

Open Data Islands and Communities

<http://tireetechwave.org/projects/open-data-islands-and-communities/>

Alan Dix, 20th June 2014

Executive Summary

Open Data has primarily been seen in terms of large entities and 'big' data: Open Government, Open Cities. However, islands and other isolated communities have a lot to gain and a lot to give from an Open Data perspective. It offers the potential for them to transform themselves from data subjects to data producers and data users.

There are four main flows of data, each with their own benefits:

- From the community to the world – giving the community control over their own data and giving external agencies and individuals access to data at a level of detail impractical by top-down means.
- From the world to the community – curating sources of external data that may be of use for within-community planning and externally facing negotiations.
- Within the community – enabling better preservation and sharing of local data and enhancing communication between disparate organisations.
- Between communities – sharing practical experience with similar communities, enhancing business through shared data publication, and enabling scholars access to unique information.

However, there are practical barriers in terms of expertise and investment, and theoretical and technical challenges in designing for true end-user use and dealing with large volumes of heterogeneous small data.

Due to limited financial and human resources in any specific community progress is likely to proceed through small pilots combined with managed programmes to create generic tools and methods.

Note: This document draws examples from a specific island community, Tiree, one of the Hebridean Isles. However, many of the issues will be common or at least similar across all islands and small communities, both rural and urban.

Background

Open Data

Various governments and governmental agencies have embraced, more or less fervently, the notion of Open Data, for its benefits in enabling commercial and non-commercial innovation and acting as a democratising force. The focus is primarily on making data collected and 'owned' by the government available to citizens and organisations.

Cities and more recently smaller authorities have taken up the same notion. As well as publishing their own data, these entities also often collate aspects of national data sets, for example, bus timetables, pertinent to their area.

Data is increasingly important in all aspects of life, exemplified in clearest quintessence by the way 'free' services (e.g. Google, Facebook) are funded by the data gathered from use. In the 21st century data is both profit and power.

Open Data Islands

Just as cities benefit from Open Data, can small communities on islands, rural villages, or city neighbourhoods also gain value from, in some way, collating, curating and using its own data?

There have been positive examples of leveraging Open Data for the benefit of poorer countries and communities across the world [e.g. Fi13, CfA]. Rural areas in the UK, while not suffering the same deprivation as many areas of the world, are still often relatively poor and underserved, often having high indices of multiple deprivation, for example, the median wage on Tiree is around £18,000 less than 75% of the UK median. While Open Data is touted as a democratising force, the reality is that it is those with the technical expertise and financial means who are likely to be able to use it. That is, rather like laissez faire economics, Open Data has the potential to deepen the existing digital divide.

So, while there are many barriers, the arguments for empowering local communities through data are at least as strong as those for larger units. Not least of these barriers is scarcity of expertise and investment, suggesting a real need for reusable and appropriately designed infrastructure to support small-scale data management. However, the potential benefits are substantial, as the Royal Society of Edinburgh report 'Spreading the benefits of digital participation' says, "*Digital technologies provide new channels for communities to connect, cohere and project themselves to the wider world*" [RSE13].

Data Flows

There are multiple forms of data related to a small community, both in terms of source and use. Figure 1 summarises some of these.

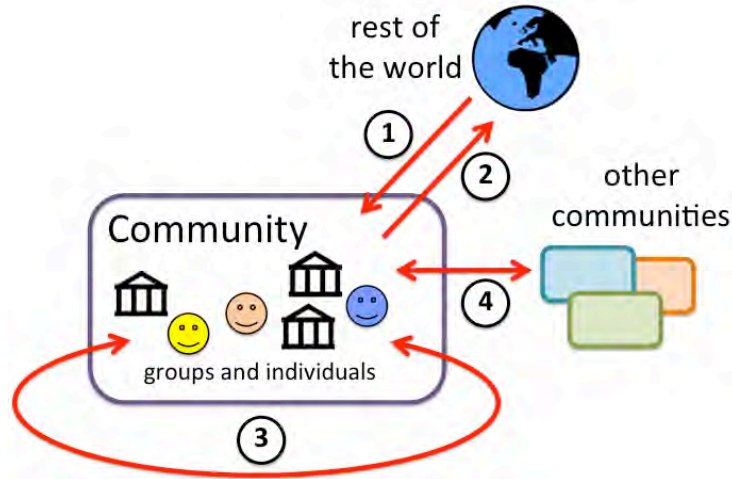


Figure 1. Data flows around a small community

1. From the community to the world

Data collected by the community is made available to the world.

This has benefits for the community in terms of visibility and control.

Visibility is important as, in a data-centric world, if you are not part of the data you are invisible and ignored.

Control is important as those at the margins, whether geographic or social, are often simply data subjects, the raw basis on which numbers are collated. Furthermore, small groups are often amalgamated into larger units where individual needs and circumstances are lost in the mean. For example, Tiree's electric meter usage data is amalgamated with neighbouring Coll and the much larger area of Mull, so that particular patterns, such as the larger than average use of storage heaters on Tiree, is lost.

Giving communities the ability to publish their own data means that they are able to make their own decisions about what is said about them, resulting in greater sense of empowerment and confidence, as Monmouthpedia's blog says, it fosters "*community identity and a powerful sense of achievement, and ownership*". Another example of this is the Google indigenous mapping project [Go14]; although this itself is potentially problematic – while giving indigenous communities a chance to influence mapping, this is done in the context of Google's own mapping engine and global infrastructure [Ge14].

Community generated data is also valuable to the world.

It may involve levels of detail that are impossible or impractical to collate in other ways. For example, in Tiree the 2011 census revealed that the population had declined by more than 10% since 2001. This has prompted action by Argyll and Bute Council (the next higher level of local government). However, close proxies to that data will have been available previously in GP

records, and postal service general knowledge, potentially leading to earlier recognition and action.

It may also involve local knowledge. For example, small fishing vessels have designs that have evolved to the particular circumstances of the area: raw materials, climate, topography. On Tiree a census of lug boats (in various levels of repair or decay) was made in 2004, including photographs and some historic details [TMT04], but this is largely in paper archives, and so not available to a craft historian wishing to compare patterns across the British Isles. In contrast when An Iodhlann, the local historic centre, obtained Nesta funding for a mobile app, Frasan [AI14, NF14], this was deliberately built based on a data-centric infrastructure, so that the geocoded data about artefacts and Gaelic placenames is available to external researchers. The value of this kind of local knowledge has been recognised for medical treatments in drugs research (even penicillin was used extensively by farriers in horse treatment long before Fleming's discovery). Although traditional medical knowledge may now have been exhausted in the UK, much cultural knowledge: stories, songs, crafts, language may be lost or at best locked in locally stored hard copy or old media inaccessible to the external scholar.

2. From the world to the community

Various government and non-governmental bodies have extensive data that may be pertinent to individual communities. There are obvious examples such as census data, meteorological data, or the mapping available from OS OpenData [OS14]. However, even when available in the public domain, sources may be complex to access even when you know whether they exist at all. For example, it is possible to obtain energy usage data based on geographic areas, but the codes for this need looking up in another standard.

This data is important to inform local decision making for both community organisations and businesses. It is also particularly important in interactions with external bodies such as grant funding applications, lobbying or negotiations. For example, when there is a problem with the electricity cable connecting Tiree to the mainland and the island is switched to the local diesel backup generators, Tilly, the community wind-turbine, has to cut back its production to tick-over mode. The island group responsible for this have had no ability to negotiate whether this is entirely necessary, as SSE simply say it must happen and there is no available information to challenge this fiat.

3. Within the community

The popular notion of a small rural community is one where gossip means that everyone knows everything about everyone else. There are elements of truth in this, the most personal details rapidly disseminate. However, despite this, the geographically distributed nature of the community means that it is very hard to let people know about an event that is on this evening. This problem is, if anything, worse in urban communities where talking to a next-door neighbour may be rare. Sharing within a community can increase cohesion and may bring immediate economic benefits. For example, some blocks of flats in deprived areas have set up shared buying schemes. Sadly, this kind of

example is rare, although mail lists such as Freecycle, Craigslist and Facebook share some of this ethos.

There is also a close link between sharing and preservation. Local information is easily lost or at best in individual memory. For example, on Tiree a few years ago there was a funded project Powerdown, which made a census of every nearly house on Tiree looking at energy use and efficiency. When OnSupply, a recent renewable energy focused project, tried to find this data, only crude aggregates were available in the final report. In the end some data was found in an old spreadsheet on a laptop that had been passed from hand to hand, but the very useful rich data had been lost, not deliberately destroyed for data protection or other purposes, but simply lost.

Local data is a community asset; preserving and sharing this could ultimately strengthen the community.

4. Between communities

Just as groups within a community gain from knowledge sharing, other communities in similar circumstances could similarly benefit. For islands, even those physically close may be distant in terms of transport links. For example Eigg is approximately 30 miles from Tiree across the sea, but a 2 day journey as ferry links go to and from the mainland; in the winter with less frequent ferries, this would be a one week trip. Even on the mainland public transport infrastructure is often hub-and-spoke, with physically close communities transport distant. However, the communities that are most closely linked in terms of circumstances may well be in completely different parts of the country; this is equally true for urban communities, although the transport barriers are likely to be less significant.

For geographically close communities there can be distinct advantages to creating shared external identity. This is probably more an issue of common web and paper media presence, for example a regional food or craft site, but shared data could undergird this, especially if this is to include online sales

Some of the community PR benefits of opening up information to the world accrue most when it is alone or rare; for example, Monmouthpedia bills itself "*the world's first Wikipedia town*". However, the main benefits for society as a whole arise when multiple communities make data available that together creates larger-scale datasets, which, while traceable and controllable by individual communities, can be used for cross-community purposes. This might include government planning, regional tourism, and cultural or historic research. The example earlier of varying patterns of local boat building is just such an example, where tracking patterns across communities could offer new insights to the cultural historian. As an example of such linking, the Curious project at the dot.rural research centre in Aberdeen has connected multiple archives across a small area in Lewis in the Outer Hebrides [Cu14].



Preserving lug boat construction

Barriers and Challenges

Expertise and resources

Finding the right expertise and funding Open Data projects is problematic enough at a regional or city level. When commenting on Australian national public sector use of Open Data, Ian Bartram global manager for analytics at Gartner said,

"I don't know if any public sector has necessarily cracked the nut on attracting the right skills and capabilities," ... "The commercial sector has, because they've got the dollars to spend." [Ho14]

This gets more problematic as the size of the unit decreases. This may often require external funding and programmes to pump-start, for example, Nesta's Open Data Scotland at local government level [NO14], which has funded technology fellowships as part of the broader Code for Europe programme [CfE].

At the level of islands, and isolated rural communities, these issues of expertise and funding are even more pronounced. The 'activists' are often retirees who may not be IT literate. In addition the number of different activities does not scale down with community size, so those who are active are often involved in several organisations. People are often willing to stop to talk and give time freely so that, at first, there appears to be a slow pace of life. However, on closer examination time pressure is a frequent problem: those of working age typically hold down several part-time jobs; those who are retired are involved in multiple volunteer activities. Ironically some of these activities may involve re-accessing, or failing to access, the same 'public' data.

The issues for urban communities are different. Time is perhaps less critical where under-employment is a major problem; but information literacy is likely to be a substantial barrier.

However, all these barriers lead to R&D challenges for Open Data.

Reusable data flows

One of the aims of Code for Europe is to "*create solutions that are easily reusable in other European cities*". In a similar way small communities need re-usable local-focused systems. For cities the problems are likely to be about how to create meaningful applications that may be published, for example, maps of crime reports or car sharing. For small communities there will be challenges about presenting data, but also in gathering it. Meta data repositories are needed that tell communities where they can go to get data about themselves (e.g. census data, historic weather reports, electricity consumption) – even knowing whether appropriate data exists is problematic. This needs to be complemented with information on how to access the data in those repositories, which may need knowledge of government coding schemes, descriptions of data fields, or even instructions on how to perform appropriate Freedom of Information requests to yield useful information that is not published, but available if requested.

Interaction design

The challenges for data presentation to the public are not dissimilar to those in cities except that in rural areas mobile and Internet connectivity is likely to be severely limited. Current UK Government policy is to discourage standalone apps, and instead use responsive web design [Gov14].

Unfortunately this is not backed by any suggested use of HTML5 offline mode, effectively promoting an information availability policy for (certain) urban areas only. Happily policies for fixed broadband infrastructure, such as the roll out of fibre to rural areas in Scotland [DS14], is improving domestic and business connectivity. However, this will always lag behind major urban centres, so there is a continuing need for design solutions that work well or at worst degrade gracefully under conditions of limited connectivity.

For data capture and use by community organisations the differences are more marked. It is not possible to rely on having IT support teams to help create reports to interpret data, and even less assume that community groups will learn RDF and place data into triple stores. For example, a group on Tiree is looking at potential for a community land buy-out, the early stages of which involves detailed mapping of croft boundaries. This has led to a suggestion of an island GIS strategy, but the complexity of this has so far proved off-putting.

Community data sharing will often need to start with familiar data management tools such as spreadsheets, or bespoke systems designed for their context. For example, archive data for Frasan, the Tiree heritage app, was geocoded using Google Earth and Google Maps, and then imported into a data store from that; other data is managed using WordPress with bespoke plugins. Solutions such as this can be generic, so can benefit from pump-priming funding to defray costs over many communities, but do need to be designed in ways that are sensitive to the technological, economic and social circumstances of remote areas.

Heterogeneous multisource data

If a data strategy for communities is to be based on representations and data that are meaningful to individuals, it will almost certainly not fully conform to redefined schemas. Platforms need to adapt to people and communities. This is not to say there are not common features in, say, heritage data, but each community will have specific aspects that are different. Whereas big data is often relatively homogeneous, for bottom-up data we need to be able to cope with large numbers of smaller datasets, which may vary in raw format and semantic details, the 'long-tail of small data' [Dx11]. In fact, a very similar requirement is found in other areas, including scientific data [Bo13, PC07]. In both the community and scientific contexts provenance is critical, so that elements of large agglomerated data sets can be traced back to the community where they originated

Conclusions

Open Data for islands and communities offers benefits to:

- The community in terms of empowerment and control, availability of information about the community, and enhanced communication within and between communities.
- The world in creating sources at a level of detail impossible to create using a top-down approach, and incorporating local knowledge and understanding.

However, there are challenges in interaction design and data management.

The limited resources of individual communities means that progress will almost certainly require external programmes of support. In keeping with the bottom up spirit, the most effective R&D strategy is likely to involve pilot Open Data communities and then generalisation to create tools and methods to help all communities curate, publish and use their own data.



<http://creates.stir.ac.uk/2013/03/17/launch-of-frasan-tiree-mobile-archive-app/>

<http://tireeliving.blogspot.co.uk/2013/03/tiree-tech-wave-5.html>

References

- [AI14] **Frasan - Tiree Mobile Archive**, An Iodhlann. <http://frasan.org.uk/>
- [Bo13] Christine L. Borgman. **Big data and the long tail: Use and reuse of little data**. Oxford eResearch Centre Seminar, 12th March 2013. <http://works.bepress.com/borgman/269/>
- CfA] **Code for Africa**. <http://www.codeforafrica.org>
- [CfE] **Code for Europe**. <http://codeforeurope.net>
- [Cu14] **The CURIOS Project**. dot.rural, Universty of Aberdeen. <http://www.dotrural.ac.uk/curios/>
- [DS14] **Digital Scotland: Superfast Broadband** <http://www.digitalscotland.org/superfast-broadband/>
- [Dx11] Alan Dix. **In praise of inconsistency - the long tail of small data**. Distinguished Alumnus Seminar, University of York, UK, 26th October 2011. <http://alandix.com/academic/talks/York-Alumnus-2011-inconsistency/>
- [Fr13] Niall Firth. **How open data empowers citizens of poorer nations**. New Scientist, 3rd Nov. 2013. <http://www.newscientist.com/article/mg22029434.400-how-open-data-empowers-citizens-of-poorer-nations.html>
- [Ge14] Grace-Yvette Gemmell (2014). **Should Google Be Mapping Tribal Lands?** Daily Beast, TECH + HEALTH, 4th June 2014. <http://www.thedailybeast.com/articles/2014/06/04/should-google-be-mapping-tribal-lands.html>
- [Go14] **Indigenous Mapping Day is August 9, 2013. Show your support & improve Google Maps!** Google, Map your World Community. <https://sites.google.com/site/mapyourworldcommunity/indigenous-mapping>
- [Gov14] **Standalone mobile apps: Government's position on mobile apps**. Gov.UK. <https://www.gov.uk/service-manual/making-software/standalone-apps.html>
- [Ho14] Brad Howarth. **Big data: how predictive analytics is taking over the public sector**. Guardian, 13th June 2014. <http://www.theguardian.com/technology/2014/jun/13/big-data-how-predictive-analytics-is-taking-over-the-public-sector>
- [NO14] **Open Data Scotland**. Nesta. <http://www.nesta.org.uk/project/open-data-scotland>
- [NF14] **Frasan: An Iodhlann**. Nesta, Digital R&D Fund for Arts and Culture in Scotland. <http://www.nesta.org.uk/project/digital-rd-fund-arts-scotland/frasan>
- [OS14] **OS OpenData**. Ordnance Survey, 2014. <http://www.ordnancesurvey.co.uk/business-and-government/products/opendata-products.html>
- [PC07] C. L. Palmer, M. H. Cragin, P. B. Heidorn and L.C. Smith. 2007. **Data curation for the long tail of science: The Case of environmental sciences**. 3rd International Digital Curation Conference, Washington, DC. https://apps.lis.illinois.edu/wiki/download/attachments/32666/Palmer_DCC2007.pdf
- [RSE13] **Spreading the benefits of digital participation**. Edinburgh, Royal Society of Edinburgh, 2013. http://www.royalsoced.org.uk/1136_FinalReport.html
- [TMT04] **Boat Survey 2004**. Tiree Maritime Trust. <http://www.tireemaritimetrust.org.uk/boat-survey-2004/>