**Emerging implications of Open and Linked Data for Knowledge Sharing in Development**
**Tim Davies and Duncan Edwards**
**September 2012**

**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 & 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 re-use.

Since 2009, over 100 open data initiatives have been launched by governments, grass roots 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)[2]. With the increased

---

[1] Public Dataset Catalogs Browser, http://datos.fundacionctic.org/sandbox/catalog/faceted/ Accessed 27 February 2012.
[2] See www.webfoundation.org/2012/02/ghana-godi-launch/ for details of the launch of a Ghana open data initiative following Grewal et al's feasibility study.

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 of government statistics, to mobile applications driven by real-time open data.

Data journalism uses open government and research datasets to identify stories and present news to the public (Bradshaw & 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 datasets 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, 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 eco-systems on the web.

Firstly, we explore the historical trajectory of data management from closed data towards open data, outlining the multi-faceted 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 several examples of linked and open data in the development sector, we consider the extent to which open and linked data, as socio-technical phenomena, have the potential to challenge, or entrench, existing power dynamics in the production and consumption of knowledge. We then discuss critical issues 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 Lopez (2011) estimate that less than one percent of global information was digitally stored, to 2007 when it is thought that 94 percent of data was digital, governments, non-government organisities (NGOs), companies and communities have adopted new technologies to generate vast new datasets 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 meta-data about publications in library catalogues. Creating datasets involves making decisions about how to encode the information and developing categories and schemas to fix its digital form (Bowker and Star, 2000). Using datasets involves turning data back into information at some point, adding context and analysis: interpreting and representing it.

Just as the default for non-digital records was often 'restricted access', early digitised datasets 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 20th Century, government secrecy in many countries came under pressure from right-to-information campaigns (Krikorian & Kapczynski, 2010). Neo-liberal 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 datasets was underlined by the emergence of large companies such as Amazon and Google who rely on near-instant calculation across vast datasets 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 datasets – open to anyone to take, re-use 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 towards an American 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 datasets;
- **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 to 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 we shall see in Section 4, 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. The table below 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].

|  | **Definitions** | **Potential implications** |
|---|---|---|
| **Big data** | Data requiring massive computing power to process (Crawford & Boyd, 2011).<br><br>Big data is often generated by merging large datasets. | 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 | Access to raw data allows |

---

[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.*

| | | |
|---|---|---|
| | manipulated, sorted, filtered, remixed. For example, rows in a spreadsheet, as opposed to summary tables.<br><br>Primary data, as collected or measured directly from the source. | journalists or citizens to check official analysis. Programmers build interactive services with raw data.<br><br>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.<br><br>Often accessed over the web as a stream of data through 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 re-use (Open Knowledge Foundation, 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 RDF[5] format using URIs[6] to identify their contents, with links made between datasets (Shadbolt et al, 2006) | A 'web of linked data' supports 'smart applications' that can follow links between datasets. This is the foundation for a semantic web. |

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'*.

### 3.1. Linked data

Consider two simple (open) datasets: 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 dataset and 'target country' in the projects dataset. 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 datasets, often playing to the lowest common denominator (for example, reducing 'climate change' and 'climate policy' to 'climate' to match across datasets); 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 datasets 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 dataset**. Instead of referring to 'Haiti' as the geographical focus, a dataset would use a web link to a linked data source which defines Haiti. If two datasets link to the same URI, computers will know they are referring to the same thing. The same approach can be used for properties the dataset 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, RDF[7], to exchange data.[8]
- **Linking to other URIs to enable people (and computers) to discover more.** Linked data encourages dataset publishers to share links. For example, the 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 meta-data
Thousands of academic papers, evaluation reports and other documents on development issues

---

[7] Resource Descriptor Framework
[8] 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'.

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 meta-data – mostly only accessible through interfaces they provide. Recent pilot projects by IDS[9] and the Department for International Development's (DFID) Research for Development (R4D) portal[10] have explored approaches to opening up their meta-data.

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 meta-data, 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.

DFID has adopted a different approach, modeling R4D 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 (triple store). Both datasets have minimal restrictions, using creative-commons compatible licenses.

### 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[11] as open (XML) data. By June 2012, 54 organisations, including the World Bank, DFID, the European Union and Aids Alliance, had published IATI datasets. Mobile-phone application and an iPad-optimised website, amongst other interfaces, have been created providing accessible ways to explore the data[12]. 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 datasets 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

---

[9] http://api.ids.ac.uk/about/, accessed 29 February, 2012
[10] www.dfid.gov.uk/r4d/, accessed 29 February, 2012
[11] www.iatistandard.org
[12] www.iatiregistry.org and www.aidinfolabs.org, accessed 29 February, 2012

four countries. With support from IKM Emergent[13], Young Lives explored how linked data could be used to communicate data and findings from the study (Powel et al, 2012). Initially, this intended to represent statistical micro-data as linked data, but privacy concerns (Ohm, 2009) shifted the focus to presenting aggregate statistics and meta-data on study publications. The resulting website[14] 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 meta-data 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 datasets. 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' 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:

---

[13] The IKM Emergent programme has explored the potential impacts of linked data in development with a workshop held in 2010 and supporting a number ofdemonstrator projects. See http://linkedinfo.ikmemergent.net.
[14] http://data.younglives.org.uk

- 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 & Steinberg, 2007). For cataloguing, this involves making structured meta-data accessible to third parties instead of creating websites or services to search for publications. For research, it may involve publishing raw datasets 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 'friction' involved in requesting data or working back from published tables and website content to underlying structure, but does not make the process frictionless. 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 & Davies, 2011) – from enriching data (4.2) to creating 'widgets' that allow visualised data to be embedded in third-party websites (4.3), sponsoring 'app competitions', or underwriting initial development of tools that make data accessible to non-technical users (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 Food and Agricultural Organisation'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 re-use, IDS has launched a grants scheme, offering funding to develop applications and plug-ins that make use of the IDS API.[15]

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, matching and collaborative support functions. Modes of open data publication impact how data

---

[15] http://api.ids.ac.uk/about/grants.shtml, accessed 29 February, 2012

publishers form relationships with those re-using 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 re-using 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 datasets. Crawford and Boyd have argued that some of the large-scale quantitative big data research techniques impact on the very definition of knowledge (Crawford and boyd, 2011), 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 datasets, can lead to the subtleties in underlying datasets 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 datasets 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 datasets 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

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 (Correa, 2010). And Gurstein, citing a programme of land-record digitisation in Bangalore justified as an Information and Communication Technologies for Development (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*' (Benjamin, Bhuvaneswari, & Rajan, 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 datasets, 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, Gigler, and Young, 2011; c.f. Surowiecki, 2005), 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[16], 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 possibile 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.

---

[16] www.thedatahub.org

The formal equality of two open datasets (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 (a linked data version of Wikipedia) URIs play a central role linking between datasets (Bizer et al., 2009). This occurs because of widespread coverage of DBPedia and because it is the place that existing linked datasets link to (in order to have a bridge to other datasets). 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 re-use of common indicators or codes from a high-profile or wide-coverage datasets 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 dataset 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.org, 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 re-use data in future and often there is pressure to adopt simplified models to allow wider re-use. Similarly, work on supporting use of IATA data has stressed the need to map taxonomies that aid administrators use (water and sanitation for example), 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, meta-data, 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'[17] requires that open datasets are shared under

---

[17] www.opendefinition.org

licenses allowing re-use, 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 license incompatibility of data from different sources. However, as we have seen, there are cases where a more gradual opening a dataset (from sensitive personal data in research datasets, 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 re-users, bridging data across linguistic and cultural divides, and attentively intervening in open data eco-systems 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 largest 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[18]) reflects the impatience of an open data movement seeking access to datasets it identifies as having a potentially powerful force for good (Berners-Lee, 2009). Given many datasets are funded by tax payers, there is little justification for keeping them closed. Yet, the 'raw data now'

---

[18] See http://www.ted.com/speakers/hans_rosling.html

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.

**Bibliography**

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

Benjamin, S., Bhuvaneswari, R., & Rajan, P. (2007). Bhoomi : "E-Governance", Or, An Anti-Politics Machine Necessary to Globalize Bangalore?

Berdou, E. (2011). Learning about New Technologies and the Changing Evidence Base for Social Science Research and Decision Making in International Development.

Berners-Lee, T. (2006, July). Linked Data - Design Issues. Retrieved from http://www.w3.org/DesignIssues/LinkedData.html

Berners-Lee, T. (2009, February). Tim Berners-Lee on the next Web. Retrieved from http://www.ted.com/talks/tim_berners_lee_on_the_next_web.html#

Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., & Hellmann, S. (2009). DBpedia-A crystallization point for the Web of Data. *Web Semantics: Science, Services and Agents on the World Wide Web.*

Bott, M., Gigler, B.-S., & Young, G. (2011). *The Role of Crowdsourcing for Better Governance in Fragile State Contexts.*

Bowker, Geoffrey C, & Star, S. L. (2000). *Sorting Things Out: Classification and Its Consequences*

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

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 & A. Kapczynski (Eds.), *Access to Knowledge in an Age of Intellectual Property*.

Crawford, K., & Boyd, D. (2011). Six Provocations for Big Data. *Computer*.

Grewal, A., Iglesias, C., Alonso, J. M., Boyera, S., & Bratt, S. (2011). *Open Government Data - Feasability Study in Ghana*.

Gurstein, M. (2011). Open data: Empowering the empowered or effective data use for everyone? *First Monday*, *16*(2).

Hilbert, M., & López, P. (2011). The world's technological capacity to store, communicate, and compute information. *Science*

Hindman, M. (2008). *The Myth of Digital Democracy*

IDS. (2005). *Institute of Development Studies Strategy 2005 - 2010*.

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., & Kapczynski, A. (2010). *Access to Knowledge in an Age of Intellectual Property*.

Kuk, G., & Davies, T. (2011). The Roles of Agency and Artifacts in Assembling Open Data Complementarities. *Thirty Second International Conference on Information Systems* (pp. 1-16).

Mayo, E., & Steinberg, T. (2007). *Power of Information Taskforce Report*.

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.

OKF - Open Knowledge Foundation. (2006). Open Knowledge Definition. Retrieved from http://www.opendefinition.org/

Ohm, P. (2009). Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization.

Powell, M., Davies, T., & Taylor, K. (2012). ICT for or against development? An introduction to the ongoing case of Web3.

Rahemtulla, H., Kaplan, J., Gigler, B.-S., Cluster, S., Kiess, J., & Brigham, C. (2011). *Open Data Kenya: Case study of the Underlying Drivers, Principle Objectives and Evolution of one of the first Open Data Initiatives in Africa*.

Shadbolt, N., Hall, W., & Berners-Lee, T. (2006). The Semantic Web Revisited. *IEEE intelligent systems*, *21*(3)

Shaxson, L., & Gwyn, E. (2010). *Developing a strategy for knowledge translation and brokering in public policymaking* (pp. 3-9).

Smith, M., & 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., & Flynn-Dapaah, K. (2008). Open ICT4D.

Surowiecki, J. (2005). The wisdom of crowds.

Wind-Cowie, M., & Lekhi, R. (2012). The Data Dividend. *Notes.*

Zittrain, J. (2008). *The Future of the Internet: And How to Stop it*

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