Open Data

Christopher Gutteridge
Christopher Gutteridge
@cgutteridge
chris@data.ac.uk

• data.southampton.ac.uk
• data.ac.uk
• data.totl.net
“The opposite of ‘open’ isn’t ‘closed’. The opposite of ‘open’ is ‘broken’.”

- John Wilbanks

... do you agree?
Some data should be private.

Open Data & Privacy go hand in hand.
Open Data?

Remove barriers to use and reuse.
Open Data?

Remove barriers to use and reuse.

• Social
• Legal
• Awareness / Discoverability
• Compatability
• Technical
• Trust
• Convenience
• Documentation
• ...
Types of Data

Factual

Statistical
Types of Data

Linked

Open

RDF
Types of Data

- Linked
- Big
- RDF
- Open
Types of Data

- Big
- Linked
- Open
- RDF
Reasons to Openly Publish

• Improve your reputation
• Because somebody makes you
• Get some sucker to write your apps for free
(equipment demo)
Reasons to Openly Publish

• Improve your reputation
• Because somebody makes you
• Get some sucker to write your apps for free
Reasons to Openly Publish

• Improve your reputation
  – all the cool kids are doing it
• Because somebody makes you
  – the ‘stick’ method
• Get some sucker to write your apps for free
  – only works if it’s in the public good
  – otherwise it’s hacksploitation!
• Because it actually achieves your personal (or business) goals
  – the ‘carrot’ method
Useful aggregations for Equipment & Facilities data

• UK Academia
• South-coast public sector
• Universities in N.Ireland + R. of Ireland
• Hampshire Businesses
Other Useful “standard” Datasets (for a university to share)

• Staff Expertise
• Vacancies
• Organisational Structure
• Buildings & Campuses
• Publications
• ... research data (a can of worms!)
Publishing models & formats
Publishing Organisation Datasets

Well known formats available for:

- Events
- Publications
- News headlines

Nothing in common use for:

- Staff Expertise
- Programmes of Events
- Vacancies
- Organisational Structure
- Buildings, Rooms
- Points of service
- Products
  - Food Menus
The Modeller

"If you're not part of the model, you're part of the problem."
RDF or XML Vocabularies don’t solve the problem by themselves.

You need:

Examples to copy.

Tools which consume and produce the format.

Online checking tools.

"If you’re not part of the model, you’re part of the problem."
A dataset should at least solve one use case.

Over modelling is fun.

Stop it.
Autodiscovery
./well-known/openorg

http://www.soton.ac.uk/.well-known/openorg

or

<link rel="openorg"
    href="http://id.southampton.ac.uk/dataset/profile/latest">
Linked Data Augmentation

• Linked datasets augment each other
• Key datasets for augmenting other data at the uni.
  – People
  – Products
  – Services
  – Places
  – Courses
  – Parts of the organisation
  – External authorities
<table>
<thead>
<tr>
<th>Product</th>
<th>Coverage</th>
<th>Download</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIDAR Composite Digital Terrain Model (DTM)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DTM at 2m spatial resolution</td>
<td></td>
<td>(40.5 MB)</td>
</tr>
<tr>
<td>DTM at 1m spatial resolution</td>
<td></td>
<td>(153.6 MB)</td>
</tr>
<tr>
<td>DTM at 50cm spatial resolution</td>
<td></td>
<td>(477.9 MB)</td>
</tr>
</tbody>
</table>
$ ./generate-world Demo
   --postcode PO381NL
   --size 250
$ ./generate-world Demo
   --postcode PO381NL
   --size 250
Publishing Open Data
Tabular Data For the Humans

• Understanding “Graph” data is a big cognitive load, compared with SQL/Excel.
• Don’t expect any small company to be able to generate good RDF.
• RDF is neat for aggregation, but make easy ways for people to get started.

• Provide tabular-data routes in & out where possible!
JSON for Developers

• SPARQL & RDF is too high a learning curve
• JSON works really well in Javascript
• Gives more structure than CSV
Open Data Publishing

Data Document + metadata → Conversion Tools → RDF → Dataset Repository
Open Data Publishing

Data Document + metadata

Conversion Tools

JSON

CSV

RDF

SPARQL Store

Dataset Repository
Open Data Publishing

Data Document + metadata

Conversion Tools

JSON

CSV

RDF

Dataset Repository

SPARQL Store

Per-Item Pages

List of Items Pages

Tools & Apps
Open Data Publishing

Data Document + metadata

Conversion Tools

JSON

CSV

RDF

Dataset Repository

SPARQL Store

URI Resolver

Per-Item Pages HTML, JSON, RDF

List of Items Pages

Tools & Apps
Open Data Publishing

Data Document + metadata → Conversion Tools

Conversion Tools → JSON → Dataset Repository
Conversion Tools → CSV → Dataset Repository
Conversion Tools → RDF

RDF → SPARQL Store

SPARQL Store → URI Resolver

URI Resolver → Per-Item Pages HTML, JSON, RDF
URI Resolver → List of Items Pages
URI Resolver → Tools & Apps

URI Resolver → Map Tile Server

Map Tile Server → Maps
Why aren’t they using our data?
“If you build it, they will come.”
"If you build it, they will come.

Nothing but lies!"
Probability of open dataset reuse =

Value of dataset to audience

×

Potential audience size

×

Ease of discovery

×

Ease of grasping the value of the dataset

×

Ease of exploiting dataset
Find yourself some open data

• data.gov.uk
  – UK Government open data
• www.hampshirehub.net
  – Hampshire council open data
• data.southampton.ac.uk
  – Our open data
• datahub.io
  – Open data catalogue
• http://sotoncitydata.org/
  – List of open data about Southampton (Needs work!)