Think beyond the data

There are several drivers may have brought you to the point of openly publishing data.

- To put a tick in a box.
- It is public data and the public should have ready access
- Cost reduction (eg pre-empting the cost of responding to FOI requests)
- Increasing the value of your data through better data sharing
- Internal processes improvement
- External collaboration and reuse
- To have an impact in the world - EA's strap line is "Making a better place..."
- Make more (and unexpected) things possible.

Whatever the driver is that gets you to this point, given the effort and expense to make the data available it is likely that you want people to use make use of it. If it languishes unnoticed and unused then at best it represents a missed opportunity.

Dataset metadata records and mechanisms for harvesting that into catalogs that can be searched will help get your data discovered. But if you are enthusiastic about the data you are publishing and want to see it used as widely as possible you need to think beyond simply surfacing the data.

Throwing CSV files over the wall and expecting good things to just happen doesn't work. Even creating a beautiful 5* data publication will only get you so far. You need to think about the likely uses to which it might be put (by others) and to how you are going to make it easy for them your data.

We've been working with the Environment Agency to help them publish information about bathing water quality as linked-data. We won't dwell on the linked-data aspects here. The thought applies whatever your publishing method.

One of the things we did encounter was an awareness of the boundaries of a department or agencies public sector remit. It is understandably out of scope for the Environment Agency to gather and mash their data with other local amenity data (car-parking, accommodation, toilets, other faclities etc.) much as the more enthusiastic might wish to do so, but openly publishing the data enables others.

Provide at least one 'iconic' Application that uses your data.

You may be looking to others to re-use/repackage your data in their applications. However, part of the discovery process for them is finding your data and seeing what might be possible with it. You need to provide them with some inspiration. Frankly, tables of data can be a bit 'dull'. They are not going to excite your management - they want to see something visual. You will need some kind of basic explorer application that helps people come to grips with your data. As well as serving to 'socialise' your data, in the best 'copy-paste' traditions of the web - it will also serve as an exemplar where folks to fire up their debugging environment and see what's going on under the hood.

Building an a data explorer will also help you understand what access functionality needs to be available to likely users of the data.

In our work with the Environment Agency - our initial 'slippy' map based demonstrator served to sell the potential of the data publication internally. It has since been enhanced to become a fully blown explorer for the both bathing water profile and bathing water quality data.

The profile pages that it produces are set to replace what was previously a PDF based publication and as a web publication - they present more up to the minute (well ok... up to the most recent weekly) bathing water assessments juxtaposed with the more static profile information.

Provide documentation (Data model and Interface)

Something inspiring will get people interested in engaging with your data. But if you really want them to take it and use it on their own, you'll need to provide them with documentation. They'll need to understand the data models and they'll need to understand the access mechanisms.

Formats

I said I wasn't going to dwell on Linked-Data, whose basic principle is "name things with URIs" (all manner of things). But I do need to acknowledge that native linked data formats (RDF) is something of a marmite experience and more often than not provokes something of an allergic reaction from more 'mainstream' web application developers - many of whom will clamour 'just give us the CSV'. The best thing to do here is acknowledge they have a point and provide data in a format (or formats) that are more readily accessible to the developer - an provide them with interfaces that don't involve them having to learn arcane query languages in order to ask questions of the data.

We use an artefact called the "Linked Data API" which provides a configurable URI driven RESTFul interface onto the data we publish and will render the data in 'object' centric (rather than RDF triple centic) JSON and XML formats as well as native RDF formats, CSV and HTML. We programme the applications we develop to the JSON formats which are much more convenient to use than 'native' RDF or query language results sets.

The message here is simple - provide data in formats that developers can readily make use of.

Think about simple uses

Not every one has access to tech savvy web developers. From a reuse point of view the EA bathing water data is likely to be of most interest what we have called 'Beach front local authorities'. Even when you have lowered the bar with respect to RESTFul interfaces and developer friendly data formats - they're unlikely to be able invest significant development effort to incorporate your data on their site. But you can lower the bar even further and potentially increase the reach and utilisation of your data by providing an embeddable widget that presents your data on their pages. Your branding can come through communicating the authority of the data source - and it's no more difficult for the consuming web site than the inclusion of some static markup on their pages.

Summary

If you are genuinely enthusiastic about the open publication of your data... you have to go much further than merely surfacing your data on the web and relying people just 'finding' it, somehow. You need to think about engaging interest in the data and inspiring potential downstream consumers/reusers/value-adders. They need to hear about your data and they need to be able to explore it. You need to provide technical information and potentially a place where questions can be ask AND answered; you'll need to invest in documentation; you'll need to lower the barriers to use by providing developer accessible formats and web interfaces; and you may need to address the needs of your least able, but possibly dominant, reuser in the form of a reusable widget that can be embedded on any web page.

If you are successful, you will be rewarded with increasing usage stats that will reinforce that your data is important/useful to someone and help build confidence to contribute more to the open data publish endeavor.

Stuart Williams
skw@epimorphics.com
@skwlilac