Public Data

Group

The Public Data Group:

supporting the National Information Infrastructure

Statement on Open Data

Summary 2014

Contents

Introduction

The Public Data Group (PDG) brings together four public sector bodies - Companies House, Land Registry, Met Office and Ordnance Survey - that collect, refine, manage and distribute data on the nation's companies, property, weather and geography. Collectively they form a substantial part of the National Information Infrastructure[1].

The Public Data Group's data is made available through a variety of channels and licences and includes both commercial agreements and the provision of Open Data[2].

The value of the data that is charged for is vast – with Ordnance Survey data widely used in the insurance sector, and the billions of pounds saved by the use of Met Office data in the aviation industry as just two examples. Equally, the value of the Open Data released by the Public Data Group is very significant and growing. The most recent estimate placed the value of Open Data released by PDG at over £900m annually[3].

Given the growing interest and importance of open data to the economy and society we feel it is important to provide in one short document a description of the kind of data these organisations already make available, and to set out PDG's commitments to future Open Data releases and activity.

The intention is to provide enough detail to inform, intrigue and inspire anyone with an interest in data and particularly those wanting to add to the growing kaleidoscope of products and services built using the platform of public data.

PDG commit to updating this document annually.

---

[1] The formal creation of which was recommended by the Shakespeare Review in 2013 with a first iteration published in October 2013: https://www.gov.uk/government/publications/national-information-infrastructure
[2] Although Open Data has many definitions one of the clearest is in the The Open Data Handbook : "Open data is data that can be freely used, reused and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike". Available at: http://opendatahandbook.org/en/what-is-open-data/.
[3] Market Assessment of Public Sector Information, Deloitte, 2013. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/198905/bis-13-743-market-assessment-of-public-sector-information.pdf

Who are the Public Data Group?

The Public Data Group was created in 2011 as an Advisory Group to Ministers at the Department of Business, Innovation and Skills. PDG has an independent Chair and brings together Companies House, Land Registry, Met Office and Ordnance Survey.

Public Data Group members perform key public roles and produce a wide range of nationally important data. The four organisations contain within them world leading expertise on both their subject matters and on the manipulation and management of data. They underpin a whole range of vital activities within the UK while also carrying the flag of British excellence globally – attending and advising numerous international bodies.

The following descriptions and statistics aim to both capture the essence of the role they play within the nation's economy as well as giving a sense of the sheer scale and complexities of their activities.

Companies House: registers the incorporation and dissolution of companies and examines, and stores and disseminates information delivered under the Companies Act and related legislation.

There are around 3.3 million companies registered in the UK, and approximately 530,000 new companies were incorporated last year. Over 130 million documents are accessible on the register of companies.

Land Registry: registers the ownership of land and property in England and Wales. It is one of the largest transactional databases in Europe.

Land Registry dealt with 26.3 million transactions in 2013/14. Almost 85 per cent of the land mass of England and Wales is registered, adding up to more than 12.9 million hectares. Just under 24 million titles are recorded in the Land Register.

Met Office: is the UK's National Weather Service.

One of the world's most accurate forecasters, using more than 10 million global weather observations a day, an advanced atmospheric model, a high performance supercomputer and meteorological expertise to create 3,000 tailored forecasts and briefings a day.

Ordnance Survey: is Great Britain's national mapping authority.

It produces the definitive digital picture of Britain's geography and holds the largest database of its kind anywhere in the world, made up of almost half a billion features. In 12/13 99.6% of significant real-world features, which are greater than six months old, were represented in Ordnance Survey's geographic data.

Is all PDG data free?

Many datasets are available for free but not all are.

The PDG organisations have a commitment to make as much data freely available as possible, but they have to balance this commitment with other requirements such as maintaining the quality of the data, covering the costs of the collection and distribution of the data, and avoiding cross subsidising one data set from another[4].

Making data that is currently charged for available for free can carry a direct cost to the taxpayer where public funds are required to replace commercial income streams. This might be appropriate where these costs are be outweighed by an increase in wider economic activity. However an assessment of a recent Open Data release by Ordnance Survey suggests that this is not necessarily the case for all datasets[5]. PDG are actively engaged in better understanding the potential economic benefits of Open Data and welcome others continuing work in this area.

Making data available for free can also have indirect costs for the organisations involved. Following its Open Data releases the Land Registry initially experienced an increase in its support costs as the number of calls querying the accuracy of the data on third party websites, increased. These additional indirect costs also need to be factored into any business case when considering where there is scope to release more data for free.

The reasons why, and the way in which, PDG organisations charge for some data reflect the different natures of these organisations. Land Registry and Companies House produce data as part of or as a by-product of their primary activities. For Ordnance Survey and the Met Office their primary activity is the collection, analysis of data and the subsequent creation of services.

Managing these different factors and balancing the potential benefits to UK economic growth of releasing data while still protecting the taxpayer requires a pragmatic approach. This approach can be summarised as making all data that can be legally released available under one of 4 pricing categories:

- **Open Data:** As far as possible and where financially feasible data is released at no cost and where necessary, under an Open Government Licence.

- **Cost Recovery:** The data is made available at a price to offset the cost of its collection, management and distribution. For many data sets these fees have reduced over time as processes have been digitised and have become more efficient. Further efficiency savings may allow data to move from this category to Open Data.

---

[4] In some cases, such as Companies House, there is a specific legal requirement to avoid cross-subsidisation but for all organisations any fees that contain cross-subsidisation risk being classified as taxation - which could have significant public spending implications. For more information see:
https://www.gov.uk/government/publications/managing-public-money

[5] An assessment of the economic impact of Ordnance Survey OpenData suggests an annual net cost to the Exchequer of between £11.7m and £15.6m "Assessing the Value of OS OpenData to the Economy of Great Britain", 2013. Available at:
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/207692/bis-13-950-assessing-value-of-opendata-to-economy-of-great-britain.pdf

- **Trial Access:** where data is available for free in either a limited form, for a limited time or for non-commercial (including research) use only. Although this data is not truly 'Open' it helps to minimise the barriers to entry for new users of commercial services.

- **Commercial Rates:** Where data is available at a price that covers the cost of its collection, distribution and any service element, and makes a reasonable level of return. Typically this data will also reflect considerable analysis and refinement of raw data and can be seen as constituting a service.

Which category each data set falls into will change over time as new IT infrastructure is created, our understanding of the impact of Open Data develops and new products emerge.

What Open Data does PDG make available?

There are numerous data sets already made available by PDG companies. To describe each in detail here would be impractical but further details and links for each organisation have been provided as an annex. The following information seeks to highlight the extent, depth and variety of Open Data that is already available.

For all registered British Companies with electronically filed accounts you can access their:

- Accounts
- Company number
- Address
- Status
- Incorporation date
- Account and annual return filing dates
- URI (Unique Reference Indicator for each company)
- Basic information about mortgage charges

For registered properties in England and Wales you can access:

- The House Price Index (the definitive tracker of UK house prices)
- Information about the number and types of applications that Land Registry receives each month
- Records of individual sale prices dating from now back to 1995
- Outlines (polygons) of the land of freehold registered properties
- The original documents from the first register in 1862

A suite of observation and forecast data in re-useable formats that include:

- UK Site Specific Forecast, either 3 hourly or daily out to 5 Days
- Last 24 hours UK Land and Marine Observations
- Map Layers showing both forecast and observations.
- Mountain weather forecasts
- Surface pressure charts
- Historic regional climate data, averages and anomalies

Detailed mapping products of Great Britain can be accessed and used that include:

- A generalised and simplified street level map
- Electoral and administrative boundaries
- All of the current postcode units in Great Britain
- Regional scale height grid and contours
- A regional overview of road networks, railway lines, cities and rural wooded areas
- A searchable gazeteer of road names
- Over 250,000 placenames and areas of interest

What formats does PDG Open Data come in?

PDG members have learnt that releasing data alone is of limited value and have therefore developed dialogues with their user communities to ensure that the data that is made available in ways that are as useable and accessible as possible. This includes releasing data in a number of formats. Two areas of particular interest are Linked data and the use of Application Programming Interfaces.

Linked Data

Linked Data is data that provides a common reference point to allow other data sets to link back to it. Connecting different datasets in this way significantly increases the value of the data overall and underpins the development of the Semantic Web[6].

As well as supporting the development of this at both a national and international level several of the PDG Open Data sets are already available as Linked Data:

- Companies House has developed a new service to supply basic company details, using a simple Uniform Resource Identifier (URI) for each company on the register.
- Land Registry make available their House Price Index with links to ONS and Ordnance Survey and their Price Paid Data is linked with Ordnance Survey data.
- Ordnance Survey offer three products as Linked Data: Boundary Line, Code-Point Open and the 1:50 000 Scale Gazetteer. Collectively these support linking between administrative boundaries, post code units and 250,000 distinctive place and feature names.
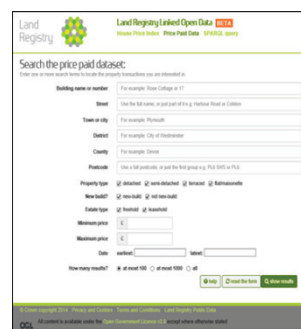
Application Programming Interface (API)

Land Registry, Met Office and Ordnance Survey all make some or all of their Open Data accessible through APIs, and Companies House will follow suit very shortly. This approach allows databases to be easily accessed and allows automation of access to the data – opening new opportunities for products to be built. Examples of the APIs include:

- Land Registry has provided an API to support access to its Linked Data.
- Met Office offers DataPoint - which provides open data in a way that allows developers to create innovative applications using real time weather information. Over the last six months there have been 40 million data requests with an average daily data download of over 50 GB.
- Ordnance Survey provides OS OpenSpace which enables developers to embed Ordnance Survey mapping in their solution. This has over 10,000 registered users and over 2,000 active sites running.
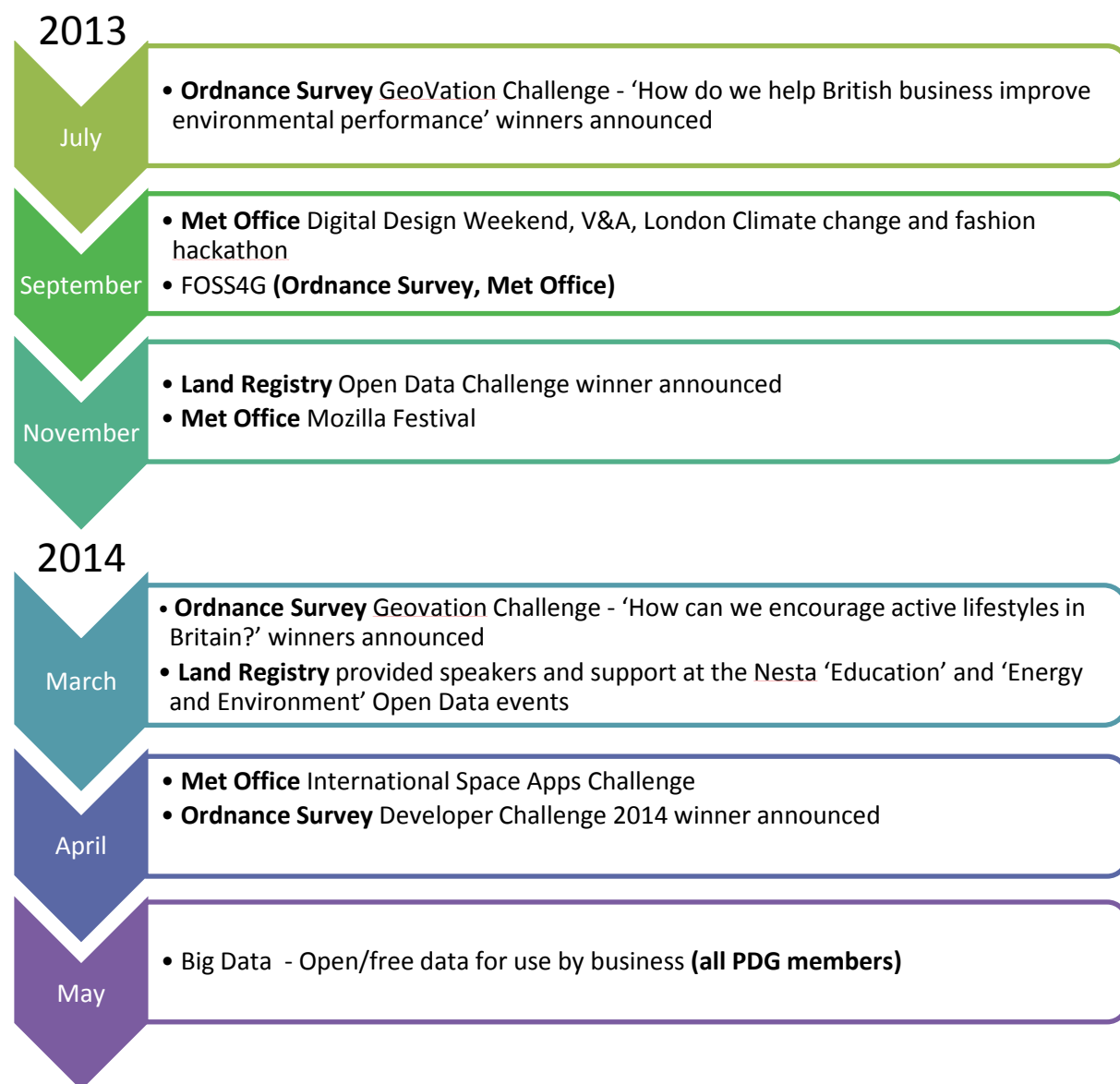
---

[6] The Semantic Web refers to the idea of an internet populated with machine readable data. More information can be found at: http://en.wikipedia.org/wiki/Semantic_Web
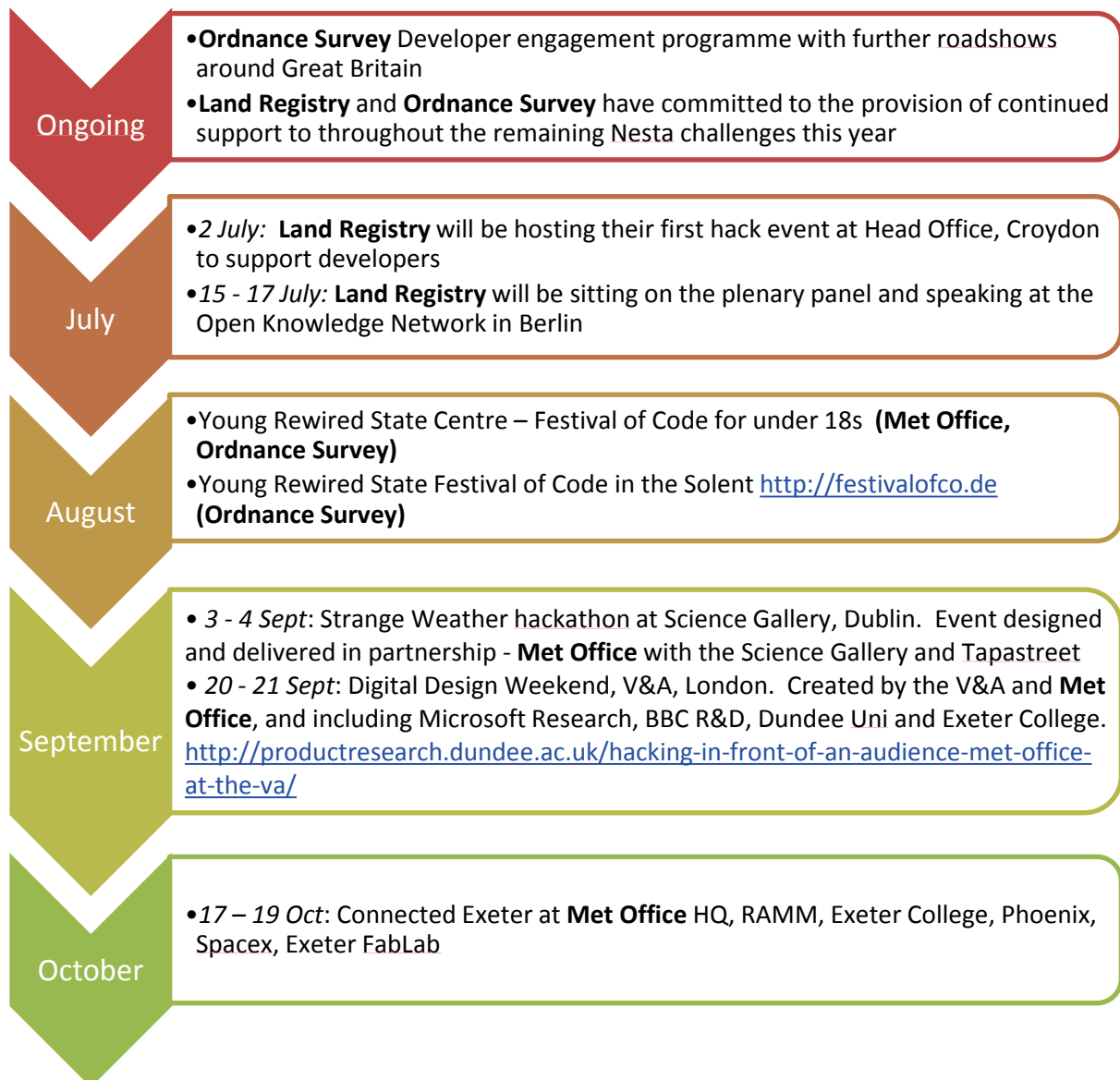
What support is available to help me use it?

In addition to providing their data in different formats PDG organisations have worked hard to communicate the availability of their data through a range of medium and events and offer direct support to users. Example of these includes:

- The 25 Ordnance Survey open data master classes run since April 2011 across Great Britain which have attracted over 900 attendees;
- A range of challenge events such as Ordnance Survey's GeoVation Challenge Series and the Land Registry's Open Data Challenge;
- Hack days such as: Met Office's collaboration with the Mozilla Festival and the International Space Agency; and Met Office and Ordnance Survey sponsoring of FOSS4G;
- Customer Focus Groups and Information Days held by Companies House around the country throughout the year;
- The Met Office set up a DataPoint Google Group in October 2012 that enables users of the data to ask questions and provide support to each other. They now have 300 members.

Recent events

2013

July
- **Ordnance Survey** GeoVation Challenge - 'How do we help British business improve environmental performance' winners announced

September
- **Met Office** Digital Design Weekend, V&A, London Climate change and fashion hackathon
- FOSS4G **(Ordnance Survey, Met Office)**

November
- **Land Registry** Open Data Challenge winner announced
- **Met Office** Mozilla Festival

2014

March
- **Ordnance Survey** Geovation Challenge - 'How can we encourage active lifestyles in Britain?' winners announced
- **Land Registry** provided speakers and support at the Nesta 'Education' and 'Energy and Environment' Open Data events

April
- **Met Office** International Space Apps Challenge
- **Ordnance Survey** Developer Challenge 2014 winner announced

May
- Big Data - Open/free data for use by business **(all PDG members)**

Still to come…

Ongoing
- **Ordnance Survey** Developer engagement programme with further roadshows around Great Britain
- **Land Registry** and **Ordnance Survey** have committed to the provision of continued support to throughout the remaining Nesta challenges this year

July
- *2 July:* **Land Registry** will be hosting their first hack event at Head Office, Croydon to support developers
- *15 - 17 July:* **Land Registry** will be sitting on the plenary panel and speaking at the Open Knowledge Network in Berlin

August
- Young Rewired State Centre – Festival of Code for under 18s **(Met Office, Ordnance Survey)**
- Young Rewired State Festival of Code in the Solent http://festivalofco.de **(Ordnance Survey)**

September
- *3 - 4 Sept*: Strange Weather hackathon at Science Gallery, Dublin. Event designed and delivered in partnership - **Met Office** with the Science Gallery and Tapastreet
- *20 - 21 Sept*: Digital Design Weekend, V&A, London. Created by the V&A and **Met Office**, and including Microsoft Research, BBC R&D, Dundee Uni and Exeter College. http://productresearch.dundee.ac.uk/hacking-in-front-of-an-audience-met-office-at-the-va/

October
- *17 – 19 Oct*: Connected Exeter at **Met Office** HQ, RAMM, Exeter College, Phoenix, Spacex, Exeter FabLab

What plans are there for doing more?

Although the Public Data Group is already making available a very large amount of Open Data and providing support to users accessing it they remain committed to making more data available under the Open Data standard where they can and working with others to understanding the economic benefits of doing so. Responding to user requests they will also publicise their release plans in advance so that the developer community is better able to take advantage of new releases. Those datasets which offer the greatest potential for broader economic growth will be prioritised for release, along with those that are most easily released.

The following section sets out PDG's key actions for 14/15 and demonstrates how they will continue to use their data and expertise to act as exemplars supporting the Open Data agenda. This commitment will be updated in one year's time.

Making a range of substantial data sets available as Open Data

- **Companies House** will be making all of their digital data available free of charge from the second quarter of 2015. Following a fundamental redesign of its digital services to create a truly open register. This will bring significant benefits to all those who use the information on the register – from members of the public, to businesses small and large – and will be a significant step forward in corporate transparency.

- **Land Registry** intends to develop a Publication Platform which will provide easy access to all their existing and future datasets – both free and charged. In 14/15 they intend to make the whole Index Map polygon layer covering England and Wales available at a cost recovery price and will release their Price Paid Data for commercially owned properties for free by March 2015.

- **Met Office** is undertaking a programme of work to make its existing data more useful, for example by transferring data available to download from the Met Office website already on to the DataPoint API and providing INSPIRE compliant formats. In addition it is developing a system to create the National Archive for the Nations Memory of the Weather – a collection of the historical observation data - and a selection of this will be available as Open Data.

- **Ordnance Survey** will continue to develop and support the Open data agenda through a combined offer of data, services and accessibility. There will be ongoing investment in the OS OpenData portfolio; including the planned addition of a new Derived River Network, an enhanced Gazetteer and enhancements to OS Street View that around depiction of roads and railways and new features added such as: car parks, major paths, major cycle routes and hill shading. Further content is being developed. In addition they are working with the Department for the Environment, Food and Rural Affairs to provide consultancy, technology and to enhance public access (through a portal) to Rights of Way data.

Setting the standard for the quality of UK open data

- In order to promote and support the increased usage of Linked Data across the public sector and beyond, PDG members, as part of the Public Sector Linked Data Working Group, will work to create a set of 'Core Reference Linked Data'. The intention is that these Data Sets would be accessible along with a range of tools to form a central hub to allow others across the public sector and beyond to build a growing network of Linked Data.

- As a first step there will be an initial user engagement process - both within and outside the private sector - to identify those datasets that would most usefully and feasibly form the spine of the Core Reference data sets. At the same time work will be undertaken to develop a single point of access to these datasets and Linked Data tools.

- To expand the range of PDG Linked Data available, we will explore the feasibility of linking core Companies House data with Ordnance Survey data to allow registered businesses to be better analysed geographically.

- Ensuring that all PDG data sets available on Data.Gov.UK have been assessed against the Open Data Certificate from the Open Data Institute by December 2014.

Further enhancing support to developers to generate economic growth

- A formal review will take place of the 'Developer Licences' (where paid for data is available for free under certain conditions) to consider how these licences could be made more developer-friendly. This review will also include resolving issues associated with the use of derived data sets. It will report in Dec 2014. This review will involve experts from across the public sector, the Open Data Institute and a range of data focused companies including Doorda, Geolytix and RM Education. The terms of reference will be published shortly.

- In addition, Ordnance Survey are implementing a try-before-you-buy scheme for new commercial partners (3 months royalty free access to premium data).

- PDG will look to combine and extend existing challenge programmes – starting with GeoVation Challenge that will be jointly sponsored by Land Registry and Ordnance Survey and launched on 10 September 2014.

- Each PDG member will offer a user group forum by 2015 ensuring that users have a space to discuss issues and access to direct advice.

Annex A: Further details on the data available from PDG members

All data is either available at cost recovery or as open data. Companies House has a range of open data products. These include basic information on companies, that is accessed over 230 million times a year, to a free bulk accounts data product:

Free Basic Company Information

The Free Company Data Product is a downloadable data snapshot containing basic company data of live companies on the register. CH's free company data product contains basic company data of all live companies on the register in reusable format. The files have been downloaded over 20,000 times since its launch.

Free Accounts Data Product

The accounts data product is a free, downloadable file containing approximately 2million accounts that have been filed digitally with CH (available as iXBRL data as a .html file, or XBRL data as a .xml file). This puts the UK in a world leading position in providing accounts data free of charge.

The accounts files contain accounts elements including:

- Directors' Report
- Auditors Report
- Profit and Loss
- Balance Sheet
- Notes to the Balance Sheet
- Footnotes (for micro-entity accounts)
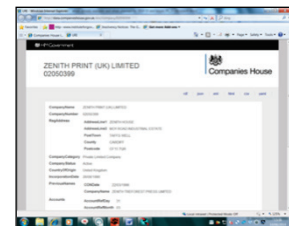
iXBRL submissions reference UK GAAP taxonomy and contains 6,654 possible elements, of which a sub-set of minimum tagging list of 1,253 elements has been determined by HMRC.XBRL submissions reference multiple UK GAAP taxonomies, which will have similar numbers of elements.

URI

The URI is a service using a simple Uniform Resource Identifier (URI) for each company on the register. The URI is a unique web address that represents the company (using the 'business.data.gov.uk' domain); and will return basic company details for that company (as per the fields above for the free basic information product). It has been accessed over 287 million times.

Mobile App

The CH mobile app allows free access to basic company details, with additional features such as statistics and frequently viewed companies. It has been downloaded by approximately over 104,000 users.
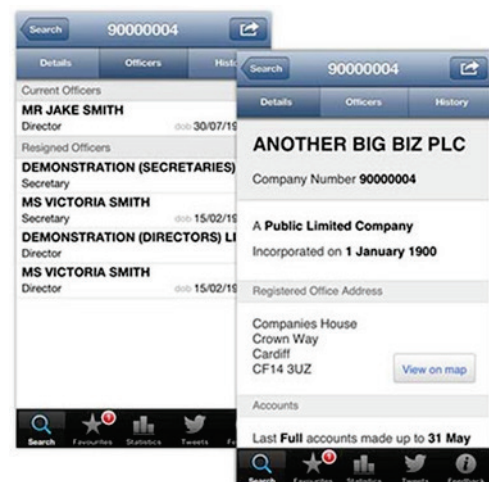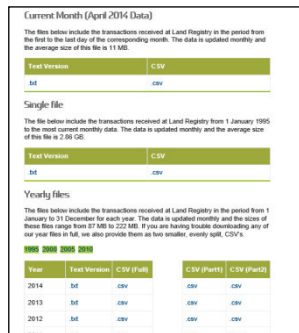
All information recorded on the Register is available through the Open Register albeit at a cost per title. Commercial Services are available which provide products and services, using Register data in a useable and accessible format.
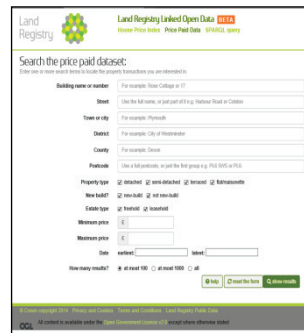
The existing open data offer includes:

Price paid data - dating back to 1995 in excel, csv and linked data formats for records. This means that there are available more than 19 million definitive records of monthly residential property price data (sold for market value). Available in txt , csv file or linked data.

Price Paid Data Download   Price Paid Report Builder

The INSPIRE dataset – which depicts the legal extent of all Freehold properties in England and Wales is one of the most popular and frequently downloaded datasets on Data.Gov.UK.

House Price Index – a comparison of the average house price today with what it was in January 1995, with the index set then at 100. It includes figures at national, regional, county and London borough level. Available as the monthly pdf doc, background data in csv/excel and linked data.
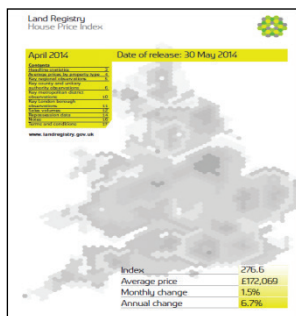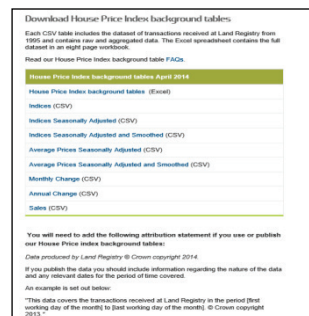
PDF doc   Search the House Price Index (beta)   Background Data Tables for download
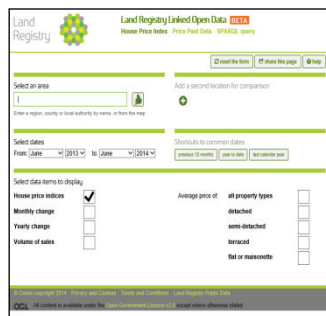
Transaction Data - provides information about the number and types of applications that have been completed by conveyancers.
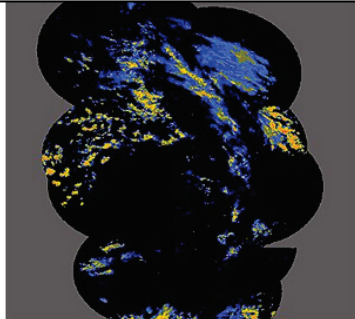
The Met Office has a large range of data available under the Open Government Licence. In broad terms it can be described as the data behind that which is published on the Met Office Public Website. Data which isn't Open is made available either through Met Office Standard Terms and Conditions or for research purposes.
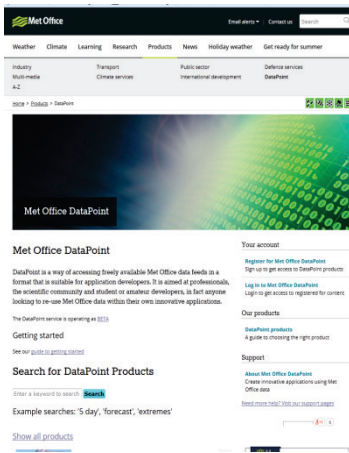

DataPoint is the Met Office API which makes available both daily and 3 hourly forecasts, updated hourly, for over 5,000 locations, as open data. The data is provided in a format that is suitable for application developers. Over the last six months there have been around 40 million data requests with an average daily data download of around 54 GB. Examples of the data accessible through Data Point include:

Rainfall Radar - a map overlay for the UK showing the rainfall radar image for the UK.	**Total cloud cover forecast map layer** - a map overlay for the UK showing a forecast of total cloud cover for the UK.
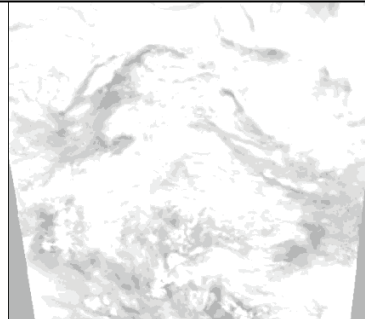
The underpinning model data used to create the Public Weather Service outputs is available through the ECOMET Wholesale Catalogue. A range of models are made available from Global atmospheric to Northwest Shelf Seas. This data is primarily used by competing weather service providers both UK and overseas.

A range of Historic data is currently available at no charge for private or research use through the Library or from a dedicated sales team for commercial use.

The UK academic community have access to both archived model data and observations through an arrangement with BADC (British Atmospheric Data Centre), free of charge.

The Ordnance Survey offers an extensive suite of commercial products which are available for free for developmental use in either a limited form, for a limited duration or under limits to their use.

OS Open Data provides over 10 quality assured, continuously updated products, from postcodes and boundaries to digital maps, providing a one-stop shop to start using their products such as:

Boundary-Line provides Electoral and administrative boundaries. (Image shows Boundary-Line over OS VectorMap District)

OS VectorMap District allows you to overlay your own information on a customisable map background.

OS Terrain 50 Regional scale height data that will improve your analysis work.

Meridian 2 - Digital maps customisable for communication and topographical themes and route planning.

Code-Point Open offers all of the current postcode units in Great Britain. (Image shows Code-Point Open over OS VectorMap District)

OS StreetView provides a generalised and simplified street level map, ideal for city-centre plans.

Ordnance Survey also provides an API (OS OpenSpace) which enables developers to embed Ordnance Survey mapping in their solution. This has over 10,000 registered users and over 2,000 active sites running.