ICT For or Against Development?
An Introduction to
the Ongoing Case of Web 3.0

IKM Working Paper No. 16

March 2012

*Mike Powell, Tim Davies and
Keisha Candice Taylor*

About IKM Emergent

In April 2007, a five-year research programme was approved for funding by the Directorate General for International Cooperation (DGIS), part of the Dutch Ministry of Foreign Affairs. The programme, Emergent Issues in Information and Knowledge Management (IKM) and International Development, is known as the IKM Emergent Research Programme. The objective of the programme is to improve development practice by promoting change in the way the development sector approaches the selection, management and use of knowledge in the formation and implementation of its policies and programs. It aims to achieve this by:

- raising awareness of the importance of knowledge to development work and its contested nature;
- promoting investment in and use of Southern knowledge production of all types and origins;
- creating an environment for innovation, supported by research on existing and emergent practice, for people working in the development sector to raise and discuss means of addressing these issues; and
- finding, creating, testing and documenting ideas for processes and tools that will illustrate the range of issues which affect how knowledge is used in development work and stimulate thought around possible solutions.

Colophon

Table of Contents

Abstract

This paper starts by exploring some of the critical questions that we need to ask of Web 3.0 technologies, based on brief reflections on the last 25 years of ICT adoption in the development sector. It then outlines some of the different aspects of the current 'data revolution', before turning to look in depth at the potential impacts of 'open data' and 'linked data' on development. Using a number of case studies it outlines how the development sector is already exploring open data and linked open data, before turning to identify other technology trends that have a role to play in constructing the future data, information and knowledge eco-system in development. The paper closes with some outline policy recommendations to guide future exploration of, and investment in, Web 3.0 technologies in the development sector.

Introduction

The development sector exists to deliver the promise that lives, which are blighted by poverty, ill-health, insecurity and lack of opportunity, can be helped towards sustainable improvement through international co-operation and various types of support. The promise can sometimes seem distant. Work in the development sector is often experienced as information overload, long and diverse lines of communication, bureaucratic and political constraints and tight deadlines. Occasionally, it may be helpful to pause and think. Are we working in ways which strengthen our capacity to reach our goals, or which undermine it? This article poses this question in the context of the development and use of new technologies, in particular Information and Communications Technologies (ICT) and in relation to a fast emerging range of technologies which may come to be known as Web 3.0.

As with many challenges in the development sector, the issues raised by this question are not new but are posed and posed again as new technologies become available or as other significant changes take place in the world as a whole to which development actors need to respond. In the 1960s and 1970s aid-supported investment in mechanized agriculture proved to be beyond the maintenance capabilities of many of the societies in which it was implanted (and often ecologically harmful as well). But it is not always just a question of the readiness of the local society for 'modern' solutions. Over the same period, the practice of spending huge proportions of the health budgets of developing countries on gleaming modern hospitals in the major cities was comprehensively debunked, not on the grounds of their ability to function but because it was proved that spending money on health rather than illness was a more efficient route to desired outcomes[1].

---

[1]Morley, David 'Paediatric Priorities in the Developing World', Butterworth 1974, represents a classic example of this argument, one which is also highly relevant to debates about the value of single issue health interventions today.

More fundamentally, definitions of 'progress' change over time. Thus in a rare and very welcome spirit of critical self-awareness, the doyen of innovation theory, Everett Rogers, was able to say in the penultimate edition of his classic book 'Diffusion of Innovations'[2]

> 'I was measuring the best recommended farming practice of that day. The organic farmer in my sample earned the lowest score on my innovativeness scale, and was characterized as a laggard. In the forty years or so since this interview.........I have come to understand that the organic farmer respondent in Iowa may actually have been the most innovative individual in my study' (p. 425)

At a meta level, the evolution over the last fifty years of more networked and globally linked societies in tandem with a range of informational developments has represented arguably the most significant change of our age. ICTs have clearly been part of such a process but whilst the growth of computing power might be following a relentlessly predictable path, the options and choices people and enterprises have had about what to do with it have been anything but. Innovation has come in waves – databases, e-mail, the world wide web, the social web – which have each offered new ways for doing things differently and each have posed financial and organisational opportunities and risks to those having to decide on whether to adopt or ignore them, or how adoption should take place.

Even so, innovation is often less instantaneous than it often appears. The term Web 2.0, 'associated with web applications that facilitate participatory information sharing, interoperability, user-centered design'[3], was coined in 1999 and was propelled into the ICT industry spotlight at a conference organised by O' Reilly Media in 2004. Although many people were by then using web 2.0 tools, the first major Web 2.0 for Development event was hosted by FAO in Rome in 2007. Web 3.0 is a term which has been in use at least since 2008 and links back to an article written by Sir Time Berners-Lee[4], the original creator of the World Wide Web, in 2001. Whilst the exact boundaries of the Web 3.0 concept are yet to be fully drawn out, an increased emphasis on data and its description is at its core: Web 3.0 technologies facilitate new forms of linkage between data (as opposed to simply links between web pages), and support new forms of manipulation and presentation of data.

The relatively long gestation period of new technologies creates an opportunity, if engagement takes place early on, for thought and experimentation on how such emerging technology can usefully be applied in any sector. Whilst the exact configuration of Web 3.0 technologies that

---

[2]Rogers, Everett M. 'Diffusion of Innovations' Fourth Edition, The Free Press, New York 1995.
[3] Web 2 entry on Wikipedia (http://en.wikipedia.org/wiki/Web_2.0#cite_note-0 accessed September 7th 2011).
[4]Berners-Lee, T, Hendler, J & Lassils, O, 'The semantic web', Scientific American, May 2001.

develop over the coming years is yet to be seen, there is no doubt that big data, open data, and linked data technologies will all have some role within the operations of international development organisations.   However, if commercial forces, academic interests, and the actions of start-ups mean that many technologies focus predominantly on data and datasets themselves as the object of interest in the bulk of the first early of Web 3.0 technologies, it is vital to bear in mind that these are not the only applications which Web 3.0 can enable. Web 3.0 technologies can also be applied to linking other sorts of information; to linking social interaction and conversations on the Internet directly to the data or documents discussions are about; to finding, accessing and playing with very specific sub-sets of information; to better connecting people, real-world resources and the knowledge required to take effective action; and other tasks which have yet to be imagined. In the past, we will argue, the development sector, or at least its most powerful organisations, have been poor innovators in the field of ICT. They have tended to let norms of ICT use emerge in other sectors and have then contorted themselves to apply these solutions to their own work, even though those solutions were created to resolve other people's problems.   By contrast, we suggest, successful innovation in ICT consists of imagining a valuable activity and then shaping the technology in order to enable it.

This paper is one outcome of a collaboration, organised by the IKM Emergent Programme, amongst people working in the development sector who are both interested in the potential of Web 3.0 technologies to improve the development information environment but also aware of the risk that they may do the opposite. As with any new technology or indeed any new investment decision, the impact of these applications on development work will depend on a number of policy decisions as to what exactly is introduced, how and with what purpose. These decisions should relate to the developmental goals towards which the investment is intended to contribute. Often however, as is the case with Web 3.0 technologies, far-reaching choices also have to be made on issues which arise in the process of implementation. Organisations must be able to identify, discuss and deal with the questions which arise. As always, thought needs to be given to possibly unintended consequences. This paper explores these policy issues, based upon a review of current projects, activities and literature, and a workshop of IKM stakeholders. A second paper[5] exploring how these policy issues may impact upon practice, specifically for organisations exploring linked data, will be available separately.

Adopting ICT in development

The G8 summit of world leaders, meeting in Osaka in 2000, which had been widely expected to try to resolve the debt crisis paralyzing many of the world's poorer countries, decided instead to

---

[5] Davies, T. 2011 in development. In the mean time, see:
http://wiki.ikmemergent.net/index.php/Workspaces:1:Linked_Open_Data (accessed 15th September 2011).

prioritize the potential of ICT as a key driver in development. We were now apparently living in 'an' information society (the singular was invariably used) in which information and ICT were central to economic prosperity. This, it was claimed, offered developing countries the chance to 'leapfrog' whole stages of development. However, despite this leapfrogging, there was also a risk of a 'digital divide' between the information 'haves' and the 'have nots', to which the entire international community was committed to taking steps to avoid. Task forces were set up, world summits convened, ICT4D programmes were launched.   Whether the whole process has been of any value whatsoever in changing the relative position between the rich and the poor is a very open question.

The focus of our argument here, however, is on the use of ICT by development organisations to manage their own operations and to communicate with others, rather than about efforts to promote ICT as tools for development in their own right. This internal use is very significant. When one considers current work and communications norms within the sector, it can be hard to remember that personal computers have only been in wide use for some twenty five years, e-mail for twenty and the World Wide Web for fifteen. Today, it is impossible to imagine functioning within a development organisation of any size, no matter where it is located, without good web and e-mail access and, almost certainly, an obligation to use ICT-based tools for numerous internal functions too. This change is significant both in terms of its impact on how we work and because it obviously represents a considerable investment.

We are not aware of any detailed empirical research on ICT spend within the sector over the last twenty years. The public accounts of most development organisations are surprisingly lacking in detail. Public criticism in the past, arguably unfair, of what were perceived as excessive administration costs encourage agencies to post as much of their expenditure as can possibly be so conceived to programme budgets rather than head office costs. In any case, money spent on ICT can cover a range of activities, such as cabling offices, buying and maintaining hardware and software, management information, training and a host of communications work, which can legitimately be accounted for under different headings. Thus even where, as in the case of the UN New York office, there is an explicit IT budget, it is unlikely that this covers all ICT spend in that office and it explicitly does not cover the IT spend of other UN stations worldwide. SOCITM, the UK association of public sector ICT managers, is quoted in a recent UK government report saying 'It is widely accepted that 3% is a benchmark of good practice in the private sector service industries for ICT spend as a percentage of total revenue'[6]. It would be surprising, given the complexity of the sector, its high demand for communications across many boundaries, its many and diverse requirements for information, if the spend in the development sector was less than this norm and

---

[6] Quoted in Government and IT- "A Recipe For Rip-Offs": Time For A New Approach - Public Administration Committee, July 2011, accessed September 15th.

indeed the figure tallies with figures provided to us informally by an IT director of an international NGO. The UK's Department for International Development[7], on the other hand, believes it will be able to meet a target of 2% of revenue to cover its entire administration costs, against what it claims is a sector average of 4%. However, as argued above, not all ICT expenditure is devoted to administration. It is also the case that substantial amounts of DFID money gets passed down a chain through multilaterals and International NGOs to local organisations in developing countries and therefore gets 'administered' and 'processed' several times over.

According to OECD[8] the total spend of all donors on Official Development Assistance in 2010 was USD 127.6 billion. If anything like 3% of this is spent on ICT, then clearly the annual ICT spend is of considerable significance and the question of the impact of this combined expenditure on the sector is a legitimate one to pose. We would argue that this spend has impacted on the sector as a whole in a multitude of ways, some of which as a result of deliberate intent and others unintentional.

Our first area of concern arises from the fact that decisions on internal ICT investment are almost invariably made with organisational priorities in mind and, if of any size, these tend to be made by senior management at headquarters level. For this cadre, many of whom will be working in parts of the organisation with little or no contact with programme staff, pressing issues in recent years have included increasing competition for funds, which may manifest itself in centralised and 'focused' objectives; a need to control costs and to be able to demonstrate cost effectiveness; and increasing demands for sophisticated monitoring systems through which to report on 'results' to donors. All of these encourage processes of centralisation and control to which certain types of ICT lend themselves and which have been followed with, it should be said, very mixed success in other sectors. We would accept that there are arguments that having such priorities and applying ICT to them may, if done well, be beneficial. However, it can also be argued that such a strategy does little to explore and much to constrain how ICT may be applied to what should be the core business of any development organisation: that is the work done, usually in intense interaction with other stakeholders, at the level that the development is supposed to take place. In this, the use of this sort of ICT mirrors and reinforces the privileging of a certain type of centralised and formalised management over the ability to work with and respond to those whose lives the 'development' is intended to improve as Tina Wallace[9] and colleagues have so fully researched and described.

Less contestable, we suggest, is the claim that the way development organisations have applied ICT has greatly widened the gap between the information rich and the information poor. Leaving

---

[7]http://www.dfid.gov.uk/Media-Room/News-Stories/2010/Spending-Review-2010/ accessed September 15th 2011.

[8]OECD, Query Wizard for International Development Statistics, accessed September 15th 2011.

[9]Wallace, T 'The Aid Chain: coercion and commitment in development NGOs', ITDG, 2006.

aside more general comparisons of bandwidth and the costs and reliability of ICT round the world, at his or her desk even a junior staff member of an international development organisation has instant access to an array of internal records and reports, external digital resources and social media which is immeasurably greater than what is available to counterparts working closer to the ground, especially in smaller organisations, especially in locations closer to the majority of the world's poor. This issue here is not that agencies are wrong to be investing in the information needs of their own staff but that they are not making similar investments in the information needs of other stakeholders in the development community on whom their work is designed to impact. Of course, exciting projects about networking, knowledge exchange or grass roots communication get funded for a period, but consistent, reliable support for the kind of information support agencies provide for their own staff is rare in the extreme. This is true both for bringing together specialist knowledge around specific themes in an organised and user-aware manner and, even more, for the enabling the sort of information environment that can offer local communities the chance to shape their own information spaces. Very few agencies even make the effort to mark up that information which they are prepared to share in a way which makes it easy for others, especially those at the grass roots, to find and use[10]. Another result of this 'us first' approach to managing digital content is the relative loss - not necessarily in terms of existence but certainly in terms of visibility - of actual content about development realities. This can even be a problem with internal systems:

> 'a large number of participants identified the need for knowledge strategies to address internal issues before addressing these broader issues... Interestingly, the focus on internal knowledge work belies the fact that all the study organisations relied on activities in the South as a key source of their most valued knowledge, and that eventually, all knowledge that is 'value generating' must by necessity be tied back to a level of Knowledge Sharing with those in the South' (Ramalingam[11])

Beyond this internal bias, lies the simple fact that the already unequal balance in the production of information between the 'developed' world and developing countries has been further exacerbated in the digital age. It may be a crude measure but a look at the 'Category: Bloggers' on Wikipedia[12] reveals 1083 American bloggers to three listed as African, two of whom appear to have European names.

Finally, attention should be paid to the ICT purchasing patterns of the development sector. We have no wish to develop an argument that large ICT companies are inherently bad or that their

---

[10]See for example, Kirimi, S. and Wakwabub, E. '<u>Learning from, promoting and using participation: The case of international development organizations in Kenya</u> ', IKM Working Paper no 6: October 2009.
[11] Ramalingam, Ben 'Implementing Knowledge Strategies: Lessons from international development agencies', Working Paper 244, ODI, 2005.
[12]<u>http://en.wikipedia.org/wiki/Category:Bloggers_by_nationality</u> accessed September 16th 2011.

products cannot offer reliability and good value. Given the tools we are using to write this paper, it would be hypocritical to do so. However, as with the previous examples of the changing balances of information visibility and wealth, what can seem sensible decisions when seen in the limited context of a particular situation or organisation, can appear very unfortunate when seen as part of the bigger picture. Is supporting near monopolistic US conglomerates to consolidate their global position the best economic model for development? More relevant to this paper is the opportunity that has been missed. The size of development sector investment could have been - and still is - sufficient to sustain a substantial community of open source developers able to produce software specifically designed with development sector realities and needs in mind. Such an approach would also allow the software produced by the expenditure of richer and more powerful organisations to be cheaply adapted and reused by less well-resourced initiatives. This process of adaptation and re-use could contribute to the growth of skills and of small businesses around the local offices of development organisations in the developing world. Such processes exist within the sector - UNESCO has notably supported a number - but they remain a poorly resourced and minority strand.

Our conclusion is that one result of investment in ICT by development organisations has been an increase in the disparities of access to and control of information and in influence and power within development discourse between the richer organisations, that have made these investments, and the less well-resourced organisations and surrounding communities which they exist to support. There is therefore an argument that very large sums of money, provided by the public and by taxpayers to promote 'development', have in fact been spent with the opposite results. This is why we suggest that it is appropriate to pause and think.

Such a reflective pause, if it is to contribute to positive future action, should pay attention to how such a situation has arisen, not least when so much of what has been said about ICT investment has concentrated on its liberatory and transformatory potential. Such a debate will have many elements. From our point of view, the most urgent argument is that choices about how and with what technologies we work are not simply technical choices but also developmental ones, choices which demand alertness and response to issues of power. They require a much deeper consideration of their strategic implications than has historically been the case.

One strand of this, which the IKM programme is in the process of elaborating more fully, is a lack of awareness of the inter-connected nature and multiple uses of development information. We have argued that the development information environment can be compared to an ecology[13]. What is

---

[13] Cummings, S., Powell, M. and Pels, J 'Development Knowledge Ecology: metaphors and meanings' Knowledge Management for Development Journal, 7:1 May 2011.

done in one part of the 'eco-system' affects the whole. Thus we in no way argue against development organisations investing in their own information systems but suggest that they should think through the impact of their decisions on the wider ecology as they do so.

A second strategic imperative is to recognise that development is based on knowledge and that knowledge, and the information upon which it is based, is needed by every person involved in any development process. One implication of this is that organisations which exist to promote development need to take all practicable steps to share their own knowledge, to contribute to the development knowledge commons. Even more important, we would argue, is the responsibility that we suggest all development organisations share to make sure all stakeholders in the processes with which they are involved have sufficient information - and by this we mean access to potentially relevant information from all sources - to participate effectively in development processes which affect them. Access to information and the ability to use it can be seen as a key capability[14] for development. Without such information, without communities being aware of the choices available and the rationale for any activity, there can be no possibility of informed consent in whatever 'development' is being done in the stakeholders' name, only various shades of imposition.

A third strategic issue is of inter-agency collaboration. In our view, it is neither possible, nor desirable, for individual organisations to meet all the information needs of all the stakeholders they work with. There is therefore a requirement for mechanisms of collaboration to help build local information resources, the value of which will develop continuously over time. In many specific geographic or thematic areas of work, such collaboration will benefit from active co-operation between those involved. However, enabling virtual collaboration - by adhering to agreed standards, by paying attention to the possible infrastructure constraints of others, by thinking of means of access and preferred language - is also increasingly important.

Finally, development cannot work if it is based purely on an understanding of the present and the past. It has to be able to look ahead and identify risks and opportunities that lie ahead and take steps to shape a more development-friendly future. While such an approach does not demand consensus, a collaborative approach offers many possibilities for mutual learning, shared risk and benefits of scale. We would recommend this approach in relation to all technologies but argue it is particularly necessary and potentially beneficial to the creation of development friendly ICT in general and Web 3 in particular. Nor can the development sector simply delegate this task to the ICT industry, its academic equivalents or its associated funding sources. These work within their own concepts and norms, which may not be appropriate to a development context. IKM's

---

[14]By this we mean to suggest that 'information literacy' should be seen alongside the other capabilities identified by Sen as being an essential pre-condition of development. See Sen, A. 'Development as Freedom' OUP.

experience of seeking to engage with them has not been entirely satisfactory. It can be hard to move the discussion from the purely technical. There seems to be an aversion to the use of exploratory and participatory research approaches in which the end products and their uses cannot be identified in advance. There is also the question of basic knowledge of the information dynamics of the societies in which the technologies may be applied. A map[15] of semantic web researchers published in 2006, shows a single respondent in South and Central America, none in Africa and none even in India. How then can they be expected to identify and respond to developmental issues without some mechanism for understanding and working with local informational dynamics?

Without informed critical input, without collaborative experiment and reflection and in a world of unequal power, the likelihood of Web 3.0 use falling into the same traps as previous waves of ICT innovation and compounding new 'data divides' on top of the information and digital divides that already work against development is very high.

WEB 3.0: A data revolution?

So what is this new wave of ICT innovation which will, we are arguing, affect the development sector, whether it likes it or not? Web 1.0, 2.0 and now 3.0 are essentially convenient labels for particular phases in the development of the World Wide Web as an information and communication technology. They are not neatly bounded, and Web 2.0 or 3.0 can often be used as marketing terms to try and show the general 'newness' of some technology. However, in the same way that the transition from a Web 1.0 world of flat pages of content published on the web, to a web 2.0 world of interactivity, social networking, rich media and online applications and services, led to many organisations having to invest in reshaping their web presence and web sites. Web 3.0 is likely to bring a similar shift in how individuals and organisations engage with the web. A common phrase used to describe Web 3.0 is 'the intelligent web'.

The latest generation of web browsers (increasingly running on mobile phones) have in-built location awareness - able to know where users are (with the users' permission). ID technologies, and social networking sites, combine to allow search engines and other services to make use of information about users to provide more personalised (and in theory, more relevant) results. And instead of just searching documents, a series of new standards, from micro-formats embedding context into web pages, to linked data, allow online applications to draw on a far wider range of content in answering users questions or providing services. Sometimes this is referred to as the 'semantic web', where services are not simply retrieving information based on keywords, but are

---

[15]MIKA, P., ELFRING, T & GROENEWEGEN, P. 'Application of semantic technology for social network analysis in the sciences *Scientometrics, Vol. 68, No. 1 (2006) 3–27 - figure 1, p. 6.*

trying to 'understand' what users want and to return the most relevant content. Beyond 'semantic search', the same standards and technologies mean different ICT systems should increasingly be able to 'talk to each other' directly, without requiring programmers to manually connect up different data sources.

Underlying this vision of a more intelligent web are a number of trends relating to data.   Data has become a hot-topic in both technology and policy circles over the last few years, and many believe we are just at the start of a 'data revolution', transforming the ways we capture, store, organise and process information. Industry has been investing in 'big data' technologies to process vast quantities of data captured in the day-to-day operation of firms, and to find economic value in it. Governments in the US, UK, across Europe, and most recently in Kenya, and international institutions like the World Bank and UN, have been publishing 'open data' to support transparency and innovation. Geek communities have been engaging with this data, to create visualisations, applications and tools driven by open data, and to create new shared datasets. A growing community, now expanding out beyond the walls of academia, is engaged in developing new ways to publish and connect all this data across the World Wide Web, working through Linked Data technologies. But what does all this amount to? And on what does the development sector need to focus?

Contemporary data revolutions

Data is fundamental to the digital world. Since computing technologies emerged they have been based on the processing of bits and bytes of data. In recent years, vastly increased computer capacity to store, process and analyze data, and vastly decreased average costs of hosting and transferring data, have opened up new ideas and approaches on how data can and should be managed and used. However, when we look at the data-related trends impacting upon society and upon organisations, specifically in the development field, we need to unpack a number of different aspects of the 'data revolution' and Web 3.0, and to explore critically some of the assumptions behind them.

Whilst some writers conflate the terms data and information, keeping a distinction between the two is important, particularly when we are concerned with knowledge management (KM). Data is encoded information whilst information is gained from a number of sources including contextualised or represented data.

Information can be captured as data in many ways. For example, an interview with a farmer about local crop management practices can be recorded using a portable video camera, encoded as video data, and published on a video-sharing website. Unless speech recognition tools are used (and these may not be available for many local languages), or meta-data tags are added to the video,

there is no way for computers to 'understand' the information contained in the video data: encoding the video as data simply allows it to be transported and shared more efficiently than recording it using analogue technology. By contrast, if a transcript of the video is shared as text, the data-encoded text could be processed in many different ways, using text-mining and forms of text data analysis to make the transcript (and the associated video) available to search, or to find connections between it and other content. Furthermore, if the 'knowledge' and information communicated in the interview is purposefully tagged or encoded as linked data using formal approaches to knowledge representation, it might become possible to carry out other forms of analysis and processing to make the information discoverable, to find connections between it and other information, or to analyze it to generate new insights and information. However, it is vital to recognise that the formal representation of the data, might miss some of the information that was conveyed in the tone of voice or turn-of-phrase used in the interview: even the video recording, if it gets compressed, can end up losing some of this richness.

Data then enables us to share, transport, retrieve and analyze information in new ways. But not all data is created equal: and the information it contains, and the information that was lost when it was encoded as data, is important to keep in mind.

We should also be aware that when we discuss disparities in access to data, and a growing data divide, we do so against a background of inequalities also in access to information. As initiatives like the World Wide Web Foundation's global open government data programme look to focus on increasing access to 'data' from southern governments, and as 'open data movements' present new data-centred campaigns, we should be mindful of the history and lessons of the Right to Information (RTI) and Access to Knowledge (A2K) movements that have struggled over recent years for political, legal and cultural change in global information environments. Access to 'raw data' (for example, the abstract facts and figures in an evaluation report), without access to the information generated from that (the report itself) can be disempowering rather than enabling. Sharing raw data can allow those with the resources to do so to reconstruct their own alternative interpretations and to set these against those provided by official information sources; but for this to happen, both processed information and raw data are needed.

Big Data

Big data is a business buzzword. It is often loosely defined as referring to "datasets whose size is beyond the ability of typical database software tools to capture, store, manage and analyze"[16].

---

[16] McKinsey Global Institute (2011), Big data: The next frontier for innovation, competition, and productivity, June 2011, p. 1.
http://www.mckinsey.com/mgi/publications/big_data/pdfs/MGI_big_data_full_report.pdf (accessed 18/09/11).

Historically we might have thought that it was only large-scale scientific projects (such as attempts to synthesize the human genome) that would collect enough data to require, and benefit from, advanced forms of statistical analysis, data visualisation and data-mining. However, in the digital age, many more organisations and settings are able to make use of big data: from companies like Tesco gathering data on shopping habits, or Facebook gathering information on people's social interaction and using this to target advertising to customers, through to search engines like Google treating all the pages on the Internet as a vast dataset from which they attempt to identify the relevance of particular web pages[17].

We say big data is a loosely defined buzz-word, as often we see it suggested, explicitly or implicitly, that fields where the datasets collected are not of similar vast sizes to those held by companies like Amazon and Tesco, or by Governments such as the UK and US, can achieve the same sorts of benefits from analyzing and 'data mining' their datasets. The UN's Global Pulse initative also plans to use some of this big data for emergency relief and development planning, in what has been coined 'data philanthropy'. Whilst for some datasets these 'big data benefits' are possible, for others, the diversity of the data, or the smaller scale of data held, means different approaches to find value in the data will be needed[18].

This paper does not consider 'big data' any further in depth beyond re-iterating the point that not all datasets are equal, and data collection or investment justified by reference to 'big data' without generating data suitable for such processing and analysis would be flawed; and noting that we need to think critically about how to use any insights generated from data-mining.

Open Data

The amount of data held by individuals, organisations and governments has grown significantly in recent years. However, much of that data is inaccessible either because it is held in proprietary formats which require expensive software to access; or because it is only released to others for a fee or not released at all; or because when it is made available the licenses under which it is published are unclear or prohibit re-use of the data. Donors, international institutions, development agencies, researchers and local communities have all built up significant (though varied) information and data resources over time, but the vast majority of this remains 'closed' - locked away, under-used and under-explored within filing cabinets and on hard-disks. Opening up this data can increase

---

[17]And, as we know the inputs to this process are biased in favour of web pages from northern organisations and individuals, we find Google's search results are 'biased' in favour of northern knowledge, a problem arguable compounded by the 'filter bubble' (REF) created when search engines such as Google attempt to use our own past searching and social online history to further customise results.
[18]And the sorts of 'value' the development sector is looking for from data may be different from that which commercial enterprises are seeking.

transparency and accountability, support collaboration, and enable innovation and the development of new approaches to problem solving. Most advocates of open data generally restrict the scope of their focus to 'non-personal' data.

Open data advocates call on owners of data to make it available so that it can be easily accessed and discovered (usually by being free to access and download online); easily manipulated (by being shared as 'raw data', rather than interpreted information, and stored in common data formats); and licensed for re-use. The widely used Open Knowledge Definition[19] says that 'open' licenses should not place any restrictions upon the re-use of data, rejecting notions of 'open' that wish to exclude free commercial re-use of information or data[20] in order to create a single open data commons that includes both commercial and non-commercial actors.

Governments and international institutions, as data holders, have been significant players in contributing to a global data commons, with Data.gov in the US, Data.gov.uk, and the World Bank's Open Data Portal all launched in the last few years with 1000s of datasets. Open data policies have often been encouraged and advocated for through hack-days[21] and competitions that take open data and use it to create websites, visualisations and applications that demonstrate the innovation potential of open data.

NGOs, local communities and other organisations have also got involved in sharing and supplying datasets, though often smaller scale data, rather than the 'big datasets' coming from government. In our own pilot mapping study of global and local open data in the development sector, we quickly found over 80 different open datasets varying in topic from environmental statistics to directories of NGOs, many of which whilst meeting open data criteria, did not self-define as 'open data', highlighting the potential breadth of open data as a data revolution[22].

When data storage and bandwidth to transfer it costs next to nothing, the argument runs, why not make data open and give everyone access to it? Whilst open data undoubtedly increases access to

---

[19]It is interesting to note that in part the adoption of the Open Knowledge Definition is down to it's accessibility - being published by a small group of people online on a dedicated domain name - and translated into different languages. It is common to see it used as a descriptive definition of what it means to be open knowledge, rather than understood as an activist intervention and a *normative* claim about how global knowledge eco-systems should operate.

[20]An exploration of the tensions in applying non-commercial terms to 'open' information, justified in some areas of the Access to Knowledge (A2K) community as a means to protect indigenous knowledge from exploitation, can be found in Krikorian, G., & Kapczynski, A. (2010). Access to Knowledge in an Age of Intellectual Property

[21]A hack-day is usually a one or two-day event where technology developers get together to build rapid prototype products using open software and data. (Note that 'hacker' is used by the technology community to refer to anyone who can program, and is not to be confused with the term 'cracker' used to refer to those breaking into systems or programming with malicious intent).

[22] Raw data at www.opendataimpacts.net/odi/2011/06/a-whole-lot-of-development-datasets/, Accessed September 2011.

data, some voices have cautioned about the creation of a new 'data divide' because of unequal access to the skills and knowledge to work with data and use it to generate information.[23]

Although some work has been done to explore the connections between open data and Right to Information movements[24], open data remains a technology-centric movement at present, a broad coalition of economic, political and technological interests much as the early A2K movement was, by contrast to the more fundamentally political or culturally centred Right to Information and local knowledge movements.

Linked data

If open data is about creating a 'commons' of datasets, published on different websites and discovered through data portals and directories, linked data is about connecting up the contents of those datasets across the World Wide Web, and building a 'web of data'.

Linked data is made up of a whole stack of technologies and conventions for publishing data. However, at its core, linked data encourages publishers of data to make use of web addresses to identify things in their datasets and, when computers attempt to access those web addresses, to return useful data to them. The companion 'practical issue' working paper outlines the technical aspects of linked data in much more detail, with an exploration of the practical and policy implications of different aspects the linked data stack. Here we pay attention to why it has been developed.

Linked data theoretically supports easier integration of different datasets: instead of every dataset using its own set of identifiers for things, linked data is based upon a simple data model (RDF), which is implemented though a mix-and-match approach to re-using existing vocabularies for describing data. This encourages data publishers to converge on common standards for representing similar data. Whilst this already happens in silos, with, for example, standards available for representing aid flows (the International Aid Transparency Initiative XML standard[25]), the linked data model makes it easier to combine different standards together, recognising that different communities might have different needs from a particular dataset. Linked data also supports smarter applications, able to follow the links from dataset-to-dataset to understand the context of the data they are using, and to discover other useful data sources. Some uses of linked data take advantage of its roots in semantic web and artificial intelligence research, using logical inference to identify implicit facts, and to better connect up diverse datasets.

---

[23] Gurstein M (2011) A Data Divide? Data "Haves" and "Have Nots" and Open (Government) Data, Gurstein's Community Informatics, 11 July 2011 (Last accessed 12 February 2012) .
[24] Krikorian, G., & Kapczynski, A. (2010). Access to Knowledge in an Age of Intellectual Property
[25] http://www.iatistandard.org, Accessed December 2011.

In practice, whilst linked data is convincingly moving beyond academia, many claims about its potential gloss over significant complexity. Tools for working with linked data are nowhere near as easy to use as most day-to-day word processing or web publishing tools, and few people have easy access to the technology and skills to properly publish or use linked data datasets. This will undoubtedly change over coming years, and there are signs of rapid change already.

Linked open data

It is possible to have open data that is not linked data, and linked data that is not open data, when, for example, linked data is used inside an organisation for its own internal data integration needs, without being publicly shared. However, it is when linked data is openly licensed that it is most useful - as without this it is far trickier for applications to integrate multiple datasets.

Visualisation

Although it relates to far more than Web 3.0 alone, and therefore goes beyond the main focus of this paper, there are clear links between the growing interest in data, the development of technologies to work with large quantities of data, and a massive growth in interest in, and tools for, data visualisation. From infographics in newspapers, to interactive displays of complex data, such as the famous bubble-chart visualisations of global developed by Hans Rosling's team at the Gapminder Foundation, the visualisation of data has the power to generate new insights, and to affect how we understand key issues. As with any other means of communication, it has the power both to inform and to mislead and is likely to be interpreted variably according to the professional and cultural background of the viewer. Visualisation[26], therefore, provides another example of a fast growing and technology-related means of communication, the rapid adoption of which within development discourse and communication could also disturb issues of access, equity and universality with the development knowledge ecology.

Early Examples from Developing Countries

There are a small but growing number of studies or fully fledged initiatives looking at the potential use of open data in developing countries. Most of these involve government data of various sorts and relate to policies or laws offering citizen rights to information. Several have been inspired by the example and, possibly, the rationale of the US and UK initiatives.

---

[26] IKM plans to publish some initial notes on how visualisation has been and might be used in development work and analysis – see:
http://wiki.ikmemergent.net/index.php/Workspaces:1._Information_artefacts#Visualisation

One of the more easily accessible is the Kenya Open Government Portal[27], launched in July 2011. Kenya – or perhaps Nairobi - has for some time had a lively community of ICT professionals, a number of whom have been involved in initiatives with social or developmental goals. Thus, in a way which demonstrates the sort of knock on value that open data is supposed to generate, there are already examples[28] of the Open Government Data being mined and reused by civil society organisations.

Feasibility studies have also been carried out for the World Wide Web Foundation on the potential for open, linked government data in both Chile[29] and Ghana[30]. They found supporters of the idea in government, civil society and academia in both places but the drivers of that interest were quite different. In Chile, freedom of information relates to memories of dictatorship and a desire for embedding democracy. In Ghana the focus on openness is more related to a history of perceived government corruption. In both countries there were serious problems with the standardisation, timeliness and accuracy of government data. In both there were constellations of politicians, activists and intellectuals who were interested in doing something with the information but not necessarily for the same reasons or with the same priorities. Both studies concentrated on the national picture, a more local perspective might have been very different.   The studies could not really be used to show any demand for or relevance of any external model. What they did show was some interest in both countries for more open information systems which could benefit certain identified audiences, interest that is being followed up in Ghana through the Ghana Open Data Initiative.

The country in the global South with probably the most experience of attempting to put government service on-line is India. There have been many initiatives at both national and state level. A number of these have been analyzed in two interesting studies. One provides a detailed case study of how a process of putting on-line land records in and around Bangalore as part of an initiative aimed at increasing government openness and the efficiency of the land market[31], ended up helping international investors and land speculators rather than the local population. The latter[32] provides a national overview and identifies a host of issues, nearly all of which would apply anywhere else. These include not only the 'elite capture of transparency', described in the Bangalore report, but

---

[27] http://opendata.go.ke/ accessed 12/02/12.
[28] For example, http://www.huduma.or.ke/ accessed 12/02/12.
[29] https://public.webfoundation.org/2011/05/OGD_Chile.pdf accessed 11/02/12.
[30] https://public.webfoundation.org/2011/05/OGD_Ghana.pdf accessed 11/02/12.
[31] Benjamin, S. Bhuvaneswari, R. Rajan, P. and Manjunatha 2007 'Bhoomi: 'E-Governance', Or, An Anti-Politics Machine Necessary to Globalize Bangalore?' A CASUM-m Working Paper . http://casumm.files.wordpress.com/2008/09/bhoomi-e-governance.pdf
[32] Wright, G, Prakash, P. Abraham, S. and Shah, N. 2011, 'Report on Open Government Data in India, Centre for Internet and Society, Bangalore,
 http://www.transparency-initiative.org/reports/open-government-data-study-india

also semantic interoperability – the difficulty of the adequate definition and translation of the multitude of terms used to describe similar things, data unreliability and timeliness, and privacy. Their main point, however, is the importance of clarifying the end purpose of work with open government data. Is it for the general citizen or a development strategy aimed at improving the lives of the poor? The answer, which does not have to be exclusive but which does need to be clear, guides all the subsequent choices as to what should be made available for whom, and the implications of that choice for which datasets are most relevant, which language is used and whether the data is for hackers or laypeople.

Data, of course, is used differently by different people. Whilst hackers, analysts and policy advocates may enjoy playing with big data, many organisations and individuals are interested in very specific data for which they have an immediate need. 'Data' also refers to a very wide range of activities. It does not consist only of the type of statistics used in World Bank or OECD reports. It also covers local price information, weather, spatial information, much of which individuals wish to access on demand.

Given that there are now over 5 billion mobile subscriptions and most of these are in developing countries, this means that the linked data web will need to be mobile, not only for the dissemination of data but also for much of its capture. Failure to incorporate the mobile in any open linked data model will fail to include the poorest, the marginalised or those located in remote and rural areas.

Data collected via mobile phones, but also tablets and PDAs can be a very effective way of gathering the data needed to build open linked data. For example the Vulnerability and Assessment Mapping Unit of the World Food Programme is using mobile technology to collect data needed to assess food security as well as food availability after disasters and conflicts[33]. They are using Personal Digital Assistants (PDAs) for long-form surveys and large data sets and mobile phones for smaller data sets and long-term monitoring[34]. It has been found that using such methods have cut down the number of errors generated as well as improved data collections speed.

The use of the 'Ethnocorder'[35], developed by Help Channel Burundi, allows monitoring information to be collected orally and tagged in a way which provides both qualitative insights and greatly improved data aggregation.

---

[33] *Mu'ammar G (2011), Gadgets Get Better Data From The Field, World Food Programme 28th April 2011* http://www.wfp.org/aid-professionals/podcast/gadgets-get-better-data-field
[34] Mu'ammar G (2011).
[35] http://www.ethnocorder.com/use.php, accessed 12/02/12. See also Horst, N 'Ethnocorder in Burundi: innovation, data collection and data use' in 'How wide are the ripples? From local participation to international organisational learning', Participatory Learning and Action no 63, IIED London 2011.

Nor does data gathering have to take place within organisational boundaries. Tapping the wisdom of the crowd for mapping & linking data that is gathered from crowdsourcing efforts has huge potential, which is still in its infancy. Studies are looking into reliability and validity issues, since participants may have varying agendas and data may not always be valid. Despite this, platforms such as Ushahadi's Swift River Platform[36] has been able to generate data crucial for helping victims of environmental disasters and some of this data has been generated via SMS.

Early Examples from the Development Sector

The IKM Emergent Programme became interested in Web 3.0 technologies, including linked open data and associated visualisations in late 2008. We are interested both in their potential development as artefacts, which could help improve the development information environment, and as a living case study of how new technologies are made known to, adopted and sometimes adapted by the development sector. Since then, we have developed, in collaboration with others, some exploratory pilot projects and tried to identify and share knowledge and ideas with others working in this field. IKM convened a workshop on Linked Open Information in Oxford in November 2010. A full document of presentations, discussions and themes from that workshop are available in a Workshop Report from the IKM Website[37].

Following the workshop, we have continued to develop our demonstrator projects and have also commissioned micro-research to map out a selection of development-related datasets available globally, and a selection specifically related to South Africa. This section of the paper presents brief details of a number of the demonstrators and pilot projects we have been involved in or have observed, and offers brief reflections on some of the issues they raise for those considering how the development sector might engage with linked and open data revolutions. A more technical summary of each of these case studies is available in the companion 'Practical Issues' working paper.

Young Lives linked data demonstrator

Young Lives is a longitudinal study on child poverty, hosted by the University of Oxford. It is following 12,000 children over 12 years in four different countries (Peru, India, Vietnam and Ethiopia) using household surveys and child surveys, inter-household data and community data related to child health, education, employment and income, family status, and welfare to understand the causes and consequences of child poverty. Young Lives has generated significant datasets and a core aim of the research programme is to make this data more accessible to policy makers, other researchers, and practitioners. We worked with the Young Lives team to explore how converting some of the data into linked open data could support the goal of increasing data accessibility and use, whilst exploring the issues involved in creating linked data.

---

[36] http://ushahidi.com/products/swiftriver-platform accessed 12/02/12.
[37] http://wiki.ikmemergent.net/files/IKM_MEETING_-_FINAL_REPORT-v1.pdf accessed September 15th 2011.

Representing a complex study like Young Lives as linked data involved considerable effort 'modelling' the data. Instead of simply creating column headings or variable names in the dataset, we needed to identify the concepts represented in the data, and to choose vocabularies, terms and identifiers to use. The decisions involved required a combination of technical knowledge (to create data which could be processed efficiently by available linked data tools), domain expertise (input from members of the study team to clarify the subtleties of different concepts, such as the meaning of India in the dataset, which really refers to a particular sample of a particular state, rather than describing statistics about the country as a whole), and knowledge of the linked data eco-system (to identify places where linkages could be made).

Whilst we created visualization tools that could graph statistics from the Young Lives data alongside comparable statistics from organizations such as the World Health Organization we found that in practice the WHO and others were not publishing linked data, and even if they were, it would need to use the same conventions as the Young Lives data (or, more likely, we would have needed to follow the conventions and data modeling decisions of the WHO) for the potential of linked data to enable easy comparison of the data to be realised.

We also encountered issues during the project about the degree to which the survey data could be 'open'. At first, we sought to publish anonymised survey results, taking care not to name the region of child respondents unless certain anonymisation criteria were met. However, on reviewing the policy around data release, we realised we could not openly publish this information online, but needed to retain the licensing arrangements of the UK Data Archive where this detailed survey data is kept to protect survey respondents. Instead we focused on publishing the questions asked in the survey as fully open data, and generating some summary statistics to publish also. During the process, we also encountered a range of different cultural approaches to open data, with data developers keen to have straightforward access to simple statistics, but academic researchers interested in ensuring statistics were always presented in context and concerned about mis-interpretation of data.

A further IKM project, building on the learning from this stage of the project is currently underway.

The Global Hunger Index as RDF

The Global Hunger Index is published annually by the International Food Policy Research Institute (IFPRI) and partners, as a statistically generated score ranking countries between 0 (no hunger) and 100 (the worst) based on the proportion of the undernourished as a percentage of the population; the prevalence of underweight children under the age of five; and the mortality rate of children under the age of five. In 2010 an Excel spreadsheet of the Global Hunger Index data, including the source data used to calculate it, was publised as open data by IFPRI alongside the

editorialised 52-page Global Hunger Index report. This open data was used by newspapers such as *The Guardian* in the United Kingdom to create their own analysis of the data, and to invite their readers to explore the data. A number of bloggers and other interested parties independently created their own visualisations of the data. IFPRI published its own custom google map visualisation of the data alongside the report on its website, and this was picked up by a number of parties and embedded in other websites and blogs.

As a simpler dataset than Young Lives, we explored how this could be converted to linked data using desktop tools. The tool we used, originally developed as 'Freebase Gridworks' has been acquired by Google and is now known as 'Google Refine'. It provides a point-and-click interface for creating RDF linked data. The linked data generated of the GHI was picked up by the FAO and integrated into their country profile linked data, taking the GHI to a wider audience and driving traffic to the Global Hunger Index website. This pilot demonstrated that for small datasets, the barriers to publishing linked data are falling as simple publishing tools become available. Where the 'linking point' between datasets is something like a country, connections between datasets are easier to envisage and make. Even so, the conversion of the data involved making choices about which identifiers to use for describing countries: as identifiers sets we chose might link to additional data on countries (from latitude and longitude and boundary data, to descriptions, flag images or political analysis) to that did not necessarily represent the publishers point of view. Selecting a set of identifiers for concepts like countries in a dataset can be both a pragmatic, and a political, decision.

The experience of IFPRI with the Global Hunger Index also shows the potential of open data for advocacy: releasing compelling data behind reports can increase the attention that the issue and the report gets, and can lengthen the life-cycle of important research work.

IKM Vines - Linked Information
IKM Vines is a demonstrator project exploring ways to combine information from different sources and to surface content from the South more prominently than it tends to be in conventional search engine results. Vines reads in textual information from articles tagged using the delicious social bookmarking platform, or shared in RSS feeds, and it then uses 'tag extraction' tools to find additional tags and key words relating to content. The tagged information, which, in the Euforic Vines prototype can include video as well as text, is made available to search and browse. For navigation, the interface presents 'leaves' on the left of the screen for the most significant categories of data uncovered, with a search box and tag-cloud on the left to help visualise the most prominent topics in the currently displayed content. Vines seeks to support both the discovery of information from the South, and enable an exploration of the particular sets of terms different communities use to discuss a subject area.

Linked data facilitates the creation of relationships between the tags used in Vines. Using data from Codesria, a partner organisation from the South, the Vines team employed terminology from a classic thesaurus. They explored adding structure to the Vines list of key words through a hierarchical organisation of terms from broader to narrower; and through more complex relationships.

Language and translation issues frequently arise in the use of development information and are likely to be just as central to attempts to link information from different sources. This is not just a question of the actual language but of the words people use. Thus Wallack and Srinivasan describe how mismatched ontologies affected development information systems for water logging in the Karnataka state of India. According to the authors, 'Bus stop conversations about water logging, and reported complaints about pipes and drains refer to the same occurrence. Yet the community's understanding of the situation and the Karnataka government record of the event label, catalogue and interpret the event in divergent manners'. This illustrates a key aspect of the digital divide, as mismatched ontology impeded the community's ability to impart information as well as the State's ability to understand the state they govern[38]. From the perspective of introducing linked data techniques to the development sector, it also illustrates the importance of paying attention to the words, ontologies and thesauri that are being used to support the new linked data architecture[39].

In this context, the Vines project has also recently turned to explore the need to map together different vocabularies, thesaurus and ontologies: connecting up legacy systems for organising information, and bridging between different fields which have their own terminology and sets of linked data identifiers for concepts and things. For example, Vines can tag content both using the commercial Thomson-Reuters 'Open Calais' system, and custom term lists of development-specific terms. Using only Thomson-Reuters Open Calais may miss key development-specific terms and lead to key information being hidden. Using only a development-focused list of terms may restrict the connections that can be made between fields. Mapping different vocabularies together can be a complex and expensive task, and often when organisations do this they do it internally and for their own purposes, not sharing the results. Creating tools to visualise the connection between different vocabularies, and to facilitate easy mapping between them by non-technical users is the next focus of the Vines project.

Vines adds meta-data to information, and in the process makes it possible to discover, sift and sort the information in different ways. However, it does not reduce information to data: you can always

---

[38]Wallack, J.S. and Srinivasan, R. (2009). 'Local-Global: Reconciling Mismatched Ontologies in Development Information Systems' Hawaii International Conference on System Sciences (HICSS).
[39] This is raised as the challenge of 'Semantic Inter-operability' in Wright et al, 'Report on Open Government Data in India, p. 37.

get back to the original articles and media, and even follow links to access them in their original context. The particular tagging services and vocabularies an instance of Vines chooses to use, and the sorts of queries that the interface chooses to make easier, will affect which knowledge is surfaced by the tool. Tools like Vines can only make use of mappings between vocabularies that are released as open data, and it is far easier to use them when they adopt standard linked data models like SKOS (Simple Knowledge Organising System) to represent the mapping.

FAO Linked Data

The Food and Agriculture Organisation (FAO) have been actively exploring the use of linked and open data over a number of years. This has included making the AGROVOC multilingual structured thesaurus, first developed in the 1980s, available as linked data; publishing country profiles and identifiers as linked data; and creating tools that support knowledge creators to take their information against these terms.

AGROVOC is a traditional thesaurus, with about 30,000 concepts, that has produced 600,000 terms, in roughly 20 languages. It is a concept-based thesaurus with ontological-based relations. Whilst Excel output has been one of the main options available in the past, AGROVOC terms are now published also as linked data. Mapping has also been carried out between AGROVOC and the European Union's EUROVOC thesaurus using a linked data model, so it becomes possible to search across information tagged in either AGROVOC or EUROVOC. In this case, FAO has taken on the task of carrying out the mapping (no small task). Mapping two vocabularies often involves careful judgments, so users have to choose whether to follow the links that AGROVOC includes, or to create their own mappings. This highlights the choice between the cost of carrying out independent mapping; or the requirement to trust the decisions made by third parties in the links they create. In our IKM Workshop this was summed up by the phrase "The economics of integration; vs the politics of delegation".

FAO have developed a range of other products to support their information management, including FAO Authority Lists for search across journals and other resources. An authority lists based approach to knowledge management, where the controlled lists define who can make assertions about what, contrast with the 'Anyone can say Anything about Anything' (AAA) principle of the web of linked data[40], but highlights that within the wider web of data, trusted authorities may remain significant players shaping choices made about which linked information sources to use and trust.

---

[40] See Allemang, D., & Hendler, J. A. (2008). *Semantic web for the working ontologist: modeling in RDF, RDFS and OWL*.

FAO have not stopped at making their open data available. They have also worked to create tools which use it - both as hosted services, and by developing and releasing open source software which other people can adopt.

The FAO experience demonstrates some of the potential of linked data to connect up disparate information sources, but also the complexity and cost involved in doing this effectively. It also highlights the value of combining open data with open source - to make it easier for diverse local groups to participate in knowledge creation and tagging, and it highlights the importance of considering the governance of key sources of authority and linkage in the web of linked data.

Critical questions about Web 3, linked data and development

Our analysis of ICT use in development organisations concluded with the identification of strategic issues for information and ICT management within the sector. These were the inter-connectedness of development knowledge; awareness of information as an essential ingredient for effective action by any stakeholder in development processes; and the advantages of inter-agency collaboration on information and knowledge issues. In addition we made the point that as well as such general developmental issues, other questions would most likely arise out of the process of adopting new technologies. This analysis offers a framework for thinking about the possible impact of Web 3.0 on development.

Impact on the Development Knowledge Ecology

Although its range of applications, at least in the development sector, is so far very limited, Web 3.0 has enormous potential to improve the connectedness of all types of information from all types of sources. Already Open Agris[41], developed by FAO, uses Web 3.0 technology (RDF) to surround the main response to a user-generated enquiry, with links to related information. The potential for breaking down the walls of the many information silos which distort the development landscape has never been greater. However, success in achieving this is not guaranteed and there is the potential for new distortions.

One potential danger is that in the excitement of discovering new insights derived from previously un-linkable permutations of data and their display, investment in the production of other forms of information will suffer from lack of attention. On the contrary, improved understandings of local realities, derived from quantitative information, needs to stimulate, not replace qualitative enquiry as to how changes might be conceived and implemented. Thus, however passionately linked data is

---

[41]This exists in a beta version. For example see http://agris.fao.org/agris-search/search/display.do?f=2010%2FJP%2FJP1014.xml%3BJP2010001379, accessed October 25th 2011.

advocated, it is vital to recall that Web 3.0 has the potential to link other types of information as well as atomized facts in datasets. It is for this reason that we prefer, when considering the impact of Web 3.0 technologies on development, to use the term 'linked information' rather than 'linked data'. IKM, with its partners, is exploring this with further pilot actions. The 'Social Life of Information' comprises an attempt to identify discussion of any data sets on the social web and to create links between the data and those discussions. Meanwhile the Knowledge Services team at the Institute of Development Studies is generating a number of topics out of the Young Lives Data, described above, to which they will link records from their unique development bibliographic database, Eldis.

Another group of problems relate to the effort that organisations are prepared to invest in creating the best balance between the openness of the ecology and the quality of the information within it. There is a danger that a group of large, 'big data' organisations like OECD and the World Bank will, as early adopters of this technology, establish norms which are the most convenient for them in terms of thesauri, authority lists and the policies on which linkage they will make, and that these norms might exclude other relevant sources of data and, unwittingly, create problem for later entrants to this world of data, posing limits on the range of data that can be explored.

Creating a world of development-related linked data within a concept of development as a purely technical activity, described exclusively in technical terms, would be to create another silo. We would instead argue that it is precisely by bringing together data and information from a range of sources - from technical experts, governments, mass media, civil society, aid transparency initiatives - that the full potential of linked data will be recognised. This will pose intellectual challenges to those trying to make good sense out of related but different data-sets, and there will be many opportunities for mis-analysis and error. It also, however, poses an operational challenge to the sector to establish norms in which different sets of information, described by different vocabularies can be brought together and made visible to one another. The process of establishing these norms, needs, we suggest, to be a collaborative one and guided by values of openness and inclusivity as well by technical and legal principles of openness.

The health and functionality of development related linked information will also depend on the policies adopted by donors. Currently, there are conflicting approaches to the ownership of and access to information. One approach seeks to use a combination of technical lock in (through the use of proprietary hardware and software) and ever more tightly defined definitions of intellectual property, often offering limited rights to use information rather than any permanent ownership of it, to secure maximum financial return on their production of information.   Another approach, characterised by the original concept of the World Wide Web, based on open and universal standards, and by the move towards the use of public information repositories and open access

journals, concentrates on making information as openly available and accessible as possible. An apparent variant of this approach is the provision of services, which are free to the user, through a proprietary system in which the provider aims to make a profit either through associated advertising revenue or from the value they can create through aggregating the data passed across their platform. In these cases, however, the permanence of the information, the option of being able to transfer it to other platforms and the arrangements for its ownership may all be subject to constraint.

Given the public or charitable sources of most of the funds used for development research and practice; the importance of information to development in a sector where many agents lack the resources to acquire information on commercial terms; pressures for transparency and accountability can be addressed by putting more information in the public domain; there are strong arguments as to why development organisations should insist on a fully open and standards-based approach. This indeed would build on existing trends of donor support in initiatives to ensure the proper communication of development research and its accessibility. However, if donors are to require that research and programme data is made freely available for use, a helpful next step would be to agree on common procedures, standards and codes of ethics. These would facilitate the potential for linkage and re-use whilst respecting the rights and safety of those to whom the data refers. This is already happening in relation to aid transparency. In a ground-breaking step in February 2011, members of the International Aid Transparency Initiative agreed on a common, open, international standard for publishing more, and better, information about aid[42]. However, it cannot be repeated too often that information about aid represents only a tiny fraction of the information that can support development.

Information support for local stakeholders

As argued above, investment in the knowledge ecology of development professionals serves little developmental purpose without an equally determined investment in the knowledge ecology of 'local stakeholders'. With some exceptions, notably around agriculture, these 'locals' have seen under-investment in their knowledge systems over many decades. Much of the this support has been led by either (or both) mechanistic, top-down, dissemination approaches or by sudden enthusiasm for either a topic (HIV or climate change) or a technology (social media or linked data). Both approaches struggle to be effective because they generally fail to connect with the dynamics of local information exchange. To IKM, a vital part of those dynamics is the need for a critical awareness of the local processes through which information from external sources will inevitably be viewed and validated and in which the people responding to the information need confidence if they are to feel empowered to act.

---

[42] http://iatistandard.org/ , accessed February 10th 2012.

We put 'local stakeholders' in inverted commas to emphasize that they do not constitute a unified and homogeneous group - 'local' can mean national, provincial, district or village; it can mean government, civil society or individual; it can mean the public or the commercial.   Information exchange needs to be effective in all these domains if development is to take place. The 'local' is, of course, heavily imbued with local dynamics of political and economic power, gender and language. Information or knowledge transmission is not a neutral process. The issue of what Wright et al., call the 'elite capture' of transparency' has already been mentioned above but there are other potential pitfalls as well. Evangelia Berdou, looking at the experience of mass data capture in Kenya, observes that "The ease with which these platforms can be deployed means that marginalised groups may be viewed simply as data sensors, cheap sources of hard to get information."[43] Furthermore, the increased visibility of vulnerable communities can also have undesired consequences[44]. There can also be issues of competing economic interests. One of the expectations of the US/UK models is that the freeing of government data will provide a host of opportunities for information businesses. However, the quantity or value of such data may not be the same in developing countries. Government data may not be collected quickly or accurately enough, which is why there may be commercially based services, such as Tradenet in Ghana, potentially competing with the 'public good' of linked government data. Apparently open mapping processes may create rights over local content which can exploited by others or, if not properly conducted may allow one set of claims to gain credence over others on the basis of first use of the technology.

It follows that for us, any new investment in local information environments should not be simply premised on some new technology but on detailed consideration of what developmental changes are desired (and by whom) and the potential role in information and knowledge in creating that change, all in the context of existing information dynamics. Once such an analysis is carried out, and it is almost certain to reveal many information gaps and communications problems, it is then possible to think about the potential value of any new technology and how it may be introduced as addressing in part the issues identified.

To our knowledge, one of the most thorough attempts to do this in the context of introducing linked open data can be found in Aidinfo's work with data about aid in Nepal. They found that although there were a number of initiatives making their information about aid available involving 'significant

---

[43]http://www.thinkinnovation.org/en/blog/2011/09/mediating-voices-communicating-realities-using-information-crowdsourcing-tools-open-data-initiatives-and-digital-media-to-support-and-protect-the-vulnerable-and-marginalised/
[44]http://www.thinkinnovation.org/en/blog/2011/09/mediating-voices-communicating-realities-using-information-crowdsourcing-tools-open-data-initiatives-and-digital-media-to-support-and-protect-the-vulnerable-and-marginalised/.

amounts of work for all those involved but, despite this, the information made available does not meet many of the needs of users or potential users'[45.]

'Making aid information accessible and useful to a wide range of people in Nepal' they argue, 'requires an information ecosystem involving many actors', a system which has three essential elements:

1. 'Data that is available to everyone: data about aid should be published on the Internet according to a common, open standard.

2. Information that is accessible to everyone: Information intermediaries are needed to make information accessible and useful for different groups of people.

3. Capacity within organisations and individuals: to use the data and information for decision-making and advocacy.'

Collaboration

The various examples given above of the actual or potential use of linked data (and linked information) technologies in development-related activities probably give only a fraction of the idea of how they will actually be being used in five years time. Yet already the vast scale of their likely use and potential impact – both positive and negative – can be foreseen.

We welcome and share Aidinfo's conclusions quoted above. We observe, however, that they imply a massive effort if their goal of building an appropriate 'information ecosystem' is to be achieved. This is especially so, if it is recognised that such a system needs to contain other development related information - about health, agriculture, water etc. from sources both local and international – as well as information about aid if it is to be a really effective resource for development. And, by implication, this effort is needed not only in every country but around every theme across the whole development knowledge ecology.

In our view, the creation of such an ecosystem even for a single field or place is beyond the capacity of any single organisation. Nor, we argue, would such an approach with the level of control it implies be desirable. Such an ecosystem, in each place and at each level, needs to be alert to and ideally led by the needs of its 'inhabitants'. It should be, we suggest, almost by definition a public good and therefore not subject to restrictive commercial or intellectual property interests, be they relating to software or content. Development information eco-systems can therefore only thrive in an atmosphere of collaboration between agencies, governments, community organisations and others.

---

[45] http://www.aidinfo.org/case-studies/better-information-better-aid-nepal-country-study accessed 11/02/12.

In the virtual sphere, as Aidinfo also argue, the agreement and adoption of common standards is an essential precondition of connectedness. Co-operation will also be needed on accepting and providing the necessary translation between heterogeneous ontologies. However, we see physical collaboration – the coming together of stakeholders to collectively discuss and understand issues – as being equally important. Overall, the creation of such ecosystems and of the collaboration needed to build and sustain them will, even if done in a gradual and incremental way, demand mechanisms of collaboration and co-financing which have yet to be created.

Pause, Reflect, Act?

The promise of linked open data is significant: smarter web applications and easier data integration. It is, however, still in its relative infancy, with many challenges to overcome, and is certainly not a panacea to all knowledge management challenges. The tools for publishing, working with and consuming linked data remain predominantly aimed at researchers and specialists, and the breadth of the linked data movement means that many people are, if not pulled in different directions with respect to its development, at least working on the basis of different understandings of its future trajectory.

There is a naïve argument which suggests that linked open data is simply a good thing. That it automatically promotes transparency and accountability and offers many potential opportunities for business spin-offs. Referencing discussions from IKM's 'Linked Open Information: Potential opportunities and pitfalls for International Development' workshop, and reflecting on how ICT has been applied by the development sector in the past, this paper examined how web 3.0 is being applied in the development sector, as advocates tout its benefits. We find that many of the same challenges that have been faced in the application of ICT tools to address development still plague the use of linked data. Inclusiveness is still a problem and as Wright et al., argue: 'the reasons that work well in the US and the UK may not work well in India' (Wright, p. 40) or, indeed, anywhere else. There is a real danger that a framework developed by the West will exclude a large part of the world's population, especially those that it is meant to help. Web 3.0 must be developed to support local information environments in diverse global settings and information work at this level will require significant investment of time, technology, skill building and money needed if it is not to be left behind.

In addition, problems with openness and standardisation persist, and this is compounded by issues of language, ontologies, authorities, which hold western bias. Data visualisation also needs to be explored with critical awareness of culture and how images are interpreted. Linked data must also incorporate ICT tools that are prominent in the developing world. The use of mobile phones in particular is vital in the development of the linked data infrastructure.

Development organisations and other actors have a choice in how they respond strategically to this emerging landscape. This choice might consist of:

- wait-and-see, leaving it to the market to shape what can and cannot be done;
- critical participation as a creator or user of linked information (with evaluation that is sensitive to the issues raised in the paper);
- pro-active action to build the development-friendly aspects of an open information data environment.

In our view, and particularly in view of the often unintentionally anti-developmental ICT policies adopted by development agencies in the past, the first choice would imply a complete abdication of responsibility. We hope instead that it will be possible to construct a shared approach that encompasses:

- A collaborative effort with common standards and, ideally common but open approaches to issues of language, ontologies and governance
- A critical awareness that the actions of individual organisations affect the wider knowledge ecology
- Support for different types of information, not just big data. In particular, we encourage attention given to effective ways to compare and connect quantitative data found in linked statistical datasets with qualitative data and information
- At least an equal investment in meeting the information needs and capacities of the poor and marginalised, appropriate to their own particular contexts, as is made in the systems of the larger and powerful organisations working in development.

Bibliography

Allemang, D., & Hendler, J. A. (2008). *Semantic web for the working ontologist: modeling in RDF, RDFS and OWL*.
        Benjamin, S. Bhuvaneswari, R. Rajan, P. and Manjunatha 2007 'Bhoomi: 'E-Governance',
Or, An Anti-Politics Machine Necessary to Globalize Bangalore?' A CASUM-m Working Paper
http://casumm.files.wordpress.com/2008/09/bhoomi-e-governance.pdf
Berners-Lee, T. (1999). *Weaving the Web: The Past, Present and Future of the World Wide Web by its Inventor*
Cummings, S., Powell, M. and Pels, J. 'Development Knowledge Ecology: metaphors and
meanings' *Knowledge Management for Development Journal*, 7:1 May 2011
Krikorian, G., & Kapczynski, A. (2010). *Access to Knowledge in an Age of Intellectual Property*. *The Teachers College Record*. Zone Books. Retrieved from
    http://www.tcrecord.org/Content.asp?ContentId=775
Wright, G, Prakash, P. Abraham, S. and Shah, N. 2011, 'Report on Open Government Data in India,
Centre for Internet and Society, Bangalore,
 http://www.transparency-initiative.org/reports/open-government-data-study-india

IKM Working Paper series

Julie E. Ferguson, Kingo Mchombu and Sarah Cummings (2008) Management of knowledge for development: meta-review and scoping study. *IKM Working Paper* No. 1, April 2008, 45pp.

D. Wickremasinghe (2008) Communicating Information and Knowledge Management: Challenges and Approaches. *IKM Working Paper* No. 2. April 2008.

Serafin D. Talisayon (2009) Monitoring and evaluation in knowledge management for development. *IKM Working Paper* No. 3, July 2009, 50pp.

Laxmi P. Pant (2009) Learning networks for bridging knowledge divides in international development: approaches and initiatives. *IKM Working Paper* No. 4, August 2009, 39pp.

Harry Jones (2009) State of the art literature review on the link between knowledge and policy in development. Joint *ODI-IKM Working Paper* No. 5, August 2009, 37pp.

Stephen Kirimi and Eliud Wakwabubi (2009) Learning from promoting and using participation: The case of International Development Organizations in Kenya. Joint *PAMFORK-IKM Working Paper* No. 6, October 2009, 44pp.

Hannah Beardon and Kate Newman (2009) How wide are the ripples? The management and use of information generated from participatory processes in international non-governmental development organizations. *IKM Working Paper* No. 7, October 2009, 29pp.

Paula Zirschky (2009) Knowledge management and multiple knowledges: A multi-case study within the Netherlands. *IKM Working Paper* No. 8, November 2009, 26pp.

Adnan Rafiq and Nazish Rafiq (2010) Good planning or benign imposition? Innovation, emergence and risk in developmental research: Learning from ICTD. Final Draft Report. JBS-IKM-BDDG Workshop 17-18 September 2009. *IKM Working Paper* No. 9, January 2010, 29pp.

Julian Jenkins (2010) 'Things can be other than they are'. Understanding the limitations of current management thinking and knowledge practice for work in the development sector. *IKM Working Paper* No. 10, July 2010, 28 pp.

Robin Mansell (2010) Power and interests in developing knowledge societies: exogenous and endogenous discourses in contention. *IKM Working Paper* No. 11, August 2010, 37pp.

Simon Hearn, Ewen Le Borgne and Valerie A. Brown (2011) Monitoring and evaluating development as a knowledge industry: ideas in current practice. *IKM Working Paper* No. 12, July 2011, 29pp.

Ewen Le Borgne, Valerie A. Brown and Simon Hearn (2011) Monitoring and evaluating development as a knowledge ecology: ideas for new collective practices. *IKM Working Paper* No. 13 July 2011, 30pp.

Robin Vincent (2012) Insights from Complexity Theory for the evaluation of development action: recognising the two faces of complexity. *IKM Working Paper* No. 14 March 2012, 47pp.

Sebastiao Darlan Mendonça Ferreira (2012) Evolution and future of the Knowledge Commons: emerging opportunities and challenges for less developed societies. *IKM Working Paper* No. 15 March 2012, 37pp.

Mike Powell, Tim Davies and Keisha Candice Taylor (2012) ICT for or against development? An introduction to the ongoing case of Web 3.0. *IKM Working Paper* No. 16 March 2012, 34pp.

IKM Background Paper series

Sarah Cummings (2008) IKM Emergent Communications Strategy, *IKM Background Paper*, December 2008, 45pp.

Joitske Hulsebosch, Mark Turpin and Sibrenne Wagenaar (2009) Monitoring and evaluating knowledge management strategies, *IKM Background Paper*, October 2009, 44pp.

M.J.R. David (2010) Storycasting on telradio: A discussion paper. *IKM Background Paper*, January 2010. 12pp.

IKM-Emergent Contact Details

Sarah Cummings, Communications Coordinator
IKM Emergent, The Netherlands
sarah@ikmemergent.net, sarahcummingswork@gmail.com
www.ikmemergent.net