Semantic IKM? Context and possible directions for IKM's engagement with emerging web technologies.

Background Paper – version 4

Introduction

Informational developments[i] are taking place in all walks of life, in all places and affecting all types of human communication. They can be driven by social or cultural changes, by business opportunity or threat or by technological development. New informational developments invariably impact on human communication in unforeseen as well as in planned ways. Of course no-one can be expected to understand all the ramifications of the changes in which they are engaged but sometimes, what is unforeseen by people in one area of work can be all too easily predicted by others with different experience or skills. However, the connections and the channels for communicating and discussing potential problems or opportunities are weak to non-existent. As a results, as argued below, information technologies which could have contributed to reducing inequalities and under-development in the world, have in fact increased them. As we look at the gathering momentum of linked data and other semantic web applications, and work on the standards and operating norms which will underpin them, how can we prevent the same mistakes happening again?

At a recent e-diplomacy workshop,organised by the Diplo Foundation , Mike Powell, director of IKM Emergent[ii], gave a brief talk on the potential of linked data to affect the information environment in which diplomats work. Although the subject was on the edge of the agenda of what was quite a hands on, practical event, several people in the audience, particularly those from smaller or less influential countries, immediately grasped the potential impact of new practices of data organisation on the politics of information sharing – on openness, inclusivity and exclusivity – in positive, negative and unforeseen ways.

The paper itself was a result of a visit by Pete Cranston, Dejan Dincic and Mike to the Online Information conference in London in December 2009, an event which focussed on developments with the semantic web. Pete's reflections on the event can be found at http://thegiraffe.wordpress.com/2009/12/04/linked-open-data-web-or-not-the-semantic-web/

and http://thegiraffe.wordpress.com/2010/01/06/not-the-semantic-web-part-two/ . We went to the conference with some scepticism as to the slow pace of progress towards the semantic web since the seminal article written by Berners-Lee and others in 2001. We left aware that developments in semantic search technologies and in the assembly and analysis of linked data represent real steps towards a semantic web and are also significant in their own right.

Background

Why might this matter to IKM?

First, the process of how new norms for knowledge, information and communication work emerge, may contain open or concealed political, cultural or economic biases and will, in any case, certainly affect how such work can then be carried out in the future. It relates directly to a fundamental argument of IKM that ICT have historically often been applied to development, and especially to the management of development organisations and development information, without proper consideration of their potential impact on development processes. This has involved, inter alia

- The use of proprietary ICT system development models which have precluded the collaborative creation, use and further adapation and re-use of development friendly open source software modules

- An uncritical acceptance of the tendency of many Management Information Systems, originating in the private sector, to centralise management controls in a way which has impacted on the capacity of local offices or grant recipients to manage their affairs in a manner sensitive to local context [iii]

- An unthinking acceptance as the inevitable 'norm'of those new models of informational behaviour, both formal and informal, which have originated in the North, with little awareness either of Southern innovation or of the impact on (two way) development communication of privileging one culture of communication over others

- The increased prominence and power of information produced by larger, and wealthier organisations vis a vis their poorer, and especially Southern, critics[iv]

- A shift in the relative 'information wealth' in terms of production of and access to even development related information, between the South and the North

- A lack of investment in off-line resources with the result that there is now less available in the form of guides and manuals for many development practitioners than there was 30 years ago[v].

Put bluntly, the introduction of ICT, Web 1 and Web 2 into the development sector has increased power and wealth imbalances within the sector, whilst arguably reducing the

availability of information to many of those who could make the most use of it. This was not inevitable. Alternative approaches and different models of technology development have been available from the start[vi] but, if noticed at all, they have received lip service rather than investment and support[vii]. The default position of senior management in the development sector has been the lazy one of either taking little notice at all or of buying into uncritical hype about innovation and digital futures. There are assumptions that ICT represent a homogenous block of technologies; that they offer no strategic choices, each with their implications for organisational and development practice; and that they can safely be left to technical departments and suppliers to sort out.

A second area is that of governance. Good governance is almost univerally seen as a necessary pre-condition of sustainable development. There is of course far less consensus as to what 'good' governance consists of, but it is widely understood to require more than good government. In particular there is interest in how government relates to other elements in society, including civil society. IKM, and many others, would argue that information flows are fundamental to the nature of that relationship and that this issue goes far beyond access to official data. Indeed we would see some dangers in any rush to promote 'open data' amongst governments and powerful multilateral bodies such as the World Bank or OECD without simultaneous action to consider how data from other, perhaps more critical, sources can be and is incorporated within the emerging information architecture and, indeed information which is necessarily not presented in the form of data at all.

The issues of information management within development organisations and of the architecture of public information about development come together in the more specific issue of the potential contribution of semantic approaches to linking development information from all potentially relevant sources. Development information, with its sources in so many disciplines, cultures and places is – and has always been – hard to categorise, hard to find. The notion of a collaboratively created development information environment or ecology has been something of a holy grail for some of us involved with IKM over many years. Despite our varied efforts –ELAND[viii], IDML[ix], assorted portals or information hubs, Euforic's sophisticated use of Delicious tagging[x] – the goal of ensuring that the right sort of relevant information from all relevant sources is available to those who need it in the form they need it has hardly been achieved in any one disciplinary branch of development, never mind its multiple transdisciplinary or cross-boundary communications needs. Attempting centralised answers never seems to catch everything and has huge resource implications. Decentralised approaches have stumbled on problems of metadata, of knowing what is

there, of finding ways to share. However if other countries follow Britain, Finland and the US in making some sort of RDF compliance standard for all types of public information, then the scope for extending this to development information is obvious. How to do it in a way which meets the needs of all users, brings information in from all relevant sources and seeks to reduce rather than augment information assymmetries is considerably less obvious and could form the starting base for some practical collaboration in this field.

Key Issues

Some of the main issues in the creation of ICT based on semantics and automatic machine-controlled exchange of information, which have development implications are:

The development process: how will new practice come into being? Will a small number of well resourced organisations see this as a new area of competitive advantage, one where they can 'take the lead' or is a more collaborative approach possible? Can a number of organisations participate in a process of sharing ideas, work and cost, developing their own applications for their own purposes but doing so in a way which also contributes to building an open and shared information environment?

Within this, is there a distinction between open and linked data? If so, this could represent more than a technical choice in how data are made available. The former,arguably, consists of simply making an organisation's data publicly available. A choice for linked data, however, implies an acceptance, even by the curators of data, of the value of a wider range of data sources and an acceptance of the impossibility (and possibly even the undesirability) of controlling how and by whom the data an organisation produces is interogated and analysed.

Permissions and source identification: There will definitely be issues about data integrity and quality. Without knowing anything of the technical means likely to be deployed, there is already some literature on access and permissions – certainly as far as what streams of potentially incoming data are accepted and possibly on restrictions which can be embedded in the provision of feeds as well. There is clearly a need for full tracability with regard to sources and for barriers against spam or deliberate disinformation. What we would hope to avoid is a two tier world where data from the World Bank, OECD etc. gets automatically accepted whilst countervailing information from small Southern organisations gets blocked by quality control firewalls.

Ontologies: a frequent criticism of the semantic web has been that it requires the acceptance of a single global ontology in order to work. This does not in fact reflect the vision of Berners-Lee in his original paper[xi] where he says 'Central control is stifling, and increasing the size and scope of such a system rapidly becomes unmanageable'. Current literature talks of 'emergent ontolgies', 'heterogenuous ontologies', 'dynamic and networked ontologies', all of which sound much better. However creating such open processes for ontology development and managing the process of translation between them in as multi-cultural, multi-layered and multi-disciplinary field as development is unlikely to be simple. It might be fun, but it won't be easy.

A rare example of what I understand to be an rdf compliant ontology in a development related field is the Knowledge model of Agropedia, hosted by the Indian Institute of technology, Kanpur[xii].

Schema[xiii]: linked data is about making statements about information in three parts or in triplets: 1 information x has 2 some relationship with 3 information y. The relationship is identified within a schema which has its own unique www location which allows the definition of that relationship to be found be any person or machine that looks for it. There is nothing to stop everyone creating their own schema, although doing so may mean that it is harder for potential linkages between information to be discovered. Alternatively, large organisations may try to become the source of commonly adopted standard schema. However, a number of problems can be foreseen. Firstly, however innocuous the schema itself – for instance it may simply offer a definition of what 'is in the same country as' means – using schema from some of the large early adopters such as the CIA World Factbook or even the BBC, may not give the desired impression of neutrality. Second, it is hard to see how amibiguity can (or even whether it should) be avoided. Starting with the term itself, 'development' is full of words which are understood in different ways, many of which may form the basis for the relationships defined by schema. For example the what is really meant by 'is a partner of' or 'is a stakeholder in' may vary between different users of the same term. At the moment, there is a mechanism whereby creators of one schema can look for other schema which appear to be the same and link them using an 'is the same as' tag. As far as I know there are not yet conventions in force which allow 'is similar to' or 'may have some resemblance to'.

Language: development ontologies and schema are complex within a single language. Given IKM's general thinking on the importance of the use of local languages with which to

discuss and manage development, it is imperative that efforts to 'simplify' the semantic web by making it an English web are resisted.

Related Developments

If we are thinking of some long term collaboration on these issues, there is also the question of whether it makes sense to focus solely on semantic web issues or to see how these will interact with other emerging, and possibly converging, ICT phenomena. These include

- The continuing development of social media
- Increasing use of visualisation as a tool for information analysis and communication (a trend which will raise a host of issues about the intellectual justifications for the linkages and analyses made as well as posing very fundamental issues about the existence – or not - of cultural variations in our understanding of visual languages)
- The less dramatic but also significant work around spatial approaches to the analysis of developmental issues[xiv]
- Orality and use of local language speech software[xv] (and whether the development of VXML standards – mainly applied in the management of call centres have any implications for this).

Future Plans

IKM plans to develop its thinking on these issues in a number of ways.

It is talking about a possible collaboration with the Diplo Foundation[xvi] and , hopefully either W3C or the W3 Foundation, to raise issues of policy and governance related to linked data and to advocate the adoption of rdf compliant systems for data handling within the UN system,

It plans to develop its own knowledge and practical experience of these issues through

- Significant further development of Information Artefacts workspace[xvii] and of topic map on our web site over the rest of this year.
- Support for a Vines 2, to add linked data from other sources (Opencalais) to the exisitng software and to explore the potential use of IKM Vines as an Ontology Comparator
- Work with the Young Lives Project and others to explore the issues in codifying real life data sets and in exploring what new can then be done with them in terms of analysis and communication

These initiatives will contribute to IKM's input to the Workshop proposed for November 2010.

It is hoped that the process of organising and holding the workshop will help us identify and develop relationships with a wider group of people with whom we might share a common interest on taking these issues forwartd in the development sector over the coming years. It will hopefully also take our ideas for what can be done and needs to be done significantly further forward as well.

Workshop

IKM plans to host a workshop, in Oxford, on the 15th and 16th of November this year. We also aim to leave the 17th open for groups of us to plan further work. It is open to the idea of co-hosting it if other organisationsare interested.

The aims of the workshop will be to explore the current .situation with regard to the use of linked data (and related innovations) in the development sector, to discuss their implications for development in general and for the development information ecology in particular and to see if there is shared interest in collaborative work in this area. Particularly if there is a parallel 'policy and governance' thread in collaboration with the Diplo Foundation, the emphasis of the workshop could be on practical opportunities and the technical and developmental challenges these pose.

The content of the workshop needs to be planned in advance by the participants. At the least it would seem sensible to ask participants to tell us about any work they are doing in this area and IKM could present its work in progress.

So far this has involved IKM working group 2, Johannes Keizer from FAO, Julia Chandler and colleagues from the technical implementation side at DFID. We would hope to display some of the work described above coming out of IKM. Johannes has offered to talk about related work he is doing on

- showing the use of AGROVOC in a collaborative environment and as linked data set that could glue together many information sets in agriculture
- talking about Drupal developments to make Drupal a "Linked Data Producer" and a "linkend data consumer"
- AgroTagger and intentions to create something like openCalais for agricultural information

We again should collectively consider who else we would like to invite.

Organising Process

We propose an open organising process in which all who express interest get added to a workshop D Group and can participate in preparation for and the planning of the event

I hope this document will get steadily added to or altered as ideas progress

---

[i] Hamelink, C 'Human Rights for the Information Society' in Girard, B and O Siochru, S (eds) 'Communicating in the Information Society', UNRISD, Geneva , 2003

[ii] IKM Emergent is a research programme which takes an overview and looks at specific aspects of the information and knowledge processes used within the international development sector in the context of what the actual information and knowledge needs of the sector are if it is to function effectively. It benefits from a variety of research methods carried out through a network of academics and development practitioners. It is supported by DGIS of the Royal Netherlands Ministry of Foreign affairs. See http://ikmemergent.net

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

[iv] Besemer – Presentation on limitations of Google Searches, Dakar (forthcoming)

[v] One example would be the range of literature available in the late 1970s on the prevention of deformities in leprosy compared to what is easily available now – on or off line

[vi] For example, Lane, G. 1990, 'Communications for progress: a guide to e-mail', CIIR, London

[vii] Mansell, R, 'Power and Interests in Developing Knowledge Societies: Exogenous and Endogenous Discourses in Contention', IKM Working Paper 11, August 2010

[viii] http://www.eadi.org/search-elnd.html

[ix] http://xml.coverpages.org/dml.html

[x] http://delicious.com/tag/euforic

[xi] Berners-Lee, T, Hendler, J & Lassils, O, 'The semantic web', Scientific American, May 2001

[xii] http://agropedia.iitk.ac.in/?q=content/knowledge-models

[xiii] Many thanks to Tim Davies for raising these issues to which I have tried to do justice here

[xiv] See for example , http://www.iapad.org/ and http://www.ifad.org/pub/map/PM_web.pdf

[xv] See for example http://www.runasimipi.org/blanco-en.php?file=manifiesto or http://www.bisharat.net/introen.htm

[xvi] As well the longstanding involvement of its Information Architect in IKM Emergent, the Diplo Foundation has played a significant role in the Internet Governance Forum

[xvii] http://wiki.ikmemergent.net/index.php/Workspaces:1._Information_artefacts