Emerging Implications of Open and Linked Data for Knowledge Sharing in Development

Tim Davies and Duncan Edwards

Abstract Movements towards open data involve the publication of datasets (from metadata on publications, to research, to operational project statistics) online in standard formats and without restrictions on reuse. A number of open datasets are published as linked data, creating a web of connected datasets. Governments, companies and non-governmental organisations (NGOs) across the world are increasingly exploring how the publication and use of open and linked data can have impacts on governance, economic growth and the delivery of services. This article outlines the historical, social and technical trajectories that have led to current interest in, and practices around, open data. Drawing on three example cases of working with open and linked data it takes a critical look at issues that development sector knowledge intermediaries may need to engage with to ensure the socio-technical innovations of open and linked data work in the interests of greater diversity and better development practice.

Our assumption is that… building openness into polices and technologies will result in greater opportunities for developing countries to transform into equitable and sustainable knowledge societies. (Smith and Elder 2010)

… for 'open data' to have a meaningful and supportive impact on the poor and marginalized, direct intervention is required to ensure that elements currently absent in the local technology and social ecosystem are in fact, made available. (Gurstein 2011)

1 Introduction

Dramatic change is taking place across the web. Institutions, from universities to national governments and intergovernmental organisations that historically restricted access to their data resources, are now placing vast quantities of data online for anyone to access and reuse.

Since 2009, over 100 open data initiatives have been launched by governments, grassroots activists, and institutions globally,[1] including the World Bank's Open Data portal and open data initiatives in Kenya (Rahemtulla *et al.* 2011) and Ghana (Grewal *et al.* 2011). With the increased availability of 'raw data' (feeding back into demands for more data) we are seeing the rapid growth of data-driven websites, tools and applications, from mapping 'mash-ups'[2] of government statistics, to mobile applications driven by real-time open data.

Data journalism uses open government and research data sets to identify stories and present news to the public (Bradshaw and Rohumaa 2011). Less visibly, citizens, researchers and policymakers are taking advantage of public data to question local state decisions, monitor trends, or produce their own independent analysis. Simultaneously, technologists are working to engineer a 'web of data', articulating technical standards for 'linked data' to make connections between diverse elements in distributed data sets in much the same way that hyperlinks on the web connect up dispersed documents (Shadbolt *et al.* 2006).

As producers and consumers of information and data, development practitioners and knowledge managers will be affected by these trends, and faced with new opportunities and challenges in mobilising knowledge to support development. Critical attention to the capacity of the sector to effectively produce open data and to make

effective use of open and linked data resources will be essential, particularly at the grassroots level. Critical engagement will also be needed, given the emerging structure of open data ecosystems on the web.

Firstly, we explore the historical trajectory of data management from closed data towards open data, outlining the multifaceted nature of advocacy for open data. We then examine the emergence of linked data as one technical approach for managing data on the web. After exploring examples of linked and open data in the development sector, we consider the extent to which open and linked data, as sociotechnical phenomena, have the potential to challenge, or entrench, existing power dynamics in the production and consumption of knowledge. We then discuss critical issues which those engaged in building and working with an open and linked data web for development need to consider.

The development of open and linked data is as much about organisational, cultural and norm changes as it is about technologies. Yet technologies play a key role in shaping possibilities, just as social and organisational forces shape technical designs. We do not shy away from including technical details, but seek to contextualise them with examples and references to further resources. We hope this broad survey of an emerging field will enable further and deeper investigation.

2 From closed to open

Management of information and knowledge has been transformed in recent decades. In addition to the shift towards digital management of information, movements adopting and advocating open approaches to share these digital resources are emerging. The journey from offline information to online open data involves several significant drivers, each of which shapes the nature and context of contemporary open data.

Technological innovation has led to total global data production and storage capacities, and internet bandwidth, growing exponentially over the last 25 years. From 1986, when Hilbert and López (2011) estimated that less than 1 per cent of global information was digitally stored, to 2007 when it is thought that 94 per cent of data was digital, governments, non-governmental organisations (NGOs), companies and

communities have adopted new technologies to generate vast new data sets and to digitise existing information as data.

Data is encoded, structured information. It can be anything from a YouTube video or journal PDF file, to statistical tables in spreadsheets or metadata about publications in library catalogues. Creating data sets involves making decisions about how to encode the information, and developing categories and schemas to fix its digital form (Bowker and Star 2000). Using data sets involves turning data back into information at some point, adding context and analysis: interpreting and re-presenting it.

Just as the default for non-digital records was often 'restricted access', early digitised data sets or information were often only accessible within the owner's institution. The specialist nature of early mainframe data processing systems and lack of bandwidth, meant that the standards and mechanisms for sharing data supported proprietary cultures. However, social, economic and technical pressures have shaped how data and information – particularly that owned by states – are understood. In the late twentieth century, government secrecy in many countries came under pressure from right-to-information campaigns (Krikorian and Kapczynski 2010). Neoliberal economic theory also turned its attention to intellectual property, extending intellectual property rights and encouraging companies, researchers, governments and NGOs to see their data as important commercial assets (*ibid*.). The potential 'value' of big data sets was underlined by the emergence of large companies such as Amazon and Google who rely on near-instant calculation across vast data sets to recommend products or web pages to their customers. This has created excitement about how 'big data' might transform businesses, research and government (see Wind-Cowie and Lekhi 2012, for example). A new movement has since emerged advocating for 'open data': the online publication, technical standardisation, and permissive licensing of data sets – open to anyone to take, reuse and remix data resources.

The open data movement is drawn from a coalition of groups across the political spectrum, including:

- **Large firms** interested in liberalised markets for public sector information and moving

Table 1 Key data trends

	Definitions	Potential implications
Big data	Data requiring massive computing power to process. Big data is often generated by merging large data sets.	Companies and researchers are exploring ways to 'data mine' vast data resources, identifying trends and patterns. For example, the United Nations Global Pulse project seeks to use big data to understand and respond to changes in human wellbeing.
Raw data	Data in a form that can be easily manipulated, sorted, filtered, remixed. For example, rows in a spreadsheet, as opposed to summary tables. Primary data, as collected or measured directly from the source.	Access to raw data allows journalists or citizens to check official analysis. Programmers build interactive services with raw data. Some raw data about people contains personal information, with privacy implications if made more accessible.
Real-time data	Data measured and made accessible with minimal delay. Often accessed over the web as a stream of data through Application Programming Interfaces (APIs).	Identifying trends in near real-time data can help development of 'early warning systems' (such as Google Flu Trends, Ushahidi).[4] 'Smart systems' and 'smart cities' can be configured to respond to real-time data and adapt to rapidly changing circumstances.
Open data	Datasets are made accessible in non-proprietary formats under licenses permitting unrestricted reuse (OKF 2006). Open government data is shared online in this way.	Third parties can innovate with 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.
Linked data	Datasets are published in Resource Description Framework (RDF) format using Uniform Resource Identifiers (URIs) to identify their contents, with links made between data sets (Shadbolt *et al.* 2006).	A 'web of linked data' supports 'smart applications' that can follow links between data sets. This is the foundation for a semantic web.

towards a US model where government data (such as mapping or weather) are not subject to copyright or charging regimes (Janssen 2011);

- **Small enterprises** and social enterprises seeking to innovate with public data sets;
- **Technological communities** inspired by decentralised and collaborative models of production and problem-solving in open source, focusing on government data, and believing in the value of open sharing of corporate data;
- **Open science advocates** believing that sharing data is essential for accountable research and solving complex new research challenges (Murray-Rust 2008);
- **Political actors** supporting the potential of open data for increased transparency and accountability;
- **Governments and development agencies** exploring the role of open data in a country's development.

All are interested in the instrumental value of open access to data and in the economic, political or social benefits that this will unlock. As Section 4 shows, the international development field is also involved in the open data movement with many projects exploring the benefits open data could bring to development.

3 Situating open data, linked data, and the semantic web
Open data is just one aspect of the 'data revolutions' taking place. Situating open data and linked data within the wider context is vital to understanding potential policy and practice

responses. Table 1 summarises key data trends discussed in the literature. Different terms are often misused or used loosely: such as when advocacy for 'open data' in general is justified with reference to outcomes specifically derived from 'big data'.[3]

Each term can be combined with the others or treated separately. It is possible to have 'big real-time raw data' and 'linked open data', as well as linked data that is not open and open data that is not linked. Some of the case studies below focus on linked open data. The next section examines the technical features of linked data, following Berdou's argument (2011) that practitioners and researchers need to '… engage with technologies themselves in order to understand the opportunities that they provide, where important points of control lie, and the choices that are encoded in their design and use' (p.5).

3.1 Linked data
Consider two simple (open) data sets: one is a table of bibliographic research information; the other a spreadsheet of information on funded development projects. Each has a thematic classification column and columns referring to geography – entitled 'geographical focus' in the research data set and 'target country' in the projects data set. There may be connections between the information contained in each, but efforts to integrate these would need: (a) manual work to interpret the column headings and identify overlaps of meaning; (b) manual, or 'brute force' computerised, matching of terms between data sets, often playing to the lowest common denominator (e.g. reducing 'climate change' and 'climate policy' to 'climate' to match across data sets); and (c) bespoke computer code to perform the integration. Linked data is a technological and organisational response to the fact that, even when open data is available in well-structured forms, making connections between data sets is challenging.

Web innovator Tim Berners-Lee proposes a linked data solution that includes the following elements (Berners-Lee 2006):

- **URIs, or web links, to identify entities and properties in the data set.** Instead of referring to 'Haiti' as the geographical focus, a data set would use a web link to a linked data source which defines Haiti. If two data

sets link to the same URI, computers will know they are referring to the same thing. The same approach can be used for properties the data set describes.
- **Providing data when people or computers look up links.** Visiting www.fao.org/countryprofiles/ geoinfo/geopolitical/resource/Haiti/ you will find data about Haiti. A linked data-aware browser visiting this URI receives the same information as structured machine-readable data. Linked data uses a standard data model, Resource Description Framework (RDF), to exchange data.[5]
- **Linking to other URIs to enable people (and computers) to discover more.** Linked data encourages data set publishers to share links. For example, the Food and Agriculture Organization (FAO) data on Haiti includes the term 'Haiti sameAs http://dbpedia.org/ resource/Haiti' which provides standardised and structured information on Haiti from another source.

4 Open and linked data in development: examples of practice
We now describe three open data projects the authors have been involved with concerning development and research communication.

4.1 Open research: IDS and R4D metadata
Thousands of academic papers, evaluation reports and other documents on development issues are published every year. Getting useful and appropriate knowledge from these publications to those who could use it is a significant challenge. Research intermediaries produce scores of abstracts and metadata – mostly only accessible through interfaces they provide. Recent pilot projects by IDS[6] and the Department for International Development's (DFID) Research for Development (R4D) portal[7] have explored approaches to opening up their metadata.

IDS has developed an API allowing third-party applications to talk directly to its database of over 32,000 abstracts and 8,200 organisation records. After a sign-up process, technically skilled third parties can build new views onto IDS metadata, providing, for example, subject-specific portals of available publications or creating mobile phone-accessible search tools. They don't need to host their own databases or transfer large databases across their internet

connections. It also allows IDS to track direct usage of its data. With VU University Amsterdam, IDS has also developed a linked data wrapper on top of the API.

R4D has adopted a different approach, modelling publications data as linked data (in addition to using the Open Archives Initiative Protocol for Metadata Harvesting that supports exchange of metadata between catalogues). R4D publishes a regularly updated 'raw' file of the data for download and hosts it in a specialist linked data system. Both data sets have minimal restrictions, using Creative Commons-compatible licences.

4.2 Open aid: International Aid Transparency Initiative

The International Aid Transparency Initiative (IATI) asks governments and other donors to publish detailed information on their aid projects and provides a technical standard for representing this information[8] as open (XML) data. By June 2012, 54 organisations, including the World Bank, DFID, the European Union and Aids Alliance, had published IATI data sets. A mobile phone application and an iPad-optimised website, amongst other interfaces, have been created providing accessible ways to explore the data.[9] Some have emerged from groups not directly involved in the IATI process; others were funded by advocacy groups to demonstrate the value of the data and secure ongoing support for the initiative. Pilot work has converted IATI data sets into linked open data, including exploring links between R4D publication records and the DFID projects that fund them.

4.3 Open linked statistics: Young Lives

Young Lives is a DFID-funded longitudinal study of childhood poverty, involving 12,000 children in four countries. With support from IKM Emergent,[10] Young Lives explored how linked data could be used to communicate data and findings from the study (Powell *et al.* 2012). Initially, this intended to represent statistical micro-data as linked data, but privacy concerns (Ohm 2010) shifted the focus to presenting aggregate statistics and metadata on study publications. The resulting website[11] provides machine-readable linked data and uses an open source platform to expose a route through the data for users. A graphing widget visualises the linked data representation of statistics as interactive graphs that can be viewed on the site or embedded in third-party websites and blogs.

4.4 Emerging practice and implications

Whilst the application of open and linked data to development knowledge sharing is in its early stages, the cases above highlight emerging practice with significant potential to alter how knowledge is communicated. Publication of metadata could enable a wider range of intermediaries to develop locally appropriate knowledge services, drawing upon raw data and APIs from existing institutions that have already invested in content; it would also enable new connections between dispersed data sets. Linked and open data could also increase the reach of statistical and operational information, supporting wider conversations, deeper scrutiny of findings, new analysis and innovation.

5 Open data implications

Enthusiasm is growing. Strong claims have been made for the potential of open data to shift power relations in development knowledge management and mobilisation. World Bank President, Robert Zoellick (2010), writes that open data is crucial to 'democratizing development economics'. The World Bank sees it as the foundation of 'a more open and inclusive model for citizen-centric development' (Walji 2011). Smith *et al.* (2008) have articulated the 'open ICT4D' (Information and Communication Technologies for Development) hypothesis that making development processes more open through ICTs, including open data, 'will generate development outcomes that are accomplished: (a) in a more efficient and/or effective manner, and/or (b) in ways that previously were not possible'.

Underlying these claims is the idea that open data will help reconfigure the range of actors and processes involved in development knowledge management, both in knowledge production and consumption. Such outcomes cannot be realised, however, in the absence of critical attention to how open data and linked data develop in practice: openness must serve the interests of marginalised and poor people. This is pertinent at three levels:

- practices in the publication and communication of data;
- capacities for, and approaches to, the use of data;
- development and emergent structuring of open data ecosystems.

5.1 Publication: creating, curating, communicating

Publishing open data requires separation between data, analysis and presentation layers of research and information (Mayo and Steinberg 2007). For cataloguing, this involves making structured metadata accessible to third parties instead of creating websites or services to search for publications. For research, it may involve publishing raw data sets alongside the analysis, allowing third parties to perform secondary research using the data, or supporting the practical realisation of open-science ideals of scrutiny and cross-checked findings (Molloy 2011). However, the extent to which findings are cross-checked, or data picked up for secondary research, depends on the availability of data and on wider social and organisational factors. Access to open data removes some of the barriers to data reuse created by the need to submit and wait for responses to data requests or to work back from published tables to underlying structured data, but it does not remove all barriers that might slow down or limit data reuse. As in the Young Lives study, publishing raw survey data may be prohibited by ethical and privacy concerns; hybrid open and non-open data management strategies will be needed (Cole 2012).

Implicit in narratives around open data is the idea that the simple act of sharing data is enough to ensure its uptake and impact. However, mobilising data resources often requires additional action (Kuk and Davies 2011) – from enriching data (see Section 4.2) to creating 'widgets' that allow visualised data to be embedded in third-party websites (see Section 4.3), sponsoring 'app competitions', or underwriting initial development of tools that make data accessible to non-technical users (see Sections 4.1, 4.2). Whilst, for example, the publication of structured linked open data from the annual Global Hunger Index (GHI) led to use of GHI figures in a wide range of locations – including the UK *Guardian* newspaper and the FAO's 'country profile' web pages – the release of this data was accompanied by a high-profile publication, pre-prepared interactive widgets, and the use of existing relationships to encourage uptake of, and integration with, the data. Recognising the need to stimulate reuse, IDS has launched a grants scheme, offering funding to develop applications and plug-ins that make use of the IDS API.[12]

The 'six functions of knowledge brokering' outlined by Shaxson and Gwyn (2010) highlight that effective knowledge mobilisation goes beyond placing information online, to include linking, matchmaking and collaborative support functions. Modes of open data publication impact how data publishers form relationships with those reusing their content. Whilst IDS requires users to register before accessing the API, open licensing means third parties can republish the data, creating downstream use that can be difficult to track. The open nature of access to IATI data (anyone can access it without identifying themselves) requires investment in building online communities to encourage those reusing the data to provide feedback and help assess the impact of the initiative, and to connect with others with common needs to avoid duplicating effort in analysing or building tools that use the data.

Publishers and knowledge intermediaries will have to consider the new roles and approaches open data requires and how to measure return on investment when value chains of open data are notoriously difficult to track.

5.2 Use: access, analysis, mobilisation

Open data takes away the need for intermediation as users go directly to data sources; instead new intermediaries are emerging, contextualising open data for particular audiences. New online analytical and visualisation tools are available for working with open data, reducing the barriers to technically savvy individuals wishing to provide their own view into, or analysis of, data. Some allow publication of interactive analysis (such as the Young Lives graphing widget), giving end users more control over what they see, making it easier to find facts and statistics relevant to their needs.

Using open data will, however, still require information and data literacy skills including basic ICT skills and the ability to select appropriate forms of data analysis. For example, in early IATI data use, users would aggregate spending figures and draw conclusions from this, even though such analysis was not appropriate for the data; and attempts to mash-up data onto a map missed showing regional or national aid projects that don't have a point location that can be mapped. Similar issues affect the use of large 'big data' open data sets. Crawford and

Davies and Edwards Emerging Implications of Open and Linked Data for Knowledge Sharing in Development

Boyd (2011) have argued that some of the large-scale quantitative big data research techniques impact on the very definition of knowledge, as statistically generated findings over partial data are taken to provide actionable facts. This quantitative turn, brought about in part by the increased accessibility of large data sets, can lead to the subtleties in underlying data sets being ignored in the face of large-scale numbers that appear to 'speak for themselves'. Crawford and Boyd argue that uncritical acceptance of 'knowledge' produced by big data analysis is misguided; that it is crucial to understand how data sets are composed, what they can and can't tell us, and the power imbalances emerging between those who have the knowledge and tools to work with vast data sets and those who don't.

Development actors must be sensitive to the existing configuration of private resources and power that mean the capacity to use and benefit from open data is not evenly distributed. Carlos Correa (2010) explores this in relation to the protection traditional knowledges might require – to avoid the situation whereby the communities who have stewarded them gain nothing, whilst corporations profit from them. Elsewhere, Gurstein, cites a programme of land-record digitisation in Bangalore justified as an ICT4D project but where digital records were 'primarily being put to use by middle and upper income people and by corporations to gain ownership of land from the marginalised and the poor' (Gurstein 2011; *cf* Benjamin *et al.* 2007). Gurstein cautions that, in practice, open data may primarily empower the already-empowered and lead to net loss for the already-excluded, particularly when formal notions of data accessibility do not take into account who has the means, technological equipment and skills for effective access and use of data (Gurstein 2011). Whilst the overall value of data being openly available will outweigh the risks, there may be winners and losers from openness. Knowledge intermediaries have a role in identifying the potential risks from opening particular data sets, and in investing in capacity-building for data use and mobilisation that mitigates those risks.

5.3 Open data ecosystems: diversity and decentralisation?
Sitting between the publication of open data, and the use of that data to drive better development outcomes are online ecosystems of data, shaped by legal, social, and technical forces. In an open and linked data world, paying attention to the nature of these ecosystems is likely to be increasingly important for those seeking to produce and mobilise knowledge for development, particularly if seeking to ensure 'decision-making… underpinned by timely and relevant information that reflects a diversity of viewpoints' (IDS 2005).

Open ICT4D advocates emphasising that digital tools, particularly mobile phones, are playing a key role in allowing new groups of individuals and communities to create (open) data through 'crowd-sourcing' (Bott *et al.* 2011), as well as supporting feedback loops that bring more voices – particularly of the marginalised – into improving development resources. Open data sharing platforms, such as TheDataHub,[13] can theoretically sit alongside large-scale institutional data, equally accessible through open data technologies. However, past experience of linking structures on the web suggests we should not be too quick to assume this will drive more effective access to diverse or decentralised content (Hindman 2008). With the reliance in linked data on hyperlinks to carry semantic information, it is possible that a small number of large institutions will become increasingly central nodes in defining the concepts and structures through which data may be published or accessed.

The formal equality of two open data sets (openly licensed, accessible online, and standardised) does not mean they are equally likely to be used. Power laws (a few information sources getting most of the traffic; a long tail of others with low use) often operate within networks of information – something already visible on the web of linked data where English language DBpedia (http://dbpedia.org, a linked data version of Wikipedia) URIs play a central role linking between data sets (Bizer *et al.* 2009). This occurs because of widespread coverage of DBpedia and because it is the place that existing linked data sets link to (in order to have a bridge to other data sets). Language is another key issue: unless connections are explicitly made between identifiers in different languages, dominant languages may shape the linked data web. Even in non-linked open data, the reuse of common indicators or codes from a high-profile or wide-coverage data sets can impact on what is

easily discoverable, and how data is expressed. For example, statistics from the World Bank's open data portal are now integrated into some Google search results (partly because of their global coverage) and are more likely to get attention than alternative data from grassroots groups.

In modelling a data set to become part of an open data commons, normative and technical judgements need to be made and balanced (Bowker and Star 2000). For example, in rendering Young Lives study results as linked data, a choice had to be made between stating that a statistic referred to India (and choosing whether to use identifiers from DBpedia, the FAO, or some other country list), or whether to publish a concept describing the area in India where the statistics were gathered, and to model the relationship of this area to India as a whole. These modelling choices impact upon complexity for those seeking to reuse data in the future and often there is pressure to adopt simplified models to allow wider reuse. Similarly, work on supporting use of IATA data has stressed the need to map taxonomies that aid administrators use (e.g. water and sanitation), to the terms that make sense on the ground (wells, toilets, etc.). This can be a technical and administrative process, but can also be carried out through participatory methodologies, supporting a degree of translation of data to become more relevant for local contexts (although constrained by the depth of the primary taxonomy chosen for the data). The IKM Emergent programme has advocated for a reframing of linked data as 'linked information' to emphasise that the linked data model can be used to connect data points to the qualitative and narrative information that gave rise to them (Powell *et al.* 2012). This, it argues, could support a heterogeneous web of data, metadata, and qualitative information linked together to support human-scale sense-making and the discovery of diverse local knowledge.

The state of the open data ecosystem is also affected by the legal conditions placed on shared data. The widely used 'Open Knowledge Definition'[14] requires that open data sets are shared under licences allowing reuse, including across different 'fields of endeavour' (commercial and non-commercial alike). This is considered particularly important for the creation of a common pool of data that can be mashed together without concerns over licence incompatibility of data from different sources. However, as we have seen, there are cases where a more gradual opening of a data set (from sensitive personal data in research data sets, to traditional knowledge) may be more appropriate. There are tensions here, in ensuring the visibility of different information and content in an open data ecosystem, and managing cases where data may not be able to become part of a commons for reasons of privacy or security. In the case of Young Lives, publishing just the summary statistics was a compromise contribution to the commons of open data, increasing the discoverability of detailed statistics, but protecting the privacy of individual study participants.

6 Looking forward

Debates about openness are not new to the research communication field, where discussion over open access have been ongoing for many years. However, the rise of open data, (linked, as we have seen, to wider shifts towards openness and the developments of data processing technologies), has introduced a new set of challenges for actors committed to increasing the effectiveness and equitability of development through research production and communication. Practices of open data provide important foundations for more localised and decentralised production of, and access to, knowledge. However, current understanding of open data is primarily from the supply-side perspective; more research and action is needed to identify the demand for open data at a local level and to explore local practices of meaning-making with open data. Although open data promises to be a force for disintermediation, a role for curators remains. Active and engaged data curation, making connections between qualitative and quantitative resources, ensuring context of data is accessible to reusers, bridging data across linguistic and cultural divides, and attentively intervening in open data ecosystems is likely to be an important future role for research communicators. Equally, the need to build the capacity of development actors to produce and consume well-structured open data and to critically assess the implications of data release should not be underestimated.

The greatest challenge, however, is in addressing the emergence of new 'data divides' from open

data releases beyond the development sector (Gurstein 2011). The call from Berners-Lee for 'raw data now' (inspired by Hans Rosling's powerful presentation of macro-level global poverty statistics)[15] reflects the impatience of an open data movement seeking access to data sets it identifies as having a potentially powerful force for good (Berners-Lee 2009). Given many data sets are funded by taxpayers, there is little justification for keeping them closed. Yet, the 'raw data now' message draws on an implicit application of the web engineering the 'procrastination principle' (Zittrain 2008): get the data online first; deal with the use of the data and the social issues second. The World Bank's study

of open data in Kenya states: 'The release of public sector information to promote transparency represents only the first step to a more informed citizenry' (Rahemtulla *et al.* 2011); and the shift towards open data is unlikely to wait until the subjects of development policy have the ICT access, skills, and information literacy needed to gain maximum benefit from newly opened data resources. Unless the investment and energy going into opening up data and building systems to manage data across the web is at least matched by investment and activity in intermediary and local level support for effective data use, open data is likely to widen, rather than narrow, economic and social divides.

Notes
1 Public Dataset Catalogs Browser, http://datos.fundacionctic.org/sandbox/catalog /faceted/ (accessed 27 February 2012).
2 In web development, a 'mash-up' is a web page or application that uses and combines data, presentation or functionality from two or more sources to create new services; see http://en.wikipedia.org/wiki/Mashup_(web_ap plication_hybrid) (accessed 6 July 2012).
3 The ability of Amazon or Google to generate recommendations using their data is a property of the scale of the data, not of openness, and thus provides little evidence about what is possible with open or linked data, *per se*.
4 See http://ushahidi.com/ and www.google.org/ flutrends/ (both accessed 2 July 2012).
5 Each title of a 'data value' on the FAO Haiti linked data page is clickable. These properties are also URIs; clicking them gives you facts about that property or other properties it is related to. Click GDPTotalInCurrentPrices, for example, and you will find it is a subPropertyOf GDP. This ability to follow the chain of definitions is why linked data is sometimes called 'self-describing data'.

6 See http://api.ids.ac.uk/about/ (accessed 29 February 2012).
7 See www.dfid.gov.uk/r4d/ (accessed 29 February 2012.
8 See www.iatistandard.org (accessed 3 July 2012).
9 See www.iatiregistry.org and www.aidinfolabs.org (accessed 29 February 2012).
10 The IKM Emergent programme has explored the potential impacts of linked data in development with a workshop held in 2010 and by supporting a number of demonstrator projects. See http://linkedinfo.ikmemergent.net (accessed 3 July 2012).
11 See http://data.younglives.org.uk (accessed 3 July 2012).
12 See http://api.ids.ac.uk/about/grants.shtml (accessed 29 February 2012).
13 See www.thedatahub.org (accessed 3 July 2012).
14 See www.opendefinition.org (accessed 3 July 2012).
15 See www.ted.com/speakers/hans_rosling.html (accessed 3 July 2012)

References

Benjamin, S.; Bhuvaneswari, R. and Rajan, P. (2007) *Bhoomi: 'E-Governance', or, an Anti-Politics Machine Necessary to Globalize Bangalore?*, CASUM-m Working Paper, http://casumm.files.wordpress.com/2008/09/ bhoomi-e-governance.pdf (accessed 9 July 2012)

Berdou, E. (2011) 'Learning about New Technologies and the Changing Evidence Base for Social Science Research and Decision Making in International Development', *IDS Practice Paper InBrief*, ILT 4, Brighton: IDS

Berners-Lee, T. (2009) *Tim Berners-Lee on the Next Web*, www.ted.com/talks/tim_berners _lee_on_the_next_web.html# (accessed 3 July 2012)

Berners-Lee, T. (2006) *Linked Data – Design Issues*, www.w3.org/DesignIssues/LinkedData.html (accessed 3 July 2012)

Bizer, C.; Lehmann, J.; Kobilarov, G.; Auer, S.; Becker, C.; Cyganiak, R. and Hellmann, S.

(2009) 'DBpedia – A Crystallization Point for the Web of Data', *Web Semantics: Science, Services and Agents on the World Wide Web* 7.3: 154–65

Bott, M.; Gigler, B.-S. and Young, G. (2011) *The Role of Crowdsourcing for Better Governance in Fragile State Contexts*, draft paper, Open Development Technology Alliance, www.scribd.com/doc/75642401/The-Role-of-Crowdsourcing-for-Better-Governance-in-Fragile-State-Contexts (accessed 9 July 2012)

Bowker, Geoffrey C. and Star, S.L. (2000) *Sorting Things Out: Classification and its Consequences*, Cambridge MA: MIT Press

Bradshaw, P. and Rohumaa, L. (2011) *The Online Journalism Handbook: Skills to Survive and Thrive in the Digital Age*, Longman

Cole, R.J. (2012) 'Some Observations on the Practice of "Open Data" as Opposed to its Promise', *The Journal of Community Informatics* 8.2

Correa, C. (2010) 'Access to Knowledge: The Case of Indigenous and Traditional Knowledge', in G. Krikorian and A. Kapczynski (eds), *Access to Knowledge in an Age of Intellectual Property*, Cambridge MA: MIT Press

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, Oxford, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1926431 (accessed 9 July 2012)

Grewal, A.; Iglesias, C.; Alonso, J.M.; Boyera, S. and Bratt, S. (2011) *Open Government Data – Feasability Study in Ghana*, http://public.webfoundation.org/2011/02/OGD_ghana_PR.pdf (accessed 4 July 2012)

Gurstein, M. (2011) 'Open Data: Empowering the Empowered or Effective Data Use for Everyone?', *First Monday* 16.2

Hilbert, M. and López, P. (2011) 'The World's Technological Capacity to Store, Communicate, and Compute Information', *Science* 332.6025: 60–5

Hindman, M. (2008) *The Myth of Digital Democracy*, Princeton: Princeton University Press

IDS (2005) *Institute of Development Studies Strategy 2005– 2010*, Brighton: IDS

Janssen, K. (2011) 'The Influence of the PSI Directive on Open Government Data: An Overview of Recent Developments', *Government Information Quarterly* 28.4

Krikorian, G. and Kapczynski, A. (2010) *Access to Knowledge in an Age of Intellectual Property*, Cambridge MA: MIT Press

Kuk, G. and Davies, T. (2011) *The Roles of Agency and Artifacts in Assembling Open Data Complementarities*, 32nd International Conference on Information Systems, Shanghai, 4–7 December 2011: 1–16

Mayo, E. and Steinberg, T. (2007) *Power of Information Taskforce Report*, http://poit.cabinetoffice.gov.uk/poit (accessed 9 July 2012)

Molloy, J.C. (2011) 'The Open Knowledge Foundation: Open Data means Better Science', *PLoS Biology* 9.12

Murray-Rust, P. (2008) 'Open Data in Science', *Serials Review* 34.1: 52–64

Ohm, P. (2010) 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization', *UCLA Law Review* 57.1701, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1450006 (accessed 9 July 2012)

OKF (Open Knowledge Foundation) (2006) *Open Knowledge Definition*, www.opendefinition.org (accessed 4 July 2012)

Powell, M.; Davies, T. and Taylor, K. (2012) *ICT For or Against Development? An Introduction to the Ongoing Case of Web 3.0*, IKM Working Paper 16, Bonn: IKM Emergent Research Programme

Rahemtulla, H.; Kaplan, J.; Gigler, B.-S.; Cluster, S.; Kiess, J. and Brigham, C. (2011) *Open Data Kenya: Case Study of the Underlying Drivers, Principal Objectives and Evolution of One of the First Open Data Initiatives in Africa*, draft, www.scribd.com/doc/75642393/Open-Data-Kenya-Long-Version (accessed 9 July 2012)

Shadbolt, N.; Hall, W. and Berners-Lee, T. (2006) 'The Semantic Web Revisited', *IEEE Intelligent Systems* 21.3, http://dx.doi.org/10.1109/MIS.2006.62 (accessed 9 July 2012)

Shaxson, L. and Gwyn, E. (2010) *Developing a Strategy for Knowledge Translation and Brokering in Public Policymaking*, workshop report, www.dfid.gov.uk/r4d/PDF/Publications/2010-11-16_knowledge-translation-and-brokering.pdf (accessed 9 July 2012)

Smith, M. and Elder, L. (2010) 'Open ICT Ecosystems Transforming the Developing World', *Information Technologies and International Development* 6.1

Smith, M.; Engler, N.J.; Christian, G.; Diga, K.; Rashid, A. and Flynn-Dapaah, K. (2008) 'Open ICT4D', working draft, International Development Research Centre, http://web.idrc.ca/uploads/user-S/12271304441Open_ICT4D_Draft.pdf (accessed 9 July 2012)

Walji, Aleem (2011) *Let's Move Beyond Open Data to Open Development*, http://blogs.worldbank.org/developmenttalk/lets-move-beyond-open-data-to-open-development (accessed 29 February 2012)

Wind-Cowie, M. and Lekhi, R. (2012) *The Data Dividend*, London: DEMOS

Zittrain, J. (2008) *The Future of the Internet: And How to Stop It*, London: Allen Lane

Zoellick, R. (2010) *Democratizing Development Economics*, Georgetown University/World Bank