IKM Discussion Note

Linked Open Information for Development: what it is and why it matters

The growth in the production of linked and open data will impact on business, politics and, more specifically, how we (both as organisations and as individuals) will use the World Wide Web. This growth has been stimulated by the commitment of both the US and UK governments to making all their data available as open data and, increasingly, in linked and open data formats and by the example of major media groups using linked open data to manage their mass of information across many media and formats and to link it with potential users[i].

The reasons people are interested in linked open data include
- A desire for greater transparency and accountability in and of governments, in the expectation of better governance
- A belief in the economic potential of easily available data as a basis for new tools and services
- The potential for scalable and standards based information management within (large) organisations or, more widely, across a sector
- The potential for new forms of interaction with stakeholders or with the wider public

Anyone interested in participating in detailed work on these issues is welcome to join an emerging community of practice at http://dgroups.org/groups/LinkInfo4Dev

What it is
Information can be open or linked or both.

Open   The openness of information has three parts: accessibility (can you access the information easily, generally over the Internet); license (are you allowed by the terms and conditions or license to access and use the information); format (is the information in an 'open' format so you can read and work with it without needing proprietary software).

Linked   Information can be published, internally or externally, using procedures which enable it to be identified, searched, questioned and connected to other relevant material[ii]

Information   For the moment, most of the interest and most of the examples of linked open information concern sets of data: series of facts recorded in a structured way. Qualitative information, such as that provided in articles or reports, can be connected to data by use of metadata, such as the forms of cataloguing long used in libraries. In future, it is likely that the computerisation of reading the meaning in texts, will allow quicker and more sophisticated connections to be made between diverse information sources and forms.

Linked Information and Development
In the development sector, linked and open information is already being produced by the US Agency for International Development (US AID) and the Department for International Development (DFID), UK, as part of the wider policies of their respective governments. The World Bank is making all its data, and the Organisation for Economic Cooperation and Development (OECD) much of its data available in open format, which means it is available for re-formatting as linked data by anyone who wants to do so. Making research data available in this form is likely to become a requirement of development research funding, much in the way as making other outputs freely accessible via public repositories is often required.

That linked and open information will have a major impact on development information and its use is increasingly beyond doubt. The real question is whether this impact will be positive. IKM Emergent argues that the result of investment by the development sector organisations in their own ICT systems over the last 20 years has been that the 'information rich' are now considerably richer in relation to the information poor: In other words this investment, seen as a whole, has been anti-developmental. We think that linked open information, especially if it is limited to handling raw data, has the same potential to priviledge certain sorts and sources of information over others; filter out the voices of the poor and marginalised and those

that live closest to them; and centralise both thought and power in ways which render development assistance less effective as well as less equitable.

In our view, information and knowledge management in the development sector, taking place in an environment of multiple disciplines, cultures and languages involving communication across barriers of class, education and status, is extremely challenging and requires constant thought, reflection and discussion. Despite claims of instant returns emanating from assorted techno-enthusiasts, this is just as much the case in relation to linked and open information as for any other aspect of development information and knowledge management. Early research has indicated both challenges and opportunities.

There is limited published work in this area. The Food and Agriculture Organisation (FAO) has been working, along with its partners, on various aspects of vocabularies (http://aims.fao.org) and the Indian Institute of Technology, Kanpur, on related knowledge models and repositories (http://agropedia.iitk.ac.in/). IKM has been conducting some pilots in this area and is also developing an annotated bibliography at (http://wiki.ikmemergent.net/index.php/Workspaces:1:Linked_Open_Data).

Some guiding principles

To achieve the potential linked open information offers and to avoid its dangers, IKM Emergent believes that the development sector needs to elaborate a set of principles to guide all its work in this area. Adherence to these principles could be a condition of development funding. These should be based on the recognition that the production of any linked open information contributes to the overall infrastructure or eco-system of development information. If this is built through a deliberate, collaborative and open process, such an infrastructure can be a global common good. But it needs to be built on an awareness of the differing agendas and resource bases of stakeholders; the multiple sites of the creation and use of development related information; and also of the dangers of misuse of information. If it is constructed as a haphazard offshoot of a mass of development organisations producing information for competitive advantage or to meet their short term goals, with no attention paid to developing connections and information linking infrastructures, it will be a chaotic mess.

In our view, any list of guiding principles should include commitments to:
- Link all relevant information, not just data
- Promote diversity of input and source in information produced and used (in particular to ensure that locally produced information and feedback is visible)
- Develop and use tools which identify the source of all information used and that help trace the re-use of that information (in the interests of transparency and to protect against misuse of information)
- Invest in creating linked open information environments relevant to actors at all levels (so that the new infrastructure can include and be informed by the whole of the development sector)
- Invest in ensuring that all stakeholders, particularly the poor and marginalised, whom the information concerns have access to and can make effective use of its content (so that development information empowers those whom development is supposed to benefit)
- Collaborate with other agencies in the development of thematic or geographic sections of the overall eco-system and infrastructure (to build a global common good)
- Agree and adhere to common standards for the licensing and technical structure of the linked open information produced (without common standards, the potential for creative re-use of material is lost)[iii]

Contact

To join the IKM mailing list: http://dgroups.org/Community.aspx?c=f7fa34a1-2794-462c-8d99-2795682119f0

Website: www.ikmemergent.net

Mike Powell   Director, IKM Emergent   email: mike@ikmemergent.net

---

[i] See for example Pete Cranston's look at connections being made using the BBC's linked and open data at http://thegiraffe.wordpress.com/2009/12/04/linked-open-data-web-or-not-the-semantic-web/
[ii] See The 'four rules of linked data' put forward by WWW inventor Sir Tim Berners-Lee
[iii] For example public domain licenses best support the combination and re-distribution of data from different sources without creating license incompatibility and the widely used suite of RDF standards (RDFa, RDFXML, N3 etc.) coupled with Linked Data conventions provide a set of technical standards likely to meet development sector needs for linked information production and use.