Benefits of an Open Data Service at the University of Southampton

Christopher Gutteridge, cjg@ecs.soton.ac.uk

The University of Southampton has run an open data service for two years. This service has focused on data about the university, rather than datasets created by researchers which is very important but requires a different approach.

Our open data service in a nutshell

- Identify a potential dataset
- Approve the data for being appropriate (Data Protection, Risk management etc)
- Create a workflow which publishes this dataset to the web in a format a computer can reuse
- Where possible, align the codes in the data with other datasets so they can be joined up and add value to each other
- Republish the dataset at an appropriate frequency
- Build a mini website, map or app using the dataset to provide value to the organisation and specifically the data provider.

The last step has proved important as it makes it possible for people to spot errors in our datasets without delving into the raw data

Getting more out of existing assets

Most of our datasets are taken from existing databases or spreadsheets so do not generate a continued maintenance cost. In some cases we've worked with the data owner to modify their working practice slightly.

When we worked with catering, they already had a spreadsheet with opening hours for all their outlets, but it was tricky to work with. We converted all the times to 24 hour and added a column to unambiguously link each point-of-service with the canonical building code, allowing us to include the information on pages about the buildings and maps. They now keep the same information in their spreadsheet but preserve our altered format. This means there is no additional work for them, and we can use the information as open data.

By publishing datasets of key structural components of your organisation, it makes it easy for other projects to use the canonical identifiers for these things, and we are starting to see the benefits of data from diverse parts of the organisation being joined up.

Find me a free computer workstation

We publish a live list of all the locations of public computer workstations, plus how many are currently free in each room. We then used this data to build a live map, which students can use to find a free machine if the room they usually use is full or there is a class being taught.
http://data.southampton.ac.uk/workstations/
This caused us to get a wonderful, informal complaint from a student that their favorite workstation room used to be quiet because nobody knew about it, but now is used as much as all the others. We've also added historical data so students can see when a computer room is busy/quiet and plan their work accordingly.

We've recently added a similar service to find unbooked teaching rooms and lecture theatres. These two datasets + services help us provide a better service to our students and get more use out of existing resources.

Where to start?

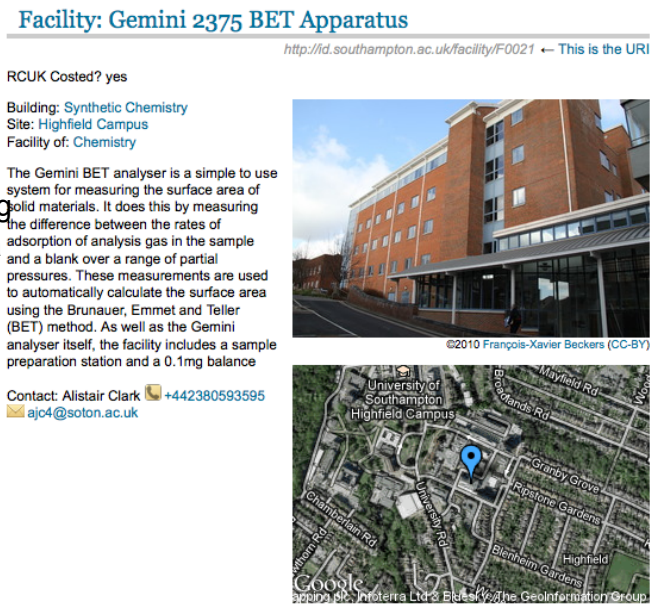
What is practical will vary between organisations. Suggested places to start:

- A profile document -- this is just like a machine readable version of the university homepage, it says things like what your logo is, the university phone number, and provides links to other datasets and data-catalogues.
- A list of buildings with the code for that building, it's name, and it's approximate latitude and longitude. Given this data you can produce apps to help people find buildings, and any dataset which links things to a building ID can now be plotted on a map.
- A list of the structure of your organisation, departments, faculties etc.
- A list of major research equipment and/or facilities
- A list of research publications (if you have a repository such as EPrints or D-Space then these are already providing open data)
- News (if you have an RSS feed of your news, then that's already a very light-weight open data set)

While some datasets may be more controversial, free reuse of the above datasets by third parties, including your own staff and students, can only provide benefit.

Linked data

The buildings & organisation-structure datasets have proved to be the “hub” datasets. Many other datasets link to one or both of them. A simple example is that some of our research facilities are linked to the building ID. This simple datum allows us to automatically add a map, and photograph of a building to the open-data page about the facility: http://data.southampton.ac.uk/facility/F0041.html and also to list the facility on the page about the building: http://data.southampton.ac.uk/building/46.html
Not all facilities managers wanted their exact building listed, in which case they are just linked to a campus, or not linked at all.



Not all the links are to internally produced data, there’s also links in the data from our courses to the students survey results, and all our catering establishments link to the government open data about their current food hygiene rating.

Research collaboration

It’s a sad fact that often researchers don’t have much idea about what goes on in other buildings or what resources and facilities they might have. Making it easy to browse and explore the information on our facilities, equipment and publications improves the chances of novel collaboration both internally and with other organisations.

Many eyes

Most of the data we publish comes from existing official databases which staff are expected to keep up to date. The reality is that this process is generally less than perfect. Once the data becomes visible on public facing apps and websites people are far more likely to spot errors and make/request corrections.

The downside of this is that our public data will have a period while it is imperfect. This is outweighed by the permanent improvements that come from opening it.

We don’t allow ‘cosmetic’ changes in the open data service; we just publish the canonical data from the data owner. This has allowed staff at the satellite campus in Winchester to discover that the estates department was using entirely different names for their buildings than the staff in Winchester, and we now have the opportunity to fix this properly and improve communication.

Amenities map

We have not had as many students write apps as we might have hoped. However the few that have been produced are extremely useful. The most notable are Southampton bus-timetable apps for each of the major smart phones, and a web-based amenities map. The amenities map combines many different datasets in a friendly map showing opening hours, bus times, free workstations and even upcoming events by building.
http://opendatamap.ecs.soton.ac.uk/

Student experience

..and staff and visitors

The services we've built on top of the open data have been very useful to the student experience and induction. It helps people get to grips with the organisation and find services and resources they might not otherwise be aware of. The open data service is especially value in induction and even includes a dataset of university jargon and acronyms.
Example, building finder: http://map.southampton.ac.uk/

When I visit another university, my key questions are where am I going? how do I get there? and where can I get a coffee enroute? While a little frivolous compared to transparency and opening up research data, I'm proud to say we can answer those questions and improve the experience for visitors!

Events calendar

We are a big organisation, and while we are significantly more centralised than many universities, it still proved impractical for the central Communications department to keep tabs on all the events which were going on.

We worked with them to produce a system which collates events listings from around 75 sources around our organisation, and augments them by linking to the datasets on campuses, buildings and the organisation structure.

http://events.soton.ac.uk/

The software which powers this has been made free & open source (FOSS), and is available for use at other organisations.

Why make it open?

Allowing the open reuse of our datasets has been of primary use to the web programmers in the IT department innovation team. We have been able to build tools quickly using the data, confident that it is approved for reuse. This can reduce the turn-around on a ‘quick-win’ idea from months to hours!

Some of our datasets are aggregated by external services. Our publications metadata has been available as open data for many years, and services which aggregate this form of data are well established.

More recently, we have been funded by EPSRC to provide http://equipment.data.ac.uk/ as a national equipment database. Open data makes this eminently sustainable as it eliminates the need to create a human-curated communications channel between the hub and 180 universities. Right now only Loughborough, Leeds and Southampton are providing data but we plan to add more organisations in the coming months. The open data also provides an excellent form of probity, as if there’s any doubt in the source of the data, anybody can verify the data at source on the institutions own website.

The service also provides useful data used in student projects. Making it openly reusable allows students to get started quickly without going through a complex process to gain access to data which was never confidential.

As more organisations provide open data, it is certain that standard methods and models will

arise for common datasets. This is exciting as it can mean that in future a tool developed for use at one organisation may be used with data from another. This can only serve to improve choice and quality. We already work closely with data.ox.ac.uk to ensure we don’t needlessly diverge.