometimes the datasets they might benefit from were aggregated from data they had originally supplied.

The project also highlighted the potential for data published by Voluntary and Community Sector (VCS) organisations to contribute to a richer analysis of needs in a local area, noting the importance of including VCS data in local data catalogues and building the capacity of organisations to strategically publish internal data as open datasets.

The full DataBridge report is available at
http://databridgeuk.wordpress.com/findings

Ivens et al. (2011)

Producing and publishing open data

As well as consuming open data, charities may also be producers and publishers of open data. In some cases charities may be obliged to publish data: for example, in the case of development charities receiving partnership funding from the UK Department for International Development who must publish project information using the IATI Standard, or in cases where charities are delivering outsourced public services and need to publish performance and cost data to support the democratic accountability of public spending. In other cases, it is to a charity's direct strategic advantage to publish open data: for example, sharing a research dataset to influence a policy debate, or demonstrating impact and reach to attract more funding. In other cases, publishing open data is part of a contribution to a shared ecosystem of data, made more valuable for everyone the more people contribute to it. The development of this ecosystem is our primary focus in the rest of this report.

The usefulness of open data depends on data reaching the public domain, most probably on the World Wide Web, being published in a standard form, and ideally being linked with related datasets. As noted above, large-scale science, social science, and government have generally been the main suppliers of the open data currently available on the web. However, it is important for the web of open data to represent a diverse range of sectors, and so open data published by charities is an important part of the ecosystem. Charities have information and insights that are not available from other sectors that, if published online as open data, will help other charities, government, and business to get a closer and more accurate representation of challenges and solutions in a particular domain, whether in a local area, or thematically in terms of health, crime, education, poverty or other social challenges.

The practicalities of publishing data online involve many of the considerations above. Publishing data also requires selecting a licence, and making data accessible to find online. We look below at the role that data portals and intermediary organisations may play in making charity data accessible, and identify possible approaches to build the infrastructures and ecosystems that can best support the charity sector to gain value from publishing open data.

Charities have information and insights that are not available from other sectors

Vignette 6
/open (SlashOpen)

Making data accessible involves making it easy to find. Whilst larger organisations may be able to maintain their own data portals, smaller organisations (and larger ones) may primarily be listing their open datasets in third-party portals (for example, one provided by an infrastructure organisation). To make sure visitors to an organisation's website can also find their open data, a number of charities are adopting the 'slashopen' approach, and creating a page on their websites at **http://charities-website.org.uk/open** which details their policy on opening up data, and provides links to available datasets.

See **http://slashopen.net** for more information and examples.

Challenges

Whether or not charities choose to consume or publish open data, they will still need to be aware of the emerging open data landscape. Open data from government already includes information about many charities, such as the amounts that charities providing services to public bodies have been paid (through the local authority spending data which includes all payments less than £500). With spend data published and accessible for public scrutiny, charities may need to look at complementing this by publishing the related data on results, so we don't end up knowing 'the cost of everything, and the value of nothing' (Clarke, 2011). Whether proactively responding to the opportunities of open data, or simply updating practices to respond to the wider landscape, charities will need to consider a number of organisational, technical and data protection challenges.

Technical challenges

Making the most of open data requires a degree of technical capacity. Whilst we can expect to see an increasing range of products and services available to support organisations wishing to use open data, or to publish structured open data on the web, engaging with open data will still require attention paid to:

— **Data repositories, access and processing tools.** Hosting or working with large volumes of data and providing interfaces to query those data can present new technological challenges, and may require access to tools not currently in use in most charities[11]. Many useful open datasets consist of very large files that cannot be easily processed using desktop software like Excel spreadsheets, requiring charities to build their capacity to work with a range of new consumer tools (see 'A selected open data toolbox'), and to develop their use of advanced data storage and access platforms where appropriate. The market for tools for working with large volumes of (open) data is rapidly developing. A number of companies are active in the area of providing solutions for data repositories (e.g. companies providing triple stores) and there are benchmarks to assess their performance. The growth of the Web of Data may add further requirements on performance, security and support of query languages. Support for distributed queries across repositories is also expected to increase in significance, and it is a topic of research work reported to be in progress (Wang et al., 2011).

— **Metadata accuracy and provenance.** The aggregation and processing of open data relies on the selection of appropriate datasets. The appropriateness can be established based on the content of datasets, the origin of the data, the authoritativeness of the datasets or the period that they cover. For example, running data analytics for a specific period in the past requires identifying the appropriate datasets that cover that particular period. Often data is available for past periods of time, but the way the data is catalogued is inconsistent and non-standard, meaning computers cannot easily locate all the available relevant data. Good processes and tools to manage structured metadata are important both for consuming and publishing open data.

Good processes and tools to manage structured metadata are important both for consuming and publishing open data

11
Fortunately, many of the tools for working with open data are Free and Open Source Software (FOSS), meaning that the capital cost of investing in these tools is low. However, selecting the right tools, and learning to use them to achieve the desired outcomes, can take a significant initial investment of time.

– **Data quality and interoperability.** The quality of data collection and management varies widely across voluntary sector organisations. Open data creates a pressure for better data management in general, and may require charities to consider the tools and support they have available for this. When thinking in addition about opening up data and integrating open data sources, then the formats in which data becomes available and the licence under which it is published will determine the extent of its reuse (Berners-Lee, 2009; Shadbolt, 2011). A star classification system was recently proposed to provide guidance on the formats and openness of published data (Berners-Lee, 2009) – see Section 3.1.2. The effective combination of datasets of different rating on the star classification system can be technologically challenging and require technical expertise and extra resources.

ICT capacity varies widely in the charity sector, with some large organisations having advanced ICT infrastructures, whilst many small organisations rely on consumer tools. The charity sector has explored a range of models of capacity building and ICT support in the past, including the ICT 'Circuit Riders' model. In this model, organisations without dedicated ICT support or staff can draw upon a mobile worker who operates across a case load of small organisations, either centrally funded, or on a cost recovery basis (Pavitt and Lord-Soares, 2006). Some existing circuit riders may be able to add open data support to their portfolio of skills, but in other cases, a full 'data science' skill set, requiring statistical literacy and a number of advanced data processing skills, may require new actors to be involved, or capacity building with ICT support organisations as well as charities themselves. Other models of accessing ICT support, including working with interns and in collaboration with socially minded technologists may also have a role to play.

Vignette 7
DataKind and Random Hacks of Kindness

DataKind and Random Hacks of Kindness (RHOK) are two models that have been working to make greater data and technology skills available to address social challenges, including working in partnership with charities.

Random Hacks of Kindness is a network of 'hack weekends' which bring together technologists, programmers, designers and organisations with social challenges to prototype and develop helpful digital tools together. Many RHOK projects involve working with open data provided by a charity. Run on a voluntary basis, RHOK events make cutting edge technical skills available to charities, and can result in longer-term collaborations between programmers and charities around open data and digital innovation projects.

DataKind adopt a similar model of weekend events called 'DataDives', but with the focus on bringing data scientists together with social-impact organisations to explore the possibilities of a particular dataset, or to look at how in-depth analysis of available data could support the organisation's mission.

More at **www.datakind.org** and **www.rhok.org**

Organisational challenges

Although charities can take their first steps into open data through agile and small-scale experimentation, over time new business processes will need to be in place to allow organisations to use, publish and maintain open data, ensuring data quality and enabling data re-use. Learning from the private and public sector suggests that proactively responding to the emergence of open data involves addressing:

— **Culture.** The culture of many organisations in business, government and the charity sector is not favourable to open data sharing, and practices are often based on pre-knowledge economy cultures of scarcity. Establishing the value of open data publishing and placing it in the context of the functions of the various departments of an organisation takes time and intentional effort. DataBridge (Ivens et al., 2011) found that often staff did not look to information from outside the organisation, highlighting the need to build a culture of accessing as well as of sharing data.

— **Capacity building and data literacy**. The more organisations get involved in the publication, maintenance and use of datasets, the more critical will be the skills to create and publish open data, and to perform and interpret data analysis. It is expected that there will be additional demand for such skills from business with the growth of the Web of Data, and so planning for staff training, recruiting specialist staff or establishing the capacity in support organisations will be important to ensure the charity sector has the skills and resources it needs to make the most of open data.

— **Strategy, governance and management.** Selecting open datasets to rely upon, or prioritising data to share involves strategic choices. Decisions that may at first seem purely technical (the choice of data standard, or choice of data formats and vocabularies for linked data) can potentially have significant consequences for an organisation, impacting on the ease of collaboration with different partners, and affecting the settings an organisation may become aligned with through its data. For that reason it is important to support technologists and managers to work together, and to ensure there are managers who feel able to be involved in technical decisions[12].

12
For an analysis of some of the techno-social-managerial issues involved in publishing RDF Linked Data, see the draft 'Creating Linked Data' paper accessible at **http://wiki.ikmemergent.net/index. php/Workspaces:1:Linked_Open_Data** (accessed 23 April 2012): Linked Data in International Development – Practical issues, Section 3. Tim Davies (forthcoming)

- **Business processes.** As with the publication of websites, organisations will have to consider the business processes that will enable them to efficiently leverage open data. Organisations may need to envisage scenarios of open data use and set up the necessary business processes. Strategic decisions will be needed on which aspects of a data portal to deploy and maintain. Depending on their size and activity, organisations will need to decide which processes they can support themselves and which will need to be outsourced. Potential cost and timeframes for business process adjustment will be a consideration. DataBridge highlighted that many organisations may turn to their infrastructure organisations to provide analysis services and that some businesses processes and support could be provided at this level.

- **Licensing.** Organisations need to determine the licences to apply to any data they publish and to evaluate the licences of data they wish to use. They will need to consider whether any third parties hold intellectual property rights (IPRs) in the data or whether there are any contractual restrictions affecting it. Sometimes organisations will need to plan to rewrite contracts with information system providers when they come up for renewal to make it possible to open up certain datasets, or to import third-party data into a system.

- **Liability and information governance.** Maintaining and using datasets online (as with other information) can make organisations liable for the contents or consequences of using the data. Charities need to understand the implications of publishing open data and the data protection, privacy and safeguarding issues that may arise, and to have good information governance practices in place. The European Union is currently considering new policies on data protection (Europa, 2012), so organisations will need to keep abreast of the developing regulatory environment.

Data protection, privacy and safeguarding

Data protection, privacy and safeguarding considerations raise both technical and organisational challenges for charities. Data protection is a matter of law, and the extent to which data can be shared is governed by regulation[13].

13
Guidance (on data protection and how to deal with data legally, for example) is available from the Information Commissioner's Office: **www.ico.gov.uk**

Privacy is a distinct concept from data protection: whilst privacy rights are legally framed in Article 8 of the European Convention on Human Rights, the protection outlined is relatively unspecific, and privacy questions are often in practice more questions of trust than of law. Safeguarding is a broader concept again, setting out the responsibilities of organisations to ensure access to sensitive data is properly managed, and to ensure that data cannot be used in ways that are to the detriment of those an organisation has responsibilities towards. Ensuring open data practices are robust when it comes to data subjects, privacy and security is essential to maintain trust both in charities that hold personal data, and to build trust in the wider development of social good through the emergence of the Web of Data.

Many of the valuable datasets charities hold will be based upon detailed personal information such as client records. Personal data is very valuable, but is at the same time sensitive. In the public sector, for example, aggregated health data can be used to develop new treatments, or to hold particular services accountable – undeniably valuable uses of data - but if not managed carefully even aggregated data could potentially allow third parties to extract sensitive information and identify individuals (Ohm, 2010). Identifying how to use and share open data, whilst managing privacy and security is a key practical challenge for individual charities.

Given the principles of open data publication are designed to support the wide use of data, traditional security measures such as 'access control' (password protection; storing data so that only certain people can access it) and 'query control' (restricting the sorts of questions particular people can ask of a database and the data they can fetch out of it) are not available for managing access to open data. Open data is based on the idea that even minimal barriers to data re-use (e.g. compulsory registration before accessing a dataset) create friction that reduces re-use and limits the value that can be added to data, particularly by smaller scale technology developers. However, when it comes to personal data, there are limitations to the open data ideology. Privacy and data protection remain important factors in open data landscape[14].

Ensuring open data practices are robust when it comes to data subjects, privacy and security is essential to maintain trust

14
In fact, for many organisations, privacy and security practices become even more important as the availability of third-party data on the web and the possibility of accidental data leakage increases. Open data is, ironically, an important factor in raising those risks.

The Data Protection Act and the Organisation for Economic Cooperation and Development (OECD) 'use limitation' principle state that personal data should be collected for a specific use, and that a data subject's consent applies only for that use. If data is to be used for a different purpose, then consent needs to be gained. For a large dataset, this will naturally impose a major overhead to data re-use – precisely the type of obstacle that open data is intended to circumvent.

One response to this challenge is to create derivative datasets from personal data by anonymising (Kavoukian and El Emam, 2011), aggregating, perturbing or pseudonymising data. For example, aggregation may take a dataset of issues faced by a charity's service users and aggregate it to highlight the number of people in different geographic areas who are facing particular types of housing issues. In this case, there are no detailed records on individuals, and the data is broadly about the experience of the service provider, rather than being personal data about individuals. Pseudonymisation, by contrast, maintains records about individuals in a dataset, but tries to remove information that would connect a record to a specific identifiable person. For example, in a health dataset we would not know who X is, but the dataset would tell us that X, a male smoker whose calorific intake for 2009 was 25% higher than the national average is the same person as the X who had a heart attack in 2012: allowing detailed research and analysis. However, even with these methods, re-identification of individuals may still be possible (Ohm, 2010; O'Hara, 2011). Hence whilst there are viable strategies for restricting the possibility that personal data will be disclosed in releasing open datasets, any dataset based on personal data should be handled carefully, and charities will need to consider the skills and support they need to do this.

There are also further considerations in aggregating data, or considering the potential impacts of an open data release. Whilst there are clear guidelines and statistical methodologies that can be followed whenever potentially sensitive data are being released, aggregating data may limit its value to a local community in a small geographic area, or might lead to the 'statistical invisibility' of certain small groups (e.g. sufferers of a rare condition) whose data is excluded from a dataset on the basis of an aggregation procedure.

One response to this challenge is to create derivative datasets from personal data by anonymising, aggregating, perturbing or pseudonymising data

It also needs to be noted that data protection only partially describes and protects privacy (it is also intended to facilitate data sharing). Furthermore, an invasion of privacy may not be the relevant harm; there are other ways in which data might be used to the detriment of individuals (see, for example, Raman 2012 on how the publication of land records negatively impacted on marginalised communities in Bangalore, India).

With careful attention, open data, privacy and security are not incompatible (O'Hara, 2011). Many datasets charities hold (such as details of the services they provide, research based on public records, or performance statistics), do not contain significant personal data, and so, once identified as such, can be more easily made available as open data. For datasets that contain personal data, with clear processes, it may be possible to create valuable derived datasets. The government has adopted the model of carrying out Privacy Impact Assessments on major open data releases, looking to weigh up whether or not the data to be released is personal data that should be protected (for example, the land registry assessments noted that the price paid for a house is public information, and so need not be protected as personal data.). This also assesses whether or not the release of data could reveal personal information and where liability for this might lie (for example, the NHS Information Centre Evaluation and Impact assessment on proposal to publish practice-level prescribing data highlights the risk that prescription data could be combined with other datasets to infer individuals identities). Where a privacy impact assessment suggests an organisation is not liable for possible inferences about or identifications of individuals, but a possibility of such remains, organisations might also consider the nature of possible harms that could result from identification.

Privacy and security continue to be key issues in the evolution of the knowledge economy and the web, and are not limited to open data. Widespread use of social network sites and social media has seen many people share large amounts of personal and private data with private companies or in public online spaces.

Individuals' general rights to determine how data they supply to an organisation is re-used are fundamental, but this does not always mean a default of restricting access applies in every case, as individuals may be willing to see data more widely used and shared as open data[15]. Trust in these cases comes down to charities being able to clearly explain the implications of including information in an open data release. As these issues continue to develop, 'openness about openness' will be vitally important to build trust in open data – sharing best practices on working with personal data, and being clear with clients, service users and supporters about how their data may be used, and steps you are taking to protect privacy.

Vignette 8
MiData

MiData is a Government-supported project to develop standards for giving citizens greater access to their personal data held by companies. It is not strictly an open data project (the data is given direct to the 'data subject', rather than posted publically online), but, as a 'Personal Data' programme, it is an important part of the changing relationship between individuals, organisations and data.

Initially MiData pilots will give individuals access to data from a number of utility companies. This data could support individuals to join group-buying schemes for power, gas or water – using data on their consumption patterns to secure a better deal, or could enable citizens to identify how changing their energy use could lead to cost savings. There may be opportunities for charities providing support to disadvantaged citizens to work with individuals to empower them to use their data (accessed through MiData) to secure a better deal on products and services.

Charities that hold personal data about their clients may also want to consider how they can return control of this data to those clients, empowering them with direct access to their own records.

15
For example, a person living with HIV may already be open about their status, and may be willing to share certain personal data openly to support a public campaign, or to support medical research.

From datasets to network effects: realising the benefits of open data at scale

We have noted that the greatest value of open data comes from connecting different datasets, and in Section 1.1 we outlined how the success of the World Wide Web was built upon network effects. Building on an understanding of how open data can be used in individual organisations, in this section we explore how open data operates across large, complex organisations and whole sectors.

Open data ecosystems

Success in adopting the use of open data depends on working practices, organisational contexts, incentives and risks, just as much as it requires the appropriate technological platform to be in place. We term this constellation of factors an 'open data ecosystem'. The term 'ecosystem' is chosen carefully to draw attention to the parallels between open data and biological systems where interrelated components are adapted to particular niche contexts and operate in interdependent ways, governed by the internal logics and requirements of the ecosystem. In such ecosystems, the issues directly pertaining to open data are never entirely independent of other issues; the relationship between two people or two organisations will also have effects on information sharing. Pre-existing networks of interaction, existing power and information asymmetries, levels of training and capacity, and attitudes towards risk in different organisations will all impact upon the development (or non-development) of an open data ecosystem in a particular setting. Open data ecosystems may also be more or less stable (much like biological ecosystems), depending on whether actors are heavily reliant upon certain data sources, applications or organisations, or whether data and capacity to work with it are distributed across actors within the setting.

Data ecosystems are the product of both emergence (as the autonomous actions of different agents co-create value on top of data), and of design (as governments, private firms, charities and informal networks act to develop new datasets, platforms and connections).

As open data spreads, information ecosystems will involve increasing numbers of intermediaries whose role is to take raw or near-raw data and to present it in creative ways. For example, figure 2 offers a schematic sketch of how data.gov.uk might supplement the mass media by presenting information indirectly to citizens. The mainstream media, social networks, applications (apps) built by technical 'infomediaries', governments and some charities can be involved in transforming data and information into presentations, tools, products or communications that put it to use.

The emergence of a range of intermediaries creating apps based on open data in different contexts is worth specifically noting. In the past, the mainstream media were one of the primary intermediaries between datasets and citizens – interpreting and presenting analysis of everything from school league tables to the sales of chart records. However, the increasing availability of both open data, and tools to create and distribute apps, has allowed new actors to create interactive tools that offer a more diverse set of 'interpretations' of data, or that support more personalised tools for citizens and consumers based on available data. For example, a number of school league table apps developed around the data.gov.uk portal, providing a range of different ways to evaluate local schools (Davies, 2010). Some apps are provided as a commercial proposition (for example, a number of apps drawing on public crime data sell a subscription service for the most detailed data analysis), whilst others are provided by their developers as public goods. It is hoped that in different ecosystems good, rigorous apps and informediary services will proliferate to provide greater and more diverse access to open data, and poor quality apps will fall by the wayside, although this at present remains an article of faith. Nevertheless, apps will inevitably become an important part of the developing ecosystems, and although currently many apps are aimed at the general public, specialist services are likely to emerge.

Apps will inevitably become an important part of the developing ecosystems, and although currently many apps are aimed at the general public, specialist services are likely to emerge
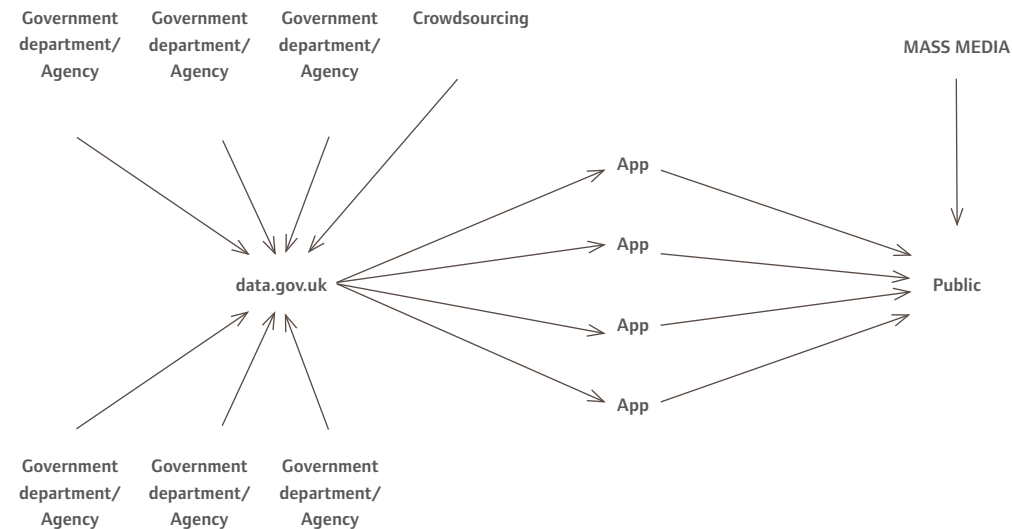
Government
department/
Agency

Government
department/
Agency

Government
department/
Agency

Crowdsourcing

MASS MEDIA

App

App

data.gov.uk

App

Public

App

Government
department/
Agency

Government
department/
Agency

Government
department/
Agency

Figure 2: An information ecosystem replete with infomediaries

The role of standards

The degree to which network effects are realised depends heavily on the licences datasets are published under, and the formats in which they are made available. Two datasets that are published to common standards under open licences are easier to combine and work with than two datasets published under difference licences and in different formats. However, imperfect open data is better than no data, and so to set out the different legal and technical levels of openness data might reach, a star classification system has been put forward.

In his five-star rating scale, Tim Berners-Lee sets out a range of assessment criteria for open data (Berners-Lee, 2009).

★ Available on the web (whatever format) but with an open licence, to be open data

★★ Available as machine-readable structured data (e.g. Excel instead of an image scan of a table)

★★★ As (2) plus: Non-proprietary format (e.g. CSV instead of Excel)

★★★★ All the above plus: Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff

★★★★★ All the above plus: Link your data to other people's data to provide context[16].

This rating scale basically codifies the value of different types and formats of data on the Web of Data. As noted earlier, open data is a public good, and so from the point of view of an individual organisation or data provider, each extra star (which marks an improvement in data quality) constitutes an overhead on that organisation. However the overheads are largely initial costs of adjustment, and so are amortisable over time, and therefore not likely to be problematic beyond the short term.

Alongside the generic five star model that primarily focuses on technical file format standards for publishing data, open data also benefits from the use of shared standards for describing particular content. When the UK Government mandated the publication of local authority spending data, it also consulted on the development of a 'Local Spending Data' standard[17], which sets out column headings to use in CSV (spreadsheet) files of spending data and provides guidance on formats and codes to use inside the files, such as using dd/mm/yyyy for dates. The guidance also sets out how this CSV (three-star data on the model above) can be transformed into linked data (four and five-stars), using a common 'vocabulary' to describe spending data, which other non-UK spending data publications might choose to use and adapt, making it possible to compare spending data across a wide range of contexts.

16
For a step-by-step interactive video introduction to what it means to use linked data standards see the first half of the embedded video at:
http://linkedinfo.ikmemergent.net

17
http://localtransparency.readandcomment.com

There are a number of aspects to any data standardisation:

— **Common file formats.** CSV is a file format for tabular data that can be accessed by most data processing systems. XML and JSON are formats often used for 'structured' data. When it comes to supporting higher degrees of openness of data, the W3C (the body that maintains many of the standards for the web) has proposed a range of standards that can be used. The Resource Description Framework (RDF) (Manola and Millereds, 2004) is an open format for publishing data and establishing links among datasets and such datasets are often referred to as 'linked data'. The Query Language for RDF (SPARQL) (Prud'hommeaux and Seaborneeds, 2008) provides a way to query linked data sets. SPARQL and RDF are two of the main technologies that have enabled the publication and interlinking of large datasets on a global scale, which are often referred to as the Linked Open Data (LOD) cloud.

— **Common fields and value formats.** A standard may specify the fields that a dataset can include. For example, the iCal standard for calendar data determines that each event recorded needs a Start Time or an End Time, and sets out the format that these should be expressed in (including time-zone, for example). Sometimes a set of 'fields' may be described as a 'vocabulary' or more broadly as a standard, as in the IATI Standard for describing aid activities.

— **Common codes.** Datasets often use codes in place of more detailed descriptions. For example, when referring to countries many datasets use ISO codes, using 'GB' in place of United Kingdom for example. Some code lists might come with detailed definitions to describe how they should be used, supporting harmonisation of datasets and comparison across datasets from different organisations.

Both technical and content standards can be complex and expensive to develop, and require ongoing governance and oversight. Standards involve making compromises between the expressiveness of a dataset, the comparability of data, and the costs of collecting detailed and granular data (Bowker and Star, 2000).

However, the distributed nature of the RDF model does allow for the use of light weight mix-and-match standardisation processes in place of top-down standardisation. This, if backed by the right social arrangements and management processes, could enable charities to better balance the desire to collect diverse data, whilst also making data comparable and taking advantage of network effects. Identifying where the sector will benefit from increased standardisation, and supporting the development of standards, are likely to be key roles for infrastructure and funding organisations.

OPEN DATA AS A PUBLIC GOOD

The open data model is based on the notion of data as a public good. Publishing data openly on the web can create value across the web, not just for the organisation publishing the data. Public goods can be prey to a 'free rider problem', where people act only as consumers – taking from the common pool of data, but not adding to it. The costs of a data release accrue to the provider, but the benefits are distributed across the whole data consumption community. The free rider issue is, however, not in itself as corrosive for digital data as it is in other economic sectors. Information and data are generally non-rival goods: one person's use of a dataset does not diminish it or use it up so that someone else cannot access and use it. Hence, with digital resources like data, individuals or organisations can use open data without contributing to the general good by adding their own data to the mix. This does not reduce the number and quality of services that other organisations can provide with respect to the data, and so the harm is relatively small. Because information is non-rival, the free-rider problem is not as severe as the so-called 'tragedy of the commons', where a commonly-owned rival resource is likely to be depleted without some kind of coordination mechanism. It needs to be recognised that by no means all organisations will wish to contribute to efforts to make data open, nevertheless, there is potential for a systemic problem if users consume open data without publishing any.

> Publishing data openly on the web can create value across the web, not just for the organisation publishing the data

Data portals: enhancing data management, discovery and access

In Section 2.3.2 we noted that open data is often published using an open data portal. Organisations might deploy their own data portals, or may publish to third-party data portals. Open data portals can provide a wide range of features. In this section we describe the different components of a complete open data portal and we discuss which parts of the portal are deployed for different organisations with different needs[18].

An open data portal is an online application that hosts metadata records about datasets (including licensing information) and may also store the datasets themselves for download. Alongside this open data, portals often host links to applications that make use of the published data, and may provide interactive and social features for users to offer feedback on datasets. Open data portals can generate information and analytics on the portal use, for example, giving statistics on which datasets or applications are accessed the most. Additional features of an open data portal that can leverage and support public engagement include tagging, search, discovery and recommendation of datasets and applications as well as sharing on social media[19]. Finally, an open data portal can offer access to both its catalogue of metadata, and potentially to data itself, via Application Programing Interfaces (APIs) – supporting federation of data that enable third party data sources and applications to access, enrich and draw upon its content.
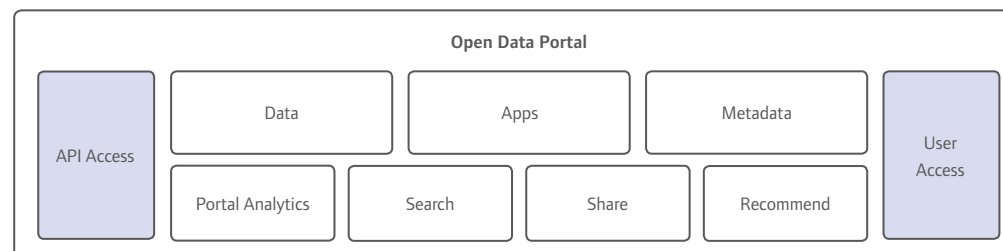
An open data portal is an online application that hosts metadata records about datasets and may also store the datasets themselves for download

Figure 3: Open data portal architecture

18
The concept of a complete open data portal is closely related to what Irving and Pollock (2012) describe as the emergence and adoption of Data Management Systems (DMS) that parallel the Content Management Systems (CMS) that many charities have implemented to run their websites and publish information on the web over the last ten years.

19
Supporting public engagement around data involves more than just technical features. A workshop at the 2012 UK Gov Camp led to a 'five stars of Open Data Engagement' model highlighting some of the activities dataset and portal administrators may need to carry out to support deeper public engagement: **www.opendataimpacts.net/engagement**

The figure above sets out a schematic model of a complete open data portal, based on best practice features observed in open data portals in government (e.g. data.gov.uk, Data.gov, Police.uk), in education (data.southampton.ac.uk) and on the features proposed to promote public engagement with open data by Socrata (Merritt and Byrum, 2011).

Vignette 9
Three data portals in practice

data.gov.uk
The UK Government data.gov.uk is a meta-data catalogue of datasets held across the public sector. Approved staff from public sector bodies submit details of their data direct to the catalogue through a secure interface, making use of tags and categories set by the central data.gov.uk team. Each dataset is presented on a page that includes a space for user comments. data.gov.uk also includes a list of applications made with the data based on details submitted by the application developers, and has recently added advanced features supporting location based search, making it possible to find datasets that relate to particular areas of a UK map. Some datasets on data.gov.uk have been converted to linked data, and are accessible through APIs and via the SPARQL query language.

Southampton University Data Portal
The data portal of the University of Southampton (**http://data.southampton. ac.uk**) hosts both datasets and applications, and aims to provide five-star data as well as data is various formats. It lists applications developed by the University or by third parties and provides query access (SPARQL endpoint) to its datasets. It thus provides a directory of datasets and applications instead of search or recommender functionality, and it enables sharing of its material using a variety of online services.

The Data Hub
The Data Hub is based upon the open source CKAN (Comprehensive Knowledge Archive Network) software, developed by the Open Knowledge Foundation. It is an open data catalogue that anyone can submit data to, and it allows third parties to maintain 'groups' of data on the portal. For example, it hosts a 'Civil Society' group that at present contains over 50 datasets. The Data Hub includes features to help users enquire about the licence of a dataset where meta-data on this is not available, and if it understands the format of a data file, it will automatically generate an API, allowing programmatic access to query and explore data without having to download it first.

Developing (or selecting) open data portals to publish data to involves addressing a number of technological challenges and questions. For example, organisations may wish to use a data repository to manage all their data (private and public) with only some exposed in a public open data portal, or they may need to find clear ways to communicate to portal users the difference between authoritative and non-authoritative datasets. In choosing to provide API access onto data, organisations may need to address technical trade-offs between supporting a range of queries, and maintaining the performance of a portal. For example, the IATI portal at **www.iatiregistry.org** splits data into 'Primary' and 'Secondary' data to indicate that which comes direct from aid organisations, and that supplied by third parties adding extra information on aid projects. IATI has also had to make choices about the sorts of APIs to provide onto the available data, because serving fast responses to some of the query operations users (who may want to add up total values of tens of thousands of projects) can require significant processing power and server capacity.

Deployed effectively, open data portals should reduce both technical and social friction for working with data: making it easy for people to discover and access the data they want, to understand the context of the data, and to engage with it. As such, they have a play a key role to play in effectively managing open data in ways that could support powerful network effects to emerge.

Open data and the cloud

Open data can generate large volumes of data that require advanced data processing and new analytic tools. It is highly likely that, just as they have provided some of the technological underpinnings for the emergence of Web 2.0 interactive networks, cloud computing platforms will play a role in the realisation of value from open data.

Cloud computing and cloud services offer the possibility to rent storage or processing power by third parties instead of making a direct investment in the deployment and maintenance of a dedicated IT infrastructure. This model is proving efficient in particular for small and medium enterprises since it reduces risks, it is more agile to varying IT demands and it is cost-efficient. Over the last few years, providers of cloud services have developed ways to provide robustness, performance and value of money, making their services appealing to larger organisations as well (Shuen, 2008). The UK Government has recently been developing its procurement frameworks to allow it to more easily make use of cloud-based services for both sensitive and non-sensitive data. Increasingly, whether or not a service is provided through 'cloud computing' is inconsequential to end users. However, access to scalable infrastructures that can respond to need with marginal costs (rather than a successful open data project suddenly requiring complex investment in new servers to meet demand) is likely to be important in ensuring 'networks' and use of open data can scale relatively smoothly.

Cloud computing is proving efficient in particular for small and medium enterprises since it reduces risks, it is more agile to varying IT demands and it is cost-efficient

Bringing the benefits: building open data ecosystems

We have already seen that the charity sector is made up of a diverse range of organisations. In order to look across the sector at how charities might develop their own open data ecosystems or how they might relate to other open data ecosystems, it is useful to develop some categories that will help us identify the possible relationships and tasks in which open data can play a part.

Charity ecosystems

Charities might be classified by their size and scale, by their geographical focus, or by their funding arrangements (e.g. primarily state funded; grant funded; funded by members or individual supporters). Kendall (2003) suggests, however, that it may be more important to classify the voluntary sector on the basis of 'vertical fields' – seeing each voluntary organisation within its field or industry, and, subject to the specificities of the charitable context set out in Section 2.1, treating charities alongside other kinds of organisations working on similar tasks, rather than taking a horizontal slice that looks only at charitable organisations. Whilst international classifications exist to classify charities according to their 'purpose'[20] in practice, identifying the field a charity is in involves looking at the operational landscape it operates within.

Alongside the relationships charities might have with other charitable, public and private actors in their fields of activity, we can also develop a rough classification of charities by their geographical scope and focus. Charities operating at a very local level may have horizontal relationships that cut across fields of activity with other local charities and organisations. They might also have specific relationships with generalist local and regional infrastructure organisations, or specialist national support services within their fields.

Lastly, we can offer a classification on the basis of the types of activities charities might undertake, from advocacy and campaigning, to operating as a membership association for the benefit of their members.

These classifications offer us three ways of looking at the ecosystems a charity might play a part in. Charities operate:

– **Within a field of activity.** For example: culture and arts; sports; health and social care; education. Charities may operate in more than one of these fields. There may be many sub-fields (e.g. theatre within culture, musical theatre within theatre), and charities may be in interaction with organisations from across all sectors within their field.

– **With a geographical scope and focus.** For example: local community level; local authority level; regional, national or international charities. Some charities may operate at more than one of these levels, or may have federated/branch structures.

– **Performing specific types of activity.** Including:
  - Service delivery
  - Infrastructural support
  - Fundraising and grant making
  - Research, campaigning and advocacy
  - Clubs and associations

Open data can have a role to play in each of these contexts, and in the following section we sketch some of these impacts of open data as they accrue beyond the context of a single organisation.

Open data in specific fields
In each field, the open data ecosystem will vary, both because of the nature of the data relevant to a field, and because of wider legal, organisational, economic, political and social factors.

For example, in the cultural domain, there are ongoing discussions at the European level about the extent to which meta-data and digital data on cultural artefacts owned by the state should be covered by open data policies. In the health field, the Government's 'Open Public Services' agenda (Cabinet Office, 2012) relies on the open availability of performance data (both for regulators, and for citizen choice to drive provision). However, the mixed economy of private, public and charitable providers can lead to very different levels of data being shared by different players, as some organisations focus on gaining competitive advantage by commercial confidentiality rather than through providing the best service on the basis of open data.

Figure 4 below offers a schematic sketch of one possible field, highlighting a number of different relationships that might exist between agencies, and between those agencies and open data. It is clear even from this simplified example how complex such an ecosystem could become.
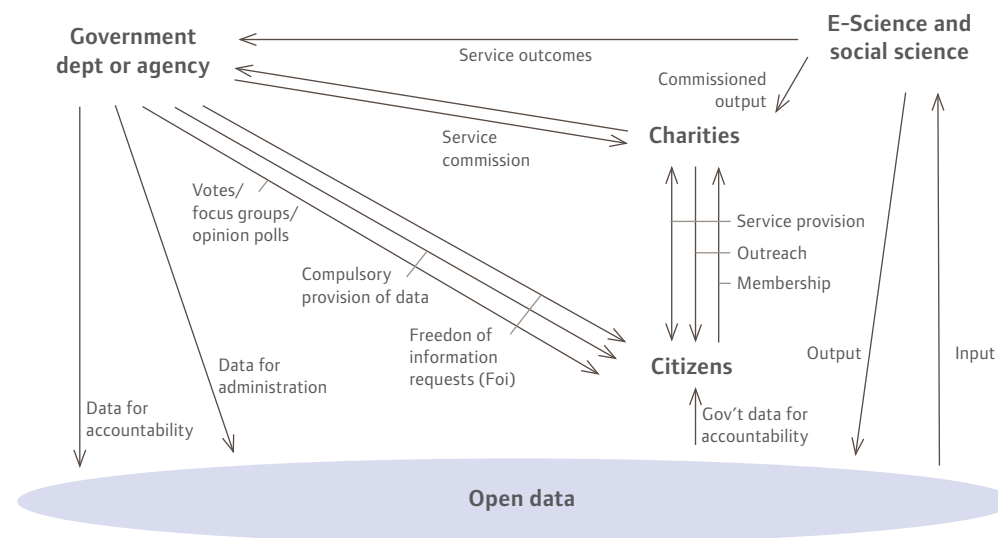

Figure 4: A sketch of the role of open data in the public sector information ecosystem

In this ecosystem, complex sets of information (and of course the diagram is only a sketch) move between government, citizens, academe and charities. A shared pool of open data is primarily provided by e-science and governments, combining data published for accountability, and data published as part of wider open data efforts (e.g. geographic data, or transport data). In this sketch, academia is also a key contributor to the stock of open data, but also draws on it for research. Citizens play a central role feeding data to government (including data voluntarily given, the data that government compels citizens to provide such as tax returns, and the data citizens provide as part of accessing public services), and in using open data (both to be informed about service provision, and to hold government to account).

Charities have strong relations with all these three groups. Government gives charities (access to) information by commissioning services from them; in return, charities must provide data about the outcomes of their interventions (this is only one of the many existing types of relationship between charities and governments). Academics may perform studies commissioned by charities, and research and evaluate the impacts of charity activities. Citizens have a complex relation with charities, depending on whether they are clients or supporters. Charities' outreach operations are intended to inform public opinion, and in return many charities receive feedback and input from members or supporters. At the same time, charities' interventions with clients facilitate a two-way flow of information stemming from service provision; charities help their clients negotiate challenging life events and issues, or seek to enrich the lives of their clients in various ways.

Governments, scientists and citizens have an interest in using the reservoir of open data online, and it is inevitable that charities will be placed in a position where the pressures to use and consume open data will increase. However, figure 4 indicates, often the routes to and from open data to charities are indirect, and there are many more connections that could be made with charities acting as direct consumers and producers of open data.

There are a number of technical and organisational underpinnings that can support data sharing in any particular field. Firstly, shared reference points and code lists to draw upon in creating and updating datasets and in authoring meta-data can facilitate the discovery and integration of data from different providers. For example, the IATI maintains a number of code lists for tagging the specific focus of an aid project, or indicating which types of organisations were involved in its delivery. Secondly, thematic data portals can bring help bring datasets from a sector together in one place, supporting different actors from the sector to discover and re-use data. Thirdly, domain expert intermediaries, able to identify gaps in the available data and to create tools which make data more accessible, play a key role in stimulating a thematic ecosystem. In some thematic areas, some of these intermediary functions will be provided by the market (for example, if cultural organisations published structured data on performances and exhibitions we might anticipate strong synergies with commercial providers who would aggregate and promote information on events). In other contexts these functions will need state support, or the infomediary role will be a charitable task: where investment is required to enrich data so that it can be used for the public good.

Vignette 10
International Aid Transparency Initiative (IATI)

The IATI was developed to support greater accountability and more effective co-ordination in international aid. It is based around a common data standard for sharing information on aid projects and the IATI Registry (a data portal) that provides access to standardised open data files published by a wide range of organisations. IATI was originally developed by governmental aid donors, but is increasingly also being used by charities to publish details of their projects, spending and results.

The development of the IATI Standard, and support for data publishers to maintain the quality of their data is provided by a small technical secretariat, drawing upon the involvement of a large network of organisations who make up the standard's Technical Advisory Group. The IATI secretariat also has advocacy and governance functions, with a steering committee overseeing the whole project, and with outreach work taking place to encourage donors to participate and publish their aid projects in the open IATI data standard.

There is a growing ecosystem of tools and applications available for working with the IATI data. These range from tools from sites like OpenAidRegister.org which make it easy for charities to publish open data on their projects, to IATIExplorer.org and AidView.net which bring together project information from a wide range of donors and charities and make it available to browse, explore and visualise. Anyone publishing open data to the IATI Standard can take advantage of these tools - allowing organisations to share the cost of creating, for example, mapping mashups for their websites, or analysis tools to find areas where organisations could work better together. The IATI Registry open data portal is being developed to include features for assessing the quality of published data.

Who is involved, where and how?

What are the basic activity details?

What are the aims of the activity and were they met?

How is the activity classified and categorised?

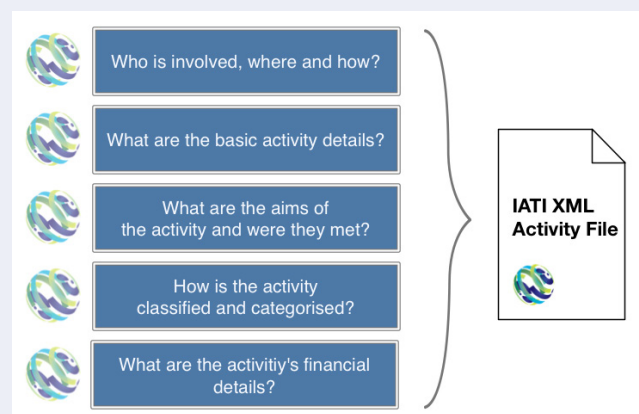What are the activitiy's financial details?

IATI XML Activity File

Image: What goes into an IATI Activity File

Linking up locally, and connecting local and national: geographical scope and focus

Small voluntary organisations often rely upon information, knowledge and support provided by local infrastructure organisations. The existence of infrastructure organisations is reliant upon their ability to show their value in improving the outcomes of charities in a local area. Where large businesses and local governments have, at least historically, had a degree of horizontal and vertical organisational integration that makes sharing data and building up a picture of a local area or a national issue easier, charities operating at the grassroots are often small-scale and independent of each other. Where data sharing does take place between grassroots organisations and local or national infrastructure organisations it can be experienced as a data sharing burden, and charities can find that the same information is being requested from them repeatedly by all sorts of different organisations. The dispersed nature of data and knowledge in local charities, and the frictions of closed data sharing relationships, can lead to missed opportunities to identify needs, access funding or develop new support and services.

Figure 5 shows this as a number of possible data sharing gaps (for the sake of clarity, the diagram categorises charities over one of the dimensions set out in Section 4.1 above, the geographical dimension; however, similar gaps re-emerge however charitable organisations are classified). If a charity has to cross each of these gaps by providing the same information over and over again to different people, then the likelihood that all gaps will be bridged is low.

The dispersed nature of data and knowledge in local charities, and the frictions of closed data sharing relationships, can lead to missed opportunities to identify needs, access funding or develop new support and service

| Umbrella organisation for regional charities | Umbrella organisation for regional charities | National charity | National charity |
| Regional charity | Regional charity | | |

Data sharing gap

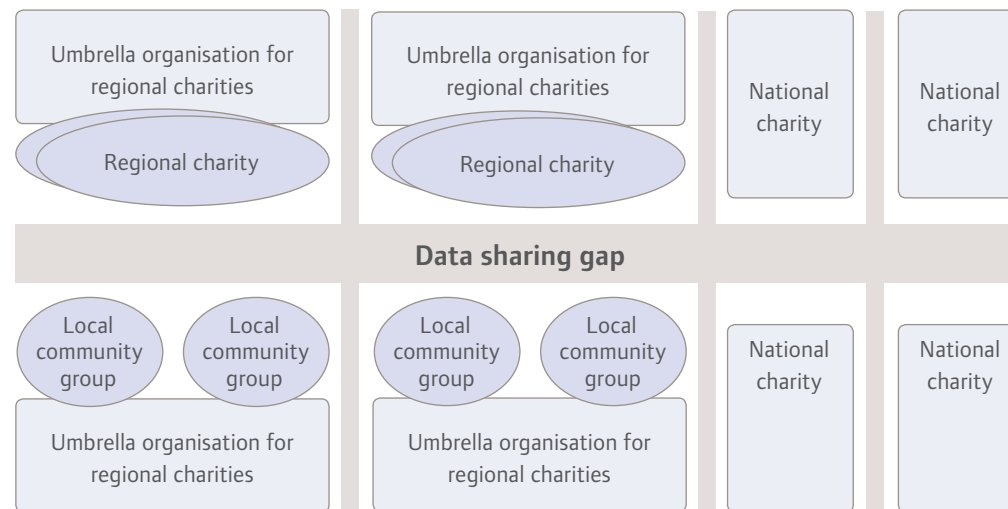| Local community group | Local community group | Local community group | Local community group | National charity | National charity |
| Umbrella organisation for regional charities | Umbrella organisation for regional charities | | |

Figure 5: The data sharing gap

Whilst some of the information charities might share with infrastructure organisations or national support organisations (e.g. the information a local branch of a health charity might share with the national charity) needs to be handled as sensitive and private data, in an open data environment charities would be encouraged to identify the information that they can share publically, and to share this as open data. Infrastructure organisations (and other third parties) can then 'pull' this data down from the web, rather than having to ask charities to constantly be 'pushing' the data to them. The social dynamics of building an open dataset are also different. If local groups feel they are contributing to a shared open resource of data, their incentives to contribute may be different from if they feel they are simply feeding the management information requirements of an infrastructure organisation or national charity. In some cases, open data approaches may even encourage the direct use of shared interactive online databases (instead of groups placing copies of their databases online to be aggregated by a third party), removing duplication and increasing the connections between organisations without requiring formal mergers and integration of organisations.

Vignette 11
Crowdsourcing the cuts

In order to get a clearer view of how public spending cuts have impacted on charities, in May 2010 the National Council for Voluntary Organisations (NCVO) launched a project called 'Crowdsourcing the cuts' in which they asked their members, and other voluntary and community organisations, to fill in a simple online form with details of any statutory funding cuts they experienced. The crowdsourcing exercise used the free Google Spreadsheets tool, meaning the contributed data was instantly available for anyone to access, so contributors could see a picture of national cuts building up, as well as seeing the picture in their local area.

NCVO were not the only organisation to have had the idea of crowdsourcing cuts data, but because the project was carried out as an open data project, when other overlapping initiatives were identified, it was easier to merge the projects – with the result that in January 2011 a new website, 'VoluntarySectorCuts.org.uk' was launched supported by over 20 infrastructure organisations (both national and regional) and building up a shared picture of reductions in statutory funding across the sector. Instead of information being collected multiple times, this open data project was able to simplify intelligence gathering for both national and local organisations.

The Voluntary Sector Cuts website provides tools to drill-down into the data by area, theme or funder, and all the data can be downloaded in a range of formats.

See **www.ncvo-vol.org.uk/cuts** and **http://voluntarysectorcuts.org.uk** for more details.

Intermediary and support organisations also provide an important role to support grassroots charities consume open data – taking and linking together national datasets and providing ways to drill down into the information relevant to a local area. For example, the housing and homelessness charity Shelter provide a Housing Databank[21] which takes government statistics and gives local groups an easy way to pull out figures and graphs they can use to inform local planning or local advocacy. Shelter could have chosen to just provide this as a proprietary product to its local offices, but by making it available as an open tool, many more users draw upon it, and it is having a significant impact shaping debates and action around housing and homelessness.

Open data in action: using open data for specific activities
The previous sections have touched upon how open data may play a role in providing services or changing the way infrastructure organisations work. The tasks a charity is involved in will impact on the nature of its participation in certain open data ecosystems, and the capacities it will benefit from developing.

For example, grant-making charities across all sectors may face some common issues with processing, managing and monitoring grants. Common open data standards for publishing basic information on grant calls, or for capturing information on grants processes could support the development of an ecosystem of tools and platforms that better get information on funding opportunities to potential applicants, and that support applicants to record the impacts of their work and return this data to funders. Funders may share data on grant calls primarily to make that data available on platforms within their field of activity, but if data can also be shared across fields of activity without additional costs to either the funder, or the user of that data, then new forms of research and analysis become possible as well.

The example on page 66 from 'Crowdsourcing the cuts' shows one way in which research and advocacy might build upon open data, and there are also increasingly opportunities for advocacy organisations to publish datasets as part of their advocacy strategies.

21
http://england.shelter.org.uk/professional_
resources/housing_databank

Datasets can support 'data journalism', gaining press interest and enabling the media to create infographics and interactive tools to illuminate a story, but also supporting a 'long tail' of niche uses of the data gathered for a report of campaign: data that is often otherwise simply filed away and left unused on a hard disk or memory stick.

PLATFORMS AND PROCESS: CHALLENGES FOR THE SECTOR

Much more work is needed to track (and theorise) the development of open data as it relates to charities. Although we have ad-hoc examples of where charities are using open data, sketching the role of charities in their own, and within wider, open data ecosystems is at present largely a speculative task. However, it is inevitable that the shift towards openness will impact on charities, and drawing on the state of the art from other sectors, and on research and anecdotal evidence about open data readiness in the charity sector, we can identify a number of key challenges that will need to be addressed at the level of particular ecosystems, or across the sector as a whole. These are challenges that involve decisions and action beyond the level of a single organisation, and are likely to require resources, collaboration and commitment from a range of actors.

A cumulative approach: realising network effects

When we consider the different relationships and ecosystems charities are embedded within, we can see how data sharing gaps might be bridged gradually by the practice of increasing data sharing between organisations, and moving data sharing that takes place to a more open model. There are many types of collaboration which make sense within the charitable sector where more open data sharing approaches might be developed, including:

– **Local to local collaboration.** Small charities working within the same local area could share data in order to provide a common, rich picture of that community.

It is inevitable that the shift towards openness will impact on charities, we can identify a number of key challenges that will need to be addressed at the level of particular ecosystems, or across the sector as a whole

– **Local to regional/regional to national.** A national or regional charity, or umbrella group, could gather, store and represent data for a range of groups covering a smaller range. The aim would be to create a detailed picture of a wider region which could be queried by those charities working at a lower level. The data moves from the smaller organisation to the larger, to exploit economies of scale in data handling. Local patterns are likely to be submerged into a wider picture.

– **Regional to local/national to regional.** The reverse operation is also possible, where wider regional data is shared in a downward direction to smaller, more local organisations, which can then link their own data into the general picture. This has the advantage of allowing the local patterns to emerge, while still allowing a wider set of data (for instance, strategic considerations of how to apply effectively for grants or benefits) to inform the local view.

– **Within fields.** Charities working in a particular sector (for instance, arts and culture; or sexual health) could pool non-private data to create a richer picture of trends and issues across that field.

– **Across fields.** Charities working in different fields, but with other connections could pool data and information. One such connection could be, for instance, geographical, such as a local group for crime victims, and a group working with single parents in the same area. Or the two sectors could have work in common, as for example with a refuge for domestic violence victims, and a housing charity, which may share concerns about changes to planning regulations.

On top of these collaborations, which should bring direct benefits to the charities involved, as noted earlier there may be pressure on charities to publish data, either via apps for supporters, or to support the current drive for government transparency.

In any of these cases, selective steps towards greater data sharing can support the emergence of reciprocal arrangements: avoiding an environment of top-down demands for data, and instead building shared open sources of data and knowledge.

The resulting shared data can provide direct benefits to those contributing to it, whilst also generating potential wider social benefits.

Such small-scale data sharing can grow fairly naturally from practices of consuming open data and linking internal data to open data sets. The advantage of small-scale steps into open data is that the wider case for putting them in place does not need to be made – the steps taken provide a direct benefit to those participating, independently of the wider social benefit they build towards. If, say, a small number of local organisations pool resources to create a richer, deeper picture of a community upon which all can draw, the result is a large communal gain for small individual effort. Furthermore the gain is clear, and not merely abstract. Different arrangements will have different advantages – sometimes a small group can learn from a larger group with greater resources, while a large group can receive a finer-grained analysis of a particular aspect of social policy from a smaller group which is more deeply involved in the lives of its clients. Starting with small scale collaborations with open data (possibility opportunistically around specific or new projects) can also make it easier to work out and develop governance structures and shared working practices.

Developing small-scale collaborations has the potential to then prompt a greater institutionalisation and acceptance of open data, which will open the door to more generalised and larger-scale data sharing operations. Such larger-scale initiatives are less likely to be tied to specific programmes, but rather will be intended to add value and to create opportunities for data users in the charitable sector. Drivers here will be less to do with charities' missions, and more to do with the opportunities presented by economies of scale, especially the possibility of employing linked data representation experts across a set of charities. It may also be that some charities of larger scope or geographical spread would find themselves with a number of overlapping data sharing agreements, which it would be valuable to rationalise by merging. Both in taking short-term small-scale steps towards open data, and in developing large initiatives across the sector, culture change, technical platforms and standardisation will play an important role.

If, say, a small number of local organisations pool resources to create a richer, deeper picture of a community upon which all can draw, the result is a large communal gain for small individual effort

Cultural change and capacity building

There are many reasons why corporations, organisations or public service agencies are nervous about engaging with open data. Perennial worries include: the risks and liabilities of releasing poor quality data; the organisational risk (for example, of being held responsible for errors or misjudgements in the past); paternalism about 'untutored' uses of data by people who are not 'data literate'; and the risk that data will be used against an organisation, or to undercut its business model.

Most corporations, the state, experts, scientists, technologists and administrators (the traditional handlers, manipulators and consumers of data) have working practices that are framed as rational, technocratic, structured, ordered and closed to outsiders. Context and representation languages are often tightly controlled, and uses of data will by and large be restricted to people in particular roles in ways that are congruent with longer-term business strategies. Open data challenges this model both in terms of consumption and production. 'The crowd' of potential data users have no necessary overarching rationality (or several competing rationalities), and is adaptive and intuitive. It is not ordered, is very open, and is interested in data that is meaningful in arbitrary contexts, languages and roles. This means (a) that carefully curated government, corporate or academic data will be used in ways never dreamt of by those collecting it, and (b) that carefully curated data will typically sit alongside a rather chaotic brew of crowdsourced data. Forms and uses of data become the interface between two very different communities, who have very different senses of relevance, authenticity and authority.

The cultural shifts involved in engaging with open data parallel many that non-profits have encountered in engaging with the social web, involving a shift in control away from top-down direction towards new strategies for working with other organisations. Champions and trailblazers are likely to be important to build the confidence of the sector in open data approaches.

The cultural shifts involved in engaging with open data parallel many that non-profits have encountered in engaging with the social web, involving a shift in control away from top-down direction towards new strategies for working with other organisations

Vignette 12
Cultural challenges: crowdsourcing and cleaning data

Following the Haiti earthquake volunteers made use of the 'Open Street Map' platform to generate an up-to-date map of the remaining Haiti infrastructure, and to keep track of activities on the ground. This 'crowdsourced' data was available to organisations alongside the existing management systems they held.

Ortmann et al.'s survey of disaster information managers revealed that engagement with these crowdsourced maps highlighted issues of uncertainty, trust, and semantic problems (i.e. interpreting the data). On the upside, disaster managers liked the idea that crowdsourced information is collected in realtime and for free. The experts surveyed saw the greatest need for improvement in filtering of information, training of volunteers, ranking of relevant information, structural compatibility and compliance with standard terminologies (for the full figures, see Ortmann et al. 2011).

As Ortmann et al put it with respect to the Haiti disaster mapping effort:
Currently, two worlds of information infrastructures exist in parallel. On the one hand, relief organizations have their own information systems. On the other hand, ad-hoc information infrastructures emerge in the social web, which are mostly fed by crowdsourced data. Integration across these two worlds is only possible manually, and in the case of Haiti a full integration did not take place. Yet, despite the flaws of crowdsourced information, many disaster managers are willing to learn how crowdsourced information can be efficiently and effectively integrated into decision making processes.

More recently, efforts to develop a digital standard for humanitarian relief data (the Humanitarian Exchange Language (HXL), based on linked data technology) have been exploring light-weight ways to integrate dispersed data from the field in disaster response situations.

The skills needed to work with open data are wide-ranging. Some uses of data rely upon sound statistical skills and the ability to draw inferences from diverse datasets; others draw upon the ability to mash-up datasets to create dashboards and operational tools that can support day-to-day decision making. Others still involve the ability to publish data in structured ways so that others can build upon it – where that data could be anything from details of cultural events, to in-depth research statistics. Different capacities will be important to different organisations. Some capacity building will be able to come from professional bodies in particular fields within the charitable world. Other capacity building will be cross-cutting, with statistical literacy, or 'big data' skills required by people in specific roles across the sector.

Technical platforms for open data
Based on experiences of deploying open data solutions for sectors such as government and science, it seems likely that the charity sector will benefit from creating shared data portals to both encourage, and to ease, the sharing of open data. Effective portals need good technical and organisational management.

As outlined in this report there can be a number of organisations as well as individuals publishing data on such portals; this is particularly the case for the portals that bring together data published in a particular field, or across a particular locality. It is important that the deployment of such portals considers both human and machine users. Humans are able to work with rough descriptions of a dataset, but can only work with a few datasets at a time. Machines are able to work with vast quantities of data, but need to have clear descriptions of what that data is. If applications are going to be able to be developed that will support the discovery and analysis of data, then for all published datasets it is important to have accurate metadata that will help address data integrity in terms of:

– **Provenance.** Who contributed the data? How were the data collected? Where a large number of charities, volunteers and individuals contribute data it is important to distinguish among datasets based on their origin. This will enable applications to establish different levels of data integrity and to provide better analytics.

– **Description.** What exactly is the data about? Which period do the data concern? A precise description of the dataset will ensure its usability in applications for the charitable sector. It will also help digital tools trying to bring together relevant data from across the (hopefully) vast collections data that are published, to find the relevant content and to prepare and use it for rigorous analysis.

– **Coverage.** Is the dataset complete? There are certain types of activity for which completeness is critical. Let's consider data on the standard of living for families in a particular region that have been receiving support from a specific charity. It is important to know whether the data concerns all the families that receive support from the specific charity or indeed all the families in the area. If not, the way in which the data was collected needs to be clearly described if the data is to be used in a meaningful way.

– **Precision.** Is the dataset based on factual data or estimated? It can be often the case that contributed data are estimations. This is particularly applicable for charities that provide periodic reports on relevant conditions in society. If the data are estimations, any indication of error margins can ensure the precision of the analysis and the precision of applications that will make use of them.

– **Authoritativeness.** To what extent is the data endorsed by the organisation that published it? Is it authoritative or non-authoritative? Authoritativeness is often perceived as another descriptor of precision. It is easy to envisage applications that provide data analysis distinguishing between authoritative and non-authoritative data. In many organisations there are even separate portals to publish these data. Another approach is to have more detailed metadata that will give this information on the level of dataset (instead of the level of portals).

Apart from precision, the licence under which data is published needs to be readily available from a portal, or else humans and machines cannot identify what re-use of the data is allowed. A portal also needs to be easily accessible by people who wish to contribute or use datasets and should provide machine-readable access to the meta-data for applications that can be programmed to pick out and process datasets published under a specific licence.

It is also important to consider the overall arrangement of portals and platforms in the sector. We have identified how most organisations are likely to participate in multiple ecosystems of open data, and if they are asked to publish details of their data to multiple portals, the data-sharing burden starts to rise again, even in the open data world. But if charities only publish to an isolated field-specific portal or local portal, potential connections with their wider ecosystems are lost. It can be possible to address this with a suitably small number of portals, and with federation of information between portals. Portal tools like CKAN support 'federation', where meta-data on one portal can be shared with another, but this requires the use of some common meta-data standards, and collaboration between portal providers to configure federation. Portal infrastructures also need to be designed with the user in mind: helping people to find the data they want without assuming they will know about all the available portals that are out there.

Developing standards
Common identifiers and standards are key to supporting the integration of datasets and to help make connections across open data sets originating from different publishers. Both within specific fields and across areas of activity there will be 'comparable' data being published that, without standards, will be very hard to use for comparisons and analysis.

Some appropriate standards might already exist in some settings, and the key challenge for charities is to identify and converge on existing shared standards. In some cases, the existing standards might need to be adapted.

Common identifiers and standards are key to supporting the integration of datasets and to help make connections across open data sets originating from different publishers

For example, a working group of charities are engaged in identifying possible reforms needed to the IATI Standard that was originally developed with government aid donors in mind, so that it can capture additional specificities of the charity context. In many cases, there will only be nascent standards, with a need to bring together charities (and other actors within the relevant ecosystems), in order to develop new standards. Bowker and Star (2000) have discussed the complexity of standards processes, and experience demonstrates that standards are rarely fixed once and for all, but need to be managed as they evolve and develop, thus identifying areas for standards development often involves making commitments to an ongoing line of activity in the sector.

Alongside standards, shared code-lists and identifiers are vital. On the Web of Data these also need to be developed and maintained, as without them many opportunities to link data up are lost. For example, let us consider one charity that publishes a dataset on unemployment and personal happiness indicators, and let's assume that it uses a certain identifier for the concept 'unemployed person'. Now, let's consider a dataset by a local community group that contains data on unemployment and quality of life indicators but which uses a different identifier for the concept 'unemployed person'. There is significant potential for applications that will investigate correlations between unemployment, personal happiness indicators and quality of life indicators based on those two datasets. If both datasets use the same identifier for common concepts, in our case that of 'unemployed person', processing would be easier.

Using common identifiers for common concepts where possible can ensure interoperability among datasets, the inclusion of more datasets in analytic tools and better applications to reflect charitable work and societal challenges. This comes in addition to the requirements for common formats and the use of open standards for publication of open data. If, for example, datasets are published in RDF, a standard that is open and provides for linking with concepts in other datasets, data processing will be more efficient. If datasets are published using an open spreadsheet format, it will be more challenging, technologically, to process concepts described in column headers in different spreadsheets.

Privacy, security, safeguarding and trust

We have already discussed at length the issues at an individual organisational level with respect to privacy, security and safeguarding when datasets are being made open. Clearly distinguishing those datasets which contain personal data and thus need careful management, from those which do not, is vital. As wider open data ecosystems develop, and the linking-up of datasets outside of conventional access control mechanisms increases, the potential for privacy grey areas to arise is only likely to increase.

There are no clear answers to the privacy problem in the open data paradigm. There are in some cases trade-offs between social value and impacts on individuals and communities, particularly in areas where the question of what is and what isn't private data is contested. The situation is complicated by differing interpretations of the law, the range of jurisdictions open data may be accessed and use within, and the fact that in a social media world, individuals play an increasingly central role in managing (or in some cases, appearing not to manage) their own privacy. In the realm of charitable organisations, where service user trust is vital, finding strategies to manage privacy concerns whilst also realising the benefits of open data is probably the trickiest issue encountered. The privacy question is also one that small organisations cannot answer alone. Instead, it will need wider discussions to identify risks, trade-offs, and the role that different organisations, have in ensuring the privacy, security and trust of the individuals and communities they serve.

In the realm of charitable organisations, where service user trust is vital, finding strategies to manage privacy concerns whilst also realising the benefits of open data is probably the trickiest issue encountered

Conclusions and recommendations

This report has presented the state of the art for open data and it has shown how open data has been successfully adopted in sectors such as government and science. The report summarises the technological aspects of open data adoption, including looking at data standards, and the role of open data portals that can host datasets and applications. It has also discussed the potential of crowdsourcing mechanisms that allow diverse groups to engage in contributing data, data quality improvement and application development. There are organisational issues to sharing open data related to making the case for the value of data sharing and putting in place the appropriate business processes. There are additional issues related to data protection and the risk of triangulating on individuals and compromising their identity based on datasets originating from different sources. Although not all of these aspects can be addressed in their full extent, it seems that reasonable measures can be taken by avoiding the publication of information that is assessed to present privacy risks. The idea of overlapping open data ecosystems and the positions of charities within these ecosystems are also discussed. The different ways in which charities can participate in these ecosystems are presented.

In conclusion then, we recommend that:

— Open data will be an increasingly important part of the information landscape. **Charities will need to understand this new landscape** and be able to operate in it. They can benefit from open data in a wide range of ways, and should look to develop their capacity to be both creators and users of open data. Organisations of any size can take the first steps to produce and consume open data.

— There are many opportunities to increase the flow and sharing of relevant datasets within the charity sector, to support engagement between charities and the public, government, funding bodies and other charities. The deployment of open data portals for the UK charities sector could help **bridge the information sharing gap** on a national or regional level, and it could provide for more intense public engagement, innovation, and impact. Infrastructure organisations have a key role to play in bridging the data gap, though much will also need to take place through gradual, bottom-up work on task-specific open data sharing in particular contexts.

- Charities may need to **invest in capacity building** in the open data world. Apart from open data portals, it is likely that they will have to invest in the development of light-weight standards to ensure harmonisation of their data infrastructures and better information sharing. This should involve a focus on web standards and open standards. Funders have an important role to play in supporting the development of standards and promoting their adoption.

- Charities need to be aware of, and to have implemented, **good information processing practices** (even in the absence of open data). At the same time, they should be aware of safeguarding issues when publishing open data and have strategies for managing related risks. They should be focused on retaining trust and maintaining the privacy of data subjects, as well as complying with their responsibilities under the Data Protection Act.

- Publishing open data will produce gains for charities whose data will help inform the decisions of others about their areas of expertise, and about the issues for which they take responsibility. As with most public goods, the tangible gain from a particular publication event will be small for a charity – the gain is less easily measurable, and over a longer term, and consists of the network effects of a growing quantity of linked data with a growing online community. Networks effects are not inevitable, and **continued attention and action to support their emergence will be important** to make sure the charity sector benefits to the maximum possible extent from open data.

In all these areas, Web Science has an important role to play in informing information sharing policies and public engagement practice, and in quantifying the benefits (and, where appropriate, the risks) of open data. The theories, analyses and syntheses of web behaviour that have been conducted under the banner of Web Science for the last few years have been instrumental in bringing about and understanding open data information ecosystems. Charities have a particularly interesting position in such ecosystems, with their public service ethos and their close connections with small communities, vulnerable groups and other sectors of society that sometimes struggle for representation. It is to be devoutly hoped for that the Web of Data will serve these sectors, in which case their input will be absolutely crucial.

References

Alani, H., Hall, W., O'Hara, K., Shadbolt, N., Chandler, P. and Szomszor, M. (2008) 'Building a pragmatic Semantic Web', IEEE Intelligent Systems, 23(3), 61-68.

Allemang, D., and Hendler, J.A. (2008) Semantic Web for the Working Ontologist: Modeling in RDF, RDFS and OWL. Morgan Kaufmann.

Benkler, Y. (2006) The Wealth of Networks. Yale University Press.

Berners-Lee, T. (2009) Linked Data - Design Issues. [Online]. 18 June 2009. w3.org. Available from: **www.w3.org/ DesignIssues/LinkedData.html** [accessed 24 June 2012]

Bowker, G.C. and Star, S.L. (2000). Sorting Things Out: Classification and its consequences (Inside Technology Series). MIT Press.

Brown, B., Chui, M. and Manyika, J. (2011) Are You Ready for the Era of 'Big Data'? McKinsey Global Institute.

Bughin, J. and Chui, M. (2010) The rise of the networked enterprise: Web 2.0 finds its payday. McKinsey Quarterly.

Cabinet Office (2012) Open Public Services 2012, white paper. Available from: **http://files.openpublicservices.cabinetoffice. gov.uk/HMG_OpenPublicServices_web.pdf** [accessed 24 June 2012]

Clarke, P. (2011) data.gov.uk one year on. Available from: **http://paulclarke.com/honestlyreal/2011/01/data-gov-uk-one-year-on** [accessed 24 June 2012]

Cole, R. (2012) Some observations on the practice of "open data" as opposed to its promise. The Journal of Community Informatics, 8:2. Available from: **http://ci-journal.net/index. php/ciej/article/view/920/917** [accessed 24 June 2012]

Crawford, K. and Boyd, D. (2011) Six Provocations for Big Data. A Decade in Internet Time: Symposium on the dynamics of the internet and society, 21 September 2011.

Davies, T. (2010) Open Data, Democracy and Public Sector Reform: A look at open government data use from data.gov.uk. Practical Participation. Available from: **www.practicalparticipation.co.uk/odi/report** [accessed 24 June 2012]

Dutton, W.H. (2011) Networking Distributed Public Expertise: Strategies for citizen sourcing advice to government. Occasional Papers in Science and Technology Policy. Science and Technology Policy Institute, Institute for Defense Analyses. Washington D.C.

Europa (2011) EU Calendar - The European Commission proposes an open data strategy. [Online]. 12 December 2011. europa.eu. Available from: **http://europa.eu/rapid/ pressReleasesAction.do?reference=IP/11/1524&format=HT ML&aged=0&language=en&guiLanguage=en**

Europa (2012) Commission proposes a comprehensive reform of the data protection rules - Justice. [Online]. 25 January 2012. ec.europa.eu. Available from: **http://ec.europa.eu/justice/ newsroom/data-protection/news/120125_en.htm** [accessed 24 June 2012]

Google Policy Europe (2012) European Public Policy Blog: Promoting small businesses and economic growth in Europe. [Online]. 24 January 2012. googlepolicyeurope.blogspot.com. Available from: **http://googlepolicyeurope.blogspot. com/2012/01/promoting-small-businesses-and-economic. html** [accessed 24 June 2012]

Hansmann, Henry (1980), "The Role of Non-profit Enterprise", The Yale Law Journal, Vol. 89, No. 5.

Hayek, F.A. (1945) The use of knowledge in society. American Economic Review, 35(4), 519-530.

HM Treasury (2011) Further Detail on Open Data Measures in the Autumn Statement 2011. [Online]. Available from: **www.cabinetoffice.gov.uk/resource-library/open-data- measures-autumn-statement-2011** [accessed 24 June 2012]

Irving, F. and Pollock, R. (2012) From CMS to DMS: C is for Content, D is for Data. Open Knowledge Foundation Blog. Available from **http://blog.okfn.org/2012/03/09/from-cms- to-dms-c-is-for-content-d-is-for-data/**

Ivens, J., Smith, T., Noble, S. and Lally, G. (2011) Data Bridge: Exploring voluntary sector use of data in Brighton and Hove.

Kavoukian, A. and El Emam, K. (2011) Dispelling the Myths Surrounding De-Identification: Anonymization remains a strong tool for protecting privacy. Toronto: Information and Privacy Commission of Ontario, Canada. Available from: **www.ipc.on.ca/images/Resources/anonymization.pdf** [accessed 24 June 2012]

Kendall, J. (2003) The Voluntary Sector: Comparative perspectives in the UK. London, Routledge.

Manola, F. and Miller, E. (eds.) (2004) RDF Primer. W3C Recommendation. Available from: **www.w3.org/TR/2004/REC-rdf-primer-20040210/** [accessed 24 June 2012]

Merritt, K. and Byrum, J. (2011) Unlock Government Data with Social Data Discovery. How a social data discovery solution promotes civic engagement, improves decision and policy making, and achieves government transparency. [Online]. Socrata. Available from: **www.socrata.com/wp-content/ uploads/2010/11/Socrata_Social_Data_Discovery- Government.pdf** [accessed 24 June 2012]

Murray-Rust, P. (2008) Open data in science. Serials Review. 34(1), 52–64.

NCVO (2010) Open data and the voluntary sector. [Online]. 3 August 2010. NCVO. Available from: **www.ncvo-vol.org.uk/ networking-discussions/blogs/116/10/08/03/open-data- voluntary-sector** [accessed 24 June 2012]

O'Hara, K. (2011) Transparency Government, Not Transparent Citizens. London: Cabinet Office. Available from **www. cabinetoffice.gov.uk/resource-library/independent- transparency-and-privacy-review** [accessed 24 June 2012]

Ohm, P. (2010) Broken promises of privacy: responding to the surprising failure of anonymization. UCLA Law Review. 57, 1701-1777.

OKF - Open Knowledge Foundation (2006) Open Knowledge Definition. Available from: **http://opendefinition.org/okd** [accessed 24 June 2012]

Ortmann, J., Limbu, M., Wang, D. and Kaupinnen, T. (2011) 'Crowdsourcing linked open data for disaster management', in R. Grütter et al. (eds), Proceedings of the Terra Cognita Workshop on Foundations, Technologies and Applications of the Geospatial Web in conjunction with the ISWC 2011. CEUR-WS 798, 11-22. Available from: **http://ceur-ws.org/Vol-798/ paper2.pdf** [accessed 24 June 2012]

Pavitt, S. and Lord-Soares, S. (2006) Circuit riders 2.0. Lasa. Available from: **www.lasa.org.uk/ict/lasa-circuit-riders** [accessed 24 June 2012]

Powell, M., Davies, T. and Taylor, K.C. (2012) ICT for or against development? An introduction to the ongoing case of Web3. IKM Working Paper No. 16. Available from: **http://wiki. ikmemergent.net/files/1204-IKM-Working_Paper_16- WEB3-Mar_2012-2.pdf** [accessed 24 June 2012]

Prud'hommeaux, E. and Seaborne, A. (eds.) (2008) SPARQL Query Language for RDF. W3C Recommendation. Available from: **www.w3.org/TR/rdf-sparql-query** [accessed 24 June 2012]

Raman, B. (2012) The rhetoric of transparency and its reality: transparent territories, opaque power and empowerment. The Journal of Community Informatics. 8(2). Available from: **http://ci-journal.net/index.php/ciej/article/view/866** [accessed 24 June 2012]

Raymond, E. (2000) The Cathedral and the Bazaar. Available from: **www.catb.org/~esr/writings/cathedral-bazaar/ cathedral-bazaar** [accessed 24 June 2012]

Shadbolt, N. (2011). Open For Business. Think Quarterly. Google. Available from: **www.thinkwithgoogle.co.uk/intl/ en_uk/quarterly/data/nigel-shadbolt-open-data.html** [accessed 24 June 2012]

Shadbolt, N., Hall, W. and Berners-Lee, T. (2006) The Semantic Web Revisited. IEEE intelligent systems. 21(3), 96–101.

Shuen, A. (2008) Web 2.0: A Strategy Guide. Business thinking and strategies behind successful Web 2.0 implementations. 1st Ed. O'Reilly Media.

Surowiecki, J. (2004) The Wisdom of Crowds: Why the many are smarter than the few. London: Little, Brown.

Vickery, G. (2011) Review of recent studies on PSI re-use and related market developments. European Commission, DG Information Society.

Wang, X., Tiropanis, T. and Davis, H.C. (2011) Evaluating Graph Traversal Algorithms for Distributed SPARQL Query Optimization. Joint International Semantic Technology Cconference(JIST2011), Hangzhou, China.

Wind-Cowie, M. and Lekhi, R. (2012) The Data Dividend. [Online]. Demos. Available from: **www.demos.co.uk/files/ The_Data_Dividend_-_web.pdf?1330693016** [accessed 24 June 2012]

Yiu, C. (2012) A Right to Data. Fulfilling the promise of open public data in the UK. [Online]. Policy Exchange. Available from: **www.policyexchange.org.uk/publications/category/item/a-right-to-data-fulfilling-the-promise-of-open-public-data-in-the-uk** [accessed 24 June 2012]

Yu, H. and Robinson, D.G. (2012) The New Ambiguity of "Open Government." Forthcoming in UCLA Law Review Discourse. Princeton CITP / Yale ISP Working Pape.Bishop, M. (2011)

About Nominet Trust

Digital technology offers a phenomenal opportunity to stimulate new forms of collaboration, to mobilise new communities of interest, and to unleash the imagination of millions of users in addressing specific local and global challenges.

At Nominet Trust we are committed to making these opportunities a reality – for as many people as possible.

Nominet Trust is a UK-based social investor that advocates the imaginative use of digital technology to improve lives and communities.

That's why we bring together, invest in and support projects committed to using digital technology to make society better.

Through our on-going research programme we identify specific areas of need and channel funding towards initiatives designed to make a significant difference to people's lives.

Since our inception in September 2008, Nominet Trust has invested in hundreds of projects, providing business support as well as financial investment, seeking to make a positive difference to the lives of disadvantages and vulnerable people.

To find out more about our work or how you can apply for funding, please visit:

www.nominettrust.org.uk

About our work

There are many ways in which the internet can bring about change. To make sure we achieve the greatest impact, our focus is on supporting projects and organisations that use digital technology to improve lives of the disadvantaged and vulnerable and to strengthen communities.

It is important to remain open to new ideas that offer a fresh perspective. Our aim is to seek out, galvanise and support innovative, early-stage projects that use digital technology to address big social challenges.

We also invest in a number of programmes that address a specific social group or issue, such as young people, local communities or health and well-being. By clustering our investment in this way we hope to increase our social impact. We regularly review the groups of people and issues we support so please check our website to find out our current focus.

Do you need support for your idea?

If you have an idea for a new initiative or would like support for an existing project then please get in touch.

We are particularly interested in projects that develop tools or models that can be replicated or scaled-up to benefit others.

To find out more about how you can apply for funding, visit us at:

www.nominettrust.org.uk

nominet1rust

Nominet Trust
Minerva House
Edmund Halley Road
Oxford Science Park
Oxford OX4 4DQ

t   +44 (0)1865 334 000
f   +44 (0)1865 332 314
enquiries@nominettrust.org.uk
www.nominettrust.org.uk

If you have printed this document, **please recycle after use**.