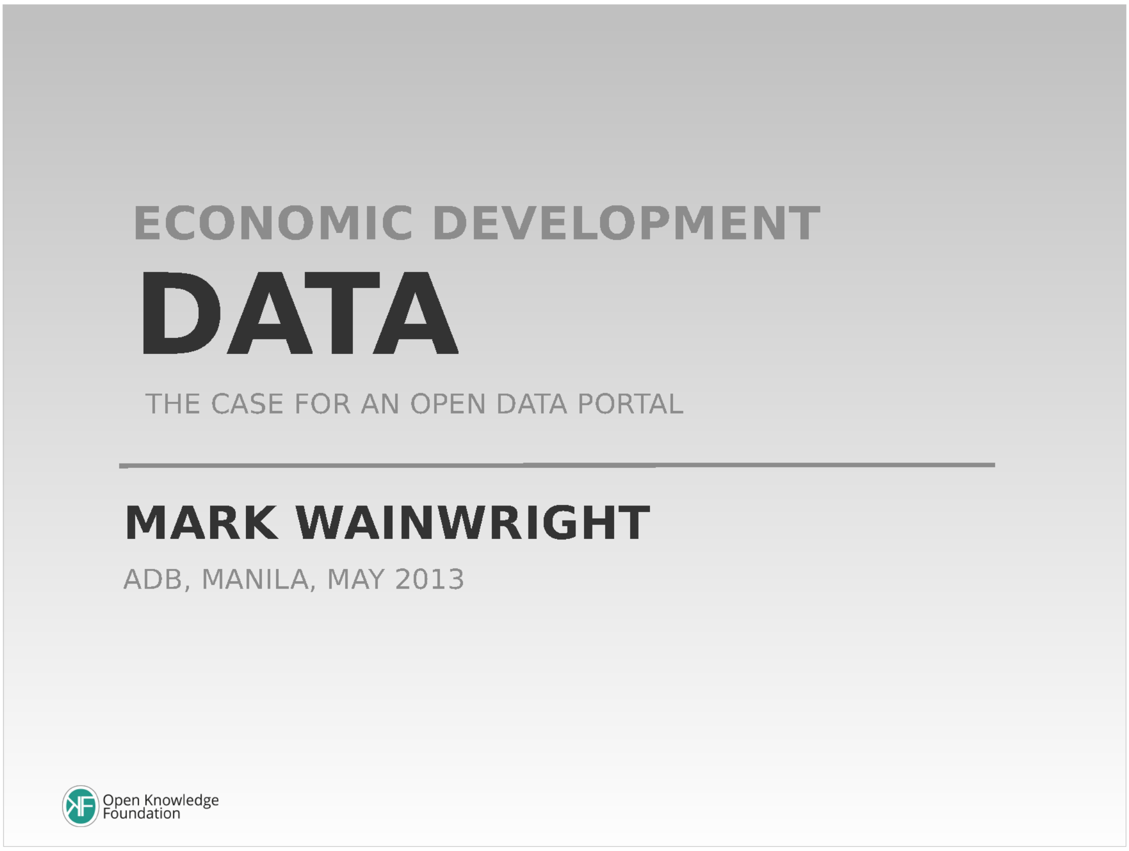


ECONOMIC DEVELOPMENT DATA

THE CASE FOR AN OPEN DATA PORTAL

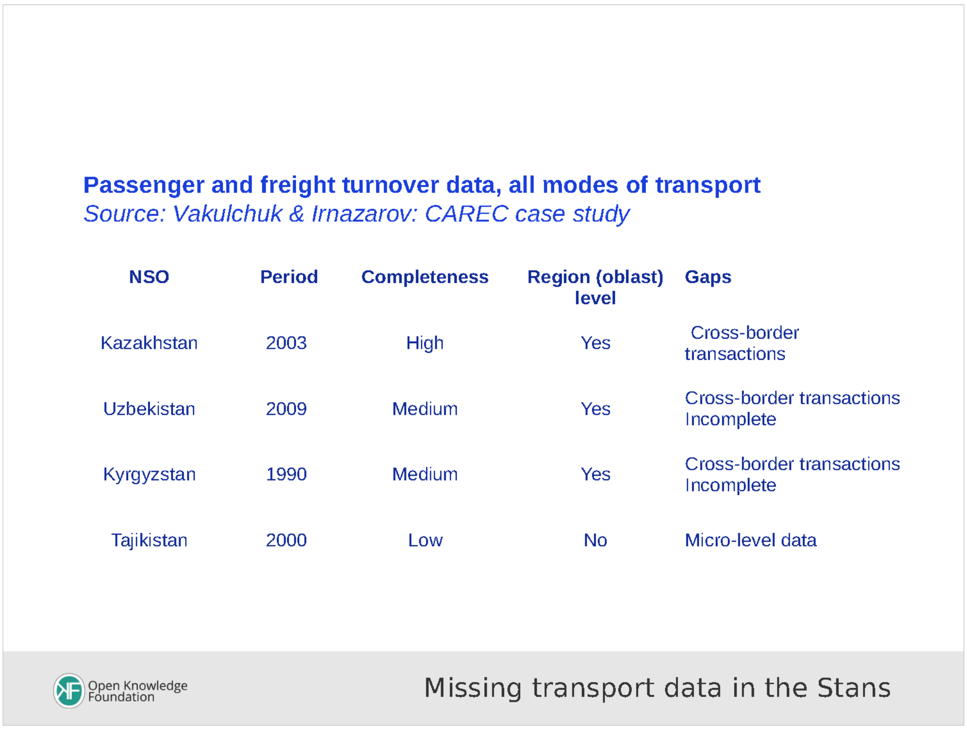
MARK WAINWRIGHT

ADB, MANILA, MAY 2013

Open Knowledge Foundation

The Open Knowledge Foundation aims to open up knowledge around the world to make it accessible and useful. It is home to many activities, including an Open Government working group and a consulting division. OKF is a non-profit organisation registered in the UK.

This is not an ADB material. The views expressed in this document are the views of the author/s and/or their organizations and do not necessarily reflect the views or policies of the Asian Development Bank, or its Board of Governors, or the governments they represent. ADB does not guarantee the accuracy and/or completeness of the material's contents, and accepts no responsibility for any direct or indirect consequence of their use or reliance, whether wholly or partially. Please feel free to contact the authors directly should you have queries.



Passenger and freight turnover data, all modes of transport

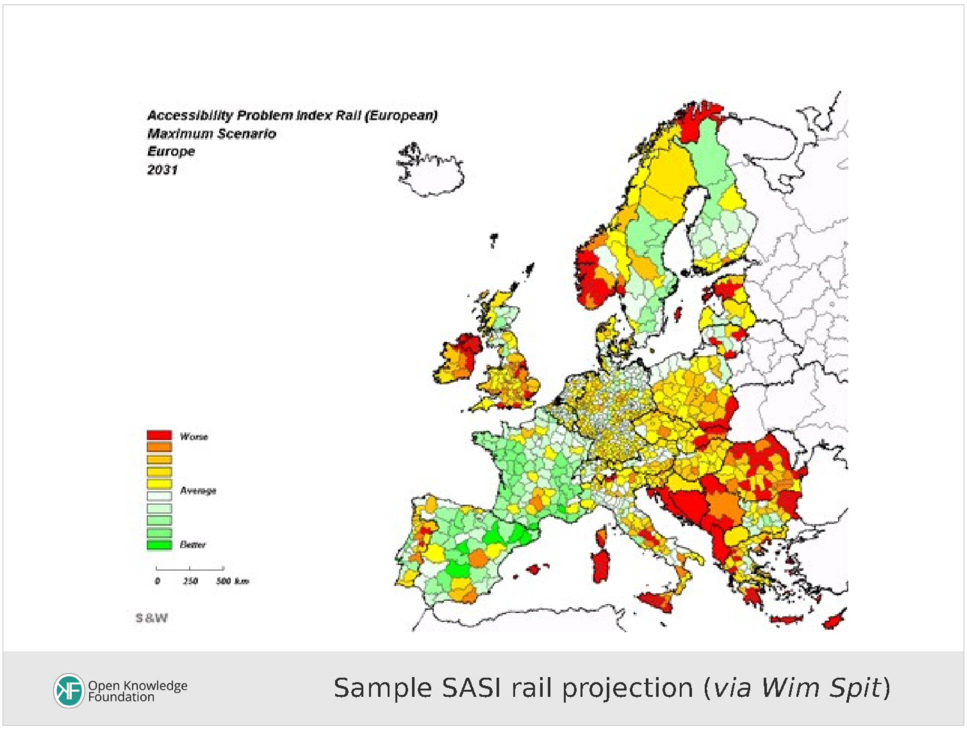
Source: Vakulchuk & Irnazarov: CAREC case study

NSO	Period	Completeness	Region (oblast) level	Gaps
Kazakhstan	2003	High	Yes	Cross-border transactions
Uzbekistan	2009	Medium	Yes	Cross-border transactions Incomplete
Kyrgyzstan	1990	Medium	Yes	Cross-border transactions Incomplete
Tajikistan	2000	Low	No	Micro-level data

Open Knowledge Foundation

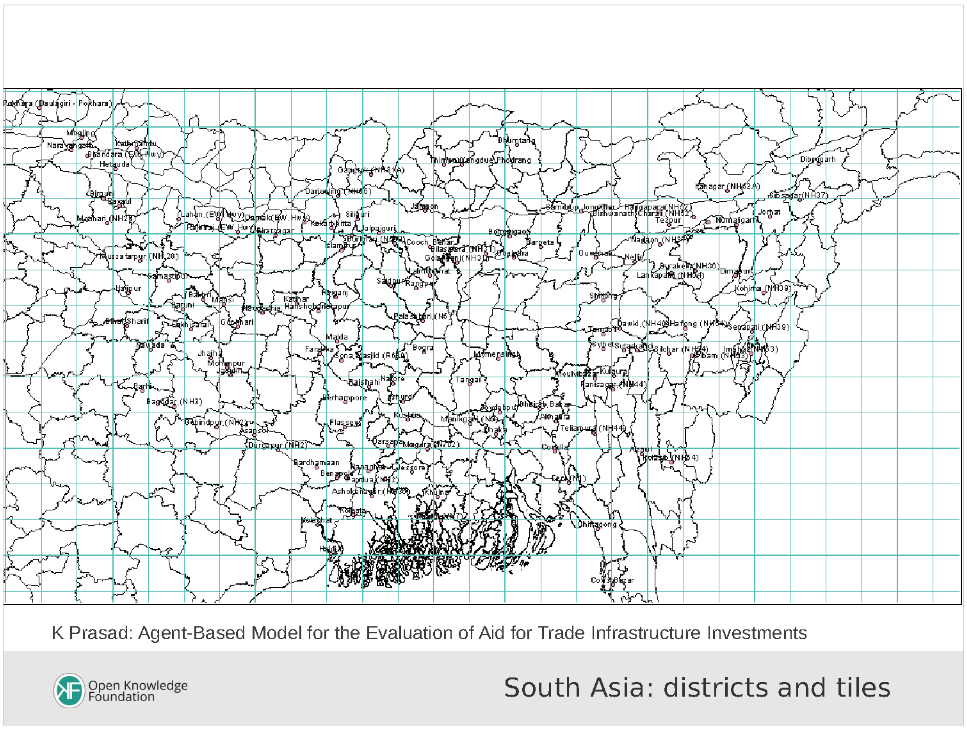
Missing transport data in the Stans

First, the bad news - a lot of data that one would like is missing, incomplete, or inconsistent, or is collected in incommensurate ways across national borders.


Accessibility Problem Index Rail (European)
Maximum Scenario
Europe
2031
Worse
Average
Better
0 250 500 km
S&W
Open Knowledge Foundation

Sample SASI rail projection (*via Wim Spit*)

This is a projection (of rail accessibility) in Europe, from the SASI model, which models socio-economic and spatial impact of different transport policy options. It's a rich model but uses base data of many variables (regional GDP, population by age and sex, labour force participation, migration, etc) over 1330 districts, as well as transport network data. This kind of analysis can provide valuable insights and guidance but requires a consistent data framework across the region.

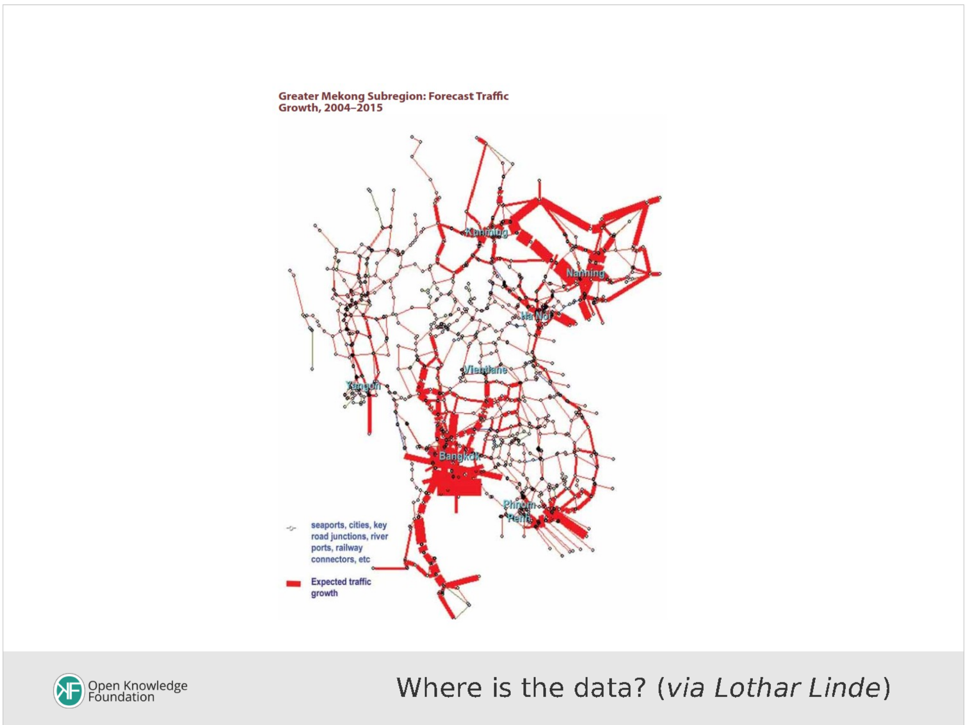


K Prasad: Agent-Based Model for the Evaluation of Aid for Trade Infrastructure Investments

Open Knowledge Foundation

South Asia: districts and tiles

While many datasets are collected at district level, some international datasets are divided into approximately square tiles covering 1 degree of longitude and latitude, particularly where the data is compiled or estimated from satellite images. This provides a further challenge when data from both types of source needs to be integrated. It's essential for data to be geo-coded so that data from different sources can be overlaid. SASEC used tile data, together with data on transport links between tiles, to model local effects of improving transport links to reduce market friction in South Asia, shown here with both tiles and districts.

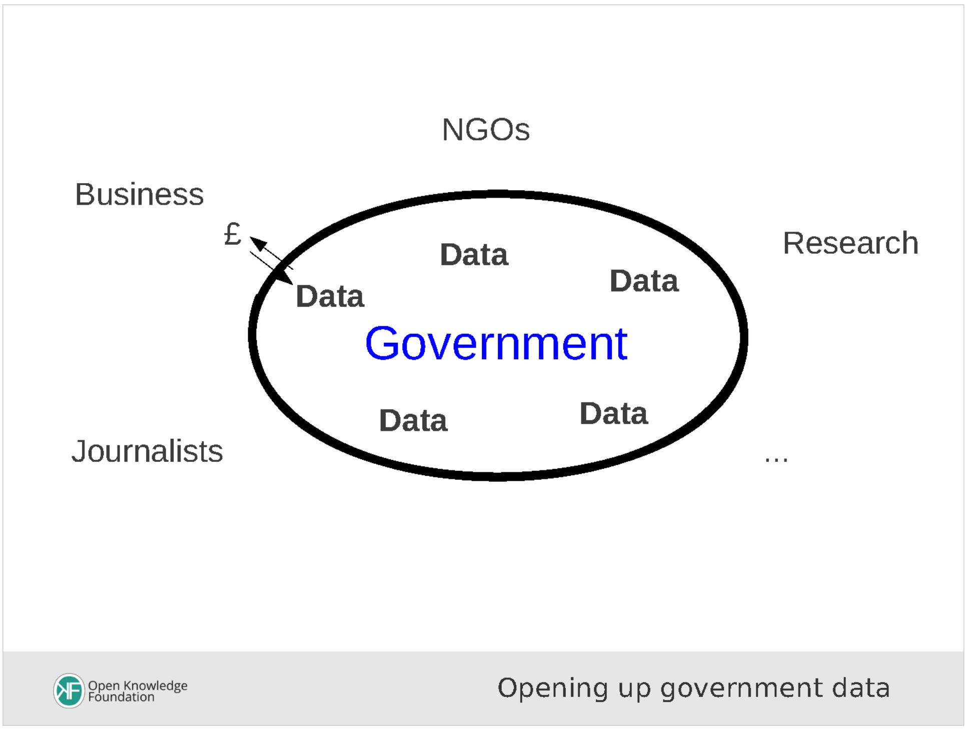


Greater Mekong Subregion: Forecast Traffic Growth, 2004–2015

Open Knowledge Foundation

Where is the data? (*via Lothar Linde*)

Speaking of transport data, this graphic was shown me by Lothar Linde of the GMS project. It shows projected traffic flows in the GMS region (thicker lines represent more traffic; Bangkok is seen to be a hub). The graphic appeared in an ADB report. Lothar wanted to use the underlying data, but he was unable to trace it despite several months of searching. Solving the problem of data that goes missing or can't be found is both far more urgent than gathering new data, and far easier.



NGOs

Business

£

Data

Data

Data

Government

Research

Journalists

Data

Data

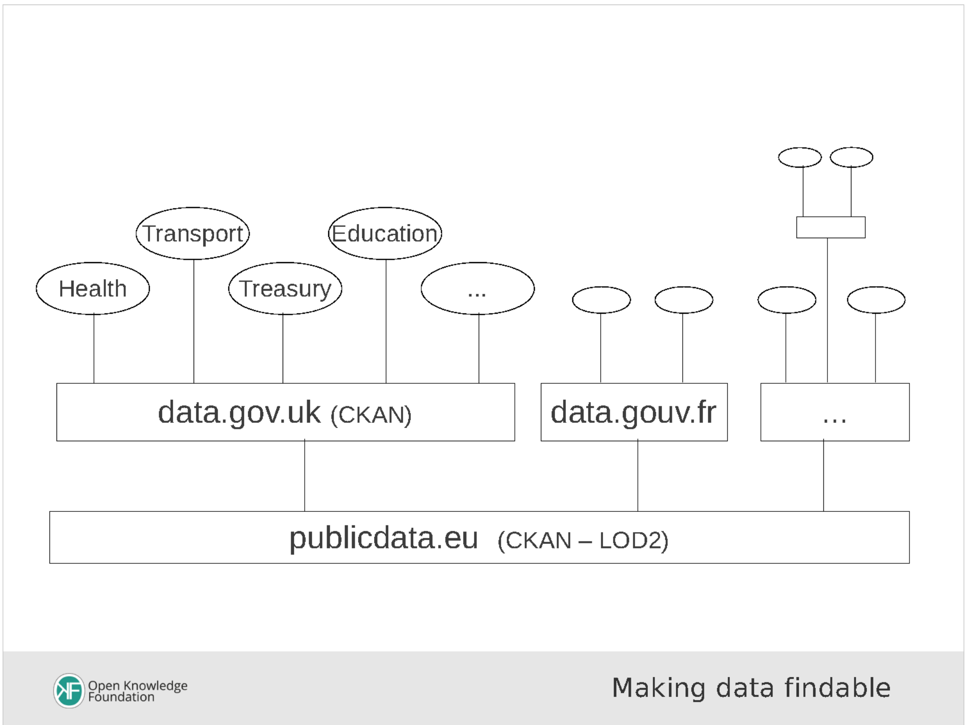
…

Open Knowledge Foundation

Opening up government data

Governments necessarily collect a lot of data, which traditionally was kept secret (or perhaps monetised) but is increasingly being released openly and freely, with Open Government Data (OGD) initiatives across the world, though very few in Asia. Benefits of OGD include transparency, increased participation and engagement, and socio-economic value. Data can be used to power new types of services and generate value and economic activity. There are also obvious benefits to research. Increasingly many governments are being convinced that the benefits far outweigh the cost of lost revenue.

Thomas Herndon is a graduate student who, in a famous recent story, found a crucial error in the data behind an influential paper of Reinhart and Rogoff (2010). To get hold of the spreadsheet he only needed to e-mail and ask for it, seemingly a small barrier. But if their data had been provided for download on the internet when the paper was published, would it have taken 3 years for the error to come to light?



Health
Transport
Treasury
Education
...

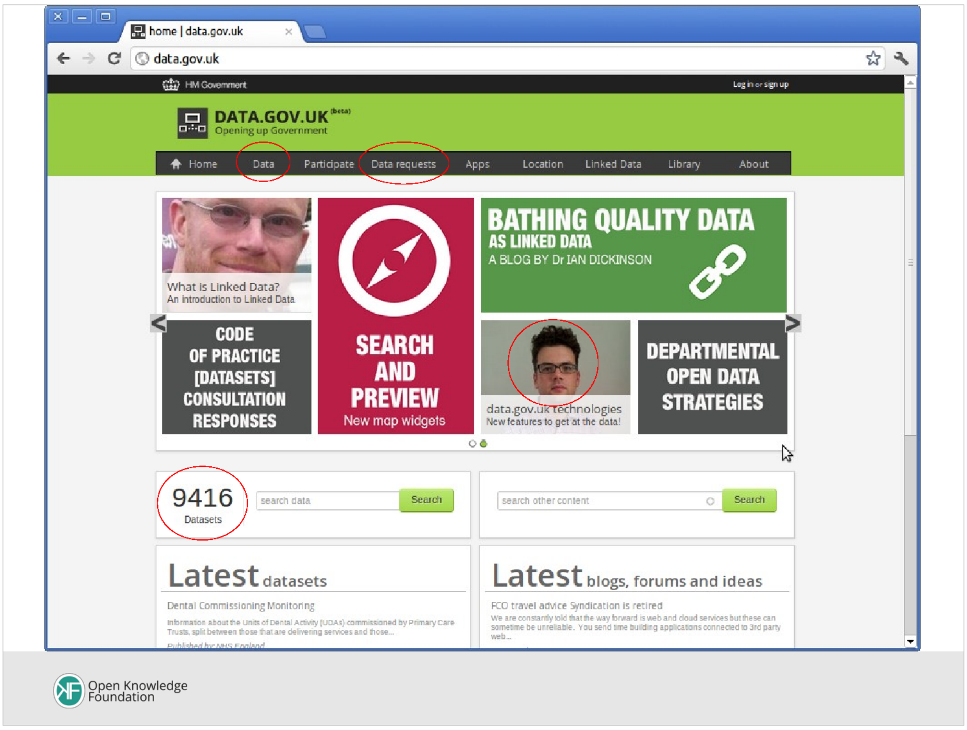
data.gov.uk (CKAN)
data.gouv.fr
...

publicdata.eu (CKAN – LOD2)

Open Knowledge Foundation

Making data findable

Even if data is available for download and can be freely re-used, it's not much use unless people can find it. UK government data is released by hundreds of different departments and bodies, each with their own website - how to find it? The answer is in data.gov.uk, a data catalogue or 'portal' where all of it is registered, and can be found easily with searching, filtering, etc. data.gov.uk currently knows of nearly 10,000 datasets. publicdata.eu is an experimental EU-wide catalogue that 'harvests' data records from a number of different national and local portals. Though these portals are principally used as a catalogue, a data portal can also be used to store the data itself. We'll look in more depth at later at CKAN, an example data portal.



Dental Commissioning Monitoring

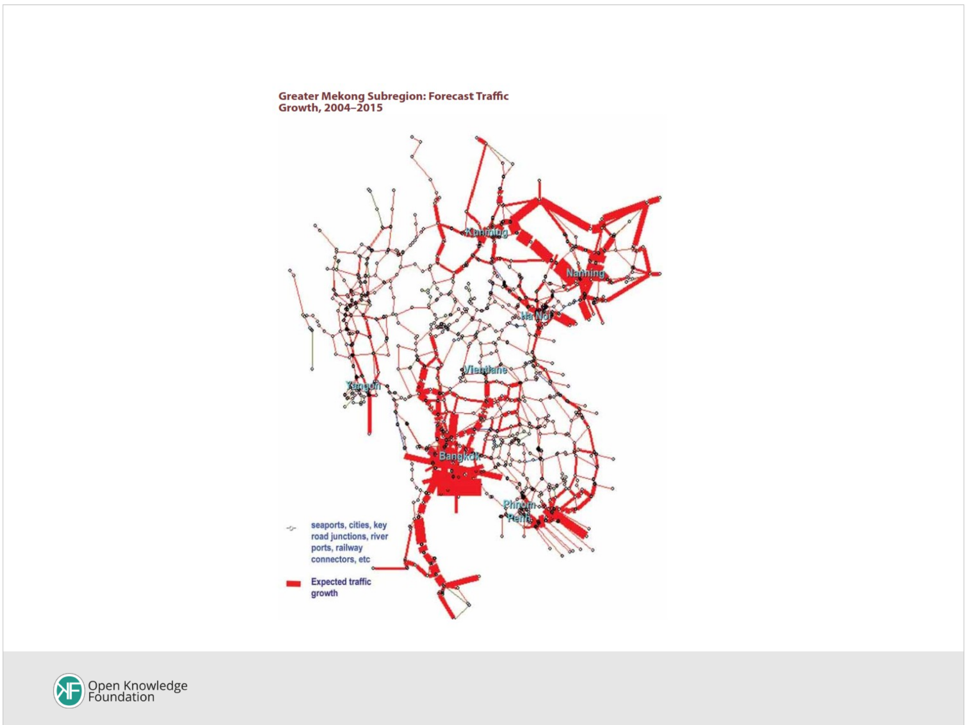
Information about the Units of Dental Activity (UDAs) commissioned by Primary Care Trusts, split between those that are delivering services and those...

Published by: NHS England

FCO travel advice Syndication is retired

We are constantly told that the way forward is web and cloud services but these can sometime be unreliable. You send time building applications connected to 3rd party web...

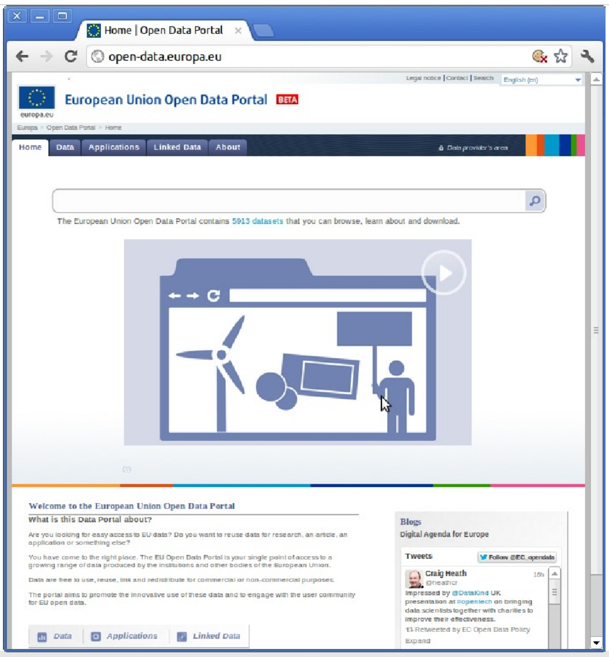
data.gov.uk was mentioned above. Note the number of datasets, the 'data' tab (giving more advanced search options), and the 'Request data' tab, allowing data users to ask for particular datasets to be made open if they are not already on the site.



Greater Mekong Subregion: Forecast Traffic Growth, 2004–2015

Kunming
Nanning
Ha Noi
Vientiane
Yangon
Bangkok
Phnom Penh
seaports, cities, key road junctions, river ports, railway connectors, etc
Expected traffic growth
Open Knowledge Foundation

If ADB had had a data portal and a policy of publishing data used in reports etc, the data behind this visualisation would probably not have gone missing. The example shows that it is valuable to publish model data as well as baseline data.



European Union Open Data Portal BETA

The European Union Open Data Portal contains 5913 datasets that you can browse, learn about and download.

Welcome to the European Union Open Data Portal

What is this Data Portal about?

Are you looking for easy access to EU data? Do you want to reuse data for research, an article, an application or something else?

You have come to the right place. The EU Open Data Portal is your single point of access to a growing range of data produced by the institutions and other bodies of the European Union.

Data are free to use, reuse, link and redistribute for commercial or non-commercial purposes.

The portal aims to promote the innovative use of these data and to engage with the user community for EU open data.

Blogs

Digital Agenda for Europe

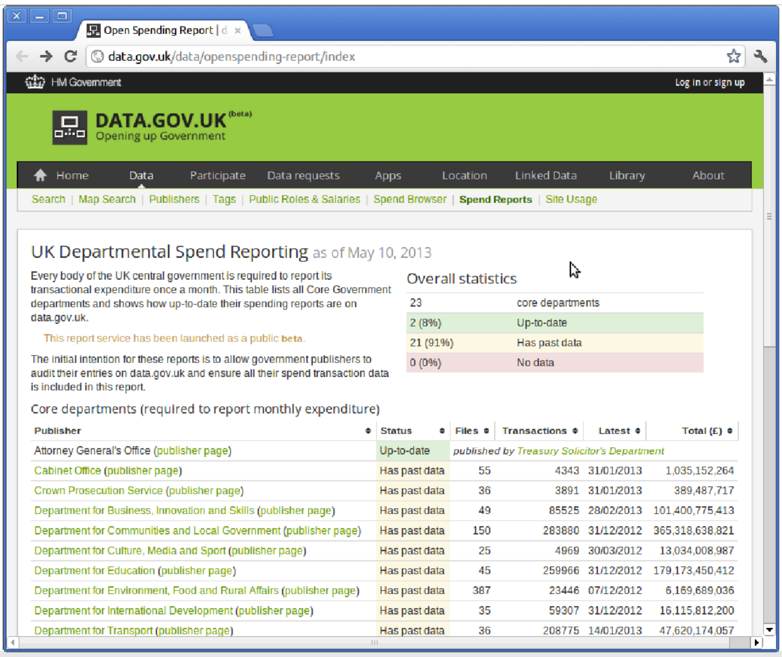
Craig Heath @heathcr 16h

Impressed by @DataKind UK presentation at #opentech on bringing data scientists together with charities to improve their effectiveness.

Retweeted by EC Open Data Policy

Open Knowledge Foundation

The EU is another major publisher of data. Note that this portal is for the EU's own data - it is not the same as publicdata.eu (mentioned above) which aggregates data records from member states. A number of EU offices and bodies publish data, including their own administrative data, but at present the great majority of the data that can be found here is from Eurostat, the EU statistics agency.



UK Departmental Spend Reporting as of May 10, 2013

Every body of the UK central government is required to report its transactional expenditure once a month. This table lists all Core Government departments and shows how up-to-date their spending reports are on data.gov.uk.

This report service has been launched as a public **beta**.

The initial intention for these reports is to allow government publishers to audit their entries on data.gov.uk and ensure all their spend transaction data is included in this report.

Overall statistics

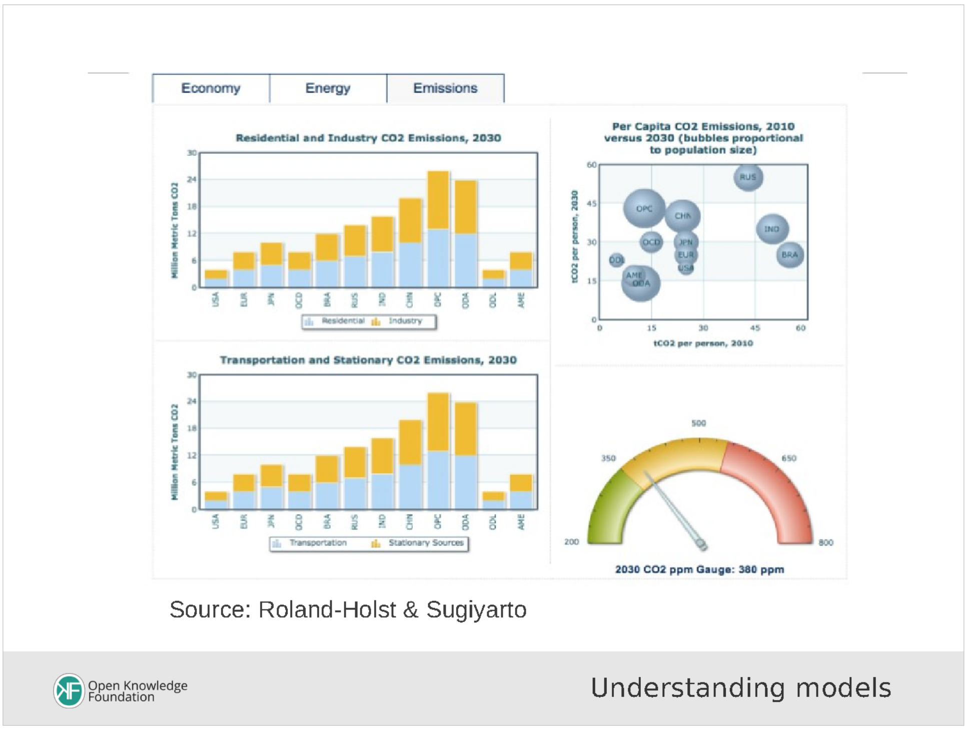
23	core departments
2 (8%)	Up-to-date
21 (91%)	Has past data
0 (0%)	No data

Core departments (required to report monthly expenditure)

Publisher	Status	Files	Transactions	Latest	Total (£)
Attorney General's Office (publisher page)	Up-to-date	*published by Treasury Solicitor's Department*			
Cabinet Office (publisher page)	Has past data	55	4343	31/01/2013	1,035,152,264
Crown Prosecution Service (publisher page)	Has past data	36	3891	31/01/2013	389,487,717
Department for Business, Innovation and Skills (publisher page)	Has past data	49	85525	28/02/2013	101,400,775,413
Department for Communities and Local Government (publisher page)	Has past data	150	283880	31/12/2012	365,318,638,821
Department for Culture, Media and Sport (publisher page)	Has past data	25	4969	30/03/2012	13,034,008,987
Department for Education (publisher page)	Has past data	45	259966	31/12/2012	179,173,450,412
Department for Environment, Food and Rural Affairs (publisher page)	Has past data	387	23446	07/12/2012	6,169,689,036
Department for International Development (publisher page)	Has past data	35	59307	31/12/2012	16,115,812,200
Department for Transport (publisher page)	Has past data	36	208775	14/01/2013	47,620,174,057

Open Knowledge Foundation

Among the data at data.gov.uk is departmental spending data - every department is supposed to publish this. The Cabinet Office team in charge of data.gov.uk put together this dashboard to show how well departments were doing. This sign of interest acted as a nudge to a number of departments that had not previously published any spending data. An Open Data policy is most effective when, as here, it comes from a high level - as with the recent US open data directive, issued by the President.

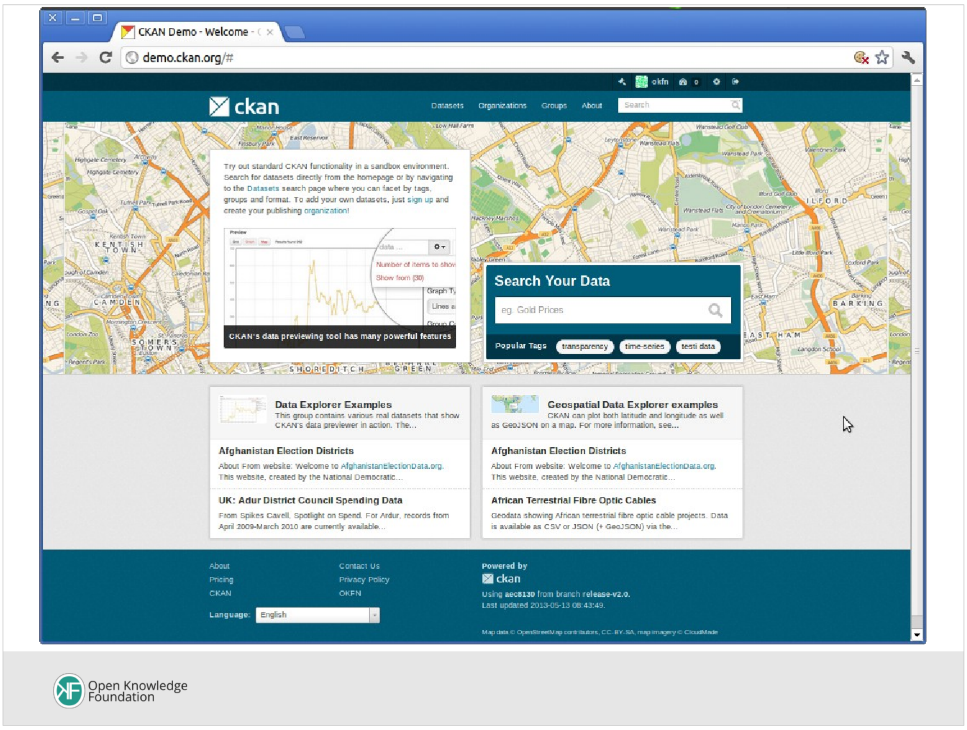


Source: Roland-Holst & Sugiyarto

Open Knowledge Foundation

Understanding models

An example of what a 'dashboard' interface to model data might look like, presenting it in a useful form to policy makers. (Of course, the data needs to be available first!) This example uses commercial data visualisation tools.



Try out standard CKAN functionality in a sandbox environment. Search for datasets directly from the homepage or by navigating to the Datasets search page where you can facet by tags, groups and format. To add your own datasets, just sign up and create your publishing organization!

CKAN's data previewing tool has many powerful features

Search Your Data

eg. Gold Prices

Popular Tags: transparency, time-series, testi data

Data Explorer Examples
This group contains various real datasets that show CKAN's data previewer in action. The...

Afghanistan Election Districts
About From website: Welcome to AfghanistanElectionData.org. This website, created by the National Democratic...

UK: Adur District Council Spending Data
From Spikes Cavell, Spotlight on Spend. For Ardur, records from April 2009-March 2010 are currently available...

Geospatial Data Explorer examples
CKAN can plot both latitude and longitude as well as GeoJSON on a map. For more information, see...

Afghanistan Election Districts
About From website: Welcome to AfghanistanElectionData.org. This website, created by the National Democratic...

African Terrestrial Fibre Optic Cables
Geodata showing African terrestrial fibre optic cable projects. Data is available as CSV or JSON (+ GeoJSON) via the...

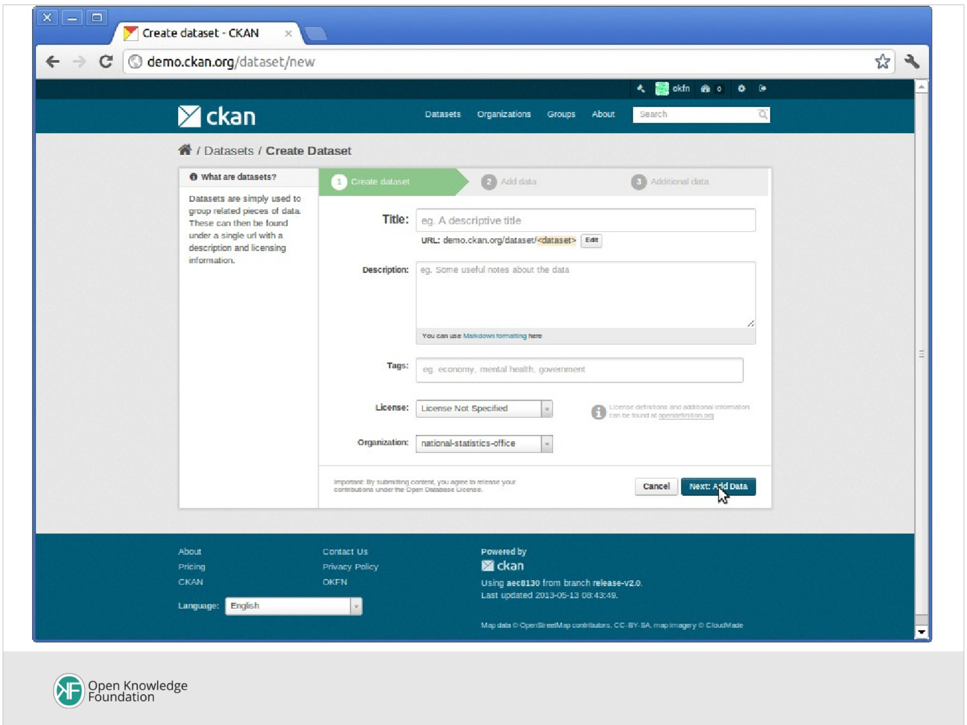
Language: English

Powered by ckan
Using aec8130 from branch release-v2.0.
Last updated 2013-05-13 08:43:49.

Map data © OpenStreetMap contributors, CC-BY-SA, map imagery © CloudMade

Open Knowledge Foundation

The default theme for CKAN 2.0, the free, open-source data portal software maintained by the Open Knowledge Foundation. CKAN has been developed since 2007 and now powers portals such as the UK and EU ones described above. By chance, the release of CKAN 2.0 was announced last week. The following slides show a very small number of its features, in particular in the area of geodata, whose importance was stressed above.



demo.ckan.org/dataset/new

/ Datasets / **Create Dataset**

What are datasets?

Datasets are simply used to group related pieces of data. These can then be found under a single url with a description and licensing information.

1 Create dataset · 2 Add data · 3 Additional data

Title: eg. A descriptive title

URL: demo.ckan.org/dataset/<dataset> Edit

Description: eg. Some useful notes about the data

You can use Markdown formatting here

Tags: eg. economy, mental health, government

License: License Not Specified

License definitions and additional information can be found at opendefinition.org

Organization: national-statistics-office

Important: By submitting content, you agree to release your contributions under the Open Database License.

About
Pricing
CKAN

Contact Us
Privacy Policy
OKFN

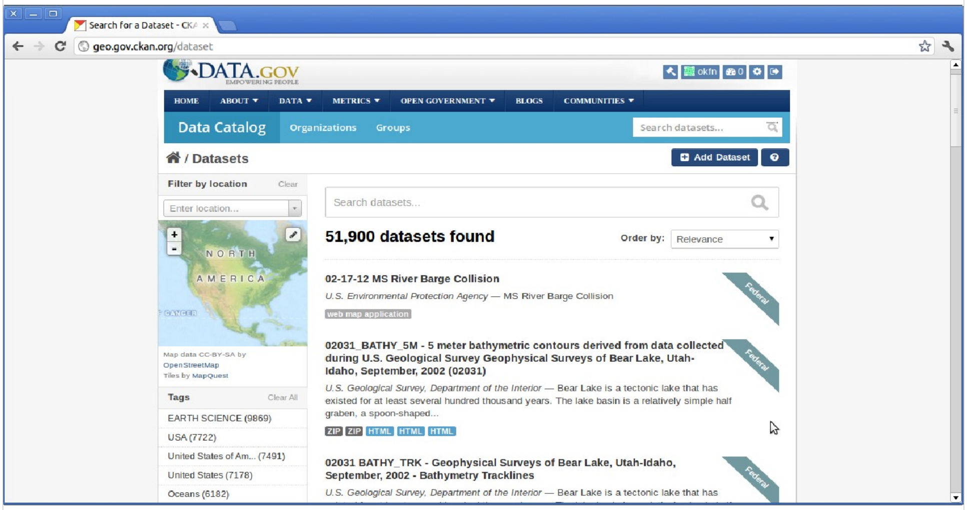
Language: English

Powered by ckan

Using aec8130 from branch release-v2.0.
Last updated 2013-05-13 08:43:49.

Map data © OpenStreetMap contributors, CC-BY-SA, map imagery © CloudMade

Data can be added to CKAN in a number of ways, including harvesting or ingesting in bulk from existing information systems. This shows part of the user interface for adding a dataset by hand. Various metadata is added such as title, description, etc (the metadata schema can be changed). Tags help users find the dataset when searching. Data should have enough metadata that users can find the data they need and know what data they are dealing with.



/ Datasets

Filter by location Clear

Enter location...

Map data CC-BY-SA by OpenStreetMap
Tiles by MapQuest

Tags Clear All

- EARTH SCIENCE (9869)
- USA (7722)
- United States of Am... (7491)
- United States (7178)
- Oceans (6182)

Search datasets...

51,900 datasets found

Order by: Relevance

02-17-12 MS River Barge Collision

U.S. Environmental Protection Agency — MS River Barge Collision

web map application

02031_BATHY_5M - 5 meter bathymetric contours derived from data collected during U.S. Geological Survey Geophysical Surveys of Bear Lake, Utah-Idaho, September, 2002 (02031)

U.S. Geological Survey, Department of the Interior — Bear Lake is a tectonic lake that has existed for at least several hundred thousand years. The lake basin is a relatively simple half graben, a spoon-shaped...

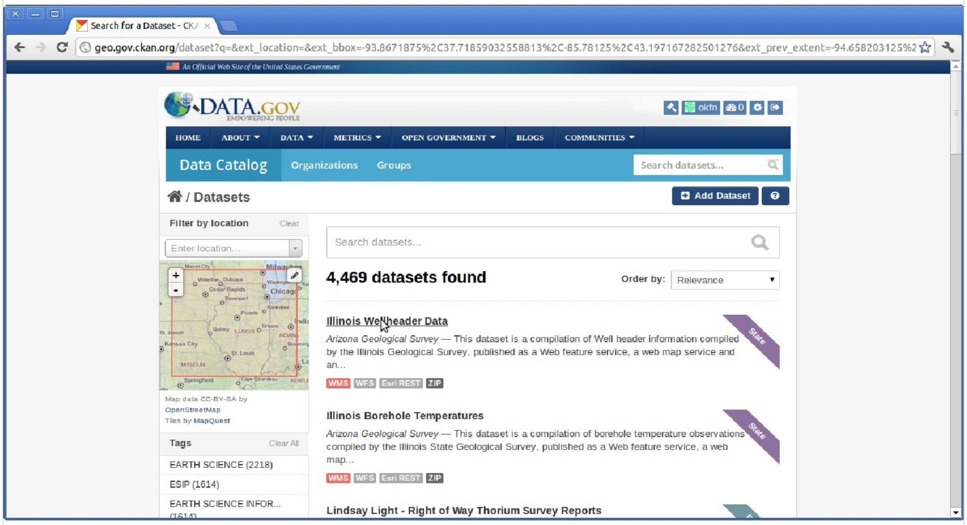
ZIP ZIP HTML HTML HTML

02031 BATHY_TRK - Geophysical Surveys of Bear Lake, Utah-Idaho, September, 2002 - Bathymetry Tracklines

U.S. Geological Survey, Department of the Interior — Bear Lake is a tectonic lake that has

Open Knowledge Foundation

data.gov, the US government data portal, is due to be relaunched (using CKAN) in a few weeks. This prototype has over 50,000 datasets. As well as searching by keywords, tags, publishing department, etc, note the map which allows searching by geographic extent where available.



geo.gov.ckan.org/dataset?q=&ext_location=&ext_bbox=-93.8671875%2C37.7185903255881 3%2C-85.78125%2C43.197167282501276&ext_prev_extent=-94.658203125%2

An Official Web Site of the United States Government

/ Datasets

Map data CC-BY-SA by OpenStreetMap
Tiles by MapQuest

EARTH SCIENCE (2218)
ESIP (1614)
EARTH SCIENCE INFOR... (1614)

4,469 datasets found

Illinois Wellheader Data

Arizona Geological Survey — This dataset is a compilation of Well header information compiled by the Illinois Geological Survey, published as a Web feature service, a web map service and an...

WMS WFS Esri REST ZIP

Illinois Borehole Temperatures

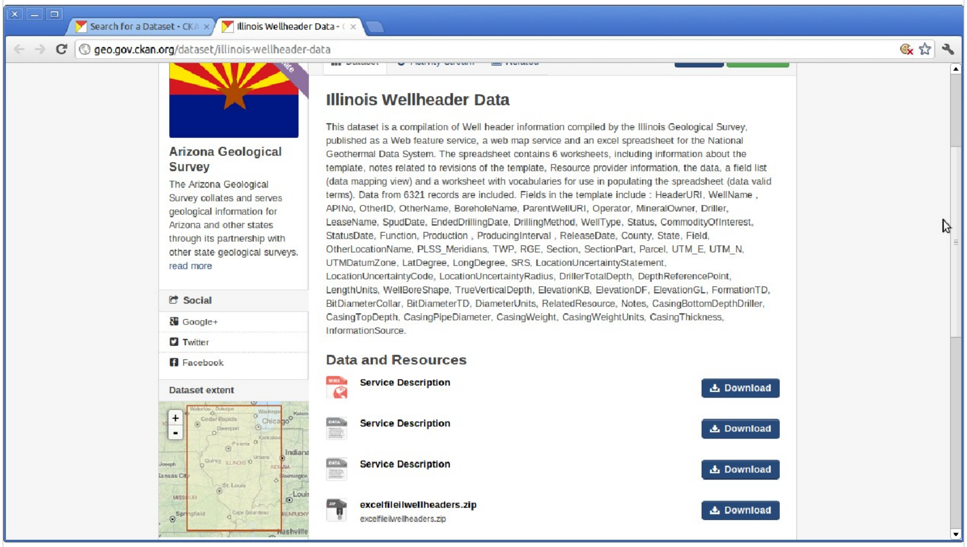
Arizona Geological Survey — This dataset is a compilation of borehole temperature observations compiled by the Illinois State Geological Survey, published as a Web feature service, a web map...

WMS WFS Esri REST ZIP

Lindsay Light - Right of Way Thorium Survey Reports

Open Knowledge Foundation

Drawing a rectangle on the map applies a filter to show only datasets that intersect with it. Note the much reduced number of results. Further searching can be carried out within the results.



Search for a Dataset - CKA
Illinois Wellheader Data - (
geo.gov.ckan.org/dataset/illinois-wellheader-data

Arizona Geological Survey

The Arizona Geological Survey collates and serves geological information for Arizona and other states through its partnership with other state geological surveys.
read more

Social
Google+
Twitter
Facebook
Dataset extent

Illinois Wellheader Data

This dataset is a compilation of Well header information compiled by the Illinois Geological Survey, published as a Web feature service, a web map service and an excel spreadsheet for the National Geothermal Data System. The spreadsheet contains 6 worksheets, including information about the template, notes related to revisions of the template, Resource provider information, the data, a field list (data mapping view) and a worksheet with vocabularies for use in populating the spreadsheet (data valid terms). Data from 6321 records are included. Fields in the template include : HeaderURI, WellName , APINo, OtherID, OtherName, BoreholeName, ParentWellURI, Operator, MineralOwner, Driller, LeaseName, SpudDate, EndedDrillingDate, DrillingMethod, WellType, Status, CommodityOfInterest, StatusDate, Function, Production , ProducingInterval , ReleaseDate, County, State, Field, OtherLocationName, PLSS_Meridians, TWP, RGE, Section, SectionPart, Parcel, UTM_E, UTM_N, UTMDatumZone, LatDegree, LongDegree, SRS, LocationUncertaintyStatement, LocationUncertaintyCode, LocationUncertaintyRadius, DrillerTotalDepth, DepthReferencePoint, LengthUnits, WellBoreShape, TrueVerticalDepth, ElevationKB, ElevationDF, ElevationGL, FormationTD, BitDiameterCollar, BitDiameterTD, DiameterUnits, RelatedResource, Notes, CasingBottomDepthDriller, CasingTopDepth, CasingPipeDiameter, CasingWeight, CasingWeightUnits, CasingThickness, InformationSource.

Data and Resources

Service Description Download

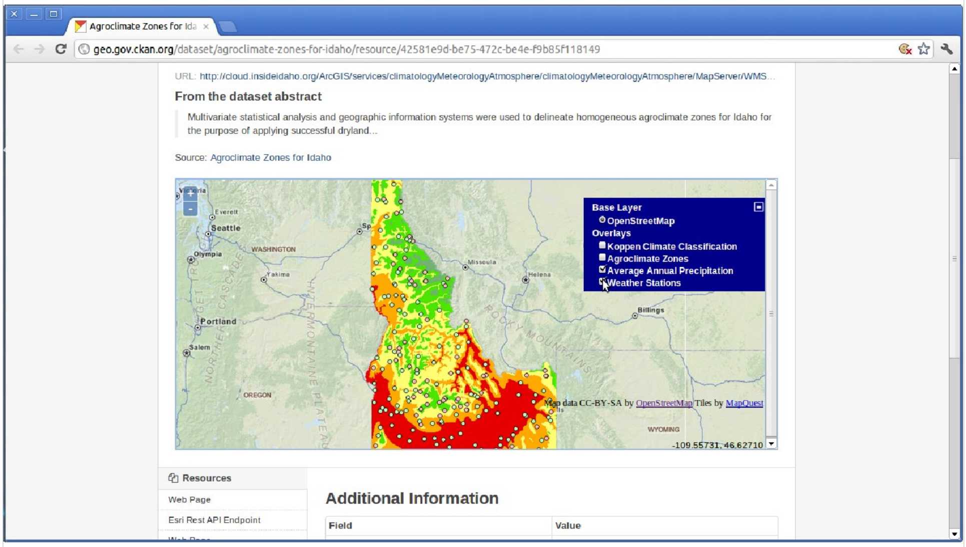
Service Description Download

Service Description Download

excelfileilwellheaders.zip Download
excelfileilwellheaders.zip

Open Knowledge Foundation

The description of the first dataset on the previous slide. Note the dataset extent map.



Agroclimate Zones for Ida
geo.gov.ckan.org/dataset/agroclimate-zones-for-idaho/resource/42581e9d-be75-472c-be4e-f9b85f118149

URL: http://cloud.insideidaho.org/ArcGIS/services/climatologyMeteorologyAtmosphere/climatologyMeteorologyAtmosphere/MapServer/WMS...

From the dataset abstract

> Multivariate statistical analysis and geographic information systems were used to delineate homogeneous agroclimate zones for Idaho for the purpose of applying successful dryland...

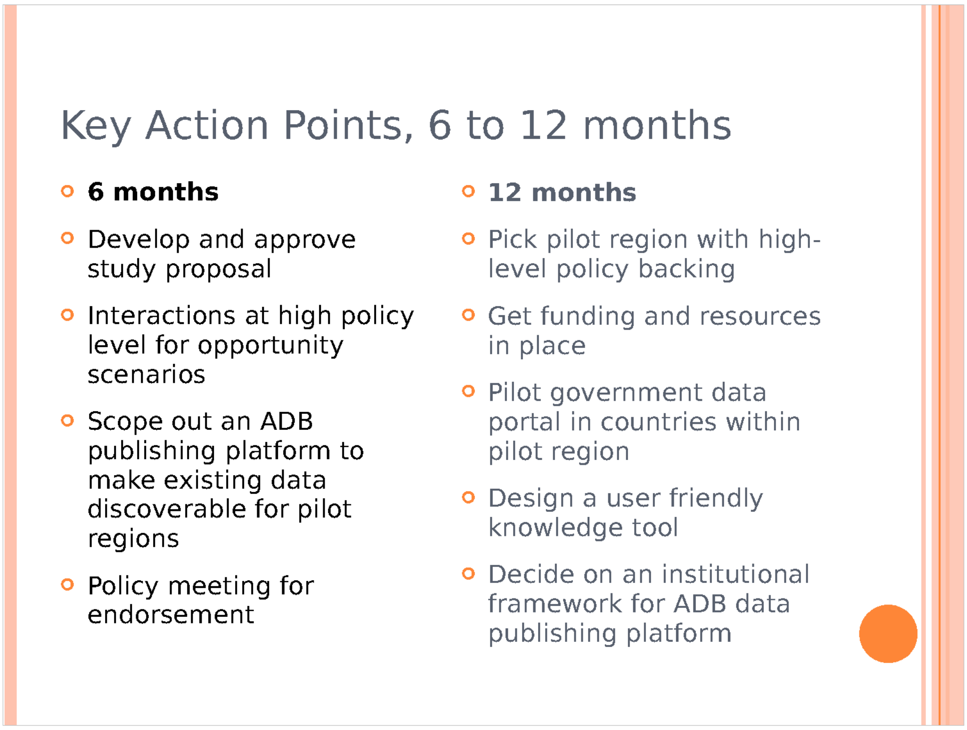
Source: Agroclimate Zones for Idaho

Base Layer
○ OpenStreetMap
Overlays
□ Koppen Climate Classification
□ Agroclimate Zones
☑ Average Annual Precipitation
☑ Weather Stations

Additional Information

Field	Value

Along with information about data found, CKAN can show previews of various kinds of data. This shows the ability to preview WMS geodata, including selecting different layers to overlaly.



Key Action Points, 6 to 12 months

- **6 months**
- Develop and approve study proposal
- Interactions at high policy level for opportunity scenarios
- Scope out an ADB publishing platform to make existing data discoverable for pilot regions
- Policy meeting for endorsement

- **12 months**
- Pick pilot region with high-level policy backing
- Get funding and resources in place
- Pilot government data portal in countries within pilot region
- Design a user friendly knowledge tool
- Decide on an institutional framework for ADB data publishing platform

A reprise of the key action points seen in the previous presentation.