

Data growth

A paper for the Funding Commission by Guy Yeomans

Introduction

“Raw Data Now! ...”ⁱ

“The functional changes that digital data facilitate—new competitors, higher-level information analysis, remixing of data, and new information dynamics—constitute a set of forces that are reshaping whole industries, governments, and communities.”ⁱⁱ

This trend considers: the growth in data volumes; the emergent open-data movement and the role of personal data. It offers a summary of why and where this might matter to the 3rd sector.

Suggesting that the volume of data we’re both individually and collectively producing is increasing seems an obvious and largely unremarkable thing to say. Indeed, as an ongoing trend, it’s one widely expected to continue based on current forecasts such as “the amount of digital information increases tenfold every five years”.ⁱⁱⁱ What may be more important is the recognition of both the ‘confluence of factors’ that have led to this and the potential changes that may now be emerging as a result.

Firstly, the range of data captured from our personal and commercial activities has expanded. Alongside the simultaneous proliferation in the number, diffusion and adoption of devices capable of creating such data, this has jointly facilitated the seemingly ever upward increase. To this ongoing digitisation of organisational and operational processes, we must now begin to add other forms of data – such as the increasing level of machine-to-machine interaction information (software logs, RFID readers, wireless sensor networks); existing ‘hidden’ data now being subject to access via open data^{iv} initiatives and also the ‘self-selecting’ personal data voluntarily submitted and shared across online social networking platforms.

Taken together, these are increasingly characterised as ‘big data’ and – according to a recent analysis by McKinsey – can be seen to represent “access to customer data from public, proprietary, and purchased sources, as well as new information gathered from Web communities and newly deployed smart assets”.^v However, it’s this confluence that has led the Aspen Institute to suggest “... as all these systems begin to interconnect with each other and as powerful new software tools and techniques are invented to analyze the data for valuable inferences, a radically new kind of “knowledge infrastructure” is materializing.”^{vi}

We’re entering an era of super-abundance, of ‘big-data’ or - as Joe Hellerstein (Professor of Computer Science at UC Berkeley) observed back in 2008 - “we are at the beginning of what

I call 'The Industrial Revolution of Data' " vii and this is the wider environment the 3rd sector needs to be mindful of over the next 5-10 years.

This surge in volume - accompanied by ongoing increases in accompanying computing processing speeds - opens up a new set of opportunities to spot trends in a whole host of fields and is leading to data becoming an economic asset in its own right. However, it also creates a range of future management issues from which item of data to record, which to track, assessing how it should be analysed or interpreted and how its storage, archiving and possible later retrieval should be organized. Also, we shouldn't forget the plethora of privacy and security issues this also creates, especially in relation to 'personally identifiable' information.

Commentator Tony Bain has tried to encapsulate the number of areas this data growth is now beginning to effect when he suggests that "...keeping in mind all the different platforms that Big Data is having an impact on (web, cloud, enterprise, mobile) combined with all the Big Data domain challenges (transaction processing, analytics, data mining, visualization) as well as many of Big Data's characteristic requirements (volume, timeliness, availability, consistency), it is easy to see how no single technology will provide a cover-all solution for the eclectic mix of needs." viii That's a tough challenge, but where does it leave the 3rd sector?

Well, most organisations may consider themselves too small or perhaps as currently producing or using too small a volume of data in their operational activities to consider this trend as relevant, however, we feel that "... what is new are the growing scale, sophistication and ubiquity of data-crunching to identify novel patterns of information and inference." Accordingly when you "commingle it with other large and very interesting secondary data sets (e.g., phone books and property records)" ix this trend will begin to affect ever more parts of our lives, roles and relationships. The potential increasingly suggested from existing commercial investigations of 'big data' is that this will create new operational opportunities based on data-led 'continuous improvement' using real-time, perhaps, automated decision making. Lastly, it's suggested this will facilitate a range of process, product and interaction experimentation which will drive using data for external change.

Open Data

"Publishing raw data, and expecting accomplished amateurs and social entrepreneurs to find uses for it, creates an eye-catching new type of partnership between citizens and the state—what the wonks sometimes call the "co-production" of public goods" x

Opendefinition.org suggests 'open data' "... is any kind of data that's freely usable, redistributable and reusable by anyone (subject only, at most, to a requirement to attribute and share-alike)". xi It may be seen as an emergent phenomenon of the last 2-3 years directed, initially, at making government data more widely available and has involved the combined efforts of both individuals and groups across '3-tiers' of society including – what a Soros

Open Society Foundation report characterises as - 'Civic Hackers'; engaged 'middle-layer' government bureaucrats and politicians offering the requisite 'top level mandate'.^{xii}

Back in May 2010 the coalition government published its approach to future public sector reform – the 'Big Society' programme^{xiii} – of which the fifth core theme centred on the intention to create a process that facilitated the ongoing publication of government data. In June, it held its first "Public Sector Transparency Board"^{xiv} meeting in which it outlined a set of over-arching 'Public Data Principles' by which public data – "the objective, factual, non-personal data on which public services run" – would be made available including outlining positions on formats, standards and licencing.

The proposed general value of open data lies not just in its apparent accessibility and transparency but in two key associated (and interlinked) activities. Firstly, the option to freely reuse and remix the data into new or novel forms (sometimes also referred to as 'mash-ups'^{xv}). This is sometimes done with the explicit intention of addressing needs or creating services for which the data may not originally have been intended. This also infers that the data will be more widely 'linked' to other data-sets.

Secondly, the 'remixing' can – and is likely - to be done by a diverse range of people with differing intentions and aims (political, commercial etc) and perhaps from outside the field from which the data was itself originally generated. A key aim is the desire to drive external innovation, an approach well represented by the 'Show Us A Better Way' website which asked users to come up with original ideas and services based on the proposition "What would you create with Public information".^{xvi}

While the emerging dual effects of these activities are likely to "influence both governmental and philanthropic policies"^{xvii}, they also raise some key short-term issues centred on organisational culture and long-term value.

The Economist neatly summarised the core 'cultural' issue underlying open data as an attempt to "... merge two cultures: the risk-averse ethos of the civil service, and the free-wheeling spirit of open-source developers, who seek continuous incremental change and see failure as a step to improvement".^{xviii} Is the 3rd sector open enough to approaches such as these?

A number of commentators have also made the point that access to data does not – of itself – lead to wider social empowerment or the creation of new 'value' and, as such, we should be wary of seeing "data-over-the-wall" as the end point of the process. As such, an important theme is recognising that "... what is needed is people who can make it make sense for the rest of us; we have the transparency but need lenses to bring out the detail."^{xix} There are also core data management issues, such as version control, sustainability of sources, accuracy and data cleansing that will need to be understood and accommodated.

However, programmes such as the International Aid Transparency Initiative offer an instructive example of how up to 18 individual donors can come together in a 'temporary

coalition' with the express purpose of creating standards-based open data for use by all stakeholders while also seeking to meet international regulatory commitments.^{xx}

We should reiterate. This process is centred on activities to create an open-data ecosystem within government; envisioned as one that eventually cascades down from Whitehall to your local town hall, perhaps even to village level. From global to hyper-local. However, that process is still nascent, is by no means a long-term 'given' nor one that can yet call upon widespread understanding or truly popular demand. Also, as noted by Christopher Csikszentmihalyi, Director, MIT Center for Future Civic Media, "...while much of the open data initiative has been about making government data public, getting citizen data to the government and to the rest of the public — whether complaints or other information — is also important."^{xxi} It is perhaps the thrust of this last comment that really begins to suggest why this trend might be important for the 3rd sector.

Key questions for the sector

The open data movement raises two clear future issues for the third-sector: i) how can it work with and leverage to best effect the range of current open data initiatives and data-sets available and ii) how will it respond to any similar imperative to become 'open' with its own individual and /or collective sectoral data?

We'll consider each question briefly in turn.

In his recent review of 3rd sector awareness and engagement on this issue NCVO's David Kane noted open data "is a new concept to much of the voluntary and community sector"^{xxii} Indeed, in an - unrelated - survey conducted by BrightOne and CharityComms into the future of charity communications trends over the next decade, only one participant out of 21 made explicit reference to the need for 'open data' across the sector in the future.^{xxiii}

As such, it's likely the sector is not actively engaged with this emerging area of potential capability or its impacts. As an initial set of thoughts, Kane offers 4 perspectives which he sees as key opportunities and challenges:^{xxiv}

Open data will:

- provide charities with new ways to engage with and assist their beneficiaries
- provide them with the means to validate and endorse their campaigning positions
- for the most part - be beyond the current skill-set and knowledge of the sector at the moment
- may subject the Government-3rd sector relationship to profound change if open data becomes a mandatory component of future public service funding processes

(NB - To extend the basis of this last opinion, Martin Brookes (New Philanthropy Capital) has recently argued for the introduction of a UK 'Charities Data Act' effectively binding

the continued award of charitable status with the creation – and adoption - of open data principles and standards by each organisation).^{xxv}

In relation to our second question – how will the sector respond to a similar open data imperative - Lucy Berholtz’s analysis of the same issue in relation to the US is also apposite for the UK, namely, “...it’s not yet known what force—third party intermediaries, regulation, market forces, leadership within the field—will drive a similar opening-up of philanthropy”. Will the 3rd sector need to follow a similar 3-tiered process? Will the core ‘sharing’ approach of open data have resonance across the sector? Can the sector conceive of a range of new, collectively influenced outcomes?

At this point, we perceive a clear lack of dialogue (let alone) consensus on what the key policy, operational, legal, technical and out-reach issues of open-data are likely to be for the sector over the next 5 years. While there are a number of organisations and commentators beginning a discussion – including the Open Knowledge Foundation, mySociety, New Philanthropy Capital and NCVO – opinions are still at the early stages of formation. We accept this could be a tough issue for any charities involved in a current operational and funding environment beset by potential cuts and short-falls to put any thought into, but how the sector captures and seeks to use its data will become of ever increasing importance. As Rufus Pollock of the Open Knowledge Foundation has said “... many institutions and communities are now facing decisions which will help to shape the future of how knowledge is shared — and which will help to determine whether we will have a plethora of poorly connected walled gardens, or a shared ecosystem that everyone can benefit from.”^{xxvi}

A note on the private sector

To help wider comparisons, a recent report on business attitudes to open data conducted by LM Research illustrated that while UK private businesses welcomed the government opening up its data they didn’t feel the process should be automatically reciprocated from within the business sector as a whole. Up to 68% of the businesses polled said “... they would not be prepared to open up access to their own data - despite recognising that sharing data could bring commercial -benefits”.^{xxvii}

The importance of individual actions

“Social data is overwhelming”^{xxviii}

There’s an increasingly individual and personal element to ‘Big Data’ that shouldn’t be overlooked and that means data which directly emanates from us all.

While being aware of, sourcing and using these kinds of data won’t surprise anyone involved in the area of prospect research – with a recent US study showing that take-up and use of

such data has increased as a result of the recession – some in the sector may be unfamiliar with the increasing sophistication with which others are using such data. As a current example, Elizabeth Crabtree, director of prospect development at Brown University, says six out of 10 donors who made a major gift (\$100,000 or more) in its current campaign were initially identified using predictive analytics and data mining^{xxix}

However, as each of us travels, shops, selects a TV channel – or takes part in any other activity that has some element of digital integration - we’re leaving behind what’s being called our ‘data exhaust’. By way of example, “... mobile devices in America are generating something like 600 billion geo-spatially tagged transactions per day”^{xxx} and while items such as this ‘space, time, travel’ data are ‘de-identified’ this kind of information is increasingly sought for use within the field of persuasive technologies, of which, a Wired magazine article warned, “... the tracking technologies have arrived and are already being tested in real situations – even if the respective databases have not yet been combined in realtime”.

^{xxxi}

We should also note the – complimentary - emergence of social analytics as a series of developing processes and technologies. These seek to derive ‘intelligent’, predictive insights out of the vast – and ever increasing - volume of social data (web browsing habits, search requests, profiles created on social networks, twitter-streams and the posting of user generated content such as pictures or videos). Underlying this, are three key approaches to how such data is used, its level of ownership and to whom any benefit should accrue. These have been characterised as ‘trust’, ‘protection’ and ‘value’. Google is the standard example of ‘trust’ in which it uses the data it gathers about usage to inform its advertising business and – in relation to privacy and ownership – then asks that we collectively trust them to ‘do no evil’. Protection approaches tend towards the use of 3rd party services, such as mozy.com, to track and manage an individual’s online data while newer services such as Bynamite^{xxxii} are seeking to create options that give-back control of data to the individual involved and, at the same time, allow them to engage in potential exchanges of value, such as making their individual profile of interests and preferences open in order to be able to qualify “for micropayments or discounts”.^{xxxiii}

Given the importance of both social media and - what’s been termed – the ‘free and public data economy’ to this sector these are important considerations going forward especially when the ‘holy grail’ is the eventual linking and integration of as many data-sets as possible. How will the sector respond to developments such as this? Does it have a position on the appropriate or ethical use of persuasive technologies for fundraising purposes?

Lastly, as Craig Mundie, head of research and strategy at Microsoft stated “the data-centred economy is just nascent. You can see the outlines of it, but the technical, infrastructural and even business-model implications are not well understood right now.”^{xxxiv}

Key next step

Consider attending the forthcoming Open Government Data Camp - <http://opengovernmentdata.org/camp2010/>

Summary

Overall, we believe that organisations that effectively collect, collate and utilise the data available to them (from whatever current or future source) will demonstrate improved operational performance from their charity actions; be better able to identify and target need (so selecting donors more efficiently) and also be able to more effectively drive improvements in fundraising values and volumes. In addition, the process of exposing and using both new and previously unobtainable data begins to open-up innovative new service delivery and business development opportunities.

Trend Analysis

Uncertainty	Generally high within 3rd sector
Rate-of-change	Big Data: Higher in business, low in 3 rd sector Open data: Higher in government, low in business and 3 rd sector Social Data: High in business, relatively low in 3 rd sector and government
Reach	All of the data areas identified have potentially sector wide impact
Adoption	Big Data: Mostly in larger enterprises Open data: Mostly in government Social Data: Mostly in commercial sector, but an increasing level of interest from 3 rd sector
Related to other trends	Cloud computing / social media / real-time web

Implications Matrix

Some initial general thoughts and questions to consider:

See below for key	LO	S-T-M	WPS
Organisational structure	Likely to consider dedicated personnel (either internal staff or external companies) with formal responsibilities and processes in place for data handling.	Likely to make this an additional area of someone's existing role or job profile particularly from within any IT-related function.	Most likely to use either 'pay-as-you-go' 3 rd party services or ad hoc freelancing staff or it just gets addressed on an 'as needs' basis.
Operational needs	<ul style="list-style-type: none"> Analytics functions Data collection and validation processes (removal of obsolete, 	<ul style="list-style-type: none"> Likely to be a far more pragmatic approach to balancing data 	<ul style="list-style-type: none"> Should aim to reduce 'downstream' burden of data

	<p>duplicated, incomplete or 'plain wrong' information)</p> <ul style="list-style-type: none"> • Error and complaints procedures • Access to 'dead data' archives and stored information • Ability to support real-time information flows • Potential for data visualisation options • Greater likelihood of 'bespoke' options being considered 	<p>needs, potential benefits, manpower availability and any external regulatory requirements.</p> <ul style="list-style-type: none"> • Processes should be thought-through carefully and tested to reduce any of the impacts noted above. 	<p>processing by determining what data is really needed (don't store crap data you don't need)</p> <ul style="list-style-type: none"> • Should seek to be as 'standards-compliant' as possible.
Infrastructure requirements	<ul style="list-style-type: none"> • Dedicated data storage options with both analytic and processing capabilities. • More likely to have disaster recovery options in place (back-up and restore etc) 	<ul style="list-style-type: none"> • Likely to have a 'mix' of hardware and software developed on an incremental basis. • May not be able to 'integrate' such a set-up successfully enough to leverage data. • May find benefit in 3rd party cloud computing services or dedicated data processing and analytic service providers (who take and use the organisation's data) 	<ul style="list-style-type: none"> • Likely to be ad hoc • May wish to consider shared facilities with another charity or cloud computing services
Recruitment and staffing	<ul style="list-style-type: none"> • Are 'C' level appointments needed? (CIO) • Who will design and implement any data handling processes? Are additional appointments needed? • How will staff be trained? • Do existing contracts or HR terms-of-reference need to be revised to accommodate new responsibilities in relation to data handling? • What 'pre-requisite' skills would you expect new staff to be able to demonstrate? 	<ul style="list-style-type: none"> • As above, this may be something you need to make an additional component of someone's existing role • A high level of overall digital awareness will become increasingly important • What 'mix' of skills will work best for your organisation and its aims and internal structure? 	<ul style="list-style-type: none"> • Anyone with dedicated 'data' skills may be too specialised to work in your circumstances, but try and look for as much exposure to the area in people's previous experience • Do you have likely or suitable internal volunteers who would be prepared to learn this area?
Education and skills	<ul style="list-style-type: none"> • What professional 	<ul style="list-style-type: none"> • Perhaps more 	<ul style="list-style-type: none"> • Difficult to

develop't	<p>qualifications would help in this field?</p> <ul style="list-style-type: none"> • What emergent skills do you need to incorporate into any internal training? • Who is monitoring standards or relevant case-studies across the sector? 	<p>incremental and relying on effective 'internal transfer' from those with skills and experience</p> <ul style="list-style-type: none"> • Likely to benefit from engaging with any wider sector community involved with this area • Useful to compare needs and responses to 'same-size' organisations facing broadly similar issues to share knowledge and best practice 	<p>address using only own resources; perhaps using volunteers to help raise awareness of specific skills</p> <p>Consider collective 'hack-days' to help generate interest and as a means to better understand how data use may help the organisation.</p>
Donors and sponsors	Enhanced ability to show impact of activities	Ditto	Ditto
Recipients and benefactors	<ul style="list-style-type: none"> • Potential for better understanding of needs • Potential to better target response or funding • Potential to provide more flexible responses that can be adapted as needed 	Ditto	Ditto
Culture and behaviour	<ul style="list-style-type: none"> • Potential to quite fundamentally change the terms of reference of existing relationships • Decision making based on explicit data / information rather than experience or personal hunches 	Ditto	Ditto

Key:

LO : Large organisations big enough to support permanent staff with functional specialism's

S-T-M: Small-to-medium sized organisations, with permanent staff but not large enough to support individual functional specialism's

WPS: Organisations without permanent, salaried staff

ⁱ http://www.ted.com/talks/tim_berners_lee_on_the_next_web.html

ⁱⁱ <http://www.scribd.com/doc/31178075/Disrupting-Philanthropy-FINAL>

ⁱⁱⁱ http://www.economist.com/specialreports/displaystory.cfm?story_id=15557443

-
- iv http://en.wikipedia.org/wiki/Open_data
- v http://www.mckinseyquarterly.com/Clouds_big_data_and_smart_assets_Ten_tech-enabled_business_trends_to_watch_2647
- vi <http://www.aspeninstitute.org/sites/default/files/content/docs/pubs/InfoTech09.pdf>
- vii <http://radar.oreilly.com/2008/11/the-commoditization-of-massive.html>
- viii http://blog.tonybain.com/tony_bain/2010/01/what-is-big-data.html
- ix <http://smartdatacollective.com/Home/25357>
- x <http://www.prospectmagazine.co.uk/2010/01/mash-the-state/>
- xi <http://www.opendefinition.org/>
- xii
- http://www.soros.org/initiatives/information/focus/communication/articles_publications/publications/open-data-study-20100519
- xiii <http://www.cabinetoffice.gov.uk/media/407789/building-big-society.pdf>
- xiv http://data.gov.uk/wiki/Public_Data_Principles
- xv http://en.wikipedia.org/wiki/Mashup_%28web_application_hybrid%29
- xvi <http://showusabetterway.com/>
- xvii <http://www.scribd.com/doc/31178075/Disrupting-Philanthropy-FINAL> (p.42)
- xviii
- http://www.economist.com/research/articlesBySubject/displaystory.cfm?subjectid=2402943&story_id=15469415
- xix <http://www.guardian.co.uk/politics/2010/jun/04/coins-treasury-public-sector-data>
- xx <http://www.aidtransparency.net>
- xxi <http://mashable.com/2010/05/28/open-data-government/>
- xxii <http://blog.okfn.org/2010/08/02/open-data-and-the-voluntary-sector/>
- xxiii <http://www.slideshare.net/benmatthews/charitycomms-2020-4499458>
- xxiv <http://blog.okfn.org/2010/08/02/open-data-and-the-voluntary-sector/>
- xxv <http://newphilanthropycapital.wordpress.com/2010/05/20/a-to-do-list-for-nick-hurd-part-4-charities-data-act/>
- xxvi <http://www.guardian.co.uk/activate/blog/speaker-interview-rufus-pollock>
- xxvii <http://www.guardian.co.uk/technology/2010/jun/29/business-data-sharing-unwilling>

xxviii <http://www.web-strategist.com/blog/2010/06/26/matrix-brand-monitoring-social-analytics-social-insights/>

xxix <http://www.smartmoney.com/personal-finance/estate-planning/are-charity-fundraisers-spying-on-you/>

xxx <http://smartdatacollective.com/Home/25357>

xxxi Wired Magazine, July 2009 “Free Choice? Don’t be so naïve”, p.99

xxxii <http://www.bynamite.com/>

xxxiii http://www.nytimes.com/2010/07/18/business/18unboxed.html?_r=2&scp=1&sq=data&st=cse

xxxiv http://www.economist.com/specialreports/displaystory.cfm?story_id=15557443