Linked Open Information: Potential opportunities and pitfalls for International Development

IKM Emergent Workshop

Rewley House, Oxford, UK

November 15 & 16, 2010

DRAFT FINAL REPORT

Prepared by Keisha Taylor and Ginette Law for IKM Emergent

Abstract/Overview

The workshop titled *Linked Open Information: Potential opportunities and pitfalls for international development* gathered twenty-one participants from a variety of backgrounds and organisations to discuss emergent related issues. Though exploratory in nature, the workshop aimed to:

- Imagine what a development information environment built upon linked information might look like.

- Conceptualised what a linked open information would ideally look

- Find out what might prevent it from evolving

- Consider the technical, policy and organisational challenges which will be faced in working towards the best imaginable outcome.

There were several informative and interesting presentations on varied topics. Confusion surrounding the definition of various linked data terms was highlighted and the difficulty of validating data was also expressed. The possibility of using aid data, micro datasets and visualisations for development was also discussed. The need for inclusiveness of communities and different countries in the linked data infrastructure was reiterated along with the need to understand information behaviour. While policy and technological issues were examined individually, the need to bridge the gap between linked data policy and technology was also emphasised and participants shared their vision of what a successful linked data infrastructure would look like.

Given the informative discussion and interesting questions raised there is a lot of opportunity for IKM to continue work in the emerging areas discussed.

Table of Contents

It was noted that there is a need to do more research to find more linked data and that useful urls should be sent to the dgroup. It was indicated that this could be a useful crowdsourced activity because the Open Knowledge Foundation (OKN) have a development working group and they want to do some more mapping. It was pointed out that the Comprehensive Knowledge Archive Network (CKAN), which has been developed by OKN, only has a fraction of data that is available and that it would be useful to also gather multi-lingual data as well. Emphasis was placed on the need for CKAN to provide more information about the vocabulary and datasets that they have worked on to have a more complete idea. Light annotation was

1) Introduction and Opening Greetings (Mike Powell)

This exploratory workshop began with opening remarks from Mike Powell, (Director, IKM Emergent), who reflected on the extent to which the development of the internet, the world wide web and related tools has been dominated by people and organisations from Northern countries. He stated that ensuring such tools could also effectively serve development goals, which require communication within and across other cultures, would not happen automatically. It requires people who are interested in information for development to speak to their colleagues and discuss related issues as new standards and practices emerge.

As part of the agenda was related to the visualisation of data, he asked participants to look at this map and reflect on how, without some deliberate intervention, we could assume that the semantic web would know about and reflect the information needs of people living in developing countries.

Figure 1. Semantic Web researchers and their network visualized according to geography[1]

---

1 From 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*

Since many participants present at IKM Emergent workshops were new, this provided a lot of opportunity for more emergent discussion. Mike Powell also reflected on the way in which the development community has come full circle on information for development issues. In the early 1990s there was a lot of excitement about this and people recognised opportunities to use information and the internet information for development. At the Okinawa G8 Summit, held on 23rd July 2000, it was said that the world did not need to worry about debt in developing countries because a new inclusive information society would be created, which would allow for the leapfrogging of whole stages of development. Mike Powell recalled speaking at the Web 2 for Development Conference, hosted by the FAO in Rome in 2007, about the necessity to examine the information needs of communities before deciding how to meet meet them, but feeling the message was drowned out in the enthusiasm to apply new technology almost as an end in itself. Mike. Powell argued that democratisation of information, and various other development related benefits promised by uncritical promoters of new technologies, has not yet happened and may likely still take time. This is something to think about as the next wave of web development, the semantic web, gathers similar steam.

Mike Powell also discussed a working paper written by Robert Mansell for IKM titled "Power and Interests in Developing Knowledge Societies: Exogenous and Endogenous Discourses in Contention". The paper speaks about the history and (lack of) balance between endogenous and exogenous models of Information Communication Technology for Development (ICT4D) applications. Having researched the speeches of various UN organisations over the years, Mansell found that there has not been any balance between the two models. She argues that exogenous models have been exported and imposed whilst that lip service was made to endogenous models of ICTD, as no real resources were allocated to make local solutions work. Despite the many millions of dollars spent on ICT in the 1990s the gap between the information rich and poor has still widened. Themes which are important to IKM's work include:

1. **Languages and translation** – These can vary between different disciplines, jargons, status etc., which are related to semantics and linked data. They also include the examination of words, ontologies and thesauri that are being used to support the new linked data architecture. An article written by Wallack and Srinivasan[2] was given to highlight the importance of this work. It examined attempts to reconcile mismatched ontologies and

---

2

Wallack and Srinivasan, 2009 Local-Global: Reconciling Mismatched Ontologies in Development Information Systems", Hawaii International Conference on System Sciences (HICSS) 2009

development information systems by looking at evidence about waterlogging 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.

2. **Examination of different linked data rules and schemas** - This is both an issue of how to support all stakeholders in the development field so that they can produce their information in linked form and the issue of how to ensure that the rules/permissions developed for linking or blocking links do not discriminate against poorer and more marginalised information sources.

3. **Encourage openness to southern voice** - Without inclusion of the voice of the people, the poor and local intellectual tradition, knowledge within the development sector will be insufficient. Including the voices of others includes thinking about the cultural contexts in which such voices are expressed and listened to.

Mike. Powell gave the story of L.J. Green[3] South African anthropologist's work in Brazil with the Palikur people in Roraima (state of Brazil). The anthropologist recorded their knowledge and perceptions of life using seven chapters or groupings of stories, which included stories of creation, clans, wars, floods, dispossession and slavery, shamans, conversion to Christianity amongst others etc. The author said:

"We organized a workshop with Palikur elders and interested parties and sought to find consensus on the chapters of the Palikur historical canon. There seemed to be at least seven 'chapters' or groupings of stories...Yet clustering these stories thematically was paradoxical: not only did several story groups overlap but also neither the categories nor the material in them lent themselves to chronological ordering. It seemed uncomfortably obvious that we were imposing a structure that did not really matter to local people. As field research progressed, we began to correlate stories with archaeological sites that we were investigating the landscape almost without realising it. After awhile, mapping the stories began to seem the obvious alternative to imposing a chronology. We began a memory-map

---

3

   Green, L.J, 2007 Cultural Heritage, Archives and Citizenship: reflections on using Virtual Reality for presenting knowledge diversity in the public sphere, Critical Arts, Volume 21, Issue 2 November 2007 , pages 308 - 320

project in order to locate stories spatially, and compiled a database of stories in relation to place names. Apparently, the approach echoed the everyday practice of telling stories about the places through which this one was moving. Yet this too ran into difficulty: where on a map do you put stories of the stars? What about the stories of the underworld? And what to do with stories with a specific seasonal relationship to a place, such as when the water was high, or when it was appropriate to hunt or to make a field? The most perplexing was where to place text where the hero was said to move directly from the underworld into the upper world of the sky - apparently not passing through this world at all. In this regard neither a horizontal mapping of the world (like a map), nor a vertical mapping of the layers of the world (as if they were stacked up from underworld to sky) provided a ready logic. In short: familiar ways of visualising space - via cartography - and familiar ways of organising time - through chronology – seemed to be blocking, not assisting, my understanding of Palikur knowledge".

Mike. Powell stated that although this does not directly relate to linked data for development, it does highlight issues that arise when trying to link information from different sources together. These are questions that should be asked as a new linked data architecture system is developed.

4. **Power elements of knowledge and diverse information sources** - Although IKM welcomes efforts by the World Bank, the British and American governments and the OECD to make their data open and potentially linked, IKM is also a bit hesitant about accepting their dominant role in the development information landscape: '*We should not only see a World Bank open data competition but there should be additional competitions to hack into and play with data and information in other countries like Brazil, India for instance.' (Mike Powell)*

5. **Recognition of unpredictability and emergence in development work** IKM believes that by encouraging people from all over the world to work on an issue, new and better ideas can emerge. It does not expect to know in advance what such conversations will produce.. This approach is somewhat different to the practce of detailed planning in most development organisations where plans, outputs, targets, activities etc. are all are mapped out in advance.

Though the workshop was organised in an exploratory way without stating specifically what should be examined, the intention behind the meeting was to map out a direction for developmentally friendly linked information as well as discoverissues of which IKM are unaware (See IKM agenda). Mike Powell asked for three issues to be considered, which included:

• What would a development oriented linked information system look like?

• Who could and should participate in contribute to it?

- What are the challenges, in all domains (e.g. policy, technical) that we face in trying to making the new linked data information environment a developmentally friendly one?

DAY 1

2) Linked Open Information: An opportunity to question or clarify the understanding of the term.

Linked open information, especially in the development sector, is still a fairly recent phenomenon. This has consequently brought rise to many questions and discussions about what exactly is should be. Will a world of linked information be one consisting of portals which are useful to the general public or will everyone have a completely different sense of how they create and access linked information? The first day of the IKM Emergent workshops tried to present a broad overview of what currently constitutes linked data in the development sector. Three presentations on the topic were presented, followed by a discussion on some of the general issues that should be addressed in linked data for development.

2.1 Presentation 1: Linked Open Data (Tim Davies)

Linked information is still to be properly defined. The term and use of *linked data* is also yet to be established. However, it is important to underline that data is not the equivalent of information. To better understand the difference and relationship between the two, Tim. Davies presented a simplified summary of the DIKW model by Rawley's (2007).

Simplified summary of DIKW model
Drawing on Rowley (2007)
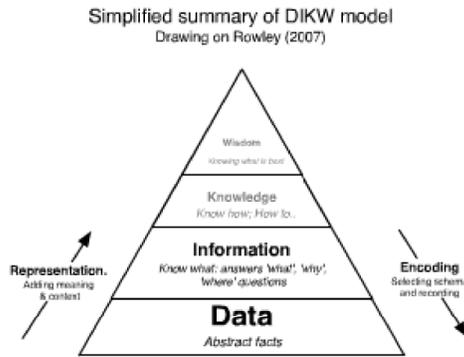
Information is generated from data by representing the data in a more human readable form. Essentially, it is the processing of data. For example, information presents data in a specific context, or uses established visual metaphors (maps; graphs; lists; tables etc.) to reveal the facts and knowledge the data encodes. Data is generated from information by encoding it in structured ways. Any encoding of information needs to decide how much of the original information to preserve, and how much of the informational context to capture and encode.

There are many kinds of data. These include:

1) Primary data

2) Secondary data

3) Meta-data

4) Operational data

RDF provides a model for representing information in data that is widely adopted for linked data. It can be 'serialised' (or written to a file) in a number of different file formats. For example, N3, Turtle, RDF/XML and NTriples. Notation 3 [N3] is optimised for human readability. RDF/XML is accessible with a common XML tool-chain. An RDF statement takes form of a *triple:*

Subject (S) - Predicate (P) - Object (O)

This model is similar to the one used in a spreadsheet. The following example illustrates how the triple model can be found in a spreadsheet or sentence.

	Location (P)
IKM Workshop (S)	Oxford(O)

The 'IKM Workshop'(S) is taking place(P) in Oxford(O)

In this case, 'IKM workshop' and 'Oxford' are just strings; they cannot be further annotated. However, when RDF triples are used to relate resources, further statements about those resources can be added.

<Oxford> <type> <City>

<Oxford> <postcode> "OX"

Any RDF representation of data is made up of the RDF mode; a file format (serialisation) and one or more vocabularies. Those vocabularies may also have linked ontologies, and schema for validating the representation.

Vocabularies provide a common list of terms to use in creating an RDF document. Common vocabularies include 'Dublin Core' for meta-data terms; 'SKOS' for describing knowledge bases; 'FOAF' for describing agents, their attributes and relationships. Vocabularies can be mixed together in an RDF model.

Ontologies are vocabularies that, in addition to this, attempt to represent the semantics of data. (For example, an ontology might articulate the relationship between the two vocabulary terms by recording that one is a sub-class of the other — or that one has a narrower scope than another. Ontologies can support reasoning across RDF data.

Tim. Davies highlighted that the openness of data depended upon three factors:

1) **Findability -** Is the data published online so it can be spidered and searched?

2) **Machine-readability** - Is the data in a standard format that machines can process?

3) **Re-usability**- Does the license the data is shared under, permit it to be re-used?

The openness of data also depends upon legal, technical and practical aspects. So what is actually meant by *Linked Open Data?* The term can be defined by examining the definition of each word:

Linked - Vocabulary/Ontology, re-use, Using URIs to identify things, connecting data across HTTP

Open - Accessible, Format, License, Re-Use

Data - Encoded information, RDF Triples Model, Many formats (N3/RDFXML, etc.), Vocabularies, Ontologies, Primary/Secondary/Meta/Operational

Open Linked Information could therefore be defined as publishing data and turning it into information or recording meta-data about information via linked data standards.

Tim Davies used the [Young Lives Open Data Demonstrator](#) as an example of modelling and representing linked open data. Young Lives is a longitudinal study on child poverty. It follows 12,000 children over 12 years in different countries using household surveys and children surveys, inter-household data and community data related to child health, education to understand the causes and consequences of child poverty. Young Lives has access to large and small datasets and aims to making them more accessible to policy makers and practitioners. However, this is difficult to do because they are statistical datasets. As a result, Young Lives got involved with IKM to try and make the data more accessible. From this project, the following desiderata was formulated for modelling:

- **Simplicity & flat structure:** to make querying the data easier and to help those unfamiliar with the graph structures possible in RDF to quickly find and use sets of the data.

- **Comparability with SPPS data:** keeping the same variable names.

- **Ease of annotation:** allowing additional information to be attached to questions.

- **Re-use of other vocabularies and ontologies**: increasing the chance of existing tool-chains being able to operate against the dataset.

- **Making linkages:** allowing data from the wider web of linked data to be used in the demonstrator.

There were a few lessons learnt from this project. First, it was important to be pragmatic about semantic modelling and finding shared vocabularies, in addition to finding the availability of linkable data. There were also some unexpected bandwidth issues. Tim Davies realised that there were many issues and complexities in doing this sort of project. Despite the fact that there was very good data and its sources were known, it was an enormous task to undertake. As Ms Caroline Knowles, from Young Lives highlighted, linked data is still in "its very early days". If there were some 'off-the-shelf' tools that existed or if collected data were already designed to be linked, it would make the task much easier. Unfortunately, this is not yet the case.

For more information:

See Tim Davies' PowerPoint presentation at: http://www.slideshare.net/timdavies/young-lives-linked-data-demonstrator

The Young Lives Comparator: http://practicalparticipation.dyndns.org/about/

2.2) Presentation 2: IKM Vines - Vines to extract knowledge (Hugo Besemer)

The Vines sites are browsers that combine information from different sources, updated from del.icio.us. Codesriavines contains references to articles to electronic Codesria journals; Euforicvines contains Euforic items tagged "South". IKM Vines combines both. For navigation there are "leaves" on the left for the most significant categories, as well as a word cloud and a search box on the left. In Euforic vines special media types (video) may appear (IKM Vines).

The first phase of the vines tried to extract knowledge as a way of making content from the South more visible. Content can be searched through vines, which have been laid down by different groups of people. The two principle goals of this project was to 1) make content from the South more visible and 2) find out if the right terminology was being used to search and extract information from existing data collection searches.

Using data from Codesria, a partner organisation from the South, Hugo Besemer's team employed terminology from an old fashion thesaurus. They learnt that if you want to put some structure in the list of key words, you might find hierarchal words from broader to narrower use of terms. Hugo Besemer underlined that there are also other ways to use a thesaurus.

Eg. *The bee is an insect.*

There is a clear hierarchical relationship between the word *bee* and *insect.* We also know that there is some sort of relationship between *bee* and *honey.* But this is not a hierarchical relation. In this case, some sort of qualifier is needed to define the relationship between the two words.

Eg. *The bee produces honey.*

The relationship between the bee and honey now becomes more evident. This is what, in some circles, is called *ontology.* In the agricultural sector, an old fashioned thesaurus is still sometimes used, and has mainly been developed for library systems. Thus, the following questions arise from the Vines project:

- Can we use these systems in the future?

- Can organisations rely on them?

- Who is going to use them?

- How will they be maintained?

Hugo Besemer presented a demonstrator using a word cloud and a more traditional system from Codesria to illustrate the different relationships between the categories and tags. He is currently working on this phase now. Extracting texts and RSS feeds from a particular source, his team is now classifying data

and news feeds items using colour. The same data can be presented from different viewpoints, as Hugo Besemer illustrated with a demonstrator. This allows different perspectives to be shown using the same data.

The next step will be to make linkages between the two points of view. Hugo. Besemer will need to develop a way to classify what is being searched in both examples to somehow connect them. Afterwards, he plans to query across different results using a standard query launch. This can be done using SPARQL, by selecting a draft where two different things are simultaneously occurring and therefore using the same URI for two distinct things. Hugo. Besemer wishes to compare the different classifications or terminologies that are being used to express a certain idea or concept.

When asked, Hugo Besemer defined the future success of this project as being able to determine patterns in the classification of linked data. This means discovering trends in the categorization of data through different perspectives and examining the way in which data and viewpoints differ.

For more information:

Visit Hugo Besemer's website: http://hugobesemer.net/codesriavines/ and http://www.hugobesemer.net/ikmsemantic

2.3) Presentation 3: Building capacity for linked data in FAO and partners (Imma Subirats)

Ms. Imma Subirats from the Food and Agricultural Organization (FAO) presented five different types of existing themes, vocabularies and models used in the agricultural sector for development. First, a description of Agrovoc and various concept schemes were presented, along with explanations on how FAO had become a producer of linked data. She then discussed the various types of tools, which allowed countries to create linked data and how they functioned.

The Agrovoc thesaurus is a traditional thesaurus, with about 30,000 concepts, that has produced 600,000 levels, in roughly 20 languages. It is a concept-based thesaurus with ontological-based relations (as mentioned in Hugo. Hugo Bessemer's presentation). The system is concept string-based, which at the

moment has an Excel output. It uses URIs with the RDF concept. Agrovoc is supposed to publish triples by the end of this month. It is possible to link vocabularies through Agrovoc with other thesaurus such as Eurovoc. This is meant to produce a thesaurus of linked data, coexisting with other vocabularies that also have the same string or concepts in a system. The outcome is a stream of data from both sources, which then can be used for linking further data.

Another online FAO product is the FAO File Description Concept Scheme (which is a librarian type system). This is an authoritarian control system based on concepts. Since a URI is provided for each concept, there is no longer an authoritarian control standard lease. There are instead concepts for each authoritarian record. The quality of the authoritarian description scheme is quite internal. The File Description Concept Scheme has a list of online agricultural journals linked to a set of useful tools.

In library sciences, an authority list contains categories and graphic records with the possibility to link to organisations (e.g. the University of Oxford), giving access to all publications from the organisation in question. It is important to note that the graphic records and the operated control lists are often systems that work in parallel. The idea behind an authority list is that there is someone who is vigorously compiling, looking and controlling the values that can be searched in the list. Who is the authority all depends on who is responsible for the list. The Library of Congress is a good example of an authority list at an international level, however at the local level there are really no standards. As Imma Subirats explained, an authority list is almost liked a forced thesaurus.

In the FAO File Description Concept Scheme, there are URI concepts and then many avenue level cities ISSNs. The URIs are linked, since systems must be linked to other open datasets to produce linked data. Links are based according to countries through specific journals and so forth. As a result, this creates linked data. For example, a FAO incubator linked their linked datasets from the FAO File Description Scheme to an RDF model. Based on FRBR[4], linked data was produced and consumed through this specific software incubator model.

Imma. Subirats also described a few tools made available by FAO. First there is the AGRIS search engine, a database service provider maintained by FAO which includes 2.6 million records from all over the world. It is free and accessible online. AGRIS can link with indexed material in Eurovoc and literally allows the linking of one site to another across FAO products.

---

[4]

     Functional Requirements for Bibliographic Records (FRBR) is a International Federation Library Association (IFLA) standard which is difficult to use outside of library catalogues. Other participants mentioned the current move from Resource Description Access (RDA) (http://www.rdatoolkit.org/) to RDF as the underlying technology in linked data.

There is also AgroTagger, a concept identification for under-structured text. It uses Agrovoc as a controlled vocabulary. It is still a prototype system that has excellent results and in practice, will produce RDF files in the future to link data similar to the way that Open Calais does. Text is pasted into the Agrotagger website and then generates agrotags, which are found in Agrovoc. This system is a bit less advanced than Open Calais. Although Open Calais is a good web service, its vocabulary is not very specific to the agriculture sector so AgroTagger often obtains better results. Ideally, FAO would have a real platform for Agrotagger, like Open Calais, so that terms can be retrieved and integrated, and so that linked data obtained from a specific platform can be maintained.

FAO also uses the Drupal platform, a content manager system, which easily manages all types of information resources. Drupal has a strong user community behind it. FOA finds version 7 the most interesting because it is based on RDF. But a semantic version is already possible with v.6. Imma Subirats' group has provided a layer in v.6 where institutions can produce rich metadata from their systems to also customise the way that they build the vocabulary behind the RDF exhibition. With AgriDrupal, FAO has provided the means to produce rich metadata through RDF but also through any data in XML.

There is also an ongoing project currently taking place in Bangladesh. The FAO had received a request from the Bangladeshi government to help find software for their website and their library. After their analysis, FAO realised that the government didn't have sufficient resources to maintain two different types of software. FAO therefore proposed to customise Drupal to allow them to upload full text documents and also describe records in a very rich format, almost the same as in a library catalogue. As a result, the Bangladeshi government now have their own website and their own document repository using the same software. They have 1,000 documents in their repository, which is currently the biggest in Bangladesh. By using Drupal, the Bangladeshi government is already on its way to producing linked data, without having to add any extra tools or going through any additional processes. In this case, the production of linked data occurred almost by default.

2.4 General Discussion and Feedback on the Presentations

These examples of the use of linked data in development have raised a few key questions:

- Do we really understand how the logical complexity of linked data actually helps program interventions on the ground?

- Is there a way to compare quantitative data found in linked statistical datasets with qualitative data?

One participant underlined that micro datasets could be used only to validate macro-economic assumptions that go against the action culture and sociological realities of families. Perhaps this would erase some of the realities related to

drought, unemployment and other factors due to definition issues. If this were to happen, we would still create alternative ways of seeing the world, which also create new biases rather than accepting what development should be, which is an affective-emotional reality as much as it is a macroeconomic one.

Others acknowledged that this was an important issue to highlight, because quantitative data presents a broad picture, but does not explain the dynamics of a specific situation. Young Lives, for example, has qualitative data, which could be coupled with their quantitative statistics, yet this data only represents a subset of children. To gather information on all children would be a big task. Consequently, looking at the statistics that are generated and then doing in depth work is a real challenge.

It was pointed out that advocacy and action data were also not the same thing. As one participant remarked, *"In today's world of social media, numbers can help attract attention to the cause or issues"*. Citing Mao, he observed that '*sometimes quantity has its own quality*'.

Others commented on the possible risks of data distortion when using these kinds of demonstrators and tools with linked data. Someone asked if this kind of distortion comes from the use of linked data or if the 'right' users can always produce the right results? Is there then a strong limitation to what can be done with linked data? If using linked data as an advocacy tool is ok, it should always be remembered that these types of data and information structures and resources can be used for other purposes. Are these the limits on the truth which theoretically can be thrown up by such information structures?

Some argue that metadata can be extracted around stories but that a story cannot be transformed into data and then changed back again into a story. This is partially the situation with the *web of data* and the *web of stories*. How is data being annotated with meaningful human narratives and stories, and not just as properties for machines readability? These are interesting questions, which are not just about getting data, but about reaching marginalised people. In perfecting machine readability, are those people treating what they receive from the data processed as representing the truth? Or are they looking for the narratives found in qualitative data as well?

Overall, major issues that were identified with linked data from this session were as followed:

- Intuitive tools need to be created to let people link data themselves, while they are engaging with linked data.

- It is important to be able to go back to the original source or dataset.

- It is difficult to marry policy and technical issues.

- We need to further examine how new data is created from linked data.

- First movers will have the advantage of being able to shape linked data and *the web of data.*

And these were the main questions raised from this session:

- How can data be made accessible to different audiences?

- How can visualisation tools be used without distorting information?

- How can other peoples' data and mash-ups be found, without consuming too much time?

- How can repeated URIs be avoided?

- What are the issues behind the complexities of pairing datasets?

- What are some of the ethical questions in making micro data available?

- How can survey designs facilitate the linking of data?

- What are the goals and objectives in linking data, from a policy perspective? How can this be used to create policy?

- What are some of the processes that can contribute to data literacy and participation?

- What will the future look like? Will URIs be more searchable and be ranked according to search?

- Authorities: who describes the world, control and maintain vocabularies?
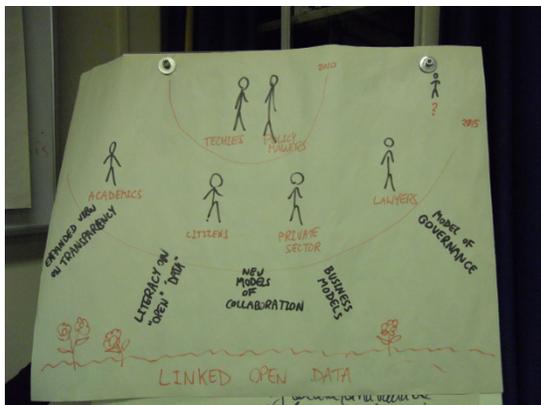
2.5) Discussion and Reflection on the issues: A policy and technical perspective

After this session, participants were divided in two groups to identify and discuss foreseeable policy and technical issues involved with linked data. Here is a summary of the principal points derived from these conversations.

Policy-Oriented Group

Participants shared insight and opinions based on previous experience with policy and development, which included work with [The International Aid Transparency Initiative (IATI)](). They found that technical issues often bogged down the discussion. Technical and policy issues were often split to avoid tension. The issue was not whether if people would meet the standards established, but if people would implement them. One solution identified was to publish data to help improve its quality. This was a decision taken by IATI who took the view that once the data was available online, the politics would eventually work itself out.

The group discussed the benefits of creating data to help put order to the system, and use data in various forms. They felt that linked data would branch out into geocoding and suggested that in five years time, everyone may geocode their data as linked data becomes more widespread. The group also discussed the possibility of using geo-names to identify data. Currently, geo-name IDs exist that open up inter-nation differences between distinct areas where unique identifiers and currency values exist. Members of the group thought that there should be standards within aid data. It was argued that the links between different countries' aid data should be just as transparent as budget transparency to allow for the alignment of aid budgets.

Some also wondered if there would be enough data made available for linking in a couple of years. It was claimed that it makes it difficult, from a policy perspective, to invest in something unpredictable like the linked data scene, since it may mean having to take a *'leap of faith'*.

The use of social media was also discussed. This was perceived as a tool that enables someone to start a conversation online around topics or subjects of interest. The role of Google was also broached. Questions were raised on how the gargantuan presence and impact Google has in online information search will affect linked data efforts. Will Google have a positive or negative impact on linked data for development?

Policy makers also need to begin looking at bottom-up interpretations of data. The group thought about ways that programmes could be designed to make knowledge more accessible to stakeholders that need to access data. How can linked data be accessible to people at the bottom of the pyramid?

From the group's perspective, the push for linked open data is a given. However, questions like: what linked open data should look like? How it is going to be funded? What kind of workshops does this necessitate? among others were raised. Nonetheless if, linked information is somewhat inevitable there is a need to contextualise, increase and broaden datasets.

There may also be a lot of pitfalls, which have the potential to widen inequalities. This also relates to questions about access, or what information sources are being used and valued within the development sector. Linked data is becoming increasingly sophisticated as new information is being created. Datasets are

linked and become part of the information architecture. However, one should take care in making assumptions about the emerging architecture must find out who and where are the drivers? *"Stuvelliance"* as it relates to linked data and geolocation technology was also discussed and it was reiterated that the North should not be the only pioneer in the linked open data movement as more diverse stakeholders are needed.

The synthesis of data across different studies, was emphasised, but so was the fact that open data is not a given. Though policy makers may want to move forward within a linked data environment, there are costs and limitations to this. Issues of ownership, policy processes and implementation also arose. For instance, technology solutions for Africa are being delivered in African countries, but do African countries have any sort of ownership for these solutions?

Furthermore, as linked data develops there may be a massive assumption that it is naturally a good thing because *'some empowered geeks'* believe they are helping the world. There is a need to involve communities and investigate what they can and cannot do with linked data.

Tech-oriented Group

Members of this group first tried identifying the purpose of information gain in linking data for development and how this purpose could be communicated. They broached various questions such as: how are people are using linked data, and how can a model be developed in order to capture the nuances of the semantic web?

Everyone acknowledged that sometimes policy doesn't match with technology and vice versa. This is why it was underlined that machine data readability is not always useful for human data readability. It is important to understand the technical difficulties. Linked data was first meant for computers to find comparable data. Now there are talks about data being human readable, with the use of a friendly user interface and interesting features or tools. Currently, machines are retrieving data from multiple sources. What is there to help us find and use linked data, both at the machine and human readability levels?

Furthermore, questions were raised on who decides what data to combine, how to link it and for what purpose. Should it be a machine process or are human technicians needed to decide this? Also, is there a clear distinction between linked data (in statistical format) and web data? Wikipedia for example, is a semantic linkage of data where there's a blurring of information and web data.

Other issues identified, were the complexities of the system. Data grows and evolves as datasets are used and new ones are created from linked data. Some wondered how to broach questions of trust. On a positive note, this could allow for the discovery of new data, by pulling linked data sources. However, there is a difference between storing knowledge (Meta data) and creating knowledge.

Storing knowledge is easy, but pulling, processing and analysing linked data, is not always a straightforward process.

It is also not yet clear how data is being integrated in repositories. Most systems still don't read RDF, making integration difficult. Maintenance of data is also a handicap of linked data. Human beings produce Meta data. It's not clear how you can maintain or keep track of the changes made in these types of datasets. No convention exists yet on how to maintain or update linked data within the created links.

Summary of Points that Need to be Taken Forward from Both Groups

The tech-based group also highlighted the need for commercial drivers and search engine infrastructure for linked data. Some predicted that in the future linked open data would be easier to integrate in search engines such as Google. Structured linked open data and description of data would be made easily available. Search engines would crawl the Internet for data as well, so the question of timing would be everything, in terms of appearing first in search results and optimising websites for linked data. At some point in the future, information managers will only be content providers and an emerging platform will take care of the technical technicalities.

It will be important to find a balance between policy and technological pragmatism. There still needs to be a line that defines where and how technology will meet policy. Questions about how data is chosen and who manages the process and the data are pressing issues. We should begin reflecting on how to make linked open data available from the bottom up rather than the top down. As focuses shift, it will be important to examine the limitations on the focuses now and what this will mean for the future.

We may begin to see a reversal in drivers for linked open data. In spite of the fact that it was once a guerrilla thing to do, governments (e.g. UK government) and big institutions (e.g. World Bank) have now begun to take the lead. Does this mean that open data be mandated from the top down?

People still agreed that the development sector is ahead of the game and that there is room to manoeuvre and drive things forward before it was still late. First movers will take the lead and become a strong node in the linked data cloud. The development community should maximise this opportunity.

3) Creating a linked information environment - Where linked data is headed and where we'd like it to go

Three differently focused presentations were given, which examined linked data as it currently exists and hopes for its future. The perspective of an international organisation, an organisation working on a particular issue (aid effectiveness) and the Southern perspective (an Indian civil society organisation) provided interesting and varied insight and views on opportunities, challenges and the way forward.

3.1) Three Perspectives on the direction of Linked Data:

3.1.1. International: Ideal linked information in an international organisation (Chris Addison)

Chris Addison gave a presentation on linked data infrastructure for international organisations, giving his perspective as a result of the work he has done with the [International Policy Food Research Institute](). He spoke about the change that the web is undergoing, citing an article in Wired Magazine that argued that ['The Web Is Dead']()[5]. The article said that some think that websites are becoming increasingly irrelevant because the top 10 websites (which include Facebook and Google) receive 80% of web traffic. As a result, we need to focus on Landing pages within a website that are featured in the main sites such as Twitter Facebook etc..

We need to provide a landing page with for example a report in context with related materials linked on the same page. Consequently, all material and reports need to be designed with the linked infrastructure in mind. Most people go to websites by searching on Google. Mr. Addison argued that everyone would still be using Google in years to come as the search engine will have adapted to accommodate this linked data scenario. He stated his view of the inevitability of a linked data future, although admitting that this might take some time as only now are Meta tags currently found in web pagesused although they have been around since the beginning of the web.

Why does this web base of data differ from the web base of full text and search?

When a Google search is done the results are based on the bias of Google's algorisms and that if everyone started putting in their own linkages, this may or may not change the order and the way information is presented. He also raised the point that if there are only northern organisations investing in the linked data infrastructure, this will reinforce the northern view of the world. This leads to the question: Is linked data better or worst than Google?

Issue of authority

Chris Addison cited the issue of authority as key in the development of linked data, pointing to questions that need to be asked, such as who determines the composition of lists (e.g. country definitions). This is not clear-cut yet though there must be reasons behind choices (for example should geo-names be used

---

5   Anderson Chris and Wolff Michael, The Web Is Dead. Long Live the Internet, August 17, 2010, 9:00 am, [Wired September 2010](), [http://www.wired.com/magazine/2010/08/ff_webrip/all/1]() **(Last accessed 29th November 2010)**

or not?) These authority lists determine control so organisations need to think about how they can be influence this.

Not only the data but the linkage

There is still a need for more diverse views and we must consider not only the data but the linkage. Some description of dialogue is necessary with a narrative to enable it to appear in search results. For example, if searching for a [UK Department for International Development's (DFID)](#) project on Google in the new semantic web, in the right hand margin there could be a video published by someone else because they have used an RDF in their description of the video. YouTube would probably automatically do this as well if there were some reference point relating to the project. It is important to be aware of how these things potentially work and we have to be clear about what is wanted from the opportunity. For instance, if one were trying to understand the impact of research funding in the UK, would this require a mash-up of different datasets with a longitudinal analysis of investments and money spent in research centres? The fact that separate datasets can be linked together, geo-referenced and provide different opportunities for visualisations seems to define ideas behind linked data.

Validating linked data

He also argued that people are wanting access to the data behind policy reports and that making this data more readily available is important as it allows you to validate the report more clearly and easily. Providing visualisations of data and the means to make comments on the data have been recently developed for the Food Secyurity Case Maps system to illustrate projected scenarios for agriculture in 2050.

Linked data approaches enable someone to browse a range of related datasets for example by country and carry out comparisons. This may be a way of bringing more awareness at the national level about the existence of data that many are still unaware of.

Levels of data entry

Chris. Addison highlighted the importance of using linked data for document discovery (to explore documents and to find related documents as well as authors, research groups and expertise, for example). He also spoke of the use of an atlas concept, which looks not only at national indicators in a world map but also at data within a country and aggregates different datasets. This could provide a final calculation of relationships between datasets that was never available before. In the same way that the introduction of HTML suddenly meant the possibility to put links to other pages in other websites, the use of linked data allows the combination of information to make new web pages.

He pointed out his interest in learning more about ways in which data from the International Policy Food Research Institute can be presented in a way that allows for interaction with the rest of the community, comprised of different

audiences. There is a need to give some guidance to technical groups so they can make more use of the data, as well as to the policy groups so they can better engage with media and intermediaries.

Producing RDF

Chris. Addison also spoke of the posibility for his own organisation's extensive databases of staff and experts, projects, publications, presentations, events and news to be produced as RDF.. With the possibility of native RDF in the Drupal 7 CMS, and a plugin for 6 it should be possible to start producing RDF.

The Digital Object Identifier

If we look to the future of the web the DOI – the digital Object Identifier becomes more and more important. This unique identifier for a report or any digital output points to a page descriing that object and ordinarily linking to related materials. At present IFPRI uses the web page for this and this is why linked data will be so important in the future.
At present we build the page from everything indexed in our Content management system, but in the future as more of our material is on other sites we will need to use RDF to combine elements of the page and link to related materials.

How does this change the way we capture data? The presentation ended with a number of questions:

- Can different things be captured if we look at these authorities?

Does this need to be processed differently?

Is there a need to think more clearly about capturing references in a different way for these authority lists?

Discussion following presentation

It was said that for a lot of this linked data work there is a need to describe the world, look at who describes it and how to describe it. The question was therefore put forward: Should there be a readymade product that enables this or should it be that every time something is linked it should describe it?

Authority was said to be a big issue as it is difficult to get an authority list of organisation and it is also very expensive to run these authority lists in totality. Chris. Addison stated that the International Policy Food Research Institute used their own keyword list for topic pages with the Agrovoc thesaurus as the core for most indexing,. It was therefore pointed out in discussion that those involved in this work would most likely be responsible for some authority lists, however hope was expressed for some of them being built by/as communities. Authority lists will be important because though it may be set by one organisation, the less a

community is involved in developing it the more likely an individual will not be able to access it.

It was also mentioned that shared identifiers are many times created because of built social relationships, through conferences etc. As a result there is a need to examine how to bridge different communities of social relationships.

3.1.2 National: Vision of a linked data world in Nepal (Bill Anderson)

Bill. Anderson spoke about the work he has done with AidInfo and the International Aid Transparency Initiative (IATI) conducting a number of pilots in various countries. One of the key aims of IATI is to deliver better information to recipients and country governments so they can better plan their own development with more forward looking and timely information on conditions. Over the last year they have done work in several countries looking at what the aid data for these countries look like by IATIs standard. They have also attempted to expand public information on aid and in particular to prove the concept that these countries can be imported into recipient countries aid management systems. He pointed out that phase one of IATI's work looks at the relationship between donors and the first port of call in aid delivery, however it does not tell much about what is actually happening in terms of aid or development. However, AidInfo is trying to take this a step further by asking: What would the impact of better information on aid mean at a local level within recipient countries? They have therefore looking at developing two pilot projects, one in Ghana and the other in Nepal. They will take all the available aid data for these countries and geocode it. However, this will need to be done locally and manually because there is no geographic information currently available in these countries.

Nepalese Project

AidInfo will examine ways to interface with civil society and add some value to that information by looking at a Nepalese province. They are investigating the possible use of several types of technology starting from a web-based system and will therefore explore the use of smart phone applications particularly using photography, GPS and phone cards that will be then link directly onto the aid database. They are also looking at the use of basic SMS technology and are trying to envision ways to gather more local and important information. Bill. Anderson stated that *'This whole idea that aid exists as a thing in itself is not real… the distinction between aid, foreign agencies and government does not exist"* Despite this being the case he noted that aid is not reflected in the national budget. As soon as one starts talking about aid at a local level the delivery of aid services is not mentioned and there is no distinction between the issue of governance locally and internationally. It is hoped that the project will start in the next 6-9 months, and it can be viewed as exciting as it may create a 'hornets' nest' as people are encouraged to start demanding their own agenda.

3.1.3 Local: Possibilities and pitfalls: a view from Indian civil society (Parminder Singh)

Parminder. Singh looked at information behaviour in small communities, which are generally the eventual target of development work. He examined the relationship between information behaviours and open linked data information structures.

He pointed out that even if a very distributive information structure exists and all voices can be heard, there is still a very real need for people to actually provide information. As a result, one needs to understand how they relate to information, why they would contribute information and whose voices should be heard in the distributed information structure. Also can openness have an adverse effect? For example, only certain people may provide funds for certain groups and development organisations, which may present a single view that is assumed to be right for lack of diversity in the linked data community. Consequently, openness could actually make things worse.

It was also highlighted that today there is a lot of talk is about the information needs of powerful development agencies, which control development aid. In general, it is agreed that it is good that the right type of information informs their decisions. However, the information needs of lower level development workers and communities should be addressed to find out what kind of distributive linked data infrastructures can be built around these needs. The same kind of technical infrastructure may be able to serve both levels, but this can only be determined once information behaviour is properly observed. Community level needs may actual differ. Analysis of this can also provide much needed insight on the general linked open data structures that are now developing.

It is often taken for granted that when information is needed, it is available. Communities and processes are being built around these ideas. However, while it is thought that information can be made easily available by going to Google, for example, some villagers do not actually think that the information they need is readily available. Therefore, information behaviour of these communities may be different than organisational needs.

Parminder. Singh spoke about his organisation's work in setting up telecentres in villages in India, and their decision to promote the idea of the telecentre as the place to go to for any information. This started a new type of information behaviour, which has many layers as it takes a lot of effort to develop that type of consciousness. Parminder. Singh expressed the need to look at the process of making information available (e.g. is information automatically web-based?). He advocated the idea of a new techno-social system for information behaviour being built around the communities, which is relevant for developing linked open data infrastructures and information ecologies for the people who work amongst these groups. Only after many months of listening to the community radio did individuals started to feel that they were and could be the source of information. It

can take a long time to build this type of information behaviour and that this must be considered when setting up linked open data systems.

Issues to highlight

Parminder. Singh argued that if you have a very sophisticated technical infrastructure then a development worker in healthcare for example could directly assess information. As a result, there is a need for information workers who are specialised in information behaviour, that can help set up such processes.

When working on large scale information systems, there is a tendency to disembody information from real people's processes so that it looks like a separate layer of information, which may not relate to information at the community level. Therefore linking open infrastructures of knowledge to the interaction models that people use is also important. Parminder Singh argued there is a lot of collective information production and consumption, which does not fit into the typical information behaviour model.
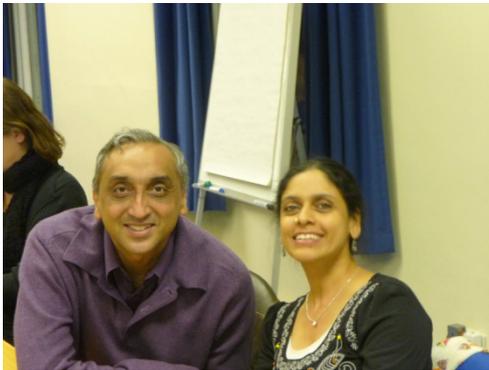
He also reminded the group that as everyone tries to influence global and national policy, there is a need to keep sight of the policies that influence technology and how they enable or constrain linked data work. In some developing countries, using the mobile Internet is highly compromised. For example in India, the Internet could be limited to Facebook, Google and a few other websites being accessed through a mobile phone. Access to the rest of the Internet is quite expensive for the average individual. This influences the kind of data architecture produced as well as the information and information services be used by people.

This links to the principle of net neutrality. Many telecommunication companies provide certain Internet services and applications, which are free or separately priced from regular Internet access. In India, an individual can get eight services (e.g. yahoo etc.) for a dollar and nothing else. Since Facebook and a few other applications are provided free of charge by the biggest telecom provider in India, this has affected people's information behaviour. It is estimated that within five years' time, the mobile phone will be the most important way to access the Internet in India.

3.1.4 Questions – Comments

Is the web dead?

Less and less of what is done online is on the web. The browser is becoming an out-dated way to access the Internet. The use of applications (e.g. on Apple's iPhone) are increasing. Any application utilising this linked data model may not be pre-approved if it does not use a commercial model. In the future there may be a need for data and not websites. However, if linked data come from certain applications on the Internet, people will be unable to access links that are not approved through controlled applications.

The open licensing of data

If Facebook is providing an application which can only be used within Facebook, Facebook will pay for its delivery which may be considered a good thing. However, linked data may then have to be licensed in a commercial way for us to use it. This idea of a commercial platform is not something that development organisations usually want to explore as they do not want to provide commercial reuse of data.

It is also possible that Facebook in India for example, may prefer one kind of data system over the other because it is more remunerative, while development organisations want it to be provided for free. In that case it may be difficult to ask them to reconsider.

An article by Tim Berners Lee titled 'Socially aware cloud storage' was referenced. This relates to each person's control of his or her data acquisition and speaks about the separation of storage from the applications that access the storage to bring about more horizontal use of the data. (This means the linking of personal data to everywhere else).

3.2) Visions of Linked Data for Development

What would the success of linked data look like in three years time? What will have been achieved and how will each of us be congratulating ourselves on accomplishments? What will things look at the organisational and individual level? Participants were divided into four groups to discuss future visions of linked data. These were the results of these discussions and conversations.

Group 1:

The first group envisioned the linked data for development movement as a broader and more open environment. Linked open data would empower people to use data but also to check information and verify the under assumptions of this data. For example, the use of open data by media organisations such as [The Guardian](#) has been adopted to validate stories.

However, the group foresaw a few possible constraints that could arise in the future. First, they believed that the control area of authorities would be contested and up for debate. The group wondered if it were possible to have a world governance model for linked data, admitting that it would be difficult as there are many stakeholders involved. This would ultimately remain an area of competition.

The second concern was that there would not be a lot of data made available to link. The development sector would have to decide what to invest in without being able to predict the outcomes. Questions about positive discrimination were also raised, in terms of deliberating making certain voices being heard. Everyone would supposedly have information because in principle anyone could obtain and use the data they want. However, this is a faux form of equality if proper efforts are not made to ensure that all voices are heard. This is why the group stresses that data alone is not sufficient. And it is important to distinguish the difference between data and information. There needs to be further exploration about building metadata with existing information.

Group 2:

The second group thought more stakeholders needed to be involved in the processes of creating, using and maintaining linked data. These should include not only technical experts and policymakers but also a broader role of actors such as academics and private sector organisations.

They also mentioned the importance of involving local communities and local lawyer networks to develop legislative framework. Their future vision saw a workable governance framework that managed the open data environment created along with an expanded view of what aid transparency means, where focus was less on inputs. This means effectiveness in a more meaningful but standardised way. They supported collaboration between stakeholders. The group also wished to see an increase in the democratisation process of linked

open data, and increase in public use. They also felt a need for more data literacy and the use of interesting stories and case studies to explain the usefulness of linked data for development.

New models of collaboration could include a peer-to-peer model of collaboration or online volunteer platforms. If development institutions opened up to new business models and services, this could stimulate innovative approaches to linking data. More models need to be developed and sought out for the local private sector in developing countries. To have effective participation, governments needs to be involved in the brainstorming of a regulatory framework for data.

One member spoke about a project from the Berkman Centre at Harvard University that created an ecosystem of virtual companies in the Boston area, using digital accountability mechanisms. The consultation of virtual communities proposed a governance framework. This highlighted how the North could potentially still be ahead of the South in addressing governance issues and other matters pertaining to linked open data. As the member of the group accentuated from this example:

'In five or six years time it is possible that we are successful with the open data initiative. A lot of rich data will be coming from around the world. The virtual Silicon Valley, can say "oh fantastic, we can actually use a lot of these datasets." Yet again the North will be several years ahead of everyone else, making use of the data that has opened up and been made more accessible. This is why it's important to involve more local stakeholders, not just in the technical aspect, but also in all aspects of the construction of the system's infrastructure at this time. When we've got all that, it will makes sense to plug into the linked data projects.'

Group 3:

According to the third group, the future would make RDF invisible to the average user. Linked open data would be equivalent to what open scholarly information is now. There would also be more user-friendly tools that make RDF omnipresent. These tools could be applications that specifically target development workers and information intermediaries. They would also help make use of linked data for contextualising information for intermediate-level users (so not necessarily end-users themselves).

Data sources would also be much more diversified. The creation of a social graph or social network for linked data would make the process more inclusive and participatory at all levels. Everyone would be able to link data, and authorities (organisations) would validate it. The social graph would allow for the annotation of different elements in the web, so that not only large officials would comment and shape data into information. Instead, stakeholders at all levels could take part in this process through social interaction around the data.

The social graph would be integrated at the web level rather than superimposing it to linked data. A thin web layer would allow annotation of data sets, and it would be structured semantically, linking to the linked web. Trust in data could increase, as people would rely on comments and annotations from their trusted contacts in their social network.

This group wishes to see the silos opening in the specifics of linked data. IKM and information specialists should be familiar with them in the future and know how to use the information it can produce.

Group 4:

The last group focused on linked data and public spaces. Members of the group used the analogy of the public park to describe the future of linked data for development. They stressed the importance for local municipalities to create the necessary spaces for the creation of linked data. Regulation would therefore be needed to enable the thriving of such types of public space. If this space were to be created, the group believed that things would happen on their own.

The need to begin with diverse forms of knowledge and then concentrate on the infrastructure linkage points was discussed. There are different types of data, such as governmental, administrative, organisational, etc. The linked web's architecture would be based on the creation of multiple data services. Local governments and public bodies should have the power to invest and create these public spaces for linked data.

The governance of data may be local or international. The group stated that data detection and dissemination were key and that sources of authority would still be influential in the trust factor of data sets. Many people have the capacity to contribute to data which affect everyday lives. The institutions that will validate this (the brokers and intermediaries in the linked data process) will become important.

The group also spoke about the inclusion of the people being studied in the process of developing linked open data. Bypassing the concern of the public

good could lead to homogenisation of sources and information. Gatekeeper accountability and the personalisation of information and data will ultimately shape the web. People should have the choice to use identifiers that are meaningful to them. Who is the linked data being produced for? How diverse will it be? How will this tie into people's worlds and realities? Essentially, everyone should have the availability to contribute and consume data and be part of the open data movement.

4) Introductory Reflections for Day 2

After a first day full discussion, people shared a few reflections on themes they wished to further brainstorm throughout the day. These included questions about local digital public architecture and models of governance for LD4D. Some commented on how to determine the black and white limitations in the development sector. What exactly constitutes development data and what does not? Participants thought it was important to know exactly what needed to be addressed. Aid agencies are often limited in what they can and cannot do. But what is currently happening in the linked data world is not necessarily limited to the development sector. Though IKM's main focus is to improve practice within the development sector, this does not mean that it is not open to other ideas and proposals to move forward on this front.

Subsequent to examining the policy and technical sides of linked data, proposals were made to prepare a document that listed the social and political issues related to the issue. Seeing that the technical side should be informed by policy, the document should not be open-ended ideals. Instead, some participants suggested that it should be something that is technically specific: What is the problem? What are the limitations? What are the real considerations for technical constraints? Such a document should therefore take into consideration real goals, and consequences, presenting something concrete for the future of LD4D.

Another participant highlighted the importance of further exploring how to obtain results in multiple languages when searching for linked information. The participant argued that local knowledge is intertwined with how a culture thinks. This is why it was necessary to have an overarching debate around linked data and development. As a flipside to the debate, the participant proposed that the notion of local digital public architecture be discussed. As the microcosm of the complex universal, in which we want to create linked and open data grows, there will be a need for specific principles to address a local digital public and architecture. The idea of open linked data should match the local egalitarian digital public architecture. Who will be responsible for this and how they will go about developing it, needs further exploration.

5) Powers and limitations of the Visualisation of Linked Data

5.1) Visualisation information: Some history and questions (Mike Powell)

Mike Powell spoke about Tom Longley's, who had hoped to come to talk about his work with Tactical Tech on visualising information for advocacy. Tom Langley had pointed out that when talking about visualisation, one should not talk just about software and graphical representation of data, as there are many other aspects of visualisation, such as using it for stories.

Mike Powell gave a demonstration of the Gapminder – software tool available, which showed income per person and GDP per capita from the1800s to present – as an excellent example of some of the new tools available for visualising information. However, he wartned that there is a lot more history to visualisation and a lot more work done on it than is reflected in existing software. Mike Powell spoke about his attendance at a conference held by the OECD and the International Institute for Information Design (IIID) on Data Designed for Decisions DD4D in Paris, on 18-20 June 2009. The conference reflected on many issues in the history of design and development during the last 50 – 60 years was explored at the conference and it bore very little resemblance to the sudden interest in software visualisation. Some of the principles between the two however should be the same.

If someone searches for literature on data visualisation, the journal articles they may come across are full of very technical information, which usually comes from California and China. A lot of it uses algorithms for generating a visual representation from data.

On the other hand a search of books, produces a very different lierature

Mike Powell cited Horn R.E[6] who talks about the integration of words, images and shapes as a Visual Language. Horn argues that formal language has its own grammar, structure and rules and that visualisation can also be considered a new language as it has all the characteristics of a language.

'Pictures and words have been combined together in documents since the invention of written language and especially in ancient Egyptian. But it is my claim that the full integration of words, images and shapes into a single unified communication unit is just now emerging as a distinct language' Horn R.E

Therefore it has to be well used for the visualisation is going to work.

6

Horn, R.E. Visual Language: global communication for the 21st century**,** Xplane Press 1998

There is also a body of work on visual thinking and cognitive processes, which can aid visualisation work. He cited the Technical Centre for Agricultural and Rural Cooperation's (CTA) work on participatory spatial planning tools, which revealed that many people think visually and spatially instead of by text and words and spoke about work being done in the Netherlands on the recent history of medical images which is being linked to arts and science. (See annotated bibliography)

The graphical presentation of information is important but so is the analysis of the many ways that it can be used. An example was given to emphasise the important of this. During a ceasefire in Liberia in the late 1990s papers were handed out giving visual information about the size of the army and the militias taking into account the centre of disarmament at that stage. When asked what the source of the data was found it was found that the question had never been raised and that everyone assumed the data was correct. Therefore this serves to point out that it is just as easy to lie or misrepresent in visualised form as in any other. This is a danger that must be acknowledged.

He also gave the example of Volker Hoffman who occasionally visits CTA to speak about his experience of visualisation in development communication. A picture of a cow was used to help him get across a very important agricultural development messages with Egyptian villagers[7]. However, there were a lot of misunderstandings about the drawing of the cow as it was said to be drawn with the wrong shaped head and that it was a German cow not an Egyptian cow. This led the Egyptians to wonder if they had to send money to help the German cow get better. This serves as a reminder that people from different cultures look at same thing differently.

There is a history of using visualised data for analysis. A graphical, visualised analysis of a map of the cholera outbreak in Soho, London by John Snow in the 1840s helped to identify an infected water pump as the cause of the outbreak. Through the use of visualisation the theory that cholera was a waterborn disease was developed.

7

Hoffmann V. 2000, Picture supported communication in Africa - Fundamentals, examples and recommendations for appropriate communication processes in rural development programmes in Sub-Saharan Africa, Weikersheim, Germany: Margraf Verlag

John Snow's Map (1854)

Mike. Powell also spoke about IKM's work in trying to build a topic map and graphical navigation system from the different themes and issues that they are looking at. IKM is trying to get away from hierarchical data links and make links through the parent, child, sibling etc. model. The intellectual basis on which this was done is still being developed.

An example of visualised linked data was illustrated using a spring graph. This was the Talis library open linked data engine, which allows for live search by linking. This showed how a network of people working together use the same data you would use in linked open data. With a little bit of programming the dataset was built statically and the links from a mock up private database was used. This was highlighted as useful for an information manager because he or she can put what they are interested in in the centre and have everything else radiate from that. Some underlying linked data would also be needed for interpretation.

Mike Powell commented that there is very little on cultural differences or on cross cultural understanding of information in all the literature on visual;isation. There is a need to examine whether visual thinking is the same for every human being or if it varies depending on the environment in which someone was brought up (e.g. is there just one visual language + French and English for example? is it the same for every human being on the planet? or are there almost as many visual languages as there are ordinary languages?). What are the cognitive processes behind visualisation and are the universal or social? These are really important questions, which the development sector needs to understand and address before they decide to visualise all their information because it has to communicate across cultures and social and professional boundaries.

He argued that intitively he anticipated there would be diffferences. Not least, it is arguable that the history of 20th Century European Art is based on learning other visual languages. One example provided was a picture of a statue from the [World Museum Liverpool](#) in the UK and some quotes displayed in the gallery. This delved into the foundations of modern art in Britain and looked at the culturally diverse ways of visualising information. Examples of quotes include:

'I have felt my strongest artistic emotions when suddenly confronted with the sublime beauty of sculptures executed by the anonymous artists of Africa. These works of a religious, passionate and rigorously logical art are the most powerful and beautiful things the human imagination has ever produced[8]' - Picasso

'I was astonished to see how they (African statues) were conceived from the point of view of sculptural language'[9] - Matisse

According to Mike Powell, it is also not just a casual case of what the colour scheme is, but rather a quite a fundamental communication and linguistic issue for if one is going to try and communicate visually in a multicultural sector, the way in which this should be done must be examined and it must be acknowledged that there is a lot to learn from the innovations that come elsewhere in the world. This is a rich but complicated field.

---

8 This quotation can be found in Samaltanos, Katia. 1984. Apollinaire: Catalyst for Primitivism, Picabia, and Duchamp. Michigan, Ann Arbor: UMI Research Press, Pp. 18.

9 This quotation is from an interview with Pierre Courthion in 1941 related to Jack Flam by Courthion in 1979. Flam, J. 1984. 'Matisse and the Fauves' in William Rubin (ed.) 1984. 'Primitivism' in 20th Century Art. New York: The Museum of Modern Art Pp. 217-218.)

5.2) Visualisation and Young Lives: Past, present and future (Caroline Knowles)

Caroline Knowles pointed out that [Young Lives](#) did very little visualisation till about one year ago. Seed funding from IKM really pushed them to think more about what they are doing and why. There have been challenges as a small research team in trying to do this. She pointed out that most people when thinking about visualisation think about Gapminder type data sets but ignore micro datasets. They however, have been working with a micro datasets of information on children that have been followed for 15 years. They thought about doing a Gapminder type visualisation but they could not because at that stage they only had 2 rounds of household survey. Though they now have 3 rounds of household surveys and longitudinal information is now available she pointed out that this work is very different from that which occurs in big well-resourced organisations.

Young Lives started doing the visualisations because of their commitment to open access and to making data publicly available from the UK archive. However, she said that even when data is made publicly available organisations do not usually promote it. Young Lives have archived data publicly and tried to get other people using the data e.g. using the Open University Research archived data. They also archived all of their information publically and make their publications more accessible in other databases.

She noted that Young Lives works closely with Save the Children as a policy partner and as a result have produced a big statistical database. However, though they want the data Save the Children cannot access it because they are not statisticians or economists. She said that one of the main reasons for visualising the data was to making small amounts of their data available for NGOs to use in their advocacy work and programme planning.

Complexity of datasets

Caroline Knowles spoke about the complexity of the data set, which is difficult to convey. (for example they are working in 4 different countries, with 2000 children and use many different types of questionnaires). Another complexity she pointed out was the issue of documentation of how the data can be used (See presentation). Due to the complexity of the data, Young Lives decided to create a virtual village to do some of the longitudinal work, which is what they planned on doing when they started. Someone could walk around the village and visit the health centre, schools and water pump to see some of the data about children lives and what it means for them. They used the quantitative data which supported some of the qualitative research and case studies from their daily lives.
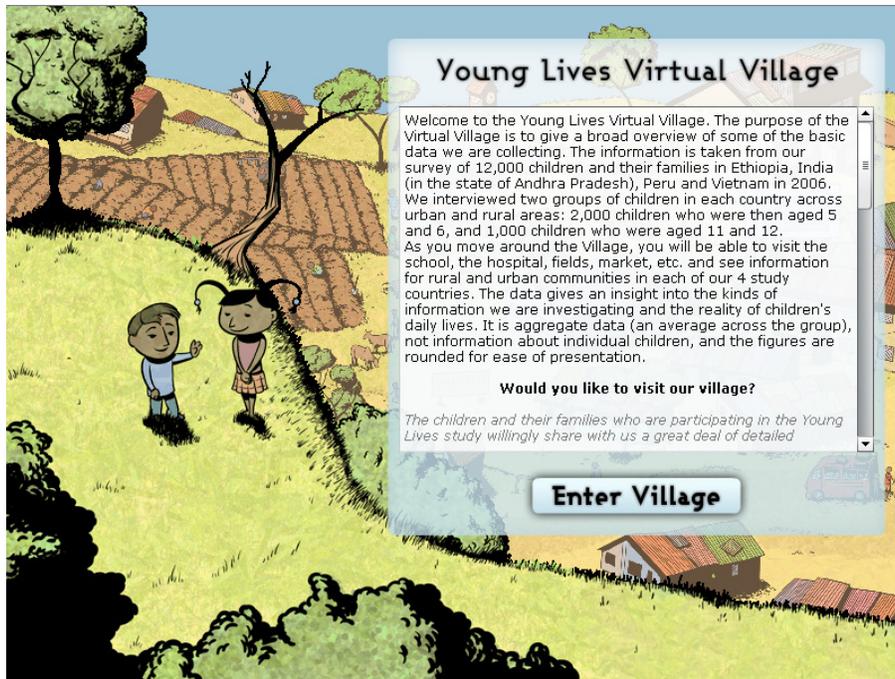
They also did a quantitative data mash-up with the ESDS data archive and produced a teaching dataset using Young Lives data. (See screenshot below). The demonstrator used interactive tables to allow for interrogation of the rest of the data and creation of your own tables. This has been underpinned with sources.

ESDS Teaching Dataset

Young Lives have also created graphs and charts as well as produced case studies of the children. They have been able to use their visualisation work for this and there are a wide range of people who are using it as well. For example, the Ethiopian government is using it for English language textbooks for children because the words come from the children themselves. Young Lives is planning to create more educational materials in the group and for the study countries.

Challenges

- Young Lives need to be clear about what definitions are and particularly how the data can be used. The Young Lives data is not a national dataset so it is not nationally representative. The organisation has to make clear what the data is presenting. This is a micro dataset and it is only for people who want to look at a very specific set of questions.

- It is challenging to present quantitative and qualitative data together.

- Ethics – The privacy of families and children must be kept. This is a long term commitment to the children and their families who place a lot of trust in those who are working on the project.

- There is a need to look at how variables are selected, as there are still messages they want to convey. A huge amount of work happens in the background before the visualisation starts. There are choices being made regarding what data is presented, how it is presented and why.

- They would like to do visualisation work with the children, which will include their comments on their research and findings.

- The data is complex and it is still early days technically. It also takes a huge amount of resources.

5.3) General discussion: What are the powers and limitations of transforming linked data with visualisation tools?

The world of graphs, pie charts and presentation of data is not anything new as this goes back to macro analysis. However, the possibility of visible language and overlapping visual cultures needs to be examined. Nancy Fraser was given as an example of someone who has done research on this.

It was argued that everyone lives in multiple knowledge societies with different codes of entry, which is a little bit like a club membership. An example was given where a famous artist in India was denied into a certain club because he was not wearing appropriate footwear as it is still the Victorian practice to allow people in only if you have shoes on. Moreover, it was said that many of the women and girls that IT for change in India works with are illiterate and at best may be able to generate an audio clip with training. On the other hand, it was also argued that they may not be able to build the capacity and the skill to do this because the

survival issues of their lives consume them so much that finding the time to participate within the public sphere is difficult.

However, it was noted that the possibility of one of these girl's video clip making it to the ivory tower club where international development decision are created was extremely slim and it may simply be seen as the woman's own voice, her own story, because she may never be able to generate mashups. As a result the importance of intermediaries was highlighted.

The need for platforms and methods of validation of linked data was also emphasised as important. For example: as more emphasis is placed on validation will some kind of democratic open space be created, which becomes the default for development and generates visual data of all kinds which are somehow more amenable to validation. Visual languages for specific counties and intermediaries must be built to help with validation.

However, the devil's advocate view on linked data and visualisation operating in offline situations as well was raised. For instance UK government data was printed on cards with pictures so people could discuss an issue based on data. Therefore linked data can be turned into conversational products and tools based on shared facts or shared identifiers, which can be taken back into the policymaking circles.

One example of the importance of the use of visualisations given was during the Haiti earthquake. Then visualisation were of more use in helping aid workers to save lives than maps, which did not make as much sense due to the devastation of the country.

It was also pointed out that there are creative ways of making visualisations available to the community. For instance applications that use voice recognition can be integrated within simple databases queries. It was also noted that text and voice messaging systems can provide a sort of question and answer type system that can help sort data, which could then be visualised in other ways as well as provide a feedback mechanism for people,

However a missing link within the conversation on visualisation was identified, which was: How will it benefit people, does it address their needs and what will be the outcome of that for someone in a community?

It was acknowledged that an organisation's message is packaged with their data. *'in fact open data is packaged data'*. Also the native format of the data may not be textual but a video, which is not easily to be linked to other things. Therefore the question was asked: How does video come into this world of text based databases and how should one convey the message to people who are not used to this?

What are we trying to talk about?

It was recognised that there is a general trend towards greater visualisation in communications and life in general. That includes everything from what people are doing with their mobile phones to digital story telling. However, it was also what happens within the ivory towers and development management issues must be examined. Though this can be very trendy area there are editorial decisions involved as *"we listen to voices we want to listen to and look at the pictures we want to see"*.

It was also pointed out that the growth in the amount of data available and the need to process this is leading to more visualising of data to try to find emerging patterns. However the issue of 'Data literacy' must be examined to ensure that that when data is presented it is not always as fact but as entry point into stories.

It was also noted out that in India the codification of modern Indian philosophy happened because of British occupation in India. German scholars like Max Müller and others came to India and interpreted Hinduism for Indians using the German and English language. When the school system was established in India all the Indian social reformers read the English text about Hinduism that was codified by the colonisers at school. Therefore it was argued that the lack of capacity to do this of type of linked data work (*"all of these skills may never be available to public institutions in a country like India, even in a more prosperous NGOs like IT for Change for several years to come")* could be detrimental for:

"Without the development of that local public and investment in the sub global, sub local and the national public and its richness, the visual cultures that are generated for development management will be codified through the eyes of the geeks of the northern world which though emphatic and sensitive, to the development needs of the south will alienate many other cultures. The development of these visual cultures, which influence the way in which people see their own world, is going to be paradigmatically different from the alienation that happens throughout the word."

It was noted that quite a lot of organisations are being funded by international agencies to make digital stories. However, none of these agencies appear to look at these stories for funding and it seems to be more of a cultural contribution, which is subsequently ignored.

It was also pointed out that the issue may not be a lack of visual capacity in India but that there may simply be a different language created through the use of different processes. For example Kailash Baariya of Anandi[10], in the State of

---

10

IT for change and IKM Emergent, 2008, Digital Story Telling, Consultation held by IT for Change in partnership with IKM Emergent Research Programme, June 5th, 2008 , Pai Viceroy Hotel, Bangalore, Consultation Report http://wiki.ikmemergent.net/files/DST_Consultation_Report_-_2008.pdf (Last accessed 3rd December 2010) Pg. 11- 13

Gujurat in India went through a multi-layered process of making a film. They spent hours developing a script, performed the script to village audiences then they discussed with these audiences what they understood from the play. They then went through several iterations improving the script and the digital story was as a tool for provoking discussion and participation. It was therefore an incredibly drawn out iterative, well-researched and sophisticated process that was being described. That helps to provide support for the importance of discussing the process with those it is supposed to help.

However, it was also argued that the equality of process also does not exist. The example was given of how stories taken with digital cameras by young people in Sheffield made a big impact on UK cabinet because it was new and their MP was keen on supporting digital work and was also situated in parliament at the time. However, if a similar project were done now, it would most likely be ignored.

It was also pointed out that it is often take for granted that the opening up of data is going to happen automatically though we are still a long way off from this. For example Young Lives is one of the few funded programmes that are doing this. As a result it was maintained that if there is not a continuing to press for openness manipulation, visualisation and integration of data *'a huge body of the world's data and information will continue to stay behind closed doors, on laptops, in people's draws and in journals people can't afford to pay for'*.

It was also noted that there is a need to keep data local and closer to the points of validation as visualisation is preoccupied with presentation while linked data seeks meaning and validity. Ease of consumption and validity of data must be kept closer together in linked data work. Video may be more easily validated, while other visual data may have many layers of interpretation, and all of this is competing for our visual attention. This is part of the new visual infrastructure ecology.

## 6) Looking Forward: Issues to Explore and Address

Finally, after much exploration and discussion about prevalent issues which currently need to be addressed in linked data for development, three principle themes were chosen for further development and consideration:

- The Socialisation Surrounding Linked Data

- Lessons to Learn from the Open Source Movement

- The Local Digital Public Architecture

Participants teamed up according to their personal interests, into one of these three groups, then reunited again to present and discuss their ideas. These are the principle ideas that were shared amongst the groups.

Socialisation Surrounding Linked Data

On one hand, there is already a place where people can talk about linked data on the web. However, there is no way of predicting or knowing where these conversations are taking or will take place. Ideas should be explored on how to enhance social interaction around linked data. Therefore, assuming that conversations are taking place, how can they be linked to the web of linked data?
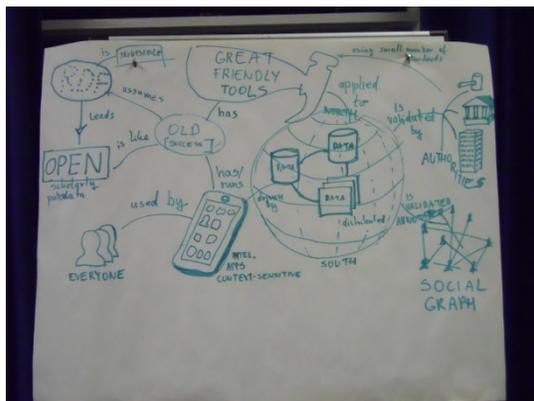
Participants suggested that geo-location be used to make it easier to discover conversations that were taking place around the datasets. Related discussions could be linked or integrated into RDFs. In order for this to occur, there needs to be Meta data on the web, so that platforms like FAO have an interface that embeds the discussions in the linked web. This would allow easy access to conversations surrounding the data in questions. References or URIs would be included for each dataset that is being discussed. Furthermore, it would be useful to combine existing vocabulary and ontologies that could be combined and published together in creating a social network of conversation around linked data.

At a technical level, this would require plugins that extend to existing platforms so that published data can be made searchable and conversations would appear in query results. Presently, references can be found when URIs are searched. In addition to this, the group suggested that conversations also be linked. This would prevent the content provider's interpretation and points of view monopolising the data, since conversations taking place would offer different perspectives and prevent information to be censored. The information produced from the data could therefore be less biased. This would not create any additional work or tasks for the author of the original dataset. Such a social graph could potentially help make data more comparable.

An example was given using the UK government and their open data to illustrate how this might work in practice. For instance, if the government hypothetically published plans for the Big Society as linked data, people could then post comments and link to its URI. If the government decided to prohibit comments being linked, alternatively a browser plugin or bookmark could exist that would be domain specific, allowing people to annotate and comment on data. On the other hand, another solution could be to create a button on the Guardian website that would link to ongoing conversations.

The group had many unanswered questions regarding how a social graph could be created around linked data. Questions such as, will such a network fit into the RDF architecture? Would it actually be adding any value to linked data? Is it helpful or necessary in addition to conversations dispersed on the Internet?

6.2) Lessons to Learn from the Open Source Movement

Two major lessons were identified from the open source movement. The first one pertained to the policy and licensing of products and the second was related to models of governance.

The open source movement has run on three main principles:

1) Meritocracy

2) A sense of community
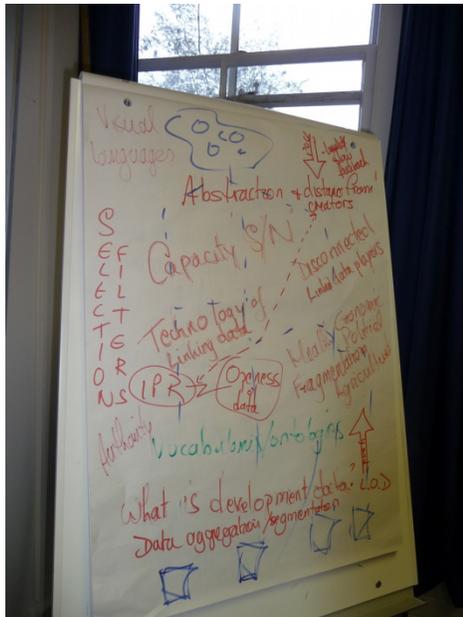
3)   Technology

There exists strong ethics surrounding the branding of data and technical contributions. Members of the community are encouraged to do this. This system works because people are recognised for their work and therefore are willing to do more. For linked data, copyright licences should be explicit about the reuse of material and products. Metadata usually refers to Creative Commons licenses (CC), when describing content. CC is thus already found in RDF that use metadata. Members of this group thought that some licenses should also allow commercial gain. In Africa for example, some governments want open data to be an economic stimulus. This could be a push forward for development.

The Debian Community was also suggested as an example for the development of the linked data community. In this community, licenses must be open sourced and there is a list which enumerates licenses that are accepted. The group

commented on the danger of being too prescriptive with the management of licenses. A lack of flexibility might close off access to unforeseeable actors. For example, someone who has a commercial interest could actually help move forward linked data. This was the case for Debian, where a commercial partner who built upon the community now helps to maintain its sustainability

The group suggested that a space with guidelines be offered and that people already engaged in the linked data scene be identified and engaged with to identify recommendations for best practices. Acknowledging that many tasks will not be done by IKM or by the participants present at the meeting, it was said that roles need to be determined.

On issues of governance, it was suggested that a similar model to the open source movement be adopted. There could be a core group that works on developing specific areas of linked data. Each area would have a manager who is responsible for overseeing the development of that area. If for some reason, there is no progress on that front, the community can decide whether to keep the work already done or not. Community, members would be able to vote and make decisions on the core principles of linked data.

This would be a self-governing model. Since it is a meritocracy, people would associate their names to the contributions they have made. This creates transparency and makes it easy to see who is participating in the discussions and processes. By creating a community similar to Debian, maintenance of relations, creation of mash-ups and standardisation to RDF format would be feasible without too much effort or burden. Furthermore, if the system is open enough, people could make money from it and then invest it back into the community.

Some however highlighted the fact that the commercialisation of open products is not always viable, citing examples such as Firefox, Apache and Wikipedia. This leads to questions about funding. The development sector has had lasting problems in collaboration and sharing files, due to the competitive environment of the non-profit sector. Often, a lot of public money is wasted because projects are being funded to do the same thing repeatedly. Could the linked data scene somehow reduce such waste of funds?

The architecture of linked data and how it is published were also discussed. Often, organisations just want to publish data as quickly as possible without taking the time to assure its quality or readability. At times, huge data sets are published but rendered useless for their lack of any consistency in format. Ckan is a suite of tools that helps create knowledge architectures. This could be used to help build the linked data architecture. Again, a list of good practices in publishing data was suggested to help improve data and format standardisation. This would propose where to publish data and how to find data.

Thus, the general idea behind this type of governance model is that instead of holding long meetings and discussing what should be done, people build and

create things first and then present them to the community, who then decides what to do next. The model is consequently based on merit, transparency and the inclusion of a whole community in the governance process.

6.3) The Local Information Ecology: developing a local digital public architecture
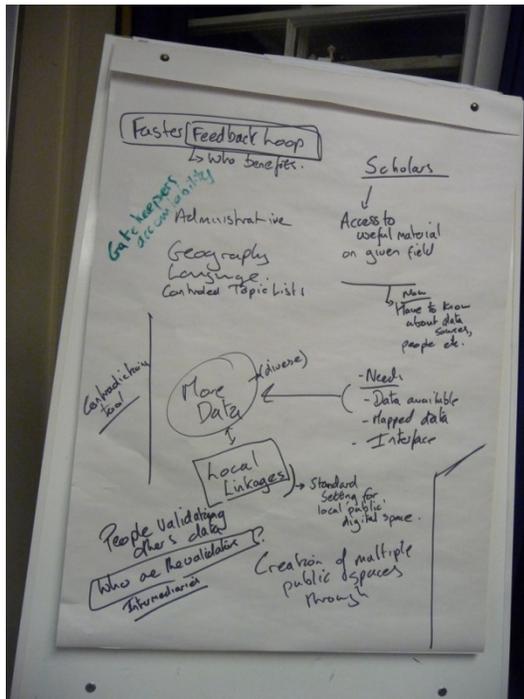
The idea behind the local digital public and architecture is to captivate the dynamics of local communities and include them in the creation and shaping of data. Taking these ideas forward, a first step would be to understand the types of communities that exist and how they can bring knowledge to the linked data sphere. Social media tools, portals and chat clients can enable better inclusion of a local public to linked data projects. The context of resources at the local level is needed to re-imagine development. If the artefacts and tools cannot be made available a locally relevant age, it is difficult to understand development in a meaningful way. For example, community radio could be done with linked infrastructure.

It is all about developing new information behaviour. Information is not relevant if it is not connected to pertinent activity. It is important to examine how people validate each other's knowledge. The perspective of an anthropologist can greatly differ from the local people being studied.

Technologists often tend to build architecture from the top down. Instead, this group proposes the building of linkages from the bottom-up. While working with a community, development workers could eliminate assumptions made about it. Standards should come from local, national and government datasets. Whatever enabling structure is built will inevitably constrain behaviours, somewhat favouring cultural assumptions and dominant powers. A bottom-up process could

at least capture real procedures and create the equivalence of a *village square* in the linked data scene. More efforts need to be made to capture the local systems and to understand how the digital ecology is serving the local informational communications architecture.

The localisation of information is important as people on the ground better relate and absorb information this way. A local architecture could also include a local search engine or local social networking sites.

The advantage of the linked data convention is that new properties can still be added. Tag sets used are important in categorising data and can be done for local purposes. Local identifiers can also help in labelling communities horizontally.

The challenge faced in creating a local digital public architecture is finding good versioning of data. Digitising information should be a policy issue, however policy should reflect what is happening on the ground already. Information should be mapped and examined according to local information behaviour. Case studies would facilitate the determination of infrastructure locations and workshops would help begin to democratise the process.

7) Closing Notes

7.1) Where does IKM stand in all this? What are the building blocks in which we will try to do more and share more? (Mike Powell)

Mike. Powell pointed out that IKM liked the idea of working in an emerging way so things sometimes happen in an unexpected way. They hope that if they bring some very knowledgeable and well-motivated people together something useful may emerge that can be of direct interest to IKM, or to other organisations. He pointed out that IKM is happy to hear of any sort of emergence, from a workshop like this even if it is not specifically IKM related. If some people get together in groups to do some work as a result of the workshop, that would also be a result.

IKM is interested in developing their programme for the remaining year as well as continuing to work funded or unfunded into the future. They have budgeted for some follow on work from this workshop for 2012. If there is something that relates to this that someone would like to work on next year, IKM can consider supporting it within the next 12 months. They would like to bridge knowledge between people and workshops in the south and develop a prototype of new tools, which can be piloted to take work to the next stage. If anyone at the workshop has suggestions for proposals they can inform IKM. Mike Powell also informed that IKM is putting together a proposal for the development of IKM 2.0 and that they would like to continue work on the development of information artefacts. Though there is no knowledge of whether IKM will secure additional funding in the future, this could be for both short term and more ambitious long-term projects.

IKM plans to write 2 substantial working papers to report on and develop the issues discussed during the workshop. This will be divided into developmental issues and technical issues. The report on this workshop would aim to develop questions not answers as well as an agenda for further research work. A dgroup has been created online to talk about the issues and people can post information there.

Common principles and community of practice

IKM will draft a brief description of what their vision of an open linked development world would look like. This would also include some common principles. A draft of this will be circulated.

This will involve building on the idea of creating a community of practice, which is defining itself by what it believes and what it is doing in practice. Some aspects of this are still to be agreed on and IKM is not aiming to develop their own standards but draw on existing standards. They will also look at the way in which existing development environment can be included.

Though there may be value in this initiative, the details are unclear as IKM is not sure if they should ask development organisations to sign up to a community of

practice and expertise or if it should be the same sort of community of enthusiastic individual (e.g. a developed floss model). However, Mike Powell indicated that it can possibly be a bit of both. This community of practice may also help to advocate for linked data and linked information to happen in the development sector. The community may be able to not only set standards but shape its development around certain lines. For example, IKM may wish to develop that community and host further meetings and this may be done in collaboration with other organisations. IKM can explore the possibility of hosting a bigger meeting, which can involve more people.

Other IKM work

IKM would also like to encourage further pilots to find out what is going on elsewhere in the development of this work. In this workshop, it has been harder to develop and demonstrate real examples linked data which clearly show additional value or insights being gained through the linkage. We have still been talking about this more as an idea rather than a practice. Mike Powell indicated surprise that there has not been a lot more to work on, especially in the development world. It was therefore suggested that IKM support work, which involves spending more time finding out what information exists in open format.

Mike. Powell also spoke about IKM provision of support for workshops being held in East Africa and Central America in the next 12 months. The possibly exists for this to be an agenda item that organisers of the workshop can think about. Support was also expressed for the idea of getting the communities involved in the Young Lives project to participate in building the digital villages.

7.2) General Feedback/ responses

It was noted that there is a need to do more research to find more linked data and that useful urls should be sent to the dgroup. It was indicated that this could be a useful crowdsourced activity because the Open Knowledge Foundation (OKN) have a development working group and they want to do some more mapping. It was pointed out that the Comprehensive Knowledge Archive Network (CKAN), which has been developed by OKN, only has a fraction of data that is available and that it would be useful to also gather multi-lingual data as well. Emphasis was placed on the need for CKAN to provide more information about the

vocabulary and datasets that they have worked on to have a more complete idea. Light annotation was put forward as a useful starting point which can lead to more information.

The view was expressed that a formal coalition is not needed at the moment because the direction that we are heading in is unknown. It can also take a long time to get organisations to sign up to anything formal as each organisation has different procedures for commiting to collaborations. However, it was agreed that it would be helpful to collectively elaborate a set of general principles and standards. If these were available, people would have a base to work together and could build an associated community of practice. People can also put such a document on their websites, as guidelines.

7.3) Next Steps

1. Circulate a description and principles within the next few weeks. Later on possibly try to create a community around it.

2. Raise awareness of the issues because they are also being followed by international organisations.

3. Link to datasets we are aware of on the IKM website. (If this scoping work and it mark up in RSS it would be useful for anyone else to use).

4. Find out more about the linked data work happening in the UK as well as well as about existing software that can be utilised.

5. Follow up with AidInfo as they are well connected in terms of communicating and promoting this. (Something can probably be posted on Owen's blog.)

6. There is some push in other countries to make information available and it would be useful to see how they are approaching the issues as well. It was noted that communities create databases and sometimes those databases are used by governments. (However, information though being made available is sometimes not available in easily standardised format (e.g. as a pdf document). Though this is openness it is not openness that lends itself to be linked so this should be further examined).

7. Be clearer about the words we use and ontologies. For example is open data just public data or linkable data. Hugo Besemer will write a short a message on this for the dgroup.

Possible projects for collaboration offered by Tim Davies

Open Data Hack Day in Oxford (& elsewhere)

A hack-day working with open 'public' data is being hosted on the 4th December 2010, and others are welcome to join either in person, or potentially remotely. Details are here: http://www.opendataday.org/wiki/Oxford,_UK (or see http://www.opendataday.org to see if there is anything happening in your city). Opportunities to work with new UN datasets and environmental datasets are being explored. The work done on the hack-day in Oxford will be informed by the discussions which have emerged from this IKM workshop.

Workshop ideas...

Interest was expressed in developing a workshop for grass-roots exploration of how open data can be useful in practical problem solving and opportunities for further conversations/collaboration around this are welcomed.

Book chapters: linking government data?

The group was asked if anyone was interested in collaborating on a proposal for a chapter for this call: http://3roundstones.com/2010/10/call-for-chapters-linking-government-data/. The proposed book is oriented towards government data; but the need to balance both technical and social implications in developing approaches to publishing development data can be articulated. The deadline for proposals is 30th November and the Chapter needs to be written by February 2012. If the Chapter was accepted a workshop may be held to help draft it.

The Open Knowledge Foundation Working Group on Open Data & CKAN

Members of this group were encouraged to join this OKN Working group. http://wiki.okfn.org/wg/development The group are also seeking to maintain the CKAN packages on International Development (http://ckan.net/group/international-development), which currently has 22 packages in of data.