Technical Innovation: The Opportunities for Increasing Transparency and Accessibility of Aid Data

June 2010

Abstract

The ease with which ordinary people can now create, contribute to and access information resources is changing the relationship between institutions and citizens. This changing set of circumstances presents the international aid community with a unique opportunity to harness the power of a disintermediated public sphere in order to close the feedback loop between donors and recipients of aid, and introduce a step-change in efficiency and agency. To take full advantage of this opportunity, the international aid community must re-imagine itself as a publisher of "open data": raw, reusable data sets published in standardised, interoperable and open formats.

About aidinfo

aidinfo aims to improve the transparency and timeliness of data on aid and other resource flows in order to increase their impact on poverty reduction. The programme is led by the not-for-profit wing of Development Initiatives, an independent UK-based organisation with an established track record of research and analysis on aid flows and poverty reduction.

You can find out more information on our websites: www.aidinfo.org and www.devinit.org.uk

Contents

Executive Summary

The ease with which ordinary people can now create, contribute to and access information resources is changing the relationship between institutions and citizens. This changing set of circumstances presents the international aid community with a unique opportunity to harness the power of a disintermediated public sphere in order to close the feedback loop between donors and recipients of aid, and introduce a step-change in efficiency and agency. To take full advantage of this opportunity, the international aid community must re-imagine itself as a publisher of "open data": raw, reusable data sets published in standardised, interoperable and open formats.

The relationship between citizens and institutions is changing.
Those institutions that mediate between the public and the public good are finding themselves under new types of scrutiny, for example as they become subject to new expectations around transparency, or as the communications environment exposes them to stakeholders who previously had no voice.

Thanks to rapid innovations in communications technologies, citizens can now "be the change they want to see".
Motivated individuals now face lower barriers to initiating change than those faced by people working within traditional institutions. And the proliferation of personal technology such as personal computers and smart phones means that everyone can gather and report information, resulting in rich feedback cycles which stretch well beyond what institutions could achieve acting alone.

Institutions respond to this change best when they recognise their strength as data providers.
Rather than appropriating the innovations of motivated individuals and attempting to reproduce them, the best contribution institutions can make is by doing the one thing citizens cannot do so easily – aggregating and publishing "open data": raw, reusable data sets published in standardised, interoperable and open formats.

Early examples of institutions that have adopted this strategy are encouraging.
Evidence from the United States, and from flagship projects in the developing world, show citizens successfully capturing information, developing new ways of presenting or visualising information, combining disparate information sources ("data mashing"), or establishing rich feedback loops. In each case, understanding and agency is enriched.

The international aid community has a unique opportunity to take advantage of these trends.
This changing set of circumstances presents the international aid community with a unique opportunity to harness the considerable power of a disintermediated public sphere to close the feedback loop between donors and recipients of aid, and to introduce a step-change in efficiency and agency.

Introduction: changing relationships

This paper seeks to explain how technological, social and political drivers have led to new opportunities for institutions, and particularly institutions within the international aid community, to harness the power of ordinary citizens in delivering their mission. It argues that, in order to seize this opportunity, institutions must re-imagine their role in the information ecosystem.

The relationship between citizens and institutions is changing. Rapid shifts in the capabilities of information and communication technology are reshaping the role of institutions in the public sphere. In particular, those institutions that mediate between the public and the public good are finding themselves under new types of scrutiny, for example as they become subject to new expectations around transparency, or as the communications environment exposes them to stakeholders who previously had no voice.

The ease with which ordinary people can now create, contribute to and access information resources has joined a set of complex social and political drivers for institutional reform to result in demands for greater transparency around all aspects of public policy. This changing set of circumstances presents the international aid community with a unique opportunity to harness the considerable power of a disintermediated public sphere to close the feedback loop between donors and recipients of aid, and to introduce a step-change in efficiency and agency. In order to take advantage of this opportunity, the international aid community must re-imagine itself as a publisher of “open data”: raw, reusable data sets published in standardised, interoperable and open formats.

The enhanced role of citizens in the public sphere

Thanks to rapid innovations in communications technology, citizens can now “be the change they want to see”. Motivated individuals now face lower barriers to initiating change than those faced by people working within traditional institutions. And the proliferation of personal technology such as personal computers and smart phones means that everyone can gather and report information, resulting in rich feedback cycles which stretch well beyond what institutions could achieve acting alone.

Innovations in the communications environment that have taken place over the past decades have enhanced the role of citizens in the public sphere, in at least two ways. Firstly, highly-motivated individuals now have cheap and accessible tools to build communications platforms of the scope and reach that, before the rise of the world wide web, could not have been produced even by extremely well-resourced institutions. Secondly, thanks to the ubiquity of personal technology, citizens can now contribute directly to information creation, dissemination and scrutiny in entirely new ways. Both these trends have emerged thanks to the mass adoption of the world wide web’s open standards, which create an evolving “many-to-many” communications environment with extremely low barrier to entry, an environment that is radically different to that of the previous, broadcast media age.

The growth of the world wide web

In the mid-nineteen nineties, the web, though growing in popularity, was still essentially "read-only". Thus although the web did allow those with the time and inclination to set up websites, gaining access to a communications medium with the power to reach audiences of a size previously only available to the mainstream media, communications remained essentially unidirectional. Those wishing to talk back had no structured way to do so beyond initiating their own web presence and broadcasting their views into the ether. The web looked like it might be an eternal cacophony.

At the turn of the century, the audience began talking back in a structured way. Sometimes, this structure was machine-led: for example, new blogging software included functionality to allow comments underneath posts, and to track and highlight mentions of blog posts elsewhere on the web. At other times, the structure was social, that is it was created by a sum of individual human actions. So, blogging software began to allow bloggers and their readers to "tag" items, such as blog posts, with free-form text, which in turn led to catalogues of content designed not around a premeditated ontology, like the Dewey decimal system, but around emerging "folksonomies".

As broadband capability expanded, and as ownership of digital devices such as digital cameras and digital video cameras became more widespread, text-based content was increasingly joined on the web by rich media – visual, audio and audio-visual content. Content-sharing platforms like Flickr (for photos) and YouTube (for videos) began to increase in popularity. Audience talkback features embedded in these platforms – particularly tagging and folksonomy – contributed positively to overall functionality by generating navigation pathways and performing editorial functions, such as identifying popular and notable content. Platforms which included these sorts of features were dubbed "architectures of participation" by Tim O'Reilly, author of the seminal 2005 paper on Web 2.0 *What is Web 2.0?*[1].

The predictions of O'Reilly – that the **websites and web services that would do best would be those driven by architectures of participation** – were borne out by the dominance in the following years of social networking and micro-blogging platforms like Facebook and Twitter. Moreover, these applications, while generating a significant amount of fear in the mainstream media around both individual privacy and the trivialisation of public discourse, bore witness to a number of significant political and social events that brought the possibilities of the new information environment tantalisingly into view.

The role of motivated individuals

The potential of the new information environment to contribute towards social and political change was realised first by grassroots groups, as opposed to institutions. Distinct patterns of development can be observed, as what start out as grassroots initiatives are gradually embraced by existing institutions to the benefit of all.

[1] (O'Reilly, 2005)

One example of this is the work of mySociety in the UK. mySociety is "a charitable project which builds websites that give people simple, tangible benefits in the civic and community aspects of their lives." Founded in 2003, the organisation develops, releases and maintains high-profile e-democracy tools, platforms and services with an emphasis on accountability and transparency, such as "TheyWorkForYou.com", "WriteToThem.com" and "WhatDoTheyKnow.com". mySociety emerged out of activities undertaken by a small group of tech-centred democracy enthusiasts who saw the potential of the nascent web to refresh democratic institutions and practices, appropriating the digital outputs of official institutions and building on them to transformative effect.

mySociety has gradually been accepted by official institutions, who have provided grants to enable the organisation to expand its activities. Although institutions within government have not been universally cooperative – mySociety continues to campaign for draft Parliamentary bills to be released in a format that would make it easier to build tools for citizens to track their progress into law – individuals within mySociety have been co-opted onto government strategy teams to devise future information policy, one outcome of which is the UK Government's own open data portal, data.gov.uk. Anecdotal evidence suggests that mySociety's social impact in the UK is considerable:

- citizens groups regularly use mySociety's Freedom of Information tool WhatDoTheyKnow, to construct national surveys of data held by local councils[2];
- campaigners regularly use interactive tools like WriteToThem to help their supporters contact elected representatives[3]; and
- MPs have themselves used TheyWorkForYou to refer to their record in Parliament[4].

TheyWorkForYou.com

TheyWorkForYou.com contains the content of the official record of UK Parliamentary activity, Hansard, enriched with search and comment functionality, as well as navigation pathways that use postcode data to take UK voters straight to the contributions of their elected representatives, presenting them with basic analysis of representatives' activities and voting record.

Why should grassroots organisations be leading the way in utilising new technologies for the public good? One answer is that **motivated individuals now face lower barriers to initiating change than those faced by people working within traditional institutions.**

[2] See, for example, Margaret and Ian Campbell's survey of allotment waiting lists data: http://www.transitiontownwestkirby.org.uk/files/ttwk_nsalg_survey_09.pdf

[3] See, for example, the Stop 152 campaign run by parliamentary reform group Unlock Democracy in 2009: http://www.unlockdemocracy.org.uk/?page_id=1934

[4] See, for example, Harriet Harman MP justifying her record following the MPs expenses scandal using data from TheyWorkForYou: http://www.harrietharman.org/my-allowances

Although there may well be technologists within institutions who are capable of seeing the benefits new technologies might bring to their mission – indeed, mySociety is made up of a significant number of former civil servants – those individuals trying to change from within will inevitably hit management barriers and the organisational syndrome of "not invented here". The flipside of this response is altogether more positive, which is that the new communications environment has empowered individuals to "be the change they want to see". That is, that the web has given skilled computer programmers and others a platform for distributing tools at very low cost, meaning that energy once spent lobbying for institutional change has been redirected towards demonstrative projects which persuade through example. A culture of cooperation between such individuals and the institutions their projects focus on therefore creates a positive feedback loop for greater innovation.

The wider role of citizens

Thanks to the ready availability of personal computers and smart phones, ordinary citizens can now contribute directly to information creation, dissemination and scrutiny in entirely new ways.

Many in the development sector remain sceptical about how likely it is that innovations in the Western world around new technologies will be mirrored in the developing world, where resources and bandwidth are lower. But worries about the "digital divide" appear increasingly misplaced: the proliferation of mobile phones in Africa and the high proportion of income the average African already spends on mobile services shows that developing world countries are more likely to leap-frog the developed world in their adoption of new technologies than to get left behind[5]. And influential examples already exist of citizens using personal technologies to contribute to the information sphere, on platforms built by motivated individuals in the developing world.

Established in 2007, following the disputed Presidential elections in Kenya, Ushahidi.com (pictured, overleaf) took advantage of the broad use of mobile phones in Kenya to solicit and publish eyewitness reports from ordinary Kenyans

[5] (Gillwald & Stork, 2008)

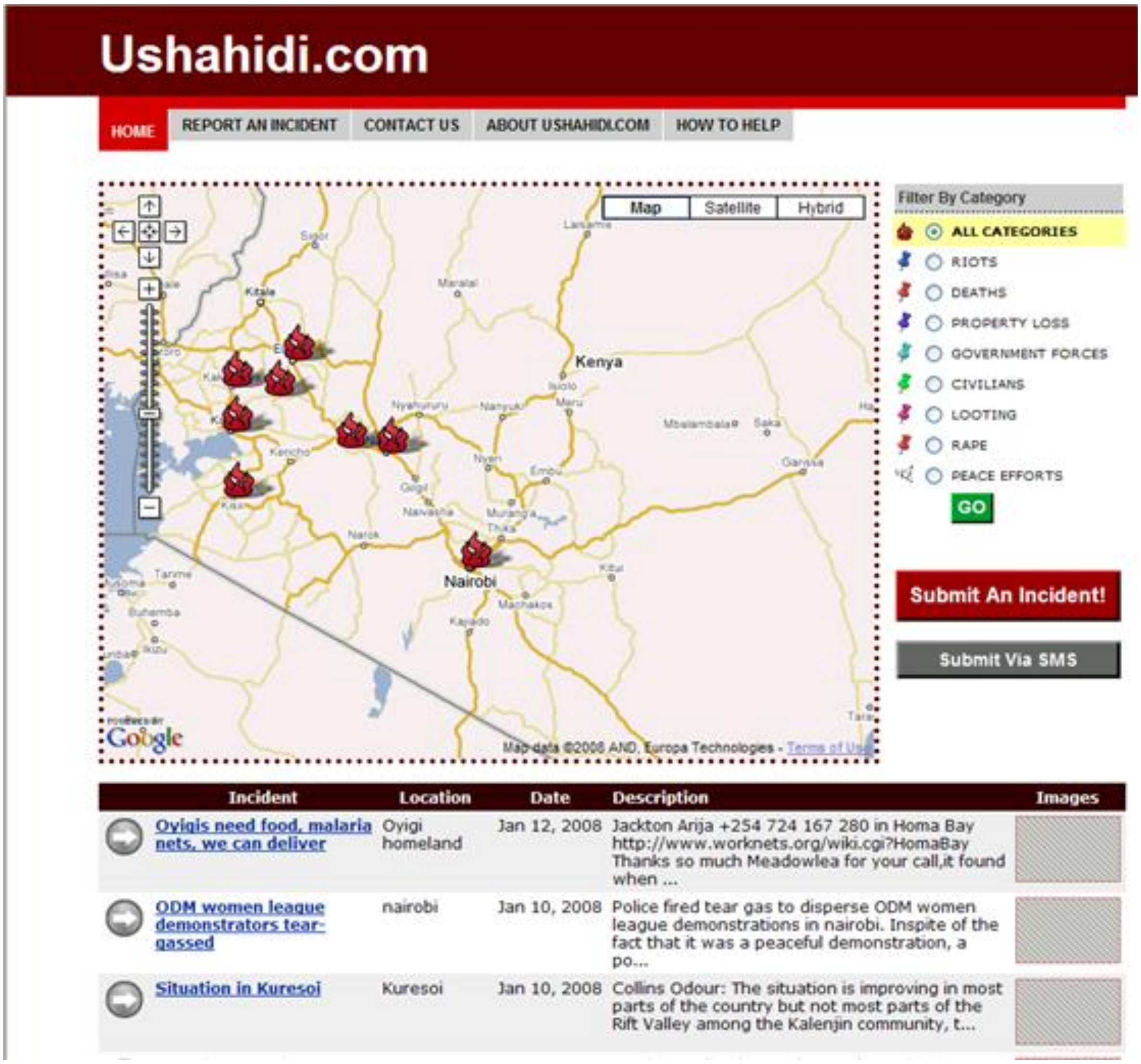
Ushahidi.com

Ushahidi.com was originally created in the aftermath of Kenya's disputed Presidential elections of 2007, by a group of Kenyan bloggers and programmers. The website solicited email and SMS eyewitness reports of violence in Kenya, which volunteers attempted to verify through an informal network of NGOs and local media, before plotting on an online map. Thus, Ushahidi uses both crowd-sourcing techniques and geo-spatial data visualisation to create an entirely new tool. The word "ushahidi" is Swahili for "testimony".

Ushashidi.com was built entirely by grassroots volunteers, who went from concept to launch in a matter of days. According to reports, just five days after launch, the site had received over 13,000 page views and published more than 20 reports. The OSCE is now using initiatives like Ushahidi to inform its conflict response work.

Ushahidi have released the website software, and it has since been used to track anti-immigrant violence in South Africa, to track pharmacy stocks across sub-Saharan Africa, to monitor elections in Afghanistan and –most recently - to aid the relief effort in Haiti.

on an interactive map. The map helped Kenyans avoid trouble areas, and also provided additional data to conflict response teams. It was developed exceedingly rapidly by a group of Kenyan bloggers and computer programmers – residents and expatriates – and used extensively once launched. The software has since been redeployed to crowd-source other location-based information, and the development, deployment and use of Ushahidi.com in Kenya has gone on to inform OSCE work on conflict response[6].



The changing role of institutions in the new ecosystem

As the previous section has attempted to outline, rapid changes in the communications environment mean that motivated individuals now face lower barriers to initiating change than those faced by people working within traditional institutions. And the proliferation of personal technology such as personal

[6] (The Organization for Security and Co-operation in Europe, 2009)

computers and smart phones means that everyone can gather and report information, resulting in rich feedback cycles which stretch well beyond what institutions could achieve acting alone. How can institutions respond to these changes?

As the following sections will demonstrate, institutions respond to this changing set of circumstances best when they recognise their strength as data providers. Rather than appropriating the innovations of motivated individuals and attempting to reproduce them, or embarking on data collection exercises on the scale achieved by citizens enabled with access to personal technology, the best contribution institutions can make is by doing the one thing citizens cannot do so easily: aggregating and publishing official open data sets. Such a response chimes not only with leading thinking on where the web will go next, but with the early experiences of institutions who are leading the way.

The future web

In 2009, Tim O'Reilly, together with journalist and publishing entrepreneur John Battelle, drew another line in the sand of web development with a paper entitled "web-squared"[7]. The authors of "web-squared" predict that the elements now contributing to Web 2.0 phenomena – software-structured architectures of participation plus the network effects of folksonomies and other human/social inputs – would soon be joined by other transformative elements, in particular, by the **proliferation of vast pools of open, linked data, published and maintained by institutions of all stripes.** These data sets, together with data derived from so-called "sensory inputs" (emerging from network-connected devices like mobile phones that were enriched with GPS-software and cameras), and from machine-learning algorithms such as image recognition technology, would result in an information environment far richer even that the environment resulting from Web 2.0. The result would be a "generative" web: that is, the web would be transformed into something capable of producing original information resources and communications outcomes not possible through the employment of any individual constituent input.

Tim O'Reilly's vision of Web Squared both complements and competes with other visions of what "Web 3.0" might look like, most notably, the vision of Sir Tim Berners Lee, a man often referred to as the "inventor" of the world wide web. Since 1999, Berners-Lee has been associated with the idea of a "semantic web", where information on the web is linked and codified such that it can be "understood" by machines, to the extent that computers could find, share and combine information in a manner at the moment only achievable by people. The idea is controversial among technologists, and the semantic web remains largely an experimental project.

Nonetheless, tools developed by the semantic web project are increasingly being put to real use in the data sphere, leading to the idea that metadata – data about data – could be codified and standardised to an extent that would allow much wider machine-directed cross-processing of data, to produce new data sets. One vision for the potential of the semantic web when it comes to linking data sets is given by the American Chief Information Officer, Vivek Kundra, as follows:

[7] (O'Reilly & Battelle, 2009)

"...let's say a web-savvy developer wants to create a mashup that visualizes and ranks various industries on revenue per employee. If one agency has published data on a designated industry's revenue and another agency has published data on its employment, then these records could be correlated if both datasets are categorized via the standard NAICS codes to produce revenue per employee for the given industry. Through reuse of these semantically harmonized and uniquely identified categories across domains, the data from multiple sources can be appropriately merged and new insights achieved."[8]

Institutions as data providers

In both the "generative" and "semantic" visions of "web 3.0", then, open data features heavily (for a definition of "open data", see box). And although non-institutional players can either appropriate official data sets (as with mySociety'sTheyWorkForYou.com) or create their own data sets using crowd-sourcing techniques (as with Ushahidi), in order to fully realise the potential of open data, institutions must enter into the information ecosystem as data providers. Several institutions are leading the way.

One of the best examples of institutions contributing to the generative web as data providers comes from the United States. Announced in the first months of the Obama administration, data.gov (see box, overleaf) is an open information portal that publishes federal-agency created datasets in open, machine-readable format. Its aim is to increase transparency, access and participation among the US and global citizenry, as well as to increase efficiency among federal government agencies. Following the launch of data.gov, several other

Open data definition

http://www.unlockingaid.info/

These recommendations for what open data might mean in an aid / development context are taken from (Gray, Hatcher, Hogge, Parrish, & Pollock, 2009):

Recommendation 1– Aid information should be legally open. The standard should require that information is published under terms which allow reuse for any purpose (including commercial). In particular material should be published under a license or legal tool that is compliant with the Open Knowledge Definition, such as a Creative Commons Attribution (CC-BY) license for content or the Public Domain Dedication and License (PDDL) or CC0 for data.

Recommendation 2 – Aid information should be technically open. The standard should require that raw data is made available in bulk with any relevant schema information. Material should be published in a format (i) which is machine readable and (ii) for which the specification is publicly available and freely usable. For example, text should be made available in HTML or XML and data should be made available in CSV or XML.

Recommendation 3 – Aid information should be easily findable. The standard should require that aid organisations add their knowledge assets to a registry with some basic metadata describing the information.

8 (Office of E-government and IT; Office of Management and Budget, 2009)

countries have announced or implemented similar projects, including Australia, New Zealand the UK and Brazil[9].

data.gov is a good example of an institution responding well to grassroots technical innovation. Rather than appropriating the innovations of the grassroots and attempting to reproduce them, the Obama administration has recognised that the best contribution it can make is by doing the one thing the grassroots cannot do so easily – aggregating and publishing federal data sets. This division of labour is a function of the new communications eco-system. As a group of Yale and Princeton academics and practitioners observe in their recent paper *Government Data and the Invisible Hand*:

> "In order for public data to benefit from the same innovation and dynamism that characterize private parties' use of the Internet, the federal government must reimagine its role as an information provider. Rather than struggling, as it currently does, to design sites that meet each end-user need, it should focus on creating a simple, reliable and publicly accessible infrastructure that "exposes" the underlying data. Private actors, either nonprofit or commercial, are better suited to deliver government information to citizens and can constantly create and reshape the tools individuals use to find and leverage public data."[10]

Encouraging early examples

Early examples of citizen-led projects resulting from institutions opening data are encouraging. New players in the public sphere really do respond to

data.gov

data.gov is a US government web portal providing the public with access to federal government-created datasets. It was created in 2009, both to allow citizen feedback and new ideas – enabling transparency, participation and collaboration between state and citizen – and to increase efficiency among government agencies. Most US government agencies already work to codified information dissemination requirements, and data.gov is conceived as a tool to aid their mission delivery.

One of the first memorandums signed by President Obama when he entered the White House was on openness and transparency. data.gov was launched by the United States Chief Information Officer, Vivek Kundra, on 21 May 2009. It initially contained 76 datasets from 11 government agencies. That number has now risen to 1,284 datasets from 170 government and related public agencies (as of April 2010).

Central to the data.gov proposition is the encouragement of data reuse, through, for example, using Application Programming Interfaces (APIs) to expose datasets to third-party web applications. data.gov does not employ data standardisation techniques associated with semantic web development, however, it does include integration with the emerging semantic web as one of its future goals.

[9] For a description of the events that led up to the release of data.gov and the UK's data.gov.uk, see (Hogge, 2010)

[10] (Robinson, Harlan, Zeller, & Felten, 2009)

institutions who release their data sets. Each of the examples given below show different features of the projects that have flowed from the US administration's commitment to open up data – search and access, visualisation, participation and feedback – although in reality they each exhibit a combination of those features. Projects derived from other data sources are also featured.

Search and access

flyontime.us is an online resource that provides information on the on-time performance of commercial flights in the US. Combining data sourced from data.gov provided by the Federal Bureau of Transportation Statistics, the Federal Aviation Commission and the National Oceanic and Atmospheric Administration, it is fronted by a simple search interface allowing users to see which operators over particular routes fly on time the most.

It also offers visualisations of air traffic data, and invites users to submit their own data on security check-in times using Twitter and smart phones. The philosophy of the individuals who created the site (which they describe as "an independent project created by private citizens") is made explicit on their "About" page:

> "FlyOnTime.us was also created to prove a broader point: when the government opens its data to the public, it saves your tax dollars! Could the Federal Government have made its own website that combines flight statistics and weather data? Sure it could have, but at a huge price. Government contracting is expensive, for a lot of reasons, but making a website like this costs a couple of volunteers like us virtually nothing. ... With just a little coding on our part - and the help of volunteers like you on your cell phones - we can create a service that is just as good or better than anything the government can do, and at a cost to the tax payers of $0!"[11]

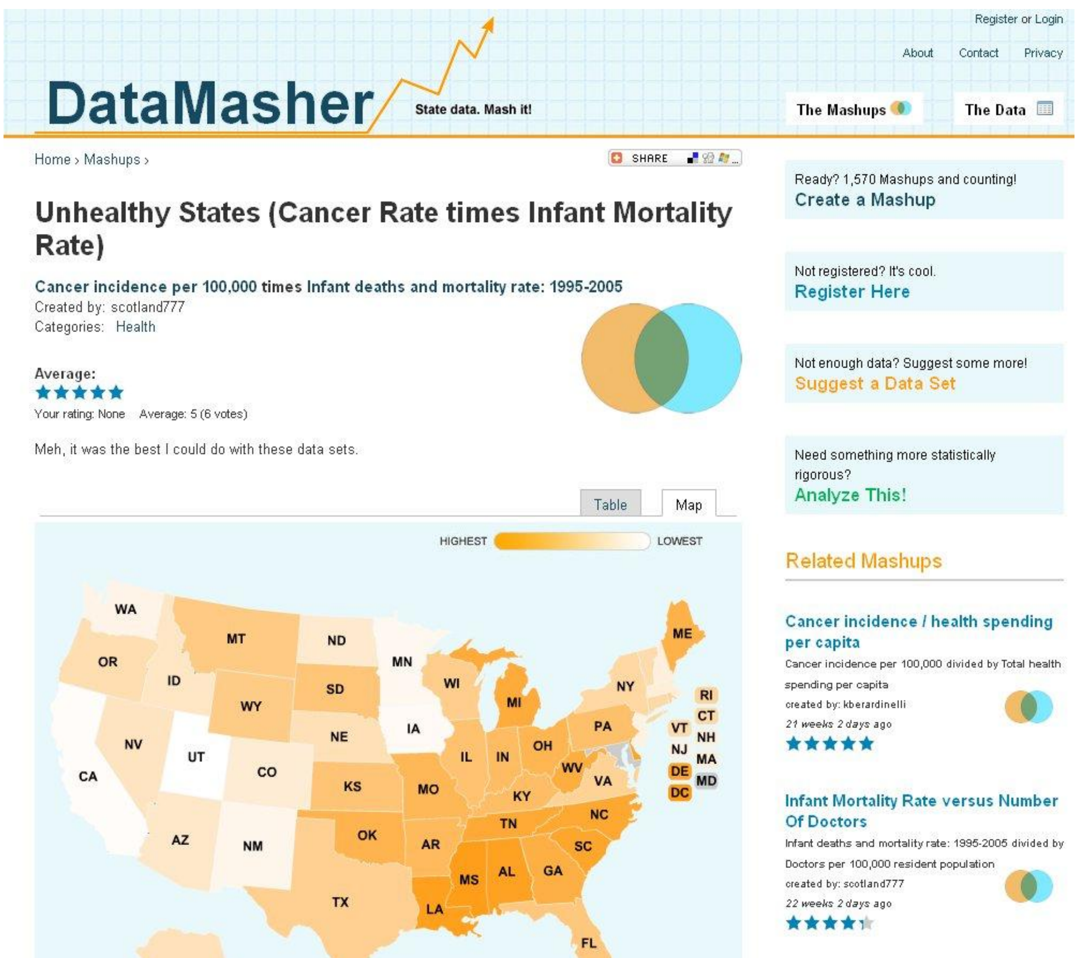
[11] (FlyOnTime.us)

Mashups & Crowd-sourcing

Data "mashups" is the popular term for services which combine two or more sets of data to produce novel information. In the example given above, flyontime.us, data from three different federal agencies have been combined. Further, the creators of the site are "crowd-sourcing" a fourth data set – information about security check-in times solicited via smart phone and Twitter from individual users – to display alongside data created by federal institutions. More examples of data mashups follow.

Mashups with other data sets

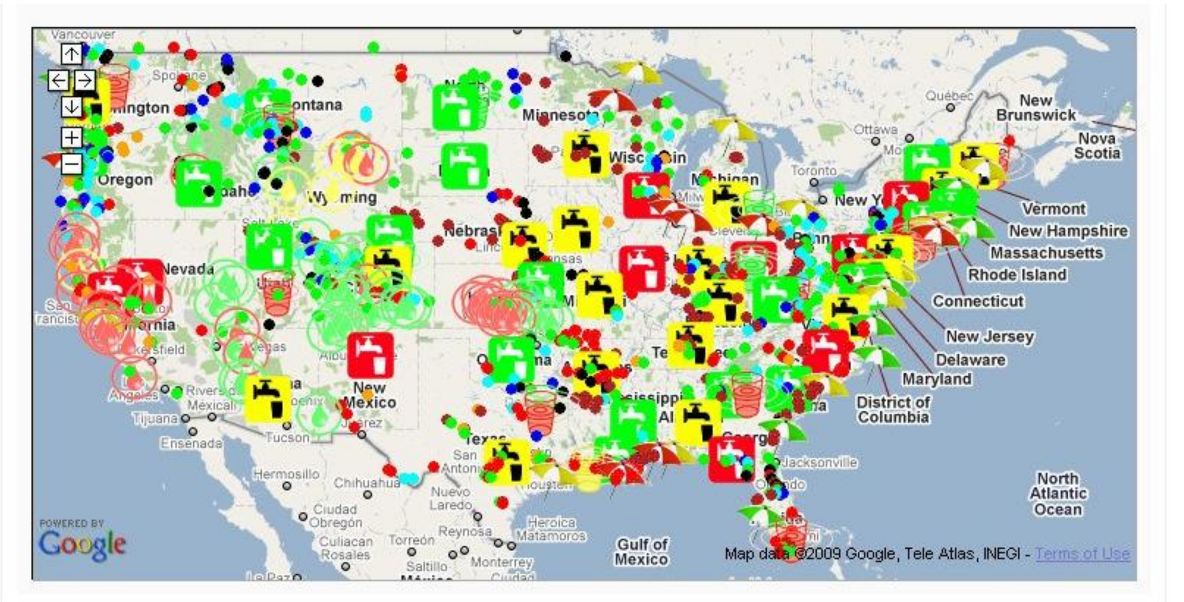
The datamasher.org service allows users to combine two or more data sets from data.gov creating maps of the US which display novel performance metrics state-by-state. It then allows users to rate and discuss these data mash-ups. So, for example, the most highly rated data mashup at time of writing combined infant mortality statistics with incidences of cancer to display "Unhealthy states", resulting in the interactive map, pictured.



The site was created by Forum One Communications, a commercial web strategy and development firm.

Mashups with crowd-sourced data

Watergoodness.org maps data about water quality across the US. The project creators use "every authoritative datapoint on water quality we can get our hands on" – including institutional data from the US Environmental Protection Agency and the US Geological Survey.

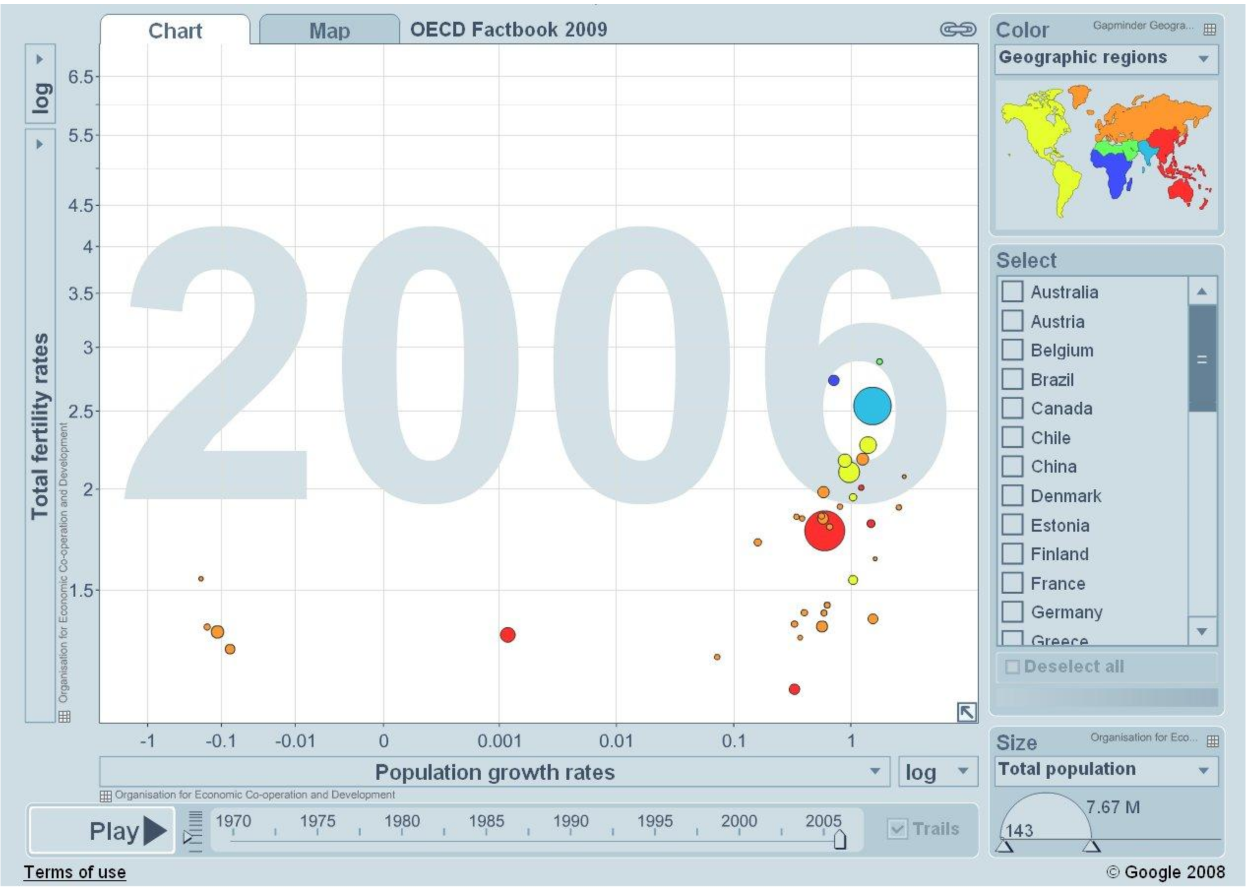


But the site also invites users to send in data about water quality using cheap water testing kits, and tries – ostensibly through a simple keyword search, that is admittedly not always successful - to show relevant activity on Twitter about water-quality issues. The result is a rich and detailed water quality map, pictured above.

Data Visualisation

Both Watergoodness.org and Datamasher.org use maps to visualise the information they create. Data visualisation – sometimes called infographics or information visualisation – is an important discipline that has developed in parallel with society's need to understand complex information, data and knowledge. Metro maps (such as Harry Beck's iconic visualisation of the London Underground, pictured, left), road signs, graphs, Venn diagrams, are all examples of infographics and information visualisation. Indeed map-making itself can be understood as an exercise in information visualisation.

Although data visualisation is not a new discipline, as technologies and societies become more complex, the basic knowledge citizens need to survive also becomes complex, and data visualisation has become correspondingly more mainstream and important. In particular, the rise of the web has witnessed a rise in the ease with which people can generate and share data visualisations. For example, tools and communities like IBM's Many Eyes project[12] and Swivel.com allow non-commercial and commercial users to create and share data visualisations[13]. The gold standard in data visualisation currently resides at Hans Rosling's Gapminder Foundation, based in Sweden. It was here that the Trendalyzer software program was developed.



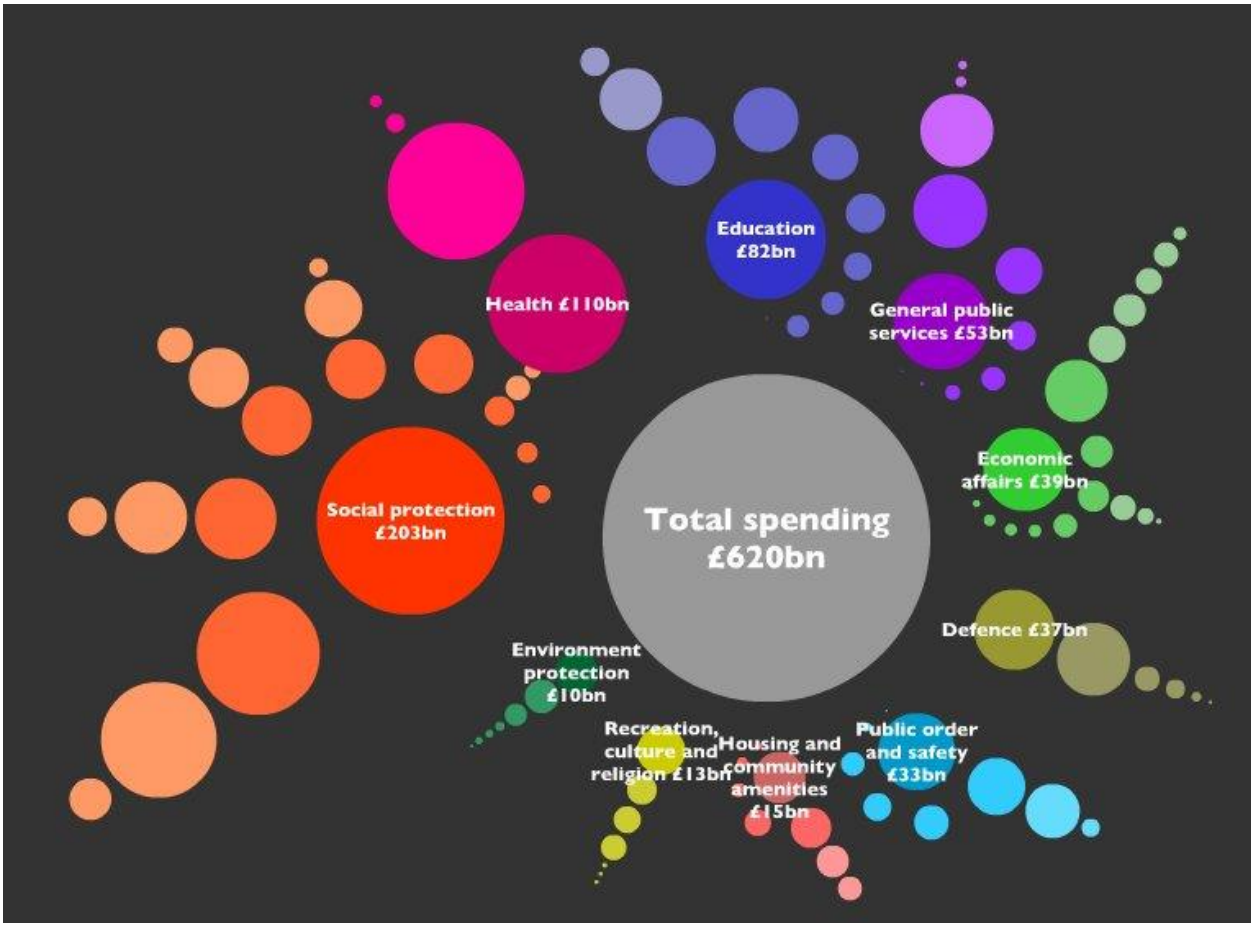
Trendalyzer is an elegant data visualisation software program that animates complex sets of statistics to make them easier to understand at a glance. Developed by the Gapminder Foundation in Sweden, it was acquired by Google in 2007. Based on the premise that "information is not understanding", the software specialises in showing country-level data as it has changed over time, and is therefore employed with growing enthusiasm by the development sector. The OECD converted their entire 2008 Factbook into a Trendalyzer portal (pictured, above).[14]

[12] See http://manyeyes.alphaworks.ibm.com/manyeyes

[13] For a guide to other visualisation tools, see Appendix III.

[14] see http://www.oecd.org/dataoecd/25/0/40679627.html)

Grassroots groups are experimenting with other data visualisation techniques, including the fledgling "Where Does My Money Go?" application that tracks UK government spending, produced by the Open Knowledge Foundation (pictured, below).[15]



The opportunity for the international aid community

As we have seen throughout this paper, institutions who respond to the changing information environment by re-imagining themselves as providers of open data, can expect to benefit greatly from the activities of the new, disintermediated public sphere. The opportunity for the international aid community in this respect is a considerable one – projects which harness the power of citizens to contribute to, access, re-use, enhance and present information using innovative visualisation techniques have the potential to close the feedback loop between donors and recipients of aid, and to introduce a step-change in efficiency and agency. Indeed, stakeholders in development issues are already calling for better information and better data to be made openly available by donors and partners in the aid space, for example in the ACCRA Agenda for Action in 2008. And a number of promising examples in this sphere are already emerging.

---

[15] See http://www.wheredoesmymoneygo.org/

Accessing information

Voice is a generally underutilised method for information access. With open technologies such as VoiceXML (a standard for defining applications where humans and computers interact with speech), SIP (an open protocol for voice over IP communication) and hosted IVR platforms (sites that host Interactive Voice Recognition applications), creating a relatively sophisticated voice application is no more difficult than building a website was a few years ago.

BBC Janala is branded as a "new way of learning English on mobile phones, the internet and television."[16] The mobile service launched in December 2009 and received over 750,000 calls in its first month. Offering regularly updated content for all levels of language learners, the BBC partnered with Bangladeshi mobile operators to ensure that each 3 minute lesson costs the user less than a cup of tea (3 pence).

With successful examples such as Janala to lead the way, and good tools to make application development easy, there appears to be broad scope for innovation. Perhaps a system that allows anyone to call a number to answer a short questionnaire about an aid project? A free number to call and find out which NGOs are operating in a particular region? As part of a multi-platform application, a mobile voice component can help to include users with lower levels of literacy and provide a different, perhaps more natural means of interaction.

Capturing information

Low cost, widely available and simple to operate, mobile phones are an ideal platform for capturing certain data. Appreciating this early on, DataDyne produced EpiSurveyor, a free application designed for data gathering in the global health and international development sectors. It allows a user with basic IT skills to set up a "worldwide, mobile-phone-based data collection system in minutes". It's strength, says DataDyne's founder Joel Selanikio, is in allowing users to "concentrate on public health and not waste time worrying about technology".

This is demonstrated by Patrick Okwen who has been leading health researchers in Cameroon using EpiSurveyor to collect data on areas including child nutritional status monitoring and investigating governance at hospitals. Patrick is a health expert, not a technologist and he's been doing this work without any training and only occasional support by email.[17]

EpiSurveyor currently boasts 1,800 users from 500 organisations and 100 countries worldwide with over 30,000 data forms uploaded in the past year alone. With a low barrier to entry, increasing support for open data standards, a developers API and facilities to export data in several formats, EpiSurveyor could become a valuable tool for capturing data on aid.

---

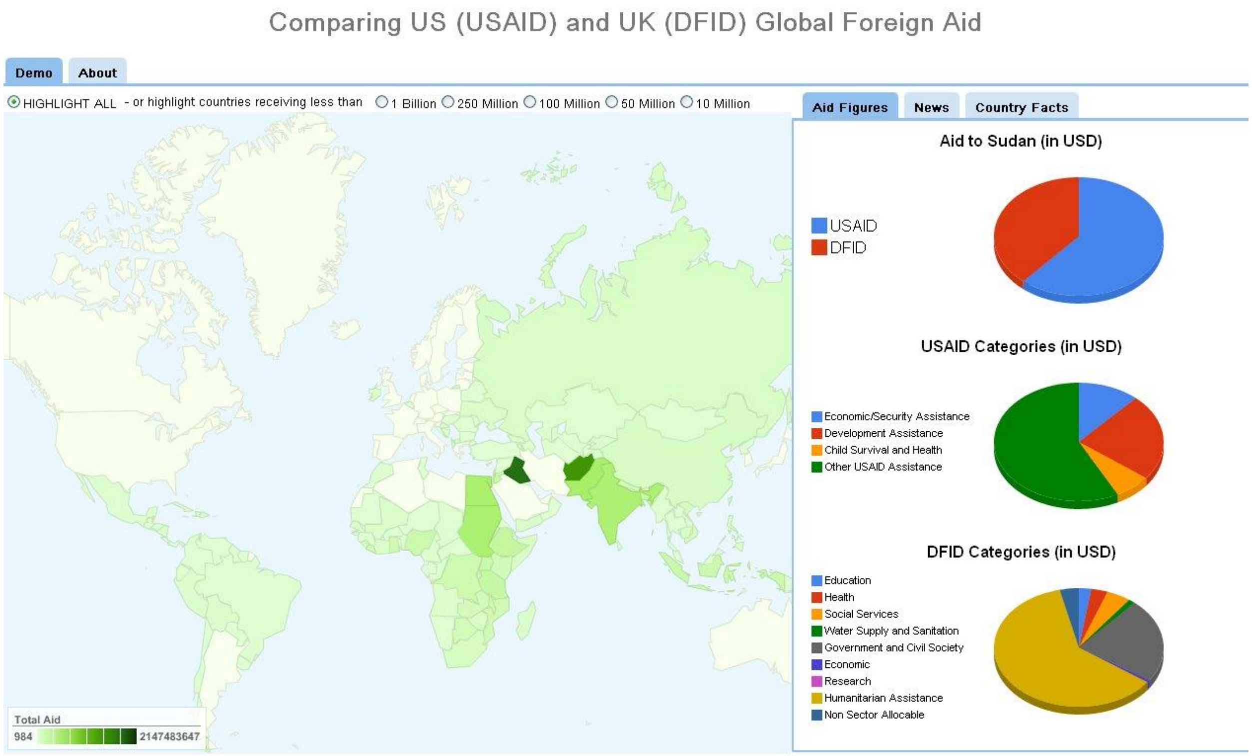
[16] See http://www.bbc.co.uk/worldservice/trust/whatwedo/where/asia/bangladesh/2010/01/100115_bangaldesh_janala_project_overview.shtml and http://www.bbcjanala.com/

[17] (Selanikio, 2010)

Visualising information

Although the number of institutions involved in international aid who have so far recognised the potential of opening their data represents only a small proportion of the development institutions that could benefit from adopting this role[18], citizen groups are already using data that has been released to create intuitive and easy-to-use mashups and visualisations.

For example, a comparative, map-based visualisation of aid distributed by the United States Agency for International Development (USAID) and aid distributed by the United Kingdom's Department for International Development (DFID) (pictured below, showing aid received by Sudan) was produced by James Michaelis, a PhD student from Rensselaer Polytechnic Institute, New York. It is derived from data released by the US and UK governments[19].



Future scenarios

The development-specific examples given above represent only the beginnings of what is possible. aidinfo have developed a number of future visions, to help stakeholders see the potential of new technologies and practices in delivering their missions. Each one of these visions combines a number of features discussed in this paper, including linked data provided by development and aid institutions, crowd-sourcing, search, visualisation and mash-ups.

---

[18] See Appendices for details of development-related organisations that have released open data so far.

[19] See http://data-gov.tw.rpi.edu/demo/linked/aidviz-1554-10030.html for more details.

- The financial management system in the Ministry of Education automatically tracks aid from donors, whether provided as general budget support, sector support, or project aid. Any conditions attached to financial aid (e.g. it must be used for text books) are coded by the donor, and read electronically by the Ministry of Education system, which ensures that transactions comply with these constraints. Reporting by the Ministry of Education is also automatic: donors can see whether and how aid has been spent by accessing a feed from their financial management system. No more laborious manual multiple reporting to donors.

- A village sits down in front of large plasma screen to agree on infrastructure plans. Where should the school be? The well? They build a “shopping list” of coordinated projects. A map of the community projects is overlaid with information about weather patterns, land use, road infrastructure, population density etc. Government and donors decide which of these investments to make – minimising the risk of duplication, or of a school with no teachers. The community then uses the same platform to report progress: what has been built, and does it meet their expectations. Taxpayers in donor countries can see on Google Earth where their funds have been used, and whether this has met the community’s needs.

- My church in Alabama would like to support an AIDS project in Kenya. Turning to AIDSPortal, I find there are 2,600 AIDS projects in Kenya. Information has been entered by NGOs working there, describing what they do and what they achieve. Other people have added pictures, videos and descriptions, and have rated the projects (e.g. responsiveness to local needs, sustainability, efficiency of administration, etc). There are anonymous comments by employees of the aid agencies, sometimes critical. My church wants to fund a project with low administration costs, so I rank the projects on this measure. Three have very low admin costs, but only one gets particularly good reports from the people who use the service. We click the “donate now” link and donate the proceeds from today’s collection plate.

- There are 10,000 water points in Tanzania, each with a unique serial number. Whenever someone goes to a water point and finds it is broken, they can send a text message. If they have a smartphone the message automatically includes GPS coordinates. A Tanzanian NGO receives the text messages and tracks which water points are well maintained. They organise the information by district, and publish league tables of effectiveness, which they send to all the locally elected representatives. They also send press releases to local radio stations. Local councillors come under pressure to ensure that the water points are properly maintained. The local NGO accesses information published by donors to discover who constructed and funded the water points. Many were provided by Oxfam: they work well. But the water points provided by an Italian NGO are twice as likely to break down. The Italian NGO comes under pressure in Italy to build water points to better standards.

- As a British taxpayer, I am sceptical of aid. I write to my MP. She writes back pointing me to a website which shows how aid money is spent. I can drill right down to how my money has been used to contribute to schools and roads and pay for teachers. I can see that in schools supported by British aid, real people have reported that the teacher is nearly always there, and that their children are getting a better education as a result.

- An NGO supported by the Open Budget Initiative in Cambodia publishes a popular guide to the budget each year, in the form of a cartoon book. It highlights how resources have been used – for example, it shows that education spending per person in one province is twice as high as another province inhabited mainly by a minority ethnic community. For the first time in 2011, the comic includes not just the spending by government, but also the spending by donors and NGOs. With the help of aidinfo, the NGO was able to get this information automatically from donors' data feeds, added up in the same way as the Cambodian budget. aidinfo showed them how to strip out the double counting, so that the aid provided by donors to government was not reported twice. The Cambodian NGO explained the budget to ordinary Cambodians and used it as the basis for a campaign for a more just allocation of resources. Ten years later, the allocation of total resources (budget and other aid taken together) converged towards equal per capita aid.

Finally, it should be noted that there is a not insignificant risk associated with development institutions' inaction on these new trends. As developed country governments warm up to their new role in the public sphere as information providers, and if motivated individuals and grassroots groups continue to demonstrate their enthusiasm for building transparency tools and services such as the ones detailed in this paper, then the citizens of the developed world are likely to become accustomed to this level of transparency and participation in public life. If the aid community cannot meet these new expectations, developed world citizens may withdraw their support for aid missions in favour of issues closer to home. The direct participation of tax-payers in shaping government policy on international aid may be just around the corner. The UK's Conservative party – now the major party in that country's ruling coalition government – included this statement in their 2009 Green Paper on International Aid:

> "We will establish a new MyAid fund, worth £40million in its first year. Every taxpayer will be able to log on to the MyAid website and view details of ten ongoing DFID (Department for International Development)-funded aid programmes, and vote for which one they think should receive the extra money. The options will include programmes run directly by DFID, as well as those run by respected NGOs. The Fund will then be distributed between the ten programmes in proportion to how many votes they receive. For example, if 25 per cent of people vote for the DFID programme in Malawi, that programme would receive 25 per cent of the Fund – £10million. Everyone who votes will be kept up to date with regular email updates about the progress of 'their' project."[20]

[20] See (Conservative Party, 2009)

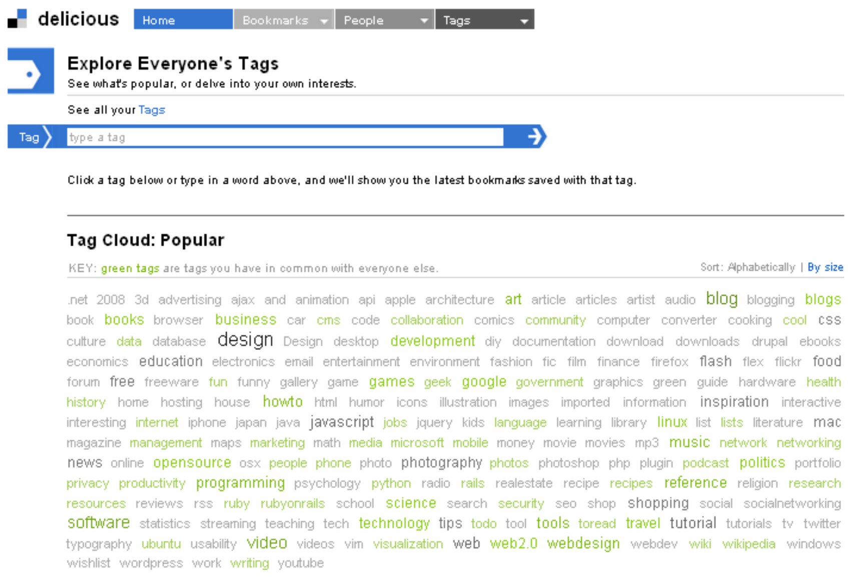
Conclusion

This paper has sought to highlight the unique opportunity that advances in communications technology present to the international aid community. It has posited a change in the relationship between institutions and citizens, and shown how rapid advances in communications technology have enhanced the role of citizens – both motivated and technically skilled individuals and a wider public enabled with new personal technologies – in the public sphere. It has shown how institutions are responding to this changing set of circumstances by adopting a new role as the providers of open data in an information environment where citizens can access, capture, reuse and remix data, as well as present data in compelling new ways. And it has looked to possible future scenarios where the international aid community have responded to this opportunity, showing how such a response has the potential to radically enrich transparency and agency in donor/recipient relationships, and mend the broken feedback loop.

Appendix I: Web development terminology and timelines

Web 2.0

Web 2.0, or the “read-write” web, describes online services and applications which are enriched by user participation. Key examples in the consumer space include social networking applications like Facebook, user-generated media sites like YouTube, Digg and Reddit, and blogging and micro-blogging platforms like Wordpress and Twitter. Commercial examples include Amazon and Google Search.



A website, web service or web application is generally accepted to be web 2.0 if users participate in a structured way, that is, if the software running the service categorises user input as it stores it, typically in a relational database. This is also known as an “architecture of participation”.

Users of web 2.0 applications and services will often be motivated by individual benefits that are distinct from the exponential benefits delivered by the “network effect” of the overall community of users. So, delicious.com offers individuals the opportunity to store their web bookmarks online, but offers its user community the opportunity to see what’s popular with lots of people on the web, either generally, or clustered around specific interests such as design, or politics (see delicious tag cloud, pictured above). And Facebook offers individuals the opportunity to stay in touch with their friends, while also compiling a dense picture of social networks that is insightful for both the wider user community, and for other stakeholders, such as advertisers.

Web 2.0 architectures of participation marry human characteristics like folksonomy with machine characteristics like relational databases. Web 2.0 can also be thought of in contrast to the traditional publishing model, with content being created on a large scale by a massively decentralised set of creators, and edited / moderated using structures built into the code layer. The term Web 2.0 was popularised five years ago by US publisher and technology guru Tim O’Reilly, who hosts an annual conference of the same name.

Web as a platform

Related to Web 2.0, the term “web as a platform” enshrines the vision of the Web not as a unidirectional information source, but as a place where people go to interact with each other, with information, with data, with images. Further, it captures the reality that just as individual web services can act as “platforms” for participation and co-creation, so the web, through the employment of interoperable standards (most notably open standards) can act as one giant platform.

“Web as a platform” also acknowledges the fact that many users connect to the web using technology other than the personal computer – for example mobile phones, or other internet-enabled devices.

Web Squared

In 2009, Tim O’Reilly, who popularised the term web 2.0, signalled an evolution in his thinking on the future of the web in a paper, co-written with John Battelle, entitled “Web Squared”. Where web 2.0 captured innovation flowing from the fusion of computer databases with unmediated user input (see Web 2.0) Web Squared added new inputs to the mix, specifically unstructured “sensory” inputs from web-enabled devices such as camera phones and structured “sensory” inputs like GPS, as well as information about external references gleaned from data libraries such as the ISBN book cataloguing system, and inputs derived from machine learning algorithms such as search or image recognition technology.

Web squared is understood to be an evolutionary leap forward from web 2.0. Where web 2.0 applications lead to “network effects”, web squared applications will go further, ultimately creating a “generative web”. Examples given by O’Reilly and Battelle include Microsoft’s Photosynth (pictured) – which can build 3D models of famous monuments using image recognition technology applied to GPS-tagged photos uploaded to the Web 2.0 photo-sharing website Flickr, and the iPhone app You R Here, which combines GPS technology with user input to help users navigate around country parks using maps they have photographed at the entrance.

Semantic web

The semantic web is a largely experimental project envisioned by Sir Tim Berners Lee that works towards a vision of information that is understandable by computers, so that they can perform more of the work involved in finding, sharing, and combining information on the web. It is based on metadata standardisation techniques that allow information to be coded with machine-readable information about its location, content and relationships with other bits of information. Although the semantic web project has so far failed to surmount the differences between human language and machine-readable language – such as human language’s tendency towards semantic vagueness and contextual inference – the toolset that the semantic web project has developed has been successfully applied to data. Using semantic web tools to enrich data sets with standardised metadata that allow them to “talk to each other” is known as “linked data”.

The semantic web toolset includes: XML (Extensible Markup Language), which governs syntactic structure; RDF (Resource Description Framework), a simple language for expressing data models which refer to objects and their relationships; OWL (Web Ontology Language) a more complex family of languages able to describe the properties and classes of an element, as well as the relationship between two or more elements.

Appendix II: Selected organisations that have opened up data

Donor Agencies

Agency	URL of Data Site
World bank	http://go.worldbank.org/0FRO32VEI0
IADB	http://www.iadb.org/projects/
EC	http://ec.europa.eu/europeaid/work/funding/
UK (DFID)	http://projects.dfid.gov.uk/

Other Development Agencies

Organisation	URL of Data Site
AidData	http://www.aiddata.org/
OECD (DAC & CRS)	http://www.oecd.org/dac/stats
OCHA (FTS)	http://ocha.unog.ch/fts/

Partner Country Systems

Organisation	URL of Data Site
AidData	http://www.aiddata.org/

Appendix III: Selected visualisation resources

There are numerous free and commercial tools that make it easier for users to visual data in common forms.

	Tool	Website
	Google Chart Tools + Visualisation API	http://code.google.com/apis/charttools/
	Free API and tools to create simple charts and more dynamic Gapminder style visualisations	
	SIMILE Tools	http://simile.mit.edu/
	Various open source tools that empower users to access, manage, visualise and reuse digital assets.	
	prefuse visualisation toolkit	http://prefuse.org/gallery/
	A Java visualisation framework providing tools for creating rich interactive data visualisations.	
	Swivel	http://www.swivel.com/
	Free version of an online tool for "social visualization of public data sets"	
	Tableau Public	http://www.tableausoftware.com/public/
	Free version of a commercial tool for "visualization of public data sets" that can then be shared.	
	IBM Many Eyes	http://manyeyes.alphaworks.ibm.com/
	Free tool to create, edit, share and discuss information visualisations.	

Bibliography

Conservative Party. (2009, July). *One World Conservatism: A Conservative Agenda for International Development*. Retrieved May 28, 2010, from conservatives.com: http://www.conservatives.com/~/media/Files/Green%20Papers/OneWorldConservatism.ashx?dl=true

FlyOnTime.us. (n.d.). *About FlyOnTime.us*. Retrieved May 19, 2010, from FlyOnTime.us: http://flyontime.us/about

Gillwald, A., & Stork, C. (2008). *Towards Evidence-based ICT Policy and Regulation: ICT access and usage in Africa*. Retrieved May 19, 2010, from researchICTafrica.net: http://www.researchictafrica.net/new/images/uploads/ria-policy-paper_ict-access-and-usage-2008.pdf

Gray, J., Hatcher, J., Hogge, B., Parrish, S., & Pollock, R. (2009, December). *Unlocking the Potential of Aid Information Version 0.2*. Retrieved April 23, 2010, from http://www.unlockingaidinfo/wp-content/uploads/2010/02/UnlockingAidInformation.pdf

Hogge, B. (2010, May). *Open Data Study*. Retrieved May 24, 2010, from The Transparency and Accountability Initiative: http://www.soros.org/initiatives/information/focus/communication/articles_publications/publications/open-data-study-20100519

mySociety. (n.d.). *About Us*. Retrieved May 19, 2010, from mySociety: http://www.mysociety.org/about/

O' Reilly, T., & Battelle, J. (2009). *Web Squared: Web 2.0 Five Years On*. Retrieved May 19, 2010, from web2summit.com: http://assets.en.oreilly.com/1/event/28/web2009_websquared-whitepaper.pdf

Office of E-government and IT; Office of Management and Budget. (2009, December 3). *Data.gov Concept of Operations: Draft.* Retrieved April 13, 2010, from http://www.ideascale.com/userimages/sub-1/736312/ConOpsFinal.pdf

O'Reilly, T. (2005, September 30). *What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software*. Retrieved May 19, 2010, from O'Reilly: http://oreilly.com/web2/archive/what-is-web-20.html

Robinson, D., Harlan, Y., Zeller, W., & Felten, E. (2009). Government Data and the Invisible Hand. *Yale Journal of Law and Technology , 11*, 160.

Selanikio, J. (2010, March 2). *Health Research in Cameroon with EpiSurveyor*. Retrieved May 28, 2010, from DataDyne.org: http://www.datadyne.org/node/533

The Organization for Security and Co-operation in Europe. (2009, November 17). *Gender and Early Warning Systems - An Introduction*. Retrieved May 19, 2010, from The Organization for Security and Co-operation in Europe: http://www.osce.org/publications/odihr/2009/11/41377_1387_en.pdf