Data.gov.uk - What's Not to Like

Story
Spotfire Dashboard
Sitemap of Data.gov.uk Web Site
Who does what in Whitehall and Beyond?
How your money is spent
Government contracts in full
Who ministers are meeting
Business plans
Applications
Data
Apps
Event
Central Government’s potential contracting opportunities
Forum
Discussion forum
Blogs
The blogs (7)
Featured blogs
Have you used our data to create an application?
Have you got a good idea for the data?
Do you want more data?
Ideas
Linked Data
Browse
Search
Ordnance Survey Ordnance Survey have released a number of their products as linked data, including postcode units and administrative areas. Browse
Datasets
Search
Transport data covers the transport infrastructure and data about traffic flow. Note: the data provided here dates from March 2010. Browse

Legislation data is provided by legislation.gov.uk. Browse

Resources
The power of government data
What is Linked Data?
Using Linked Data
SPARQL tutorials
History of project
What is the Semantic Web?

About
What's data.gov.uk all about?
What's available?
Is data just public information?
How can I use the data?
How can I add my organisation's data?
How can I find out more?

Terms and conditions
Code of conduct
Project Info
1. What is data.gov.uk?
2. What is our approach?
3. How can I get involved?
4. More questions?
5. How can I contact the project?
6. Who is involved in the project?
7. Awards

FAQ
Q. What is the project doing?
Q. What is the status of this site?
Q. Will personal information be at risk?

Updated: Sat, 19 Sep 2015 03:02:35 GMT
Powered by mindtouch™
Q. What is a mashup?
Q. What is Linked Data and how does it fit into the Semantic Web?
Q. How can people submit applications and visualisations?
Q. How can people submit ideas
Q. How can people use data.gov.uk?
Q. How were the datasets in data.gov.uk selected?
Q. Why is a particular dataset not available yet, and when will it be?
Q. Why is a particular dataset not available through an API yet, and when will it be?
Q. How can people make contact with the project?
Q. Is there a place for developers to discuss ideas, applications and using of the data?
Q. Will more public data become available in the future?
Q. What are the commercial use rights if people have commercial ideas?
Q. Under what licence is the data available?
Q. Are there any “do”s and “don’t”s?

Contact us
Accessibility
Moderation policy
Privacy
Consultation

Research Notes
Jim Hendler

Data.gov.uk URLs
Making Open Data Real: A Public Consultation
This online consultation has now closed, but you can still review comments within each section of the consultation. We will now review consultation responses and update on the outcomes as soon as they are available.
Performance data on Government ICT projects (31 July 2010)
Forum
Energy Performance Certificate Ratings
XML data & direct download links.

COINS as Linked Data
Introduction
What is COINS?
The COINS data release
Hypercube structure of COINS
For illustration, an example entry from COINS is shown below
COINS as linked data: overview
The data cube vocabulary
COINS graph structure
Data Sharing Between Government Departments and Report on Public Acceptability
Office of National Statistics

Data.gov.uk - What's Not to Like

Source: [http://gov.aol.com/2011/12/05/data-g...s-not-to-like/](http://gov.aol.com/2011/12/05/data-g...s-not-to-like/)

By Brand Niemann

Published: December 5, 2011

This week at the SemTech Biz DC Conference, Jim Hendler, advisor to Data.gov, explained the history of the “friendly competition” between the US data.gov and data.gov.uk and said that the latter had about 6000 data sets that were in better shape than the former. So I decided to take another look and was very impressed.

Hendler also said that the UK Government has designed and made great use of standard Web address practices in their linked data and moved even further ahead of the US in open data with creation of the Open Data Institute.
First I decided to help myself and my readers by building a [sitemap of the data.gov.uk](http://semanticommunity.info/AOL_Government/Data.gov.uk_-_What's_Not_to_Like) site because the topics it includes are just what people have been asking the US data.gov to provide, namely: meetings, business plans, contracts, forums, blogs, etc.

Then I downloaded a few of their key datasets to see how easy it was to work with and understand them. I downloaded their Data Catalog, COINS, and Performance Data and created a dashboard shown elsewhere.

I thought the introductory explanation and [video](http://semanticommunity.info/AOL_Government/Data.gov.uk_-_What's_Not_to_Like) from their Prime Minister David Cameron (above) was especially good at explaining the benefit intended from data.gov.uk.

**What's data.gov.uk all about?**

The Government is releasing public data to help people understand how government works and how policies are made. Some of this data is already available, but [data.gov.uk](http://semanticommunity.info/AOL_Government/Data.gov.uk_-_What's_Not_to_Like) brings it together in one searchable website. Making this data easily available means it will be easier for people to make decisions and suggestions about government policies based on detailed information. Hear more about the Government's Transparency agenda from the Prime Minister

**What's available?**

There are currently over 5,400 datasets available, from all central government departments and a number of other public sector bodies and local authorities.

**Is data just public information?**

Not really. From data.gov.uk, you can access the raw data driving government forward. This can then be used by people to build useful applications that help society, or investigate how effective policy changes have been over time. Previous on Data.gov limitations - done at Data.gov.uk

**What is COINS?**

In June 2010, HM Treasury, the principal custodian of financial data for the UK government, released previously restricted information from its Combined Online Information System (COINS). Until then, access to COINS by journalists and members of the public had been very difficult, and COINS itself had been the subject of various Freedom of Information requests. The release of COINS data therefore represented a step change in the Treasury's willingness to share data openly.

While the release of the COINS data is very significant, it does not by itself make the government's finances scrutable. COINS is a highly technical information source, requiring detailed knowledge of the encoding schemes used to make sense of it. Moreover, the data files are too large to load into widely available analysis tools such as Microsoft Excel, a common tool-of-choice for many data investigators.

Some enterprising teams have stepped up to the challenge of providing better access to the COINS data, notable among them the [Where does my money go?](http://semanticommunity.info/AOL_Government/Data.gov.uk_-_What's_Not_to_Like) project, and the [Guardian datastore](http://semanticommunity.info/AOL_Government/Data.gov.uk_-_What's_Not_to_Like).

We believe, however, that more can be done to open up the COINS dataset for exploration and analysis. [Linked Open Data](http://semanticommunity.info/AOL_Government/Data.gov.uk_-_What's_Not_to_Like) is the name of a distributed, global initiative to provide better means to open up datasets, and, crucially, to publish links and connections between data. Each datum in COINS is identified by seven key indices and the metadata and data dictionary are provide on my [Semantic Community](http://semanticommunity.info/AOL_Government/Data.gov.uk_-_What's_Not_to_Like) site.

So my goal was to integrate all of this (three data sets and their data dictionaries and metadata) into one dashboard. The dashboard I created is shown elsewhere on my [SemanticCommunity](http://semanticommunity.info/AOL_Government/Data.gov.uk_-_What's_Not_to_Like) site and convinced me that it was very easy to access, use, and understand their Data Catalog, COINS, and Performance Data Sets. This is what we need the US data.gov and other administration open government web sites to be.
Story

In a [pending story](http://semanticommunity.info/AOL_Government/Data.gov.uk_-_What's_Not_to_Like), I talked about the difficulty I had in using and understanding a featured data set at data.gov, namely the GSA's Energy Usage Analysis System (EUAS).

This week at the SemTech Biz DC Conference, Jim Hendler, advisor to Data.gov, explained the history of the "friendly competition" between data.gov.us and data.gov.uk and said that the latter had about 6000 data sets that were in better shape than the former so I decided to take another look and was very impressed. Hendler also said that the UK Government has designed and made great use of standard Web address practices in their linked data and moved even further ahead of the US in open data with creation of the Open Data Institute.

First I decided to help myself and my readers by building a sitemap of the data.gov.uk site because the topics it includes are just what people have been asking the US data.gov to provide, namely: meetings, business plans, contracts, forums, blogs, etc. Then I downloaded a few of their key datasets to see how easy it was to work with and understand them. I downloaded their Data Catalog, COINS, and Performance Data and created a dashboard shown elsewhere.

I thought the introductory explanation and video from their Prime Minister was especially good so I included it below.

### What's data.gov.uk all about?

The Government is releasing public data to help people understand how government works and how policies are made. Some of this data is already available, but [data.gov.uk](http://data.gov.uk) brings it together in one searchable website. Making this data easily available means it will be easier for people to make decisions and suggestions about government policies based on detailed information. Hear more about the Government's Transparency agenda from the Prime Minister

Media, iframe, embed and object tags are not supported inside of a PDF.

### What's available?

There are currently over 5,400 datasets available, from all central government departments and a number of other public sector bodies and local authorities.

### Is data just public information?

Not really. From data.gov.uk, you can access the raw data driving government forward. This can then be used by people to build useful applications that help society, or investigate how effective policy changes have been over time. Previous on Data.gov limitations - done at Data.gov.uk

### What is COINS?

In June 2010, HM Treasury, the principal custodian of financial data for the UK government, released previously restricted information from its Combined Online Information System (COINS). Until then, access to COINS by journalists and members of the public had been very difficult, and COINS itself had been the subject of various Freedom of
Information requests. The release of COINS data therefore represented a step change in the Treasury's willingness to share data openly.

While the release of the COINS data is very significant, it does not by itself make the government's finances scrutable. COINS is a highly technical information source, requiring detailed knowledge of the encoding schemes used to make sense of. Moreover, the data files are too large to load into widely available analysis tools such as Microsoft Excel, a common tool-of-choice for many data investigators. Some enterprising teams have stepped up to the challenge of providing better access to the COINS data, notable among them the Where does my money go? project, and the Guardian datastore. We believe, however, that more can be done to open up the COINS dataset for exploration and analysis. Linked Open Data is the name of a distributed, global initiative to provide better means to open up datasets, and, crucially, to publish links and connections between data. Each datum in COINS is identified by seven key indices and the metadata and data dictionary are provide elsewhere.

So my goal was to integrate all of this (three data sets and their data dictionaries and metadata) into one dashboard. The dashboard I created is shown elsewhere and convinced me that it was very easy to access, use, and understand their Data Catalog, COINS, and Performance Data Sets. This is what we need the US data.gov and other administration open government web sites to be.

Spotfire Dashboard

For Internet Explorer Users and Those Wanting Full Screen Display Use: Web Player Get Spotfire for iPad App

Media, iframe, embed and object tags are not supported inside of a PDF.

Sitemap of Data.gov.uk Web Site

Source: http://data.gov.uk/
Enter your postcode to find tools and information about your local area using government data.

**Who does what in Whitehall and Beyond?**

Source: [http://data.gov.uk/organogram](http://data.gov.uk/organogram)

Browse around the organisational structure for each Government department. Details are shown for all Ministers and key Civil Servants.

Here you can see the most comprehensive organisational charts of the UK Civil Service ever released online, taking another step towards the Government’s goal of being the most transparent government in the world and opening up the structure of the Civil Service to public scrutiny.

These structure charts, known as "organograms" include the names, job titles and salaries of all civil service directors and more senior civil servants. They also include details of the numbers, grades, pay ranges and professions of staff within each team, as at 31 March 2011. They are available in a consistent, user-friendly style, giving you a view of the whole of central government in one place.

**How your money is spent**

Source: [http://data.gov.uk/openspending](http://data.gov.uk/openspending)

Explore the searchable database to view Government expenditure, which can be sorted by department, recipient or value.

View all Government expenditure. You can browse by date, spender, recipient or amount, by selecting from the headings in the blue bar below. You can view spending data for a specific department by using the ‘Filter by Department’ drop down menu. Use the Search box to find all data related to an individual recipient.

**Government contracts in full**

Source: [http://data.gov.uk/contractfinder](http://data.gov.uk/contractfinder)

Search and read about all Government opportunities, tenders and contracts worth more than £10,000.

Contracts Finder is a free new service for businesses, central government buyers and the public. It is now the main source of government procurement opportunities with a value greater than £10,000. People can also use Contracts Finder to access and search for closed tender and contract documentation published by central government departments and agencies. Contracts Finder is accessible from the Business Link website, enabling users to easily click through to guidance on public sector procurement, as well as wider business support information and resources.

**Who ministers are meeting**

Source: [http://data.gov.uk/whoslobbying](http://data.gov.uk/whoslobbying)

Find out which organisations and industries have met with Ministers from across Government departments.
Browse the lists below to find information about which Ministers have met with representatives of organisations and industries. The information is provided by the whoslobbying.com website, which uses public data to display the details in a clear format.

**Business plans**

Source: [http://data.gov.uk/business_plans](http://data.gov.uk/business_plans)

Check on the Government’s progress on implementing its policies. See how every department is performing against its priorities.

Select a department to find out how it is performing in delivering on the priorities set out in its business plan. Browse the individual actions under each priority to see the colour coded status for each individual activity. The information is displayed by using data provided from the Number 10 website.

**Applications**

Source: [http://data.gov.uk/apps](http://data.gov.uk/apps)

Use your iPhone to view neighbourhood information and crime stats for the area.

This wide range of apps can provide you with information on everything from local services information, to managing finance, and even environmental issues to name but a few. Search the apps list to find out more.

**Data**

Source: [http://data.gov.uk/data](http://data.gov.uk/data)

Search over 7,800 datasets. You can also access this catalogue via the API or download it as a CSV/JSON dump. CSV

EDITOR’S PICK: COINS – HM Treasury’s database of public spending CSV

**Apps**

Source: [http://data.gov.uk/apps](http://data.gov.uk/apps)

See Below

**Event**

Source: [http://data.gov.uk/government-procurement](http://data.gov.uk/government-procurement)

**Central Government’s potential contracting opportunities**

For the first time, Central Government acting as a single customer is publishing information on its future contract pipeline. This pipeline will give an indication of potential ICT and FM opportunities, and potential Government Procurement Service frameworks including the Public Sector Network (PSN), which may be competed over this
Parliament and with an estimated value of £5m or more.

Whilst every effort has been made to ensure this data indicates the best overall picture of the future pipeline, we cannot at this stage give a definitive indication as to how these contracts might be tendered. The data will be refined over time as Departments align more closely with Government’s ICT and FM procurement strategies or defer or stop projects that are no longer required.

The pipeline is based on unaudited administrative data and should not be considered as national statistics or as official statistics. Values for the GPS frameworks are estimates over the contract duration.

The Cabinet Office was responsible for collating this data but it is owned by Departments. Plans for updating the data will be released shortly.

Please note the following about the FM data:

- It is based on 2009/10 Departmental FM data return actual spend, extrapolated over the contract term to calculate Total Contract Value
- The majority of contracts ending 2011/12 are already encompassed within departmental strategies and therefore the opportunity is no longer available
- Land Registry data submitted in 2009/10 was within Ministry of Justice data - in 2011 Land Registry transferred to BIS and for the purposes of this exercise are reported as part of BIS.
- Excludes estate rationalisation plans and therefore size of price could be significantly smaller
- Excludes contracts that did not have start/end dates submitted
- Excludes IPD CA Category (Rent an Rates) but includes unitary charges for PFIs in total where FM cannot be identified separately
- Additional spend data from departments is expected shortly. Anticipated Ministry of Defence ICT spend through the Emporium and Grapevine projects is significant, but is not available in detail at this time.

Note: CSV Files to Download

---

Forum

Source: [http://data.gov.uk/forum](http://data.gov.uk/forum)

**Discussion forum**

<table>
<thead>
<tr>
<th>Forum</th>
<th>Topics</th>
<th>Posts</th>
<th>Last post</th>
</tr>
</thead>
<tbody>
<tr>
<td>General discussion</td>
<td>70</td>
<td>440</td>
<td>Spam comments by chetanpatel 1 day 15 hours ago</td>
</tr>
</tbody>
</table>


Updated: Sat, 19 Sep 2015 03:02:35 GMT
Powered by mindtouch

10
<table>
<thead>
<tr>
<th>Forum</th>
<th>Topics</th>
<th>Posts</th>
<th>Last post</th>
</tr>
</thead>
<tbody>
<tr>
<td>Using data</td>
<td>24</td>
<td>61</td>
<td>Using ... by exstat 05/08/2011 - 17:30</td>
</tr>
<tr>
<td>Publishing data</td>
<td>10</td>
<td>35</td>
<td>Publishing Data by Anonymous (not verified) 18/08/2011 - 09:25</td>
</tr>
<tr>
<td>Tools and Applications</td>
<td>4</td>
<td>23</td>
<td>Manage My Chain by jrobinforall 03/10/2011 - 05:18</td>
</tr>
<tr>
<td>Team building</td>
<td>3</td>
<td>12</td>
<td>Open up ... by Jacky Ron 25/11/2011 - 02:06</td>
</tr>
<tr>
<td>Open Data</td>
<td>2</td>
<td>6</td>
<td>Finding the ... by Opmo jerry 02/10/2011 - 03:43</td>
</tr>
</tbody>
</table>

Forum Contains New Posts

Forum Contains No New Posts

Forum is Locked
Blogs

Source: [http://data.gov.uk/blog](http://data.gov.uk/blog)

Does data matter? Why should it be open to everyone? Find out more from the team behind this website and add your own thoughts.

The blogs (7)

**Leadership**
48 posts - Last post 08/11/2011

**Site**
47 posts - Last post 15/06/2011

**Events**
3 posts - Last post 20/01/2011

**Labs**
1 posts - Last post 15/10/2010

**Data**
18 posts - Last post 07/07/2011

**Real Time Energy**
5 posts - Last post 18/03/2011

**UK Location**
1 posts - Last post 14/03/2011

Featured blogs
- [Sir Tim Berners-Lee](http://data.gov.uk/blog/author/sir-tim-erners-lee)
- [Professor Nigel Shadbolt](http://data.gov.uk/blog/author/professor-nigel-shadbolt)
- [Number 10](http://data.gov.uk/blog/author/number-10)
- [Digital Engagement](http://data.gov.uk/blog/author/digital-engagement)
- [Digigov (COI)](http://data.gov.uk/blog/author/digigov-coi)
- [Jeni Tennison](http://data.gov.uk/blog/author/jeni-tennison)
Have you used our data to create an application?
Let us know...Share your app

Have you got a good idea for the data?
We can help you find a developer....Share your idea

Do you want more data?
Request new data

Ideas

Source: http://data.gov.uk/ideas

Does Britain need a map showing which roads have been gritted when it snows? Read and rate over 360 bright ideas about how the data on this website could be used.

EDITOR’S PICK

A database of energy performance certificate ratings so consumers can compare ratings of homes they are considering buying or renting.

Linked Data

Source: http://data.gov.uk/linked-data

Linked data is data in which real-world things are given addresses on the web (URIs), and data is published about them in machine-readable formats at those locations. Other datasets can then point to those things using their URIs, which means that people using the data can find out more about something without that information being copied into the original dataset. This page lists the sectors for which we currently publish linked data and some additional resources that will help you to use it. Most sectors have one or more SPARQL endpoints, which enable you to perform searches across the data; you can access these interactively on this site. Reference Reference data covers the central working of government, including organisational structures where these have been made available as RDF.

Browse

- Visualisation
- Government Departments
- Other Public Bodies
- Ministers
- Members of Parliament
- Members of the House of Lords
- Companies House

Updated: Sat, 19 Sep 2015 03:02:35 GMT
Powered by mindtouch™
Ordnance Survey have released a number of their products as linked data, including postcode units and administrative areas. Browse

Datasets

- 1:50,000 Scale Gazetteer
- Code Point® Open

Transport data covers the transport infrastructure and data about traffic flow. Note: the data provided here dates from March 2010. Browse

Datasets

- NPTG
- NaPTAN

Guides

- Guide to SPARQLing Transport Data
Legislation Legislation data is provided by legislation.gov.uk. Browse

- UK Legislation
- Scottish Legislation
- Welsh Legislation
- Northern Ireland Legislation

Search

- Advanced Search
- SPARQL Endpoint provided by TSO

Guides

- Legislation.gov.uk API
- Background on legislation.gov.uk
- Developer Documentation

Finance COINS is available as linked data in a beta version. We are working on providing an API onto this data, but for now only the SPARQL endpoint is available. Search

- SPARQL Endpoint provided by TSO

Datasets

- COINS

Further Information

- SPARQL page on data.gov.uk
- SPARQL Wiki page on data.gov.uk
- Using SPARQL Endpoints provided by Talis
- Using SPARQL Endpoints provided by TSO
- SPARQL by Example
- SPARQL Tutorial

Resources

Source: http://data.gov.uk/resources

Here’s where you can access tools to help you analyse, adapt and share the data. You’ll also find video interviews and other features to help you get more out of this website.
The power of government data

Professor Nigel Shadbolt gave a talk about the power of government data and the work that he and others have been doing to open it up at The Guardian’s Activate 2010 conference. Below is a video of his talk:

This video is made available under the Creative Commons Non-Commercial No Derivatives licence. See the licence’s deed page for more details.

What is Linked Data?

Here is Sir Tim Berners-Lee on this at TED: http://www.ted.com/talks/view/lang/en/id/484

Michael Hausenblas's introductions:

- 3 minutes about Linked Data
- 6 minutes about the Semantic Web

Yves Raimond and Michael Smethurst's, BBC, A skim-read introduction to linked data

Tom Heath, Open University, technical tutorial on how to publish Linked Data

Using Linked Data

Jeni Tennison's excellent practitioner's blog
Semantic Web Development Tools
The Talis wiki (link to introduction to SPARQL)

SPARQL tutorials

- SPARQL By Example
- SPARQL Tutorial
- Data Extraction & Exploration with SPARQL & the Talis platform

History of project

- Power of Information Taskforce report
- Sir Tim Berners-Lee and Professor Nigel Shadbolt - scope of the appointment
- Design note from Sir Tim Berners-Lee on how to put government data online
- PM letter to Cabinet
- Creation of the Transparency Board & Publication of Public Data principles
What is the Semantic Web?

Combining different data sources has never been easy, but the Semantic Web will enable data to be joined easily across boundaries. Find out more on this page

About

Source: [http://data.gov.uk/about](http://data.gov.uk/about)

What's data.gov.uk all about?

The Government is releasing public data to help people understand how government works and how policies are made. Some of this data is already available, but [data.gov.uk](http://data.gov.uk) brings it together in one searchable website. Making this data easily available means it will be easier for people to make decisions and suggestions about government policies based on detailed information. Hear more about the Government's Transparency agenda from the Prime Minister

What's available?

There are currently over 5,400 datasets available, from all central government departments and a number of other public sector bodies and local authorities.

Is data just public information?

Not really. From data.gov.uk, you can access the raw data driving government forward. This can then be used by people to build useful applications that help society, or investigate how effective policy changes have been over time. General public information - such as how to find out if you are entitled to tax credits, or how to tax your car - can be found at [DirectGov](http://direct.gov.uk)

How can I use the data?

You can use the data in all sorts of ways. This may be simply to analyse trends over time from one policy area, or to compare how different parts of government go about their work. Technical users will be able to create useful applications out of the raw data files, which can then be used by everyone.

How can I add my organisation's data?

There is a simple how-to guide that takes you through the steps needed to include your data in data.gov.uk's index.

How can I find out more?

You can find out more about the data.gov.uk project by following these links: [FAQ](#), [Project Info](#), and [Resources](#)
Terms and conditions

Source: http://data.gov.uk/terms-and-conditions

The data and information available through www.data.gov.uk are available under terms described in the "licence" or "constraints" field of individual dataset records (meta-data). Except where otherwise noted this is the Open Government Licence.

All dataset records (meta-data) published on www.data.gov.uk are licensed under the Open Government Licence.

Code of conduct

Source: http://data.gov.uk/code-conduct

The data is provided to you the developers for you to develop applications which will make the world a better place, providing more or improved utility for the general public based on public sector datasets. We are here to help you do the great things we know you’re itching to do, but, as part of this being a trusted space in which we work together, there are some basic rules.

You are encouraged to:

- use provided live sources for the datasets as your primary data source, where possible – much of the data provided is dynamic and fluid (although in many cases over an annual or longer period of time), and basing your application on the keys in the data we are providing will ensure that your application is as up to date as possible;
- where using our data feeds, please if possible cache the data over a short period of time to reduce load both on your application and our servers;
- where you cannot use our live sources as the primary data source, please still use the keys contained within the data as the primary or foreign keys within your application, as this will enable you to migrate as simply as possible to up-to-date data and information when it is made available;
- where using your own long-term caches of our data, please refresh the data within your application in a timely manner when a new set of the data is made available;
- engage with us when something’s wrong – point out any inconsistencies in the data you find, make suggestions about how to improve the service, request new data sources to be added, and tell us about what you have built so we can help you to promote it!

You should:

- refer back to and cite the origin or source of the data when using it;
- clearly state where possible about which data is derived from official sources, which is from other sources, and which data is from any calculation within or outwith your app.

You must not:

- make an application that pretends to be an official government service;
- present the data in a misleading or incorrect manner or to misrepresent or change the data;
- use this site for party political purposes, as the site paid for with public money so it is not appropriate to engage in party-political activity here;
- use these materials in or to support a criminal or illicit activity;

Updated: Sat, 19 Sep 2015 03:02:35 GMT
Powered by mindtouch™
• use these materials on an application to inflame or make comments that are racist, sexist or homophobic, or which promote or incite violence or illegal activity.

**Participation Guidelines**

When posting comments about ideas or applications, or posting to the site forum, please observe our site participation guidelines:

• Be respectful of other users of this site
• Stay on topic and do not upload content that is unrelated to the purpose of this site.
  Do not use language that is offensive, inflammatory or provocative (this includes swearing and obscene or vulgar comments)
• Do not select a user name that is offensive, inflammatory or provocative
• Do not break the law (this includes libel, condoning illegal activity and contempt of court)
• In the course of using the community areas of this site, please do not post personal information – addresses, phone numbers, email addresses or other online contact details – relating either to you or other individuals
• Do not register more than one user account per person
• Do not impersonate or falsely claim to represent a person or an organisation. Do not attempt to log on using another user’s account
• Do not make any commercial endorsement or promotion of any product, service or publication not relevant to the discussion
• Do not post in a language other than English

If you are aged 16 or under, please get your parent’s/guardian’s permission before participating. Users without this consent are not allowed to participate or provide us with personal information.

**Project Info**

Source: [http://data.gov.uk/project](http://data.gov.uk/project)

1. **What is data.gov.uk?**

This site seeks to give a way into the wealth of government data. As highlighted by the [Power of Information Taskforce](http://data.gov.uk/project), this means it needs to be:

• easy to find;
• easy to license; and
• easy to re-use.

We are drawing on the combined expertise and wisdom of the Transparency Board to publish government data in a way consistent with the Public Data Principles.

2. **What is our approach?**

• Working with the web;
• Keeping things simple: we aim to make the smallest possible changes that will make the web work better;
Working with the grain: we are not looking to rebuild the world. We appreciate that some things take time; others can be done relatively quickly. Everything has its own time and pace;

- Using open standards, open source and open data: these are the core elements of a modular, sustainable system; and
- Building communities, and working with and through them (both inside government and outside).

Here is Sir Tim Berners-Lee on this at TED: http://www.ted.com/talks/view/lang/en/id/484

Formally launched in January 2010 the site will continue to develop over the coming months as we continue to incorporate your feedback and thoughts.

3. How can I get involved?

There are a number of ways of getting involved in the project, dependent on your background or interest. For example:

- Ideas: let us know any ideas you would like to see someone build as applications? Tell the community via the Ideas page and someone might just pick up the challenge!
- Applications: have you built something cool with our data? Share it with us through the Apps page. You never know who might get excited!
- Visualisations: one of the challenges is making data come to life. Have you found a great way of displaying our data? Does it combine with other data to give a new insight? Show us your talent?
- Architecture: would you like to help us make this site better? Get in touch!

4. More questions?

Have a look at our frequently asked questions page.

5. How can I contact the project?

There are a few ways to get hold of us: Email: Contact page Twitter: http://twitter.com/DirDigEng IRC: #opendata on freenode http://www.freenode.net

6. Who is involved in the project?

Data.gov.uk is a key part of the Government’s work on Transparency which is being lead by the Transparency Board. Data.gov.uk implementation is being led by the Transparency and Digital Engagement team in the Cabinet Office, working across government departments to ensure that data is released in a timely and accessible way. This work is being supported by Sir Tim-Berners Lee & Professor Nigel Shadbolt. There are a number of technical partners involved in the project to date. These include the Comprehensive Knowledge Archive Network (CKAN): CKAN stores the catalogue behind data.gov.uk and a growing number of open data registries around the world. It is a project created by the Open Knowledge Foundation to provide make it easy to find, share and reuse open content and data. The CKAN software provides its own web interface, programmer’s API, feeds notifying of changes, and a browsable history of all changes. The API is available at: catalogue.data.gov.uk
7. Awards

Data.gov.uk has been shortlisted for the following awards:

- Nominet Internet Awards 2010 – Empowering Young People & Citizens
- ISPA Awards 2010 - Internet Hero

FAQ

Source: [http://data.gov.uk/faq](http://data.gov.uk/faq)

Q. What is the project doing?

A. The government is opening up its data for other people to re-use. This is only about non-personal, non-sensitive data – information like the list of schools, crime rates or the performance of your council.

To make sure that we do this properly and work with the web, this work is being overseen by the Transparency Board.

This preview site is the first step in creating a network of re-useable government data.

Q. What is the status of this site?

A. Data.gov.uk launched publicly with a beta version in January 2010. Since then we have launched a number of enhancements which you can read about via the blog. Developments are continuing across the site so do check back regularly and of course please tell us what you think.

Q. Will personal information be at risk?

A. The data we publish here, and on related websites, will not identify or provide ways to identify individuals, unless that information is already published (like head teachers of schools).
Q. What is a mashup?

A. A mashup is a web page or, more usually, a web application that combines data or code from two or more sources. It provides information or functionality beyond that designed or envisaged by the original producers of the data. The attraction of data mashing lies in the ease and speed with which new web applications can be launched with limited resources. People also create visualisations – pictures that show the data in clear, imaginative ways and tell a story about the underlying information.

For example, a mashup combining data sources such as school addresses and school league tables to display the results using a mapping tool could show where high-achieving schools are. This would be of interest to parents of school-age children who are considering moving house.

We expect that new and exciting mashups will be the main product of collaboration through this website: be inventive and make things.

Q. What is Linked Data and how does it fit into the Semantic Web?

A. The Semantic Web is an evolution of the World Wide Web that, rather than just linking from one document to another, focuses on their meaning in relation to each other. Linked Data is a set of technologies to achieve this for data, creating a web of data.


While the technical details are complicated, and very well explained by resources like Jeni Tennison’s practitioner’s blog or the Talis platform wiki, the important thing is how it enables people to exchange links to information and the context for that information. For example, you might point to Bartholomew School, and it will have attributes such as “head teacher”, and be part of the collection “All schools in Oxfordshire”.

Q. How can people submit applications and visualisations?

A. Go to the Apps page, where you’ll find information on all applications and visualisations that people have submitted so far, a link for submitting your own application and facilities for joining in discussions.

Q. How can people submit ideas

A. This site is not just for people with technical know-how; it’s also for people who might have a good idea for an application, but aren’t in a position to build it themselves. Go to the ideas page where you’ll find all the ideas that people have submitted and the facility to submit yours.

Q. How can people use data.gov.uk?

A. Data.gov.uk uses a search engine to offer several ways of finding the data you want. You can view all the datasets to see everything that is currently available, or search by keyword, category or department / agency. Each dataset provides guidance on accessing its data. You can also browse datasets by public body, by nation, a-z listing, as well as subject tags.

Q. How were the datasets in data.gov.uk selected?

Updated: Sat, 19 Sep 2015 03:02:35 GMT
Powered by mindtouch
A. Excluding personal and sensitive information, all information created by public sector bodies is, in principle, available for re-use. In the past, different approaches were adopted by local and regional authorities and individual agencies. The government is now widely encouraging all previously inaccessible public information to be made accessible through this website. For more details on what is being prioritised you should have a look at Putting the Frontline First: Smarter Government

Q. Why is a particular dataset not available yet, and when will it be?

A. Given the vast volumes of data, it will be some time before everything can be made available. If there are particular datasets that you believe should be made available more quickly, please flag them to the OPSI unlocking service.

Q. Why is a particular dataset not available through an API yet, and when will it be?

A. The W3C guidance on opening up government data suggests that data should be published in its original raw format so that it’s available for re-use as soon as possible. Over time, we will covert datasets to use Linked Data standards, including access through a SPARQL end-point; this will provide an API for easy re-use.

Q. How can people make contact with the project?

A. We will provide developer community members who register through our Google Group with project updates. You can also follow the project team on Twitter and by reading the blogs page. To keep up with the latest progress and developments, check the posts on the apps page and ideas page discussion boards.

Q. Is there a place for developers to discuss ideas, applications and using of the data?

A. Yes, the Discussion forum is for that purpose. We realise that you’ll want to ask us and each other questions and we’ll try and help as best we can. There is also a wiki where you can work together to share techniques, ideas, problems and tools, using data from the UK Government and other sources to help people understand and get more from their area, issues of interest, and the world we live in.

Q. Will more public data become available in the future?

A. Yes, this is just the start. As the project matures, more data will flow on to the site.

Q. What are the commercial use rights if people have commercial ideas?

A. The Open Government License enables you to use all our data for commercial purposes. We are delighted that you see a commercial opportunity in using our data.

Q. Under what licence is the data available?

A. In general, the data is licensed under the Open Government License. You can read more details on what that means on our licence page.

Q. Are there any “do”s and “don’t”s?

A. We have a simple code of conduct relating to your use of the data in your application. It covers how to relate your application data to the data we provide, and simple common-sense usage of the data.
Contact us

Source: http://data.gov.uk/contact

You can leave a message using the contact form below.

• Please send us your username if you have an account query.
• Provide email addresses for anyone you want cc'd in our response.

Accessibility

Source: http://data.gov.uk/accessibility-statement

Data.gov.uk is committed to making this website accessible to all users, regardless of circumstances. We are actively working to increase the accessibility and usability of the site; if you have any questions or suggestions regarding the accessibility of this site, or if you have difficulty using any part of it, please contact us.

Standards compliance

We are working towards ensuring that all pages on this site comply with priority 2 guidelines of the W3C Web Content Accessibility Guidelines. Whilst data.gov.uk strives to adhere to the accepted guidelines and standards for accessibility and usability, it is not always possible to do so in all areas of the website. We are continually seeking out solutions that will bring all areas of the site up to the same level of overall accessibility.

How do I make the text bigger so that it is easier for me to read?

Some users might find screen text easier to read by increasing the size at which text is displayed. To do this you should simply adjust your browser’s settings.

• Using Internet Explorer, go to View > Text size > and select the size you want
• Using Firefox, go to the View menu, select Zoom and then either Increase Ctrl + +, Decrease Ctrl + – or Normal Ctrl + 0.
• Using Opera, View > Zoom > and select the size you want

What are PDFs and how do I read them?

PDF stands for Portable Document File. You need a plug-in called Adobe Acrobat Reader to view these files which you can download for free from Adobe’s website.

Users who have concerns about accessibility should visit Adobe’s accessibility website. Recent versions of Acrobat Reader have a Read Out Loud facility, which can be found under the View menu.

How can I view a Word file if I don’t have Word software?

A free Word viewer is available from Microsoft at the Microsoft Download Center. You’ll find Word, Excel and PowerPoint viewers here too. Open source versions of similar software can be accessed at http://www.openoffice.org/
Accessibility software

1. **JAWS** is a screen reader for Windows. A time-limited, downloadable demo is available.
2. **Lynx** is a free text-only web browser for blind users with refreshable Braille displays.
3. **Opera** is a visual browser with many accessibility-related features, including text zooming, user stylesheets and image toggle. A free downloadable version is available. It is compatible with Windows, Macintosh, Linux, and several other operating systems.

Accessibility services

1. **HTML Validator** is a free service for checking web pages conform to published HTML standards.
2. **Web Page Backward Compatibility Viewer** is a tool for viewing your web pages without a variety of modern browser features.
3. **Lynx Viewer** is a free service for viewing what your web pages would look like in Lynx.

Please note that these links are placed here only for your convenience only and not as an endorsement of either product or software.

Moderation policy

Source: [http://data.gov.uk/moderation-policy](http://data.gov.uk/moderation-policy)

Comments added to the site by registered, logged-in members will appear on the site without moderation.

Comments added to the site by unregistered or not logged-in users will not appear until they have been moderated.

Ideas and applications submitted are always moderated before appearing on the site.

The Data.gov.uk Team will sometimes edit content before it appears on the site. Such editing will take place only if we feel that it will make the content more understandable (for example, amending typos or moving content into correct fields) or easier to view (for example, if images uploaded are too large for the site or a title is too long). We will not change the substance or meaning of a contribution to the site.

The Data.gov.uk Team will always strive to moderate content added within a 24 hour period during the working week. Comments added over the weekend or bank holidays will be moderated on first subsequent working day.

If your submission to the site does not appear on the site, or is removed, it means that the Data.gov.uk team feel that the contribution does not comply with the [code of conduct](http://data.gov.uk/moderation-policy) of the site.

Where possible or appropriate we will contact you to explain why your contribution did not fit with the guidelines, with suggestions of how the content could be re-phrased so that it can appear on the site.

If you have a query about why your contribution has not appeared on the site, our participation guidelines or approach to moderation, please contact us at team@data.gov.uk

Tagged:
- forum
Privacy

Source: http://data.gov.uk/privacy

Your personal data

From time to time, you may be asked to submit personal information about yourself (e.g. name and e-mail address) in order to receive or use services on our website. Such services include e-mail updates, website feedback, or suggesting an idea / application.

By entering your details in the fields requested, you enable the Crown/Government (Cabinet Office) and its service providers to provide you with the services you select. Any personal information you provide to us will only be used by us, our agents and service providers, and will not be disclosed unless we are obliged or permitted to by law to do so.

If you post or send offensive, inappropriate or objectionable content anywhere on this site, we may use whatever information about you is available to us to stop such behaviour.

We will hold your personal information on our systems for as long as you use the service you have requested, and remove it in the event that the purpose has been met or when you no longer wish to continue your subscription.

Accessibility statement

This website is run by a team drawn from Cabinet Office, The National Archives and Central Office of Information. It is our intention that the website is usable and accessible to all users. This website’s objective is to conform to the Guidelines for (UK Government Websites), which supports Web Content Accessibility Guidelines of the World Wide Web Consortium. A full accessibility statement is now available.

Feedback

We very much welcome and actively encourage your feedback on this site. The site is being developed in a phased approach so we encourage you to provide feedback to ensure that future iterations meet your needs. This project is seen by us as a collaborative one between government and yourselves so do get in touch and help shape future direction.

We will not pass on any of your personal information when dealing with your enquiry, unless you have given us permission to do so. Once we have replied to you, we keep a record of your message for audit purposes.

The Data Protection Act

Under the Data Protection Act 1998, we have a legal duty to protect any information we collect from you. We use leading technologies and encryption software to safeguard your data, and keep strict security standards to prevent any unauthorised access to it.

We will not pass on your personal details to any third party without your permission.
If you wish to see our records of any correspondence you have sent to us, or if you have a query or complaint about this privacy policy or about the site, you can contact us using the contact us form.

**Changes to this privacy policy**

If this privacy policy changes in any way, we will place an updated version on this page. Regularly reviewing this page ensures you are always aware of what information we collect, how we use it and under what circumstances, if any, we will share it with other parties.

---

**Consultation**

Source: [http://data.gov.uk/opendataconsultation-closed](http://data.gov.uk/opendataconsultation-closed)

Making Open Data Real: A Public Consultation

The Open Data consultation closed on 27 October 2011, but you can review the feedback submitted here. We will now review consultation responses and an update on the consultation outcomes will be provided when available.

---

**Research Notes**

Jim Hendler

---

**Data.gov.uk URLs**


[http://data.gov.uk/dataset/coins](http://data.gov.uk/dataset/coins)

Making Open Data Real: A Public Consultation

Source: [http://data.gov.uk/opendataconsultation](http://data.gov.uk/opendataconsultation)

This online consultation has now closed, but you can still review comments within each section of the consultation. We will now review consultation responses and update on the outcomes as soon as they are available.

*August 2011*

#opendata
#openuk

The Open Data consultation paper sets out Government’s proposed approach for Transparency and Open Data Strategy, which is aimed at establishing a culture of openness and transparency in public services. We want to hear from everyone – citizens, businesses, public services themselves, and other interest groups – on how we can best embed a culture of openness and transparency in our public services. The consultation closes on 27 October 2011.

You can access the consultation by clicking on the links below, or by going direct to the links under “Further Info” (on the right hand side).

**Table of contents**

0. Foreword
1. [Glossary of key terms](http://data.gov.uk/opendataconsultation#glossary-of-key-terms)
2. [Table of abbreviations](http://data.gov.uk/opendataconsultation#table-of-abbreviations)
3. Executive Summary
4. Vision
5. Scope and Objectives
6. [Current landscape and scope for improvement](http://data.gov.uk/opendataconsultation#current-landscape-and-scope-for-improvement)
7. The Six Opportunities of Open Data
8. Policy Challenge Questions
9. Next Steps
10. Annex 1 – Best Practice and Emerging Evidence

Performance data on Government ICT projects (31 July 2010)

Source: http://data.gov.uk/dataset/performan...s-31-july-2010

A spreadsheet of details on Government ICT projects as at 31 July 2010 and valued at over £1million. Departments own the data. ERG collated the data. All the data is published apart from that redacted for reasons of commercial sensitivity or because it is confidential. These are not official statistics; since collection some of the data will have changed; as a result of the freeze on new ICT project spend and the Departmental reviews of their ICT Project portfolios, some of the projects listed here will since have been stopped or significantly re-scoped; the data will continue to change in the future; this publication will be updated regularly, however the specific form of future publications has not yet been determined.

Data on UK Government ICT projects over £1million, as of 31/07/2010 (74.37 KB)
Data on UK Government ICT contracts as part of projects over £1million, as of 31/07/2010 (137.31 KB)
Data on UK Government ICT projects over £1million, including a tab on contracts, as of 31/07/2010 (415 KB)

Full Description: http://interim.cabinetoffice.gov.uk/govtalk/data.aspx Data Dictionary

Forum

http://data.gov.uk/forum: COULD BE A DATA SET

Energy Performance Certificate Ratings

Posted by Nigel Farren on 04/06/2010

Source: http://data.gov.uk/ideas/energy-perf...ficate-ratings
Database of EPC ratings so consumers can compare ratings of homes they are considering buying or renting. REQUEST FOR DB

XML data & direct download links.


Provide the data as XML & directly downloadable from this site.

Providing the data is useful, but most seem to be in XLS format, which isn't really much use when building an application as the data has to be manually edited & converted prior to use. Additionally using a standard format would allow data to be compared & combined more easily.

Also, as far as I can tell, none of the data is available directly from this site. Some of the download links require new searches on associated sites before a variety of files are available. Datasets do not need embedded graphs.

I think this would help people use more of the data, as the main gripe in the forum seems to be the inability to find the data .... on the site that seems to extol itself as the gateway for .gov data.

COINS as Linked Data

Source: http://data.gov.uk/resources/coins: READ THIS!
Introduction

In common with other governments around the world, the UK Coalition Government has made a commitment to increasing government transparency. See, for example, Chapter 16 of the Coalition Agreement. A key initiative under the general umbrella of government transparency worldwide is open data – the release to the public of government data, unencumbered by restrictive licenses or fees.

Unsurprisingly perhaps, one of the topics of greatest interest has proved to be how the government spends its money. In June 2010, HM Treasury, the principal custodian of financial data for the UK government, released previously restricted information from its Combined Online Information System (COINS). Until then, access to COINS by journalists and members of the public had been very difficult, and COINS itself had been the subject of various Freedom of Information requests. The release of COINS data therefore represented a step change in the Treasury’s willingness to share data openly.

While the release of the COINS data is very significant, it does not by itself make the government’s finances scrutable. COINS is a highly technical information source, requiring detailed knowledge of the encoding schemes used to make sense of. Moreover, the data files are too large to load into widely available analysis tools such as Microsoft Excel, a common tool-of-choice for many data investigators. Some enterprising teams have stepped up to the challenge of providing better access to the COINS data, notable among them the Where does my money go? project, and the Guardian datastore. We believe, however, that more can be done to open up the COINS dataset for exploration and analysis.

Linked Open Data is the name of a distributed, global initiative to provide better means to open up datasets, and, crucially, to publish links and connections between data.

At its heart, Linked Open Data uses a collection of data encoding standards maintained by the W3C, with the resource description framework (RDF) at its core. Linked Open Data, often abbreviated LOD, is founded on a small number of key principles:

- naming things with uniform resource identifiers (URI's), which can be resolved to get both human-readable and machine-readable descriptions of the resource
- a uniform data model based on the assertion of a named relationship between two resources, or between a resource and a data value (such as a number or a date)
- extensibility, or the ability to add extra annotations to a resource without first requiring a prescriptive data schema to be updated

For more background information on linked data, see linkeddata.org and the Wikipedia page on linked data.

This article explains how we converted COINS to RDF in linked-data style, so that it can be more easily explored and investigated. We will explain some of the principles involved in the representation, and illustrates some of the queries that can be performed against the RDF data.
**What is COINS?**

COINS is a large database system at the centre of operations in HM Treasury. The [Treasury itself describes COINS](#) as:

> COINS – the Combined On-line Information System – is used by the Treasury to collect financial data from across the public sector to support fiscal management, the production of Parliamentary Supply Estimates and public expenditure statistics, the preparation of Whole of Government Accounts (WGA) and to meet data requirements of the Office for National Statistics (ONS). Up to nine years of data are actively maintained: five historic (or outturn) years, the current year and up to three future (or plan) years depending on the timing of the latest spending review. COINS is a consolidation system rather than an accounts application, so it does not hold details of individual financial transactions by departments.

More detailed descriptions of COINS and its contents can be found on the [Treasury's web site](#).

**The COINS data release**

The currently released data from COINS contains snapshots of the contents of the database on 14th June, 2010. Not all spending lines have been included: data from the Ministry of Defence and numbers relating directly to security and intelligence have been omitted. Nonetheless, the five data files released contain between 3.3 and 4.9 million rows of data. Each row represents one aggregate transaction, reported by one of the "spending departments" – for example the [Department of Health](#). These transactions are not individual items of expenditure (other datasets are becoming available which do provide that level of detail), rather these are figures that the departments report to the Treasury to allow overall government spending to be monitored, understood and reported accurately.

Some of the data rows in the published data files are *zero-valued* or dummy entries. These are typically entered during the process of preparing to make a finalised entry into COINS. We have omitted these values from our linked-data presentation of COINS. The COINS release contains two sets of tables: the *fact tables* present a snapshot of the state of the COINS database when the data was released, while the *adjustment tables* catalogue the various additions and updates to the database that lead to the snapshot state. In our work to date, we have concentrated exclusively on the fact tables.

**Hypercube structure of COINS**

Each datum in COINS is identified by seven key indices. The combination of values of these indices is unique for every data entry in COINS. By analogy to the way in which three values, \(x, y, z\) can be used to identify a point in three-dimensional space, we refer to these seven indices as the *dimensions* of the COINS dataset, and the corresponding structure a *hypercube* (since it has more than the three dimensions we normally experience). This terminology is shared with OLAP cubes. The seven identifying dimensions in COINS are:

1. **time** – contains the year or month to which the data relates.
2. **data-type** – distinguishes Plans (i.e. budgets), and Outturn (i.e. actuals)
3. **data-subtype** – the status of the reported transaction (draft, submitted, approved) and the reason for entering the data (e.g. corrections, reserve claims).
4. department – identifies the department responsible for the spending. Note that for budgeting and forecast outturn, the "arms-length bodies" are consolidated into their parent department.

5. account – contains the chart of accounts, identifying the economic nature of the data.

6. programme object – records the function associated with the data, similar to a cost centre. The programme object codes and descriptions are maintained by departments.

7. counterparty – records the second party in some transactions between departments.

In addition to these dimensions, there thirty-three further fields which provide some additional insight into the transaction being reported. These include, for example, the type of department identified by the department ID, whether expenditure falls under resource or capital budgets, which National Account Code (NAC) it is classified under, etc. Some fields in the COINS release are annotated as no longer being in use, so we have omitted these from our linked-data translation.

For illustration, an example entry from COINS is shown below

| value     | 1119900 | this is the value of the transaction in thousands of pounds sterling |
| dataType  | plans   | this amount is planned expenditure                                    |
| time      | 2009-10 | the time at which the plan was recorded                              |
| accountCode | 51611300 | account code for Other Grants to Local Authorities (CAP)           |
| counterpartyCode | cpid.na | no counterparty recorded                                            |
| dataSubtype | pdiraa  | Plans Dept Internal Restructuring Approved Adjustments              |
| departmentCode | des022  | Department for Children, Schools and Families                      |
| programmeObjectCode | p0110002 | labelled as P01 Devolved Capital LA                                |
| accountingAuthority | aa101    | Central Government, supply: all voted payments and receipts by central government departments |
| accountsCapitalCurrent | a-cap    | capital expenditure                                                |
| activityCode | act0201 | the activity code is "Used to allocate expenditure to Departmental Annual Reports and Supplementary Budgetary Information publications." |
| budgetBoundary | del      | department expenditure limit                                        |
| budgetCapitalCurrent | b-cap    | capital budget                                                      |
Estimate line is "Used to identify specific sections in Supply Estimates in the current financial year"
COINS as linked data: overview

Our goal is to make the COINS data available as linked data. There are several reasons why doing so will, we hope, be useful to users of the COINS data:

- using open standard representation makes it easier to work with the data with available technologies, such as the SPARQL query language and the linked data API. As time goes on, we can expect more products and open-source software to make use of these technologies to provide easy-to-use explorer tools;
- individual transactions and groups of transactions are given an identity, and so can be referenced by web address (URL), to allow them to be discussed, annotated, or listed as source data for articles or visualizations;
- cross-links between linked-data datasets allow for much richer exploration of related datasets

To achieve a linked-data presentation of the COINS data, we need a vocabulary that specialises the very general notations used in RDF to the particular needs of COINS. For some other datasets, this requires authoring a new vocabulary (sometimes referred to as an ontology), or adapting an existing one to the purpose at hand. Fortunately, there is already a suitable vocabulary for encoding statistical data that has a hypercube structure: the data cube vocabulary.

The data cube vocabulary

The data cube vocabulary allows a statistical dataset, on the web in linked-data format together with descriptions of its dimensions and metadata. The vocabulary was developed in 2010 as a joint activity among a group of academic research groups and linked-data companies. It derives some of its basic principles from the work of the Statistical Data and Metadata Exchange (SDMX) Initiative.

The vocabulary specification document describes the basic cube organization as follows:

A statistical data set comprises a collection of observations made at some points across some logical space. The collection can be characterized by a set of dimensions that define what the observation applies to (e.g. time, area, population) along with metadata describing what has been measured (e.g. economic activity), how it was measured and how the observations are expressed (e.g. units, multipliers, status). We can think of the statistical data set as multi-dimensional space, or hyper-cube, indexed by those dimensions. This space is commonly referred to as a cube for short; though the name shouldn't be taken literally, it is not meant to imply that there are exactly three dimensions (there can be more or fewer) nor that all the dimensions are somehow similar in size. [source]

For the COINS data, the dimensions and attributes of the cube are outlined above, although they are not in linked-data form. We will also require the data-set descriptor (DSD), which describes the arrangement of dimensions and attributes to consumers of the data, and we will need to convert the encoded values used in the data into RDF resources. Finally,
given the importance of tracing statistical data to its origins, we need to show the provenance of the data in linked data form.

**COINS graph structure**

The RDF encoding of data links nodes, or resources via named predicates to other resources or to data literals (such as a string or a number). Mathematically, this structure is a directed acyclic graph, and we use the term named graph to group together related RDF resources so that we can refer to them by URI.

For each of the years of the COINS release, we will have an authoritative graph which contains the main data transcribed from the Treasury COINS data, a metadata graph which contains provenance and other information describing the authoritative graph, and a supplementary graph which contains additional links we are able to derive from the main data but which is not part of the core data release. Figure 1, below, shows the relationships of these graph structures, together with an overview of the data cube vocabulary applied to COINS:

![Figure 1: The basic COINS graph structure](Image)

Each row of data in the COINS release, omitting the zero-valued dummy entries, will become a single qb:Observation resource. As we show below, the URI for that resource is composed from the code values of the dimensions. Values for each of the dimensions, and for the other non-null valued columns in COINS become properties attached to the observation. Observations may be grouped together as "slices" of the dataset. While these may provide helpful means of navigating the structure of the dataset, they are not fundamental to the encoding. Note that a given observation may be grouped into more than one qb:slice of the data.

Individual column values in COINS are represented as code lists. Some narrative documentation on the code lists used was released with COINS, but definitive catalogues of the code lists is not currently available. As we show below, we have created code lists from the data values in the release, augmented with descriptions from narrative sources. This makes the linked-data presentation of COINS a complete and self-contained unit, but limits the amount of cross-linking to other datasets – a key value of the linked data approach. Where possible, we provide links in the supplementary graph which annotate the code list entries with relationships to other vocabularies. For example, the Ordnance Survey maintain a list of geographic entries, which are clearly related to the territory codes used in COINS.
COINS datasets and named graphs

COINS is an example of one of the more complex statistical datasets being publishing via data.gov.uk. Part of the complexity of COINS arises from the nature of the data being released. The published COINS datasets cover expenditure related to five different years (2005–06 to 2009–10). The actual COINS database at HM Treasury is updated daily. In principle at least, multiple snapshots of the COINS data could be released through the year. In the current, somewhat experimental, phase, we only have one snapshot which was released in June 2010.

In conceptual terms, we have at least four perspectives on what we mean by “COINS” data: the abstract notion of “all of COINS”, the data for a particular year, the version of the data for a particular year released on a given date, and the constituent graphs which hold both the authoritative data translated from HMT’s own sources, and additional supplementary information which we derive from the data, for example by cross-linking to other datasets.

Figure 2, below, shows these conceptual objects, together with the naming schemes we have used and the RDF predicates which link these different views:

**Figure 2: The various perspectives of COINS data and their relationships to RDF resources and named graphs**

The URI patterns we use are summarised in the table below

<table>
<thead>
<tr>
<th>URI pattern</th>
<th>RDF type</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://semanticommunity.info/AOL_Government/Data.gov.uk_-_What's_Not_to_Like">http://semanticommunity.info/AOL_Government/Data.gov.uk_-_What's_Not_to_Like</a></td>
<td>Updated: Sat, 19 Sep 2015 03:02:35 GMT</td>
<td>Powered by Mindtouch</td>
</tr>
</tbody>
</table>
The top-level, most abstract concept in the domain is the whole of COINS.

A particular identified collection of data, such as the COINS data for 2009 – 10

Identifies a particular version of the dataset by date of release

Container for the authoritative, supplementary or metadata triples for a dataset version. This identifier is both a dataset resource and a graph name.

When the data is published via a SPARQL endpoint, the default graph will contain a union of the current versions of the subset graphs, making it possible to query older versions of the data by name while also being straightforward to query the most up-to-date data.

Converting COINS to RDF

In this section, we describe the process of converting the released COINS data to RDF linked data. Our aims are primarily to share experience and lessons with other developers engaged in similar tasks, and to provide a more detailed description of the structure of COINS in RDF to guide consumers of the data.

Downloading and preparing

The COINS release files are available via data.gov.uk, as either straightforward or BitTorrent downloads. The files, once downloaded, are .zip archives containing CSV files (notionally comma-separated values, though in this case the comma character is replaced by the @ character). Listing 1, which was run under Linux but should work on Windows with cygwin, prepares the files for the next stage of the process. The primary goals are to convert the data from UTF-16 to UTF-8, and compact the CSV’s slightly by omitting unnecessary strings such as NULL.

It would be possible at this point to process the files directly from the CSV format. However, we found it much more convenient to load the data into a database, so that the conversion code could use SQL queries to, for example, efficiently collect the lists of code values used for each column. The script in listing 2 loads one of the above generated CSV files into a PostgresQL instance, adding an index column to identify each row of the table with an index value:

Defining the cube structure: the COINS DSD

The data structure definition, or DSD, specifies the structure of the qb:dataset – the dimensions and attributes available. For dimensions, the order in which they are applied is also specified explicitly, since the underlying RDF representation is not ordered: the sequence of statements in an RDF source file does not control the order in which triples are extracted from a triple store containing that source file.
While values are normally attached to observation resources directly, the DSD also allows some information to be attached at the dataset level for brevity. In the case of COINS, all of the values are in thousands of pounds sterling. In principle, this information could be attached at the dataset level, and become implicitly part of each observation. In general, avoiding repetition is a good thing. However, one of our use cases for the linked data version of COINS is to allow others to link to individual observations, which suggests that these observations should be standalone and self-contained – and should therefore have explicit multipliers and units on each observation. The cube vocabulary community has not reached a consensus on this issue as yet; one suggestion is to author data without the duplication, but have the data publication tools "flatten" the compact representation into standalone observations during the publication process. The complete COINS DSD can be seen in listing 4.

**Code lists**

For each of the dimensions and attributes in the DSD, with the exception of the value property itself (sdmx:obsValue), we define the range of permitted values as a code list. For some datasets, code lists are available as separate resources, which can be converted into linked data form. In the case of COINS, we do not have such standalone code lists. We therefore generate controlled vocabularies to represent the code lists by querying the released COINS data in the PostgresQL database. Specifically, for each column in the database, we construct the set of unique values that are in use, then create a code list resource programmatically. For example, the only values in the territory field of the database are:

- ENG
- E&W
- GB
- NI
- SCO
- UK
- WAL

We create a code list from this set of values, by defining each to be a member of a SKOS concept scheme. So that we can specify the range constraint on the RDF property denoting the territory attribute of the DSD, we also create an RDFS class to denote the entries in the code list. The final listing is shown in full in listing 5. Currently, these code lists are generated automatically from the COINS data using some custom Ruby code. In future, we expect that commercially available or open-source tools with this capability will become the norm.

**Observations and slices of the cube**

An individual data point in the data cube vocabulary is an observation, so we represent each transaction from the COINS dataset as an observation, even though this use slightly stretches our real-world understanding of that word. Each observation has one RDF property for each of the dimensions, and a further property for the other fields in COINS, which the data cube vocabulary terms attributes. In addition, we attach properties to denote the value of the observation, its units and multiplier. For COINS, we know from the narrative description of the dataset that the values are in thousands of pounds sterling, so the units are pounds sterling and the multiplier is $10^3$ (encoded using the SDMX code value sdmx-code:unit-mult-3).
Each observation is given a URI, so that it can be referred to by name. For linked data, it is important that the URI actually resolves to a representation of the observation, so that consumers of the data can easily retrieve the details of a given value, and follow links from there to other observations in the dataset. From one point of view, the actual structure of the URI as a sequence of letters and words is not important: for machine processing, it is the use of consistent identifiers that allows resources, such as observations, to be linked together to create context and meaning. However, as human readers of the data, and especially as software developers, it is useful to have the URI's follow a meaningful and useful pattern.

Our COINS observation URI's are composed of a base name, identifying the resources as part of COINS, followed by a compound name sufficient to uniquely identify each observation. The pattern is:

\[
http://{base}/data/coins/{year}/{version}/{key}
\]

The version allows us to distinguish different releases of COINS data: the only snapshot provided by HM Treasury to date was released on the 4th of June 2010, then updated on the 14th. We therefore use 2010-06-14 as the version field, though this will change with future releases. The year field denotes the government year which the data applies to (e.g. 2005-06), while the key is used to uniquely identify each observation in the dataset. Since each observation has a unique combination of values for the seven dimensions used in COINS, we use the short-form name of the code value to form the URI. Thus:

\[
http://source.data.gov.uk/data/coins/2009-10/2010-06-14/snapshot30/bcaa/mod017/31070000/segment.na/mod017.cpi
\]

With respect to the base URI, we use two different subdomains of data.gov.uk: source and finance. The team's strategy here is to distinguish between stable long-lived identifiers, such as the names of statistical concepts or of the COINS DSD, and the outputs from data conversion. Ontology terms, concept schemes and other concepts will be maintained under domain.data.gov.uk, where domain is appropriate to the dataset (e.g. finance for COINS). Data conversion, which is currently being handled by the central Linked Data Kernel Project team, will in time be managed from the data-owning departments. At that point the departments may elect to change the data URI's to a domain that they control. Should this happen in future, URI's under source.data.gov.uk will be automatically redirected to the new identities, so data-consuming applications should see few, if any, disruptions.

**Data slices**

A slice through a dataset is a view of the data in which some dimensions are fixed. For example, in a sales volume dataset organized with dimensions of time against product line, fixing the time dimension to one of its values, say July 2010, would create a slice showing the volume of sales in each product line for that month. In SDMX, it is common to create slices in which all dimensions except the time dimension are fixed, creating a time series. In COINS, while there is a time dimension, it does not record a particular series of values, such as the months of a year. Rather, the meaning of the time dimension is bound up with the datatype dimension: for example, a forecast of expenditure expected in a forthcoming time period.
We therefore do not create a slice specifically to represent a time series. In fact, it is somewhat unclear at this point which slices through the data will be useful to COINS-RDF users. Currently, we create a set of slices in which the time and datatype are fixed, but we will review this design choice over time. While the datacube vocabulary links datasets to slices, and slices to observations, it also includes a reference from the observation to the dataset in which it appears (see figure 1). The flexibility of SPARQL as a query language makes it straightforward to exploit this link to query observations in a dataset without needing the associated slice:

```
```

### Metadata and provenance

An important benefit of linked data is that we are able to annotate data, at a fine-grained level of detail, to record information about the data itself. This includes where it came from – the provenance of the data – but could include annotations from reviewers, links to other useful resources, etc. Being able to trust that data is correct and reliable is a central value for government-published data, so recording provenance is a key requirement for the COINS data.

We use the Open Provenance Model vocabulary (OPMV) to record the provenance of the COINS data. OPMV allows us to record not only the source of the data and its publisher, but also the stages of transformation that were undertaken to convert the COINS CSV release files into RDF on the web.

### Accessing the data

So far we have shown how the data conversion from raw CSV to linked data has worked, but not how to access the data once it is published. To a large extent, we would not expect end users of COINS data to need to access the data in its raw form. Rather, exploration an analysis of the data will be driven by user-friendly tools and web widgets. However, it is not appropriate to speculate on what tools will be created in the future, so in this section we will illustrate some of the more basic means of accessing COINS as linked data. In particular, we will show some sample SPARQL queries, which is the most flexible means of accessing the data. We will also show a configuration of the linked data API, a middleware layer designed to make RDF data more amenable to applications using a JSON-based HTTP application program interface (API). The linked-data API does not offer quite as much flexibility as the full SPARQL end-point, but is considerably easier to use for web applications.

### Example SPARQL queries

Before showing some example queries, we should re-iterate some of the guidance from the Treasury on the use of COINS data. First, COINS contains a mixture of tentative and firm data. It also contains a mixture of forecast and outturn data. A transaction value may be entered in draft form by a department, reviewed by Treasury officials, perhaps adjusted, and finally entered in accepted form. Consequently, we must be careful to pick out consistent data, or risk double-counting some values. The Treasury advises various criteria when selecting actual spending. Quoting from the document:
Defining the data required for a given aggregate can be very complex. The example below shows the “fields” and settings needed for those fields to produce a DEL resource outturn number for a given department and year from “live” data (e.g. not from snapshots).

Using [the COINS fact table] you need the following settings:

- **Data_type**: set to ‘Outturn’
- **Department_code**: set to required departments(s);
- **Time1**: set to required year;
- **Budget_Boundary**: set to DEL;
- **Resource_capital**: set to Resource (on 2010-11 budgeting basis);
- **Data_subtype**: set to all entries with “approved” or = submitted_outturn (both conditions required).

[From *Understanding the COINS data*, pp8–9]

On the last point, it is inefficient to have to query all the time for a string match (i.e. does the data-subtype contain the string "approved"?) A small illustration of the flexibility of the open, schemaless data format we are using is that we can augment the code list for data-subtype with a property denoting the status of the transaction. Since at its core the distinction is between transactions that have been or have not been accepted as final by the Treasury, we re-use the terminology approved and draft from the (very simple) status ontology at [http://reference.data.gov.uk/def/status]. Thus a typical data-subtype code entry becomes:

```xml
a <http://finance.data.gov.uk/def/coins/DataSubtype>, skos:Concept;
rdfs:label "AME Forecast Approved Adjustments"@en;
skos:prefLabel "AME Forecast Approved Adjustments"@en;
skos:notation "afaa";
```

**Example query: spending for one department**

The following query shows the outturn spending for the Department of Children, Schools and Families (DCSF) for 2007-08:

```
select ?boundaryLabel ?rcLabel (COUNT(?amount) as ?n) (SUM(?amount) as ?total) 
where 
{
  ?obs
```

[http://reference.data.gov.uk/def/status]
This query produces the following result:

<table>
<thead>
<tr>
<th>boundaryLabel</th>
<th>rcLabel</th>
<th>n</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;DEL&quot;@en</td>
<td>&quot;Resource&quot;@en</td>
<td>264</td>
<td>44725725</td>
</tr>
<tr>
<td>&quot;DEL&quot;@en</td>
<td>&quot;Capital&quot;@en</td>
<td>82</td>
<td>5225818</td>
</tr>
<tr>
<td>&quot;AME&quot;@en</td>
<td>&quot;Resource&quot;@en</td>
<td>14</td>
<td>11396</td>
</tr>
<tr>
<td>&quot;Not DEL/AME&quot;@en</td>
<td>&quot;Not Resource/Capital&quot;@en</td>
<td>15</td>
<td>4416790</td>
</tr>
</tbody>
</table>

Time: 2.013 sec

Example query: Total spending by department

The following query shows capital and resource spending, by department, for 2007–08. Note the use of a union operator to restrict the reported outcome results to just those categories:

```sparql
select ?deptLongName ?rcLabel (COUNT(?amount) as ?n) (SUM(?amount) as ?total)
where
{
  ?obs
  coins-dimension:dataType <http://finance.data.gov.uk/def/coins/data-type/outturn>;
  coins-dimension:departmentCode ?dept;
  coins-measure:amount ?amount;
  coins-attribute:resourceCapital ?rc
  .

  union
}
```
A sample of the results from this query is shown below:

<table>
<thead>
<tr>
<th>deptLongName</th>
<th>rcLabel</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Armed Forces Retired Pay, Pensions etc&quot;</td>
<td>&quot;Resource&quot;@en</td>
<td>27</td>
</tr>
<tr>
<td>5474562</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;Army Base Repair Organisation&quot;</td>
<td>&quot;Capital&quot;@en</td>
<td>4</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;British Waterways Board&quot;</td>
<td>&quot;Capital&quot;@en</td>
<td>6</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;Cabinet Office&quot;</td>
<td>&quot;Capital&quot;@en</td>
<td>40</td>
</tr>
<tr>
<td>33072</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;Cabinet Office&quot;</td>
<td>&quot;Resource&quot;@en</td>
<td>187</td>
</tr>
<tr>
<td>359887</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;Cabinet Office: Civil superannuation&quot;</td>
<td>&quot;Resource&quot;@en</td>
<td>42</td>
</tr>
<tr>
<td>7479267</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;Central Office of Information&quot;</td>
<td>&quot;Resource&quot;@en</td>
<td>5</td>
</tr>
<tr>
<td>336</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;Charity Commission&quot;</td>
<td>&quot;Capital&quot;@en</td>
<td>2</td>
</tr>
<tr>
<td>968</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;Charity Commission&quot;</td>
<td>&quot;Resource&quot;@en</td>
<td>25</td>
</tr>
<tr>
<td>32458</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Example query: retrieving all data for a given observation

Pulling back data on individual observations in RDF using SPARQL is straightforward. In the simplest case, where we know the identity of the observation (perhaps through some user interactions on the UI), we can simply ask:

```
```

The output from the above query will appear similar to listing 6. If the specific identity (i.e. the URI) of the observation is not known, we can still use describe with a `where` clause to select the observations of interest. For example, the following query selects the same observation:
describe ?obs
  where {
  }

Other means of accessing the COINS data

While SPARQL provides a rich and feature-complete means of accessing the COINS data, it is not a notation that is familiar to many developers. The linked data API (LOD API) provides a simplified entry point to a large subset of SPARQL capability against some RDF data store, in a way that is optimised for access by web applications. In particular, the LOD API provides a REST-style interface to both resources and collections of resources from a triple store, and can return results in JSON format, which makes integration with other web applications easier.

We also set up HTTP-level redirections, via the web server, so that directly resolving the URI of a resource from COINS will directly return a description of that resource from the triple store. This means that performing an HTTP GET on, for example, the identity of an observation will return a description of that observation. Depending on the MIME type requested, this description may be in JSON, XML, RDF or human-readable HTML format. Since the URI of an observation, or dataset, or statistical concept can be used both to identify and to access a resource, it becomes straightforward to add additional information, such as annotations, to the data. As long as there is a persistent store which can contain the extra assertions, and user-contributed information is distinguished from authoritative source data, it is easy to see how, for example, a collaborative annotations service could be created to facilitate "crowd-sourced" dataset investigation.

Both the direct-access approach and the LOD API approach to inspecting the data can be very useful in layering richer, more comprehensible user interfaces on top of complex datasets, such as COINS.

Conclusion

In this article, we described the translation of the first release of HM Treasury's COINS data as linked-data using RDF. We have outlined the process used to convert the data, the various elements of the encoding using the Data Cube vocabulary, and illustrated some sample queries that can be made against the data using the query language SPARQL. This is a first release of the linked data version of COINS, and we anticipate that some clarifications or refinements may be needed in future releases, though we will endeavour to keep the URI structure as stable as possible. Feedback, questions and suggestions for improvement are welcome – please email the author at the address at the head of this article.