



EIS LAB Summer Semester 2014

# Exploiting Data From Gov. Data Portals

Supervisor: Prof. Dr. Sören Auer

Mentor: Judie Attard

By: Group F

Group Member: M. Usman Asghar Bajwa:2383436

Email: husmanab@gmail.com

Date: 15-Sep-2014



# Table of Contents

1	Introduction .....	5
2	Requirement Specification.....	6
2.1	Problem Definition.....	6
2.2	Functional Requirements.....	6
	• Import from government portals.....	6
	• Export to enterprise information system .....	6
	• Analyses and comparison .....	6
	• Different format supporting .....	6
	• Add and modify portals or domains .....	6
	• Automatic data update .....	7
	• Mining and visualization .....	7
2.3	Non-Functional Requirements.....	7
	• Real-time Performance .....	7
	• Robustness to portals (reliability, availability and maintainability).....	7
3	Use Case Specification .....	8
4	Architecture .....	9
5	Government Data Portals .....	10
5.1	Data.gov.uk .....	11
5.1.1	Status of this site? .....	11
5.1.2	How to use data.gov.uk?.....	11
5.1.3	Conversion of Dataset to linked data.....	11
5.1.4	Data Commercialization and Licensing .....	11
5.1.5	Who is involved in the project? .....	11
5.1.6	Data holders.....	12
5.1.7	Data.gov.uk DATA (dataset) Access .....	12
5.1.8	Site Analytics .....	21
5.1.9	Applications:.....	21
5.1.10	Technical Details .....	22
5.1.11	Issues to Access Data from data.gove.uk.....	23

5.2	Germany (www.govdata.de).....	24
5.2.1	Who is responsible for GOVDATA?.....	24
5.2.2	How to access Data .....	24
5.2.3	Provision of Data to data.gov.de .....	26
5.2.4	Free and Non–commercial Data Usage .....	27
5.2.5	Data Exclusion (Un-Published Data) .....	27
5.2.6	Data Reusability .....	27
5.2.7	Data Formats.....	28
5.2.8	Data Providers.....	28
5.2.9	Websites for public administration.....	28
5.2.10	Metadata Structure.....	29
5.2.11	Issues in govdata.de to access data .....	30
6	France (www.data.gouv.fr) .....	30
6.1	Data Access .....	31
6.2	Etalab .....	37
6.3	Issues.....	37
7	Data Access Issues and Proposed Possible Solutions .....	38
7.1	Multi-linguistic Data Portals.....	38
7.2	Single Data Access Point for Data Sets.....	38
7.3	Open and Machine Readable Data Sets Formats .....	39
7.4	Heterogeneity of API's for Data Access .....	40
8	CKAN-CMS Integration (API Integration) .....	40
9	ExGovtDataPortalExcel2SQL Data Import Tool.....	42
10	Data Analysis and Visualization Tool.....	44
10.1	Home.....	44
10.1.1	Population.....	44
10.1.2	Population Density.....	44
10.1.3	Growth Rate % .....	44
10.1.4	Birth Rate (Birth/1000 Population).....	44
10.1.5	Death Rate (Death/1000 Population) .....	44
10.2	Summary .....	45
10.2.1	Population Germany, France and UK since 2000.....	45

10.2.2	Population Density Germany, France and UK.....	45
10.2.3	Population Growth of Germany, France and UK .....	46
10.2.4	Birthrate of Germany, France and UK.....	46
10.2.5	Death Rate of Germany, France and UK .....	47
10.3	Data Analysis and Visualization of Germany.....	48
10.3.1	Population of Germany .....	48
10.3.2	Population Density of Germany .....	48
10.3.3	Growth Rate in Germany .....	49
10.3.4	Birth rate in Germany .....	49
10.3.5	Death rate in Germany.....	50
10.4	France Data Analysis and Visualization.....	51
10.4.1	Population of France .....	51
10.4.2	Population Density of France .....	51
10.4.3	Growth Rate of France.....	52
10.4.4	Birth Rate of France .....	52
10.4.5	Death Rate of France .....	53
10.5	UK Data Analysis and Visualization .....	54
10.5.1	Population of UK .....	54
10.5.2	Population Density of UK.....	54
10.5.3	Growth Rate in UK.....	55
10.5.4	Birth Rate in UK.....	55
10.5.5	Death Rate in UK.....	56
11	Recommendation and future work.....	56
12	References .....	58

# 1 Introduction

Information is the raw material of the 21st century. The open and structured access to freely available data held by the public sector is an important contribution to the development of a knowledge society. It strengthens the trust between government and civil society, between government, business and the media. Open government data according to a report prepared for the European Commission study in the amount of up to € 40 billion for the EU therefore not only increase the acceptance of public decision-making processes - Moreover, the opening of such "data resources" the administration has enormous economic potential.. They also make it possible to gain new knowledge and to facilitate their use by innovative applications in everyday life.

Whether the open provision of data for each manager is associated with additional cost depends on the individual case. In some cases there are records already available in formats and under conditions of use, which allow open deployment very easy. In other cases, for example, questions about the authorship of the records or the use by third parties are not well enough understood. In other cases, technical questions must be answered, for example, the question of appropriate file formats.

In general: The provision of open government data is a process that can be performed by employees of individual public bodies - of course only if the establishment of such a publication has agreed.

Open Data is a gradual process. Begin therefore best with the provision of individual 'simple' records on your own websites and develop such a sense, as you can open additional data useful. To make their data known to an even wider group of users, it makes sense, then in the data catalog of GovData, the data portal to enter. Records for citizens, science, and business and last but not least, other administrative bodies are easier to find.

Give free access to data is to have more public services reforms. On **govt. Data portals data can** be find, review and information about countries and communities can easily be used e.g. to web application or mobile application development etc.

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 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 Government.

## 2 Requirement Specification

### 2.1 Problem Definition

The scope of this lab project is to exploit Govt. Data from the govt. Data portals of the Europe. Initially by selecting the specific portals, downloading the data and analyze the data, exporting this data to an Enterprise information System like SQL or Oracle and then Visualize the data both in graphically as well as in tabular form.

### 2.2 Functional Requirements

- **Import from government portals**

**Description:** Download structured data based on API or web page rules from manually configured government websites or portals

**Priority:** High

- **Export to enterprise information system**

**Description:** Export structured data from the system into enterprise information system such as database/warehouse

**Priority:** High

- **Analyses and comparison**

**Description:** Compare features, tools, available data, data reusability etc. for each government portals

**Priority:** Medium

- **Different format supporting**

**Description:** Support different formats for import e.g. .txt, .csv, .xls, .xml, .json file etc.

**Priority:** Medium

- **Add and modify portals or domains**

**Description:** Manually add or modify portal address or specific domains for government portals

**Priority:** Low

- **Automatic data update**

**Description:** After the first time download for initialization, automatic data update from the specific government data portals is available

**Priority:** Low

- **Mining and visualization**

**Description:** Implement data mining and analyses based on the government data and make user decision-oriented visualization

**Priority:** Low

## 2.3 Non-Functional Requirements

- **Real-time Performance**

**Description:** Real-time update from the government portals, supplying the latest data, news and information on specific domains from governments

**Priority:** Low

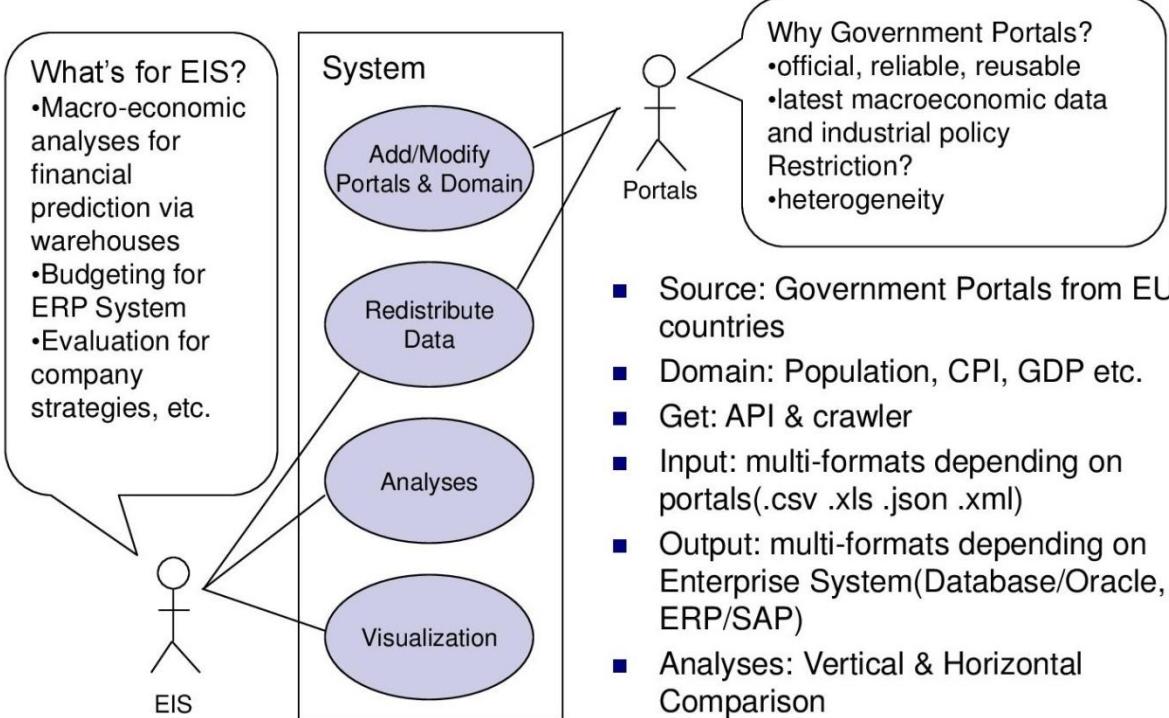
- **Robustness to portals (reliability, availability and maintainability)**

**Description:** Check out downloading efficiency, data correctness and completeness and detect whether portals or APIs have been modified or not

**Priority:** Low

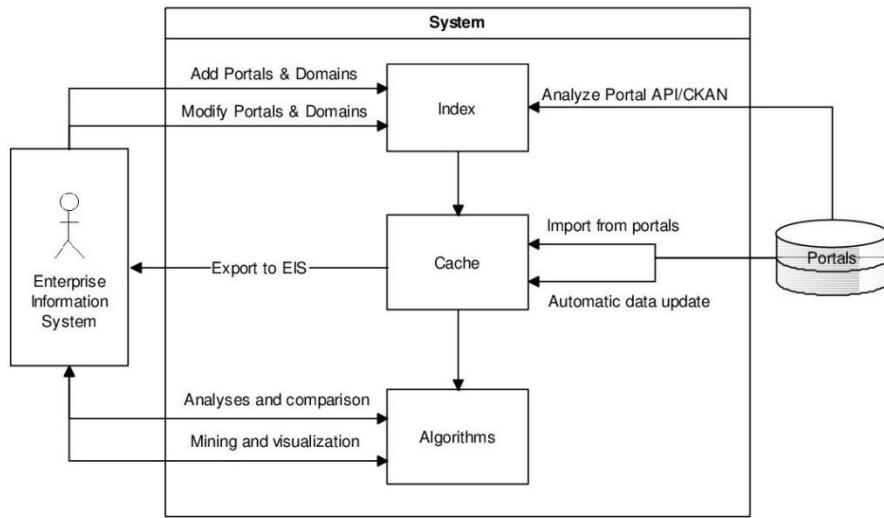
### 3 Use Case Specification

## Exploiting Data from Government Data



## 4 Architecture

### Exploiting Data from Government Data



- Portal: Government Portals from EU countries
- Domain: Population, etc.
- Access: API, CKAN & crawler
- Import: multi-formats depending on portals (.json .csv .xls.xml)
- Export: multi-formats depending on Enterprise Information System (Database/Oracle)
- Analyses: Vertical & Horizontal Comparison

## 5 Government Data Portals

All EU Govt. data portals follow the Open Data Charter principles to publish and release data sets for public. The Open Data Charter sets out 5 strategic principles that all G8 members<sup>[57]</sup> will act on. Open Data by Default. These include an expectation that all government data will be published openly by default, alongside principles to increase the quality, quantity and re-use of the data that is released.

G8 members have also identified 14 high-value areas – from education to transport, and from health to crime and justice – from which they will release data. These will help unlock the economic potential of open data, support innovation and provide greater accountability.<sup>[57][58]</sup>

Furthermore all Government data portals are Mashup. They not only combine data or code from at least two sources but also provide 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.

Currently we have selected Govt. data portals of 3 EU countries including United Kingdom, Germany and France and we have also selected the “Population” as our domain of data sets to exploit data from these portals. During this exploitation process we found various issues in these portals to access data and how it is difficult for general public to get the right information for them as well as for technical users. In each portal we also address the various ways for data set accessibility like through SKAN API of manual downloads, heterogeneity of language, number of steps, searching of specific data etc.

There are also various advantages of these sites for general public to have a close eye on Government and produce various web and mobile apps etc on basis of this data.

# 5.1 Data.gov.uk

## 5.1.1 Status of this site?

Data.gov.uk a beta version was launched in January 2010. Since then there are many enhancement in it and it is continuous improving with the passage of time.

## 5.1.2 How to use data.gov.uk?

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, themes, file format or department/agency. Each dataset provides guidance on accessing its data.

## 5.1.3 Conversion of Dataset to linked data

The W3C guidance on opening up government data [53] 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. The site does have catalogue (metadata) APIs

## 5.1.4 Data Commercialization and Licensing

The Open Government License enables to use all data of data.gov.uk for commercial purposes. The data is licensed under the Open Government License [54]

## 5.1.5 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 Open Data 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 runs the catalogue at data.gov.uk/data as well as a growing number of open data registries around the world. It is a project created by the Open Knowledge Foundation to make it easy to find, share and reuse open content and data. The CKAN software provides a web interface, programmer's API, feeds notifying of changes, and a browsable history of all changes. The API is documented here: <http://data.gov.uk/data/api>.

## 5.1.6 Data holders

Crown Copyright is the default copyright applied to all government department published documents. [51]

## 5.1.7 Data.gov.uk DATA (dataset) Access

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.

There are following three ways to access data from [www.data.gov.uk](http://www.data.gov.uk)

### 5.1.7.1 Dataset

- a. Open [www.data.gov.uk](http://www.data.gov.uk) and go to data and click on dataset
- b. Search a specific domain
- c. Click in specified search result
- d. Download dataset as its detail will also be shown
- e. Redirected to a particular data website like <http://neighbourhood.statistics.gov.uk> and

Note: it depends upon the redirected website how many further steps are involved to download the data

- f. Check download and next
- g. Select a specified file of a particular year in the list of year or months and next
- h. Select a particular file to download available in specific format like \*.xlsx or CSV etc. in a zip format.
- i. Unzip data file and analyze data.

**a- Open [www.data.gov.uk](http://www.data.gov.uk) and go to data and click on dataset**

The screenshot shows the DATA.GOV.UK beta homepage. At the top, there's a navigation bar with links to Home, Data, Apps, Interact, and a search bar. Below the navigation is a green header bar with links to Datasets, Map Search, Data Requests, Publishers, Public Roles & Salaries, Spend Reports, Site Analytics, and Reports. The main content area features several cards with images and titles:

- A card with a portrait of a man and the text "Transparency and Open Data: Priorities for 2014".
- A card with a location pin icon and the text "New UK Location Metadata Editor".
- A card with an image of a building and the text "The Linked Open British National Bibliography".
- A card with a woman's portrait and the text "Strengthening the National Information Infrastructure (NII)".
- A card with a person at a desk and the text "Public servants, open data wants you!".
- A card with a house icon and the text "Housing Land Availability Assessment – Data Constraints".

At the bottom of the page, there's a footer with links to FAQ, Code of conduct, Accessibility, Cookies policy, Privacy, Moderation policy, Contact, Terms and conditions, API, and About.

- b. Search a specific domain**
- c. Click in specified search result**

The screenshot shows the DATA.GOV.UK search results for the query "population". The search bar at the top contains "population". The results page displays 979 results. On the left, there are filters for "NII DATASETS", "LICENCE", and "THEME". The results are listed in a grid format:

- Population Clock** (Northern Ireland Statistics and Research Agency) - A clock which updates every minute to produce an estimated population total for Northern Ireland, at the current time and date UPDATED on 21 Feb 2011: at present the Population Clock..
- Wales's Population** (Welsh Government) - Presents a comprehensive overview of demographic statistics for Wales. Source agency: Welsh Government Designation: National Statistics Language: English Alternative title: Wales's...
- Population Trends** (Office for National Statistics) - Covers population and demographic information. It contains commentary on the latest findings and topical articles on relevant subjects. Source agency: Office for National Statistics...
- Population density** (Office for National Statistics) - Population density: Persons per hectare Source: Census 2001 Publisher: Neighbourhood Statistics Geographies: Output Area (OA), Lower Layer Super Output Area (LSOA), Middle Layer...
- Ageing Population**

d. Click on Download dataset (Data Set detail can also be seen)

The screenshot shows the DATA.GOV.UK homepage with the 'Population density' dataset selected. A context menu is open over the dataset details, with 'Download' being the highlighted option. The dataset details include:

- DATA RESOURCES (1)**
- In People and Society/Population and Migration/Key Regeneration Related Statistics.
- ADDITIONAL INFORMATION**
- Added to data.gov.uk: 09/02
- Theme: Society
- Geographic coverage: England, Wales
- Update frequency: Every 10 years
- Mandate: No value
- Temporal coverage: No value
- Date added computed: No value
- Date updated computed: No value

e. Redirected to a particular data Set website like  
<http://neighbourhood.statistics.gov.uk>

Note: it depends upon the redirected website how many further steps are involved to download the data set.

f. Click on download and next

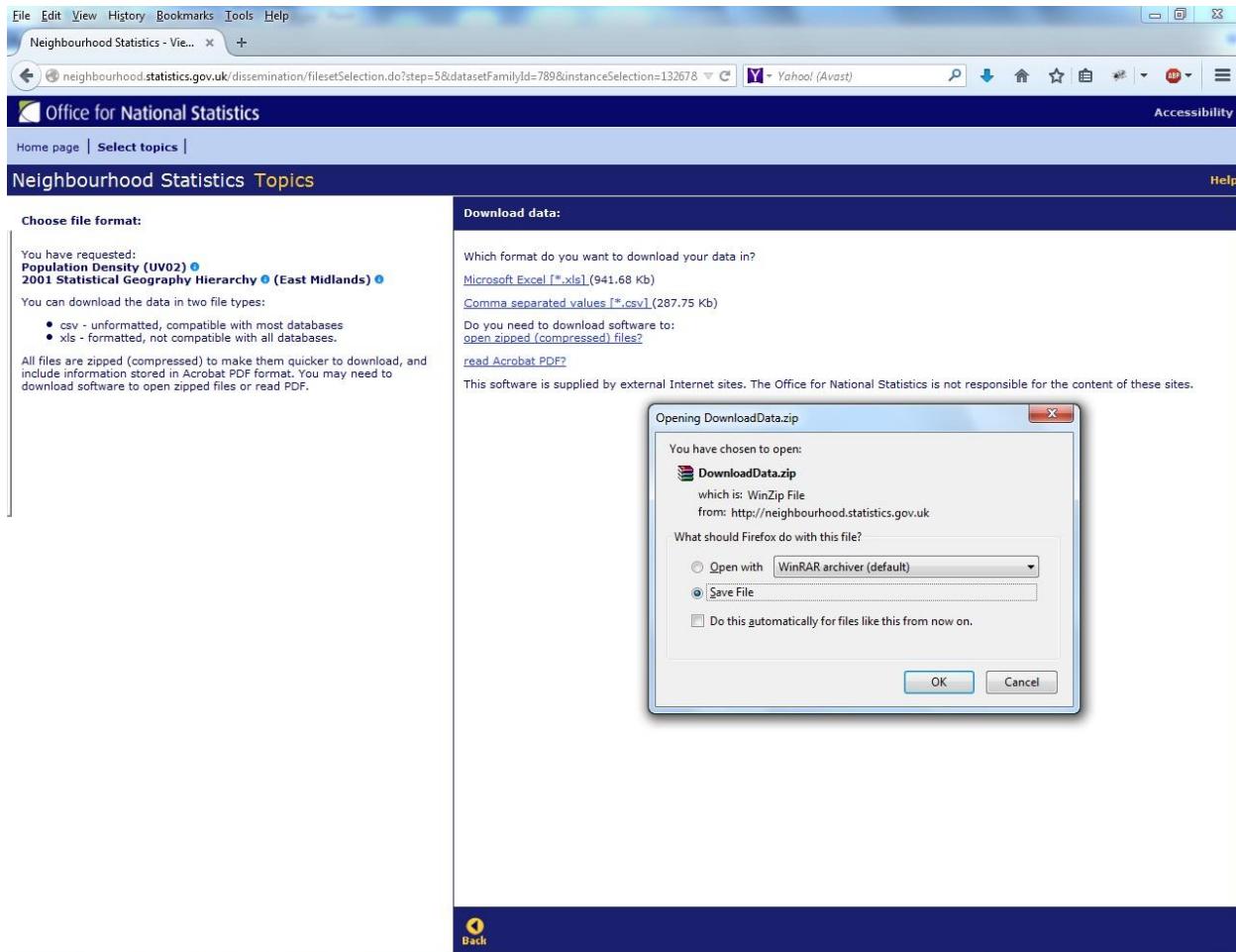
The screenshot shows the Neighbourhood Statistics Topics page. The 'Population Density (UV02)' dataset is listed with a 'Download' link next to it.

- g. Select a specified file of a particular year in the list of year or months and next

The screenshot shows a web browser window with the URL [neighbourhood.statistics.gov.uk/dissemination/instanceSelection.do?JSAllowed=true&Function=8%24ph=61&CurrentPage=1](http://neighbourhood.statistics.gov.uk/dissemination/instanceSelection.do?JSAllowed=true&Function=8%24ph=61&CurrentPage=1). The page title is "Neighbourhood Statistics Topics". On the left, there is a sidebar with "Selecting sets of data:" and a list of selected items: "Population Density (UV02)". Below this, it says "This data is organised according to geographic area. You can download the data by area, for example London or East Midlands, or by area type, for example local authorities or health areas. Some sets of data have been broken up into smaller files to make them easier to download." On the right, there is a main content area titled "Files available to download:" with a "Select" button. A list of files is shown, each with a radio button and a link. The files are: 2001 Statistical Geography Hierarchy (East Midlands), 2001 Statistical Geography Hierarchy (Greater London), 2001 Statistical Geography Hierarchy (North East), 2001 Statistical Geography Hierarchy (North West), 2001 Statistical Geography Hierarchy (South East), 2001 Statistical Geography Hierarchy (South West), 2001 Statistical Geography Hierarchy (United Kingdom), 2001 Statistical Geography Hierarchy (Wales), 2001 Statistical Geography Hierarchy (West Midlands), 2001 Statistical Geography Hierarchy (Yorkshire and The Humber), and 2002 Administrative Hierarchy.

- h. Select a particular file to download available in a specific format like \*.xlsx or CSV etc. in a zip format

The screenshot shows a web browser window with the same URL as the previous screenshot. The page title is "Neighbourhood Statistics Topics". On the left, there is a sidebar with "Choose file format:" and a list of selected items: "Population Density (UV02)" and "2001 Statistical Geography Hierarchy (East Midlands)". Below this, it says "You can download the data in two file types: csv - unformatted, compatible with most databases; xls - formatted, not compatible with all databases. All files are supplied (compressed) to make them quicker to download, and include information stored in Acrobat PDF format. You may need to download software to open zipped files or read PDF." On the right, there is a main content area titled "Download data:" with a "Back" button at the bottom. It asks "Which format do you want to download your data in?" and lists "Microsoft Excel (\*.xls)(941.68 Kb)" and "Comma separated values (\*.csv)(287.75 Kb)". It also asks "Do you need to download software to open zipped (compressed) files?" and provides a link to "read Acrobat PDF". At the bottom, it states "This software is supplied by external Internet sites. The Office for National Statistics is not responsible for the content of these sites."



- i. Unzip data file and analyze data.

### 5.1.7.2 Map Search

- Find particular map/find area of Interest
- Draw the map or box around particular area or draw search area
- Submit to find Data of selected area
- Select the domain from search results dataset
- Download data
- Fill the form and send request to access data and wait until its approved [52]
- Find particular map/find area of Interest

The screenshot shows the DATA.GOV.UK beta website's 'Map Based Search' feature. The interface is organized into three main sections:

- STEP 1 (OPTIONAL):** A text input field containing "london" with a "Find" button next to it. Below the input field is a dropdown menu listing various London boroughs, with "Barnet London Boro, London Borough" currently selected.
- STEP 2:** A section titled "Draw search area" with a "quick tip" link and a checkbox for "Display administrative areas". It includes "Draw" and "Clear" buttons.
- STEP 3:** A section titled "Find data for selected area" with a "Submit" button.

A map of the United Kingdom is displayed in the background, with a small red square marker indicating a specific location. At the bottom of the page, there is a footer with links to FAQ, Code of conduct, Accessibility, Cookies policy, Privacy, Moderation policy, Contact, Terms and conditions, API, About, and various system status indicators like EN, 12:23, and 07/09/2014.

- b. Draw the map or box around particular area or draw search area
- c. Submit to find Data of selected area

**STEP 1 (OPTIONAL)**  
Locate area of interest  
Example: London or EH99 1SP or 54.58,-5.93  
Place name, postcode or coordir

**STEP 2**  
Draw search area [quick tip]  
 Display administrative areas

**STEP 3**  
Find data for selected area

Map showing administrative boundaries and major roads in the London area. A red rectangle highlights a specific region.

- a. Select the data set domain from search results dataset

**PUBLISHER**  
British Geological Survey (226)  
Centre for Ecology & Hydrology (217)  
Office for National Statistics (129)  
Marine Management Organisation (120)  
Scottish Government Spatial Data Infrastructure (119)  
Environment Agency (90)  
... more ...

**OPENNESS SCORE (BETA)**  
★★★★★ (151)  
★★★★★ (1156)

**BROKEN LINKS**  
OK (1115)  
Partially broken (48)  
Broken (18)

**UK LOCATION DATASET TYPE**  
UK Location Dataset (1307)  
Dataset (1077)  
Service (192)  
View (134)  
Download (54)  
Series (33)  
Non-Geographic Dataset (4)

**Various development zones**  
London Borough of Brent  
Growth Areas are designated in the London Plan as specific areas for new residential development to accommodate future population growth, as outlined in the Government's Sustainable...

**Air Quality Management Areas**  
London Borough of Brent  
The Government set air quality standards for the protection of health. Under the Environment Act 1995 Brent Council is required to designate Air Quality Management Areas (AQMA) where...

**Town Centre Boundaries**  
London Borough of Brent  
Town Centre Boundaries include the primary shopping areas and areas predominantly occupied by main town centre uses. The Town Centre hierarchy is designated in the London Plan

**Smoke Control Zones**  
London Borough of Brent  
Smoke Control Orders came into operation pursuant to the Clean Air Act 1956 (both the Clean Air Acts of 1956 and 1968 were replaced by the Clean Air Act 1993). It is an offence for an...

**Environmental Permit Sites**  
London Borough of Brent  
Under the Environmental Permitting (England and Wales) Regulations 2010, local authorities must regulate certain types of industry and other activities such as dry cleaners. This is to...

d. Click on Inspire download service and direct link

The screenshot shows a web browser window with the URL [data.gov.uk/dataset/smoke-control-zones3](http://data.gov.uk/dataset/smoke-control-zones3). The page content includes:

- DATA RESOURCES (2)**
  - INSPIRE View Service
  - INSPIRE Download Service (highlighted with a green border)
- ADDITIONAL LINKS (2)**
  - Source GEMINI2 record (XML)
  - Source GEMINI2 record (formatted) (HTML)
- ADDITIONAL INFORMATION**

Added to data.gov.uk	04/02/2014
Temporal coverage	1/1/2013 - 31/12/2016
Harvest URL	<a href="http://inspire.misoportal.com/metadata/files/brent_council">http://inspire.misoportal.com/metadata/files/brent_council</a>
Harvest Date	04/02/2014 17:30
Harvest GUID	384b312a-3404-41bd-89b6-d9e10e693c35v2
Extent	Latitude: 51.603689° to 51.52707° Longitude: -0.33658° to -0.1889°
Spatial reference system	urn:ogc:def:crs:EPSG::27700
Metadata date	31/1/2014
Dataset reference date	16/7/2010 (creation)
Frequency of update	unknown

e. Fill the form and send request to access data and wait until its approved [52]

The screenshot shows a web browser window with the URL <http://misoportal.com/portal/api/request/licensed-data/432>. The page content includes:

### restricted data access request

The data you have requested is subject to restricted licensing. In order to access the data, this must be agreed by the owner of the data. Please complete the request form below and we will contact you as soon as they have accepted or revoked your request.

You are requesting access to the **Town Centre Boundaries** dataset uploaded by **Brent Council**.

**\* Important Information \***

Completing this form does not guarantee you access to this data. The data owner holds the right to either grant or decline your request. If your request is granted, the data will be restricted to the IP address of the network you are currently filling this form from.

**Your Details**

Name \*

Email Address \*

Phone Number \*

**Your Company's Details**

Company Name \*

Street \*

### 5.1.7.3 Data request

User has to register and after logging to the data.gov.uk, he can send request to get access to a particular data. He can also see the data request by other users as under

- 1- Search Data
- 2- View Request

The screenshot shows the DATA.GOV.UK Beta website interface. At the top, there's a navigation bar with links for Home, Data, Apps, Interact, and a search bar. Below the navigation is a green header bar with links for Datasets, Map Search, Data Requests, Publishers, Public Roles & Salaries, Spend Reports, Site Analytics, and Reports. The main content area displays a list of recent data requests:

- ATM Locations** - Submitted by raustine 4 days 22 hours ago
- Attendance allowance (AA) claimants by postcode** - Submitted by arm10 6 days 2 hours ago
- Broad Rental Market Areas** - Submitted by CambisInsight 1 week 2 days ago
- Traineeship Success Data** - Submitted by Ian Taylor 1 week 4 days ago
- WFD Catchment Management Information Reasons for Failure (AfA318)** - Submitted by plaindrain 1 week 5 days ago

To the right of the request list, there's a sidebar with a search bar showing "538 Dataset requests + 302 confidential requests", a "Login to request new data" button, and a "PROGRESS ON REQUESTS" section featuring a stacked bar chart and a "See Dashboard" button. At the bottom of the sidebar, there's a "ODUG MEMBERS" section listing Robert Barr, Dominique Lazanski, Gesche Schmid, and Jacqui Taylor. The bottom right corner of the screen shows system status icons and the date/time: 16:41 07/09/2014.

## 5.1.8 Site Analytics

Site analytics feature is available to analyze the data.gov.uk by itself. Like the usability and accessibility of the portal etc.

The screenshot shows the 'Site Usage' section of the DATA.GOV.UK website. At the top, there's a navigation bar with links like Home, Data, Apps, Interact, and Site Analytics. Below that is a sub-navigation for Site Analytics. A sidebar on the left lists categories such as Datasets, Map Search, Data Requests, Publishers, Public Roles & Salaries, Spend Reports, Site Analytics, and Reports. On the right, there's a 'Jump To...' dropdown menu with options like Publisher Usage Statistics and Dataset Usage Statistics. The main content area features a 'Graph Legend' section with the message '(No graph is loaded)'. Below it is a table with various site metrics:

Name	Value	History
Total page views	12592009	
Total visits	3921720	
Pages per visit	3.25	
Bounce rate (home page)	32.88%	
New visits	76.49%	
Average time on site	00:02:12 (132.98 seconds)	

At the bottom of the page, there's a footer with links to FAQ, Code of conduct, Accessibility, Cookies policy, Privacy, Moderation policy, Contact, Terms and conditions, API, and About. It also mentions 'Powered by: CKAN'.

## 5.1.9 Applications:

Different mobile and web application created and developed by various vendors for the easiness of society and community to access and visualize the data can also be accessed and searched in this portal.

The screenshot shows the 'Latest apps' section of the DATA.GOV.UK website. At the top, there's a search bar with the placeholder 'Search apps...'. To the right of the search bar, it says '349 Apps'. Below the search bar, there are three cards featuring different mobile applications:

- Scope Nights**: Astronomy Weather Reports. The card shows a blue icon of a telescope and stars.
- Walkonomics**: How walkable is your street? The card shows a hand holding a smartphone displaying a map with walking icons.
- Floodalerts API**: FloodAlerts API. The card shows a red triangle warning sign with a house inside, with water waves at its base.

Below these cards, there are more cards for other apps like 'Latest apps' and 'Top rated apps'. The bottom of the page includes a footer with the DATA logo and a copyright notice: 'Exploiting Data Government Data Portals'.

## 5.1.10 Technical Details

Data.gov.uk runs open source software. All code is available at [38] under an open license.

### 5.1.10.1 Overview

Data.gov.uk runs on a mix of Drupal and CKAN. Data.gov.uk proudly announces: 'Data.gov.uk to Go' - a package containing the software for the well-known open data website Data.gov.uk. This allows other governments and open data communities to quickly install and customize a full open data website and develop it further in partnership with the worldwide community.

Although the central components CKAN, Drupal and data.gov.uk's custom components have each been open source for several years, few have used them in combination, due to the complexities of set-up. Data.gov.uk to go provides an organized way to configure these components to quickly launch a fully-featured open data portal along the lines of data.gov.uk.

Data.gov.uk To Go builds on the CKAN's data catalogue to provide extra features such as:

- Blog posts
- App listing
- Joint log-in (CKAN+Drupal)
- Data request mechanism
- Spam protection
- Data quality reports
- Flexible content using Drupal's CMS features

The UK Government has contributed Data.gov.uk To Go to the world-wide community to quick-start the use and shared development of this open data platform. The UK wants to develop it in partnership with other providers of Open Data portals, through the usual Open Source / GitHub model of forking, pull requests, issues etc. that everyone is encouraged to contribute to.

### 5.1.10.2 CKAN Software

The 'Data' section of data.gov.uk runs on the CKAN software, which was designed as an open data catalogue by the Open Knowledge Foundation[39]. Customizations (such as visual branding) for data.gov.uk are developed in a CKAN extension called 'ckanext-dgu'. Other CKAN extensions have been adapted and developed for the site to provide spatial metadata harvesting, map search & preview, data caching, five stars of openness, analytics reporting etc.

These are the CKAN modules used in data.gov.uk (last updated Jan 2014):

- CKAN version 2.2 (with minor modifications) [40]
- CKAN data.gov.uk extension [41]
- CKAN Archiver extension [42]
- CKAN Data Preview extension [43]
- CKAN Google Analytics Report extension [44]
- CKAN Harvest extension [45]
- CKAN Hierarchy extension[46]
- CKAN Import Library extension [47]
- CKAN ODI Certificates extension
- CKAN Ordnance Survey extension [48]
- CKAN QA extension [49]
- CKAN Spatial extension[50]

### 5.1.11 Issues to Access Data from data.gov.uk

There is no any single point and easy access to data; user has to follow at least 6 to 8 steps to download a simple and small piece of information. Most of the dataset are available only in MS excel or CSV format while technical users mostly required in xml or linked or other format etc. Moreover, there is a simple code of conduct to use data in individual application It only covers how to relate your application data to the data of data.gov.uk , and simple common-sense usage of the data.

## 5.2 Germany ([www.govdata.de](http://www.govdata.de))

The data portal for Germany provides a single, centralized access to administrative data from federal, state and local governments. The goal is to make this data in one place to find and make it easier to use. To increase the supply of machine readable raw data and to promote open license usage term “Open Data” is used. Currently, there is a public beta version of the portal, which is gradually expanded and optimized within the possibilities. And until 2015 it will be available in full control mode.

Data from all sectors from government are publically and centrally available and from 2014 the portal is operated for testing purposes and is gradually extended and optimized.

The goal of this portal is to make public sector data as comprehensively as possible and open for further use. GovData provides central access to reusable data from federal, state and local governments.

One of the goals of the Data Portal is to make data easier to use and all records and documents is clearly defined the conditions under which they may be used.

GovData offer to public sector entities at federal, state and local data management. In particular, administrative officials, citizens, businesses and scientists should be given the opportunity to access just across a single point of entry to data and information of the public administration in Germany. The aim is that these "data resources" from management be better used and re-used, so that new ideas and combining and analyzing new knowledge from existing data and new applications can be developed. The portal is currently part of a public beta-phase operated as a prototype.

### 5.2.1 Who is responsible for GOVDATA?

The portal was part of a research project commissioned by the Federal Ministry of the Interior implemented. Development and operation in the current public beta phase is carried out by Fraunhofer FOKUS. To ensure cross-level coordination, it is being closely monitored by the federal-state interdepartmental working group "Open Government" and is closely related to the corresponding control project of the IT Planning Council toothed.

### 5.2.2 How to access Data

The metadata catalog of GovData is accessible via a programming interface (API): information about the API of GovData metadata catalog and Technical information on the development of the portal is available on GitHub<sup>[15]</sup>.

### 5.2.2.1 Operators and functionalities of GovData search

#### 5.2.2.1.1 Combining search terms

If multiple search terms are entered into the search, results are displayed by default, which *listed all terms* included. The same effect is achieved if the search terms using the operator AND or + are linked, e.g. Berlin Brandenburg AND or Berlin + Brandenburg .

#### 5.2.2.1.2 Search for specific data fields

In GovData, OGD-metadata structure may be used to limit the result. In the form <JSON identifier>: <Keywords> can be specified in a simple fashion as a special filter already in the search box.

In this way, all the records can be found, for example, easily, the "Data License 1.0 Germany - Attribution" use ( license\_id: dl-de-by-1.0 ). Currently, the following fields can be used in this way:

- title : Title of the data set
- name : The human-readable unique identifier of the record
- url : The original address of the data set, for example, to the appropriate authorities providing site
- notes : description text to record
- author : the "publication point", ie the institution from which the data originated
- author\_email : The E-mail / contact address of the published site
- maintainer : The "data responsible party", if not identical to the "published site".
- maintainer\_email : The E-mail / contact address of the data responsible entity
- license\_id : Short name of the license of a data set (corresponding list of OGD metadata structure)
- tags : Free percussion or key of the record
- groups : Categories for the records and documents
- res\_description : description of a data resource
- res\_format : File format of a resource record
- res\_url : Web address of a data resource

#### 5.2.2.1.3 Fuzzy search

The search for specific data fields offers another feature: When searching for a specific field, the fuzzy search can be activated. This is not wanted for the exact wording of the specified term, but also for similar terms. For this purpose, the desired keyword to the tilde character ("~") and a "blur-degree" added between 0.0 and 1.0. In this way, the

search returns `author: Bärlin ~ 0.5` also results in which "Berlin" was entered. A blur of 0.0 degrees allows it the largest and the smallest 1.0 blur the search term.

**Note:** fuzzy search only works if the searching for the above-mentioned specific data fields.

### 5.2.2.2 How to download data

- Search specific domain
- Download a specific data
- Data is only available in .xlsx and .csv format
- Click on free
- Download in .xlsx or csv format

### 5.2.3 Provision of Data to data.gov.de

Public or researched can be provided to in the following ways after user or institution registration and confirmation as data provider through editorial team.

#### 5.2.3.1 Manual- providing web form

This is the easiest, but with a larger number of records the most laborious provisioning variant. It will be only used if users are just starting only with the opening of their data or want to provide a few data. In the medium term, but they should try to choose one of three automatic variants.

For use of the web form, you must have registered as an accredited provider data on the portal (see above). On the pages of "data "or" documents "then find the link" record / add documents ". An input form is set to be interrogated about the all the necessary information. Once form is completed, click" Submit". Your details are then checked again by the writing team and then published.

To edit the information on previously published records afterwards, visit the detail page of each data set and click on "Edit". Here you can change any details or add additional resources.

#### 5.2.3.2 (Automated) provision by CKAN API

If you own a private data catalog, you can put your data here on automated available. CKAN, which also forms the basis of GovData, provides a programming interface to (English "application programming interface" API) that can be exchanged directly through the new records or updates. You must write for your catalog for a script that this interface responsive.

The URL of the interface of GovData is <https://www.govdata.de/ckan/api>. The records have the metadata structure of corresponding GovData. In addition, you need to use a so-called API Key and editorial team may be contacted for any help.

### **5.2.3.3 (*Automated*) by providing "JSON dump"**

If you have registered your records in a different catalog, but does not want to use the CKAN API, it is possible that you provide portal team to export your catalog via a static URL. In this way they can then import data at regular intervals in GovData.

The prerequisite is that your data catalog export in JSON format and according to the metadata structure of GovData done. In the description of the metadata structure can be found including information on how you can check whether your export meets the requirements of GovData metadata structure.

### **5.2.3.4 (*Automated*) provision by CSW or CKAN**

Finally, there is the possibility that your data GovData of a so-called "harvesters" are collected. This assumes that your records are registered in another CKAN-based or CSW-compliant catalog. This method lends itself if your metadata structure does not match the structure of GovData.

## **5.2.4 Free and Non-commercial Data Usage**

For the data provider providing data on the portal is free of charge. In general, if any usage restrictions (costs incurred, prohibition of commercial exploitation, or the like) are present, this is clearly marked in each record. Data and documents under the "German Data License - Attribution - noncommercial" stand may not be used for commercial purposes. But data usage for scientific or educational purposes is permitted.

## **5.2.5 Data Exclusion (Un-Published Data)**

Personal data are not published for privacy reasons. Safety data are excluded, as are dates on which third parties have rights, unless the respective author a corresponding publication has not consented. Data providers should ensure that their data cannot be de-anonymized. Moreover, only such data are published, their descriptions exist sufficiently comprehensive; so the lack of a clear example of use for a data set is an exclusion criterion.

## **5.2.6 Data Reusability**

Some data formats restrict simple further processing. Other data formats are doing an automated re-use much easier. For the provision of administrative data as open data must be provided in open formats as possible. For example, tabular data are indeed in Portable Document Format (PDF) good for people to read, but difficult to interpret for machines.

This data can be usefully re-used by third parties, they should be provided as possible in a fine-grained resolution, but at the same time no existing data protection rules should be violated.

### 5.2.7 Data Formats

Preferably, data should be provided in such formats based on open standards and enable machine readability. For tabular data, these are, for example, CSV or XML. Etc. PDF files, images (JPG, TIF, GIF, BMP, etc.) are not well suited, because they prevent it, that the data are processed directly and automatically, Excel files (XLS, XLSX) are suitable limited; Although they are generally processed further, however, is not the standard open format.

### 5.2.8 Data Providers

Data providers are in principle public authorities at all levels of government. To start different data following public bodies are available:

- Federal Statistical Office / GENESIS-Online
- Federal Agency for Cartography and Geodesy
- Environmental Information Portal of Lower Saxony
- Land Baden-Württemberg
- Bavaria
- Land Berlin
- City of Bonn
- Free Hanseatic City of Bremen
- Free and Hanseatic City of Hamburg
- City of Cologne
- City of Moers
- City of Münster
- Hanseatic City of Rostock
- City of Ulm
- Community Wennigsen (Deister) etc.

### 5.2.9 Websites for public administration

- ***European Union Open Data Portal[32]***

The EU data portal with data from all EU institutions

- ***Publicdata.eu[31]***

Similarly GovData nationally collected publicdata.eu open data from all EU Member States and makes it available via a central portal.

- ***GENESIS-Online[30]***

GENESIS is the Statistical Offices of the country and the Federal Statistical Office offered data portal for German statistical data.

- **Geodata Portal Germany[29]**

Geoportal.de is a search engine for spatial data and provides access to maps, aerial images and thematic maps on various topics - from energy to nature conservation.

- **PortalU[28]**

PortalU is the federal-wide data catalog for environmental data.

- **WebSKM [27]**

WebSKM is the online database of the Federal Statistical Office for administrative costs in Germany

- **Environmental Information Portal of Lower Saxony[20]**

It is the Lower Saxony environment portal and provides official information on the subject of Environment.

- **Berlin Open Data portal[21]**

Here you will find open administrative data of the Federal State of Berlin and its districts.

- **Bremer Open Data Portal[22]**

Here you will find open administrative data of the Federal State of Bremen.

- **Open Data in Baden-Württemberg[23]**

Here you will find open government data of the State of Baden-Württemberg. The offer is still in a test phase.

- **Open Data Portal Bavaria[24]**

Here you will find open government data of Bavaria

- **Open Data portal Hamburg[25]**

Here you will find open government data from Hamburg

- **Open Data Portal Rhineland-Palatinate[26]**

Here you will find open government data from Rhineland-Palatinate

## 5.2.10 Metadata Structure

### 5.2.10.1 CKAN

CKAN is a web-based system for storing and distributing data and is used in a variety of open data projects worldwide. It forms the basis of this portal.

### 5.2.10.2 Liferay

Liferay is an open source content management system and provides a basis for this portal. Metadata schema "Open Government Data Germany" can be found at github [19]. This metadata schema is the basis of the data catalog in this portal.

### 5.2.11 Issues in govdata.de to access data

- a. The govdata.de portal is not multi-lingual and its only in German Language. Due to this it's really difficult to use or get data access if you don't know German
- b. There are too many and unnecessary steps to download the specific data which makes it very difficult to find and download the data of specific domain automatically without human intervention.
- c. Most of the data is available in .CSV and .XLSX but not in the XML format which needs more effort to make it usable for technical users, developers, scientist etc.

## 6 France ([www.data.gouv.fr](http://www.data.gouv.fr))

The "data.gouv.fr" Platform enables utilities to publish public data and civil society to enhance, modify, and interpret, to co-produce information of general interest.

Everyone! Citizens, utilities or any organization registered under the identity of an individual or a corporation can share data and registration and use of the platform are free.

"Data.gouv.fr" also pursuing the ambition to be a tool available to citizens to enable them to produce, enriches or supplements such data or share information of public concern, e.g., the inventory of a public library, Population, or environmental performance of a company.

Therefore the personal data and those whose publication is against the law (medical secret, secret national defense, statistical confidentiality, trade secrets, etc.) are not published for public.

For a public service: the opening and sharing of public data helps account for the functioning of institutions and foster innovation. Data.gouv.fr enhances the visibility of the data by their national SEO and provides tools that facilitate conversations with each community revolving around the public service (associations, journalists, researchers, students, citizens).

For an individual, a company, a researcher, an association produces data of public interest to publish in Open Data and share interpretations with the public to strengthen the freedom of information.

## 6.1 Data Access

Data can be downloaded in the following sequence.

- a) Go to France GovData website and Search the specific domain by entering domain and area where you want to search
  - b) Click on specific result of your choice from researched results
  - c) Click on particular resource.
  - d) Click on download (you will be redirected to particular website instead of download data set)
  - e) Click on unfold all results on the redirected site
  - f) Click on particular unfolded result
  - g) Select year and particular region and submit
  - h) Select particular result
  - i) Select year range for dataset
  - j) Select particular data format of your choice and download the result dataset for analysis.
- a. Go to France GovData website and Search the specific domain by entering domain and area where you want to search

## b. Click on specific result of your choice from researched results

The screenshot shows the data.gouv.fr website interface. At the top, there's a navigation bar with links for File, Edit, View, Favorites, Tools, Help, and a sign-in option. Below the header, the URL https://www.data.gouv.fr/en/search/?q=population is visible. The main content area has a search bar containing 'population'. To the right of the search bar are buttons for Topics, CONTRIBUTE!, and a sort dropdown. Below the search bar, there are tabs for Datasets (111), Reuses (61), Organizations (6), and Users (0). A sidebar on the right lists various organizations contributing data, such as Eurostat, Banque Mondiale, Région Île-de-France, and Ministère de l'Intérieur. The main content area displays five search results, each with a thumbnail, title, description, and metadata like date range and location. The first result is from INSEE, titled 'Population'.

## c. Click on particular resource.

This screenshot shows the detailed view of the 'Population' dataset on data.gouv.fr. The top navigation bar and search bar are identical to the previous screenshot. The main content area is titled 'Population' and includes a note that it comes from a certified public service. It was published on August 23, 2013, by Institut National de la Statistique et des Etudes Economiques (INSEE). The description explains that the dataset contains population census results from 1962 to 2013, providing data on population diversity and evolution across France. It also mentions the INSEE's role in collecting, producing, analyzing, and diffusing economic and social information. On the left, there's a 'Resources' sidebar with links to 'Séries chronologiques actives issues de la BDM', 'Les résultats des recensements de la population', and 'Publications et statistiques'. On the right, there's a 'Producer' section featuring the INSEE logo and slogan 'Mesurer pour comprendre', along with a 'FOLLOW' button. A sidebar on the right provides 'Informations' such as license (Licence Ouverte / Open Licence), date range (1962 to 2013), location (Unknown, Town, France), and categories (DÉCÈS, DÉMOGRAPHIE, MARIAGES, NAISSANCES, POPULATION, RECENSEMENT).

- d. Click on download (you will be redirected to particular website instead of download data set)

The screenshot shows a web browser window with the URL <https://www.data.gouv.fr/en/datasets/population/>. A modal dialog box is open, titled "Séries chronologiques actives issues de la BDM". The dialog contains the following information:

- URL:** <http://www.bdm.insee.fr/bdm2/choixTheme.action?code=44>
- Format:** HTML
- Created on:** March 7, 2014 4:40 AM
- Modified on:** December 10, 2013 3:53 PM
- Published on:** September 4, 2014 7:24 AM

Below the dialog, the main page content is visible, including the Insee logo and some text about population statistics.

- e. Click on unfold all results on the redirected site.

The screenshot shows a web browser window with the URL <http://www.bdm.insee.fr/bdm2/choixTheme.action?code=44>. The page is titled "Insee - Databases - BDM - D...". The navigation bar includes links for Home, Topics, Databases, Publications and Services, Regions, Definitions and methods, Publics, and INSEE and official statistics. The "Databases" menu is expanded, showing categories like ALISSE, enterprise data, Statistical indices and series, Bulletin statistique, Macro-economic database (BDM), The population census, Local statistics, Microdata files, and SIRENE Register. The "Demography" section is also expanded, showing sub-categories like Census population, Registry office and demographic trends, Couples - Families - Households, and Foreigners - immigrants. At the bottom of the page, there are links for RSS, Newsletter, Sitemap, Buy publications, and legal notices.

### f. Click on particular result

The screenshot shows the Insee Databases website with the URL <http://www.bdm.insee.fr/bdm2/choixTheme.action?code=44>. The page is titled "Demography". The left sidebar has a "Databases" menu with "Macro-economic database (BDM)" selected. The main content area shows a tree view of demographic data categories under "Demography", including "Census population", "Registry office and demographic trends", and "Couples - Families - Households". A search bar at the top right contains "Search the website" and a "Google Übersetzer" button.

### g. Select year and particular region and submit

The screenshot shows the Insee Databases website with the URL <http://www.bdm.insee.fr/bdm2/choixCriteres?codeGroupe=27>. The page is titled "Population of metropolitan France at the census, by region". It displays a "Criteria for selecting series" section with instructions and a list of available series. Below this is a "Regions (23 items)" dropdown menu containing a list of French regions. At the bottom is a red "Submit" button. The status bar at the bottom right shows "EN" and the date "27/09/2014".

The screenshot shows a web browser displaying the Insee Databases - BDM website. The URL is <http://www.bdm.insee.fr/bdm2/choixCriteres?codeGroupe=27>. The page title is "Population of metropolitan France at the census, by region". A sidebar on the left lists various statistical series, and the main content area shows a dropdown menu for selecting regions. The dropdown menu lists 23 regions, including Champagne-Ardenne, Corse (since 1970), Franche-Comté, Haute-Normandie, Ile-de-France, Languedoc-Roussillon, Limousin, Bourgogne, Midi-Pyrénées, Nord-Pas-de-Calais, Poitou-Charentes, Picardie, Rhône-Alpes, and others. At the bottom of the dropdown, there are "Select all" and "Deselect all" buttons, and a "Submit" button.

## h. Select particular result

The screenshot shows the same Insee Databases - BDM website, but now the "List series" section is visible. It displays a table of selected series, each with a checkbox in the "Select" column. The table includes columns for Heading, Click to consult a serie, Series identifier, Beginning date, Completion date, Periodicity, Update, and Select. The heading for the table is "Demography - Population without double counting at the census - Ile-de-France- Backcasted series 1851-1962 - Stopped series". The table contains 7 rows of data, each corresponding to a different region or series type. At the bottom of the table, there are "Select all", "Unselect all", and "OK" buttons, along with a "Top of page" link.

### i. Select year range for dataset

**Population of metropolitan France at the census, by region**

Selection criteria > List series > 6 series selected out of 6.

Last update	2004-03-05	2004-03-05	2004-03-05	2004-03-05	2004-03-05	2004-03-05
Year	000067420	000067423	000067428	000067430	000067430	000067439
1962	6.02.910	1.385.159	2.190.216	925.485	730.675	1.546.153
1954	3.717.258	1.274.198	1.056.039	856.345	739.929	1.449.101
1946	6.597.930	1.162.033	1.682.051	804.119	779.556	1.430.125
1936	6.785.913	1.219.457	1.664.147	838.170	798.176	1.514.284
1931	6.705.746	1.211.066	1.679.839	883.492	807.884	1.539.010
1926	6.146.374	1.193.744	1.785.779	850.437	839.748	1.529.039
1921	5.682.768	1.183.830	1.683.923	837.040	852.287	1.497.761
1911	5.335.595	1.201.034	1.931.810	911.979	960.644	1.530.203
1906	4.960.529	1.194.019	1.843.330	915.825	977.256	1.553.459

### j. Select particular data format of your choice and download the result dataset for analysis.

**Population of metropolitan France at the census, by region**

Selection criteria > List series > 6 series selected out of 6.

**Download file**

File of values and characteristics of the series

All the series  
 The [0] last periods  
 Include flags  
 Sort values in chronological order (by default values are sorted from newer to older values)

File format: Excel (XLS)

## 6.2 Etalab

Etalab publishes an open platform that hosts the contributions of all. Its responsibility is that of a host. "Data.gouv.fr" promotes documentation, reliability and data enrichment as well as the identification of novel rather than definitive or political spin correlations. It belongs to the community members to ensure the quality of content they share. So the platform provides moderation tools (voting record) and certification content.



Individuals, associations and companies publish data in Open License and reuse (applications, visualizations, and web services) in the license of their choice, including allowing commercial uses.[59,60]

## 6.3 Issues

The full web site [www.data.gouv.fr](http://www.data.gouv.fr) is not in English which make it harder for international users or researchers to access the specific data easily. 2<sup>nd</sup> there are too many steps to download data, not only in the main website, but also in redirected website which make really hard for the user to download data easily. 3<sup>rd</sup> this data is only available in .xls and .csv formats only. Although datasets are in machine readable formats but they are not standard data format for technical users or for research purposes. The major issue with this site only available in French and partially available in English which make it difficult to find the specific data or information on the data.gouv.fr

# 7 Data Access Issues and Proposed Possible Solutions

There are many issues in Govt. data portals to access data but major are

- Multi-linguistic Data Portals
- Single Data Access Point for Data Sets
- Machine Readable Data Sets
- Heterogeneity of API's for Data Access

## 7.1 Multi-linguistic Data Portals

The mostly Government data portals are on single language of their specific country either in English or French or in German language etc. and available data set formats are also in their native language . Both make it hard for international researcher and scientists etc. not only to access data but also to find and understand data as well as data access policies etc. on the specific portal if they don't know the portal or data language properly.

This issue can be solved while providing both data sets and data portals in English and it will be better if they are available in other major continental or well known languages like Arabic, Chinese, Urdu, Hindi, French, German and Persian and Spanish and African etc.

## 7.2 Single Data Access Point for Data Sets

There is no single point access to download the data for public. Users have to first pass through many steps to search the specific data of their interest and then they redirected to particular website for that data and then there are also at least 5-6 steps at least to access the particular data or user came to know that he need to send data request to particular authority to get this data. It makes the users panic and as well as its wastage of both effort and time. It not only challenges the transparency of data portals or government policies but also leads to lose their interest and they will not be able to suggest new ideas.

Data should be available in a single point access and on the main data portal web site not only to ease the general public but also for international researchers and scientist as it is also easy to provide one data portal in multi-linguistic manners instead of all sub-portals.

## 7.3 Open and Machine Readable Data Sets Formats

Currently, the mostly available dataset format either is in PDF (Portable Document Format), XLS (Microsoft Proprietary Excel Spreadsheet) CSV (Comma Separated Values), HTML (Hyper Text Markup Language) etc.

**PDF** is used to represent documents with System independent layout and it is mainly human readable file format. But its tool dependent to connect metadata or extract text from it to make it machine readable which needs lot of effort and time which also leads to data validity and consistency issues.

**CSV** is open format and used to store and organize tabular data (only text and numbers). But to convert it to database, or other machine readable formats will also create new data validity and consistence issues and required some tools and efforts as well.

**HTML** is used to display web pages and other info in browser, it mainly focus on how info is displayed only and without metadata or improper form of tagging do not make it make it machine readable in a useful sense.

It is therefore recommended for data portals to and all others to stop Excel like formats for data exchange and may move to truly open and machine readable format. Machine readable simply mean to readable by a very simple program and lot of commonly preferred formats do not fit to this. E.g. Ms Excel, XML, quoted CSV etc.

A few format that fit to this are SQL dumps, JSON, RDF, data etc. and it will be more good to provide data in open linked format i.e. 5-Star Data. The UK GOVT. Data Portal has an option to provide public data to public in linked format in future. The some govt. data Portals offer data in JSON format like Germany but others do not. Here is brief info about some of recommended formats.

- **JSON:** (JavaScript Object Notation) it is derived from Java Script and used for used to represent simple data structures and associative arrays. It is language and machine independent format. It's also an open format RFC4627. It is also used by Face book to publish public social networking info and its usage is increasing in data sharing and syndication apps.
- **Metadata:** Currently, metadata is not provided with data in downloaded files. Machine readable data is most useful when information about who produced it, when, etc. is associated with the data. Both documents and structured datasets require metadata to be searchable, sharable, etc.
- **Five-star (linked open) Data:** Tim Berners-Lee, inventor of the World Wide Web, proposed a rating scheme for open government data. He proposed that making data available online is one star, making it machine readable is two-stars, using an open-format is three stars, using RDF format is four stars, and using RDF with links to other datasets is five stars. There for it is highly recommended that Govt. Data Portals data may be in linked format for its use and reuse

- **RDF:** (Resource Description Framework). It is a data language used to represent data and information as a web resource to link them together. It is getting popularity as a dataset release language because it allows common terms to be linked between datasets. It is an open standard from W3C.

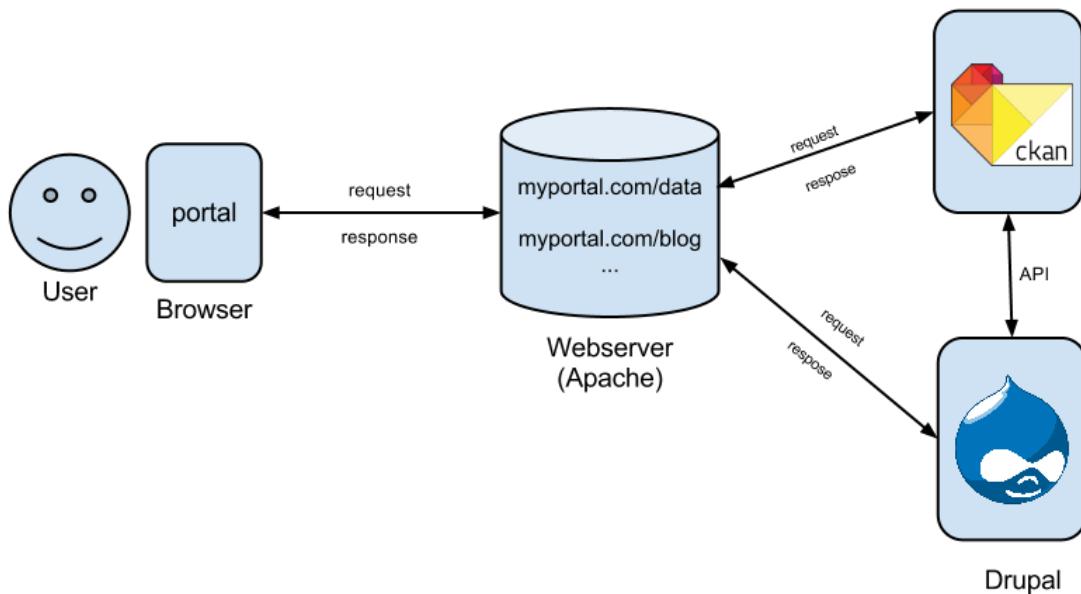
## 7.4 Heterogeneity of API's for Data Access

Currently, data portals have their own API's to access data or provide data to them. There is no any common API for open data used by all of them, it make scientists and international researchers and all other technical users harder to access all of data portal at single point. And also it needs lot of efforts and time to learn all APIs and configure them properly. It is therefore recommended that all open Govt. may use or publish a common API for all users to publish, socialize and re-usage of data.

## 8 CKAN-CMS Integration (API Integration)

The CKAN API is used some of data portals to access their data and provide data to Govt. data portals. Here is its brief description how to integrate it with your own portal to access or provide data to Govt. Data portals.

CKAN has a very extensive and customizable API and there are various ways to integrate a CMS e.g. Drupal and CKAN etc. It is recommended to take a fairly loose integration approach, where CKAN and the CMS live “side by side” and both show different parts of the portal. It is recommended over ‘facade’ or vertical approach, in which one application is ‘hidden’ behind the other as this results in a lot of work to reproduce the features of one or the other system – e.g. it is a lot of work to reproduce all the web interface features of CKAN in Drupal and vice versa, even though you have access to rich APIs on either side. On the other hand, even in a side-by-side integration there is nothing to stop you using the APIs to share information, so you get the best of both worlds.

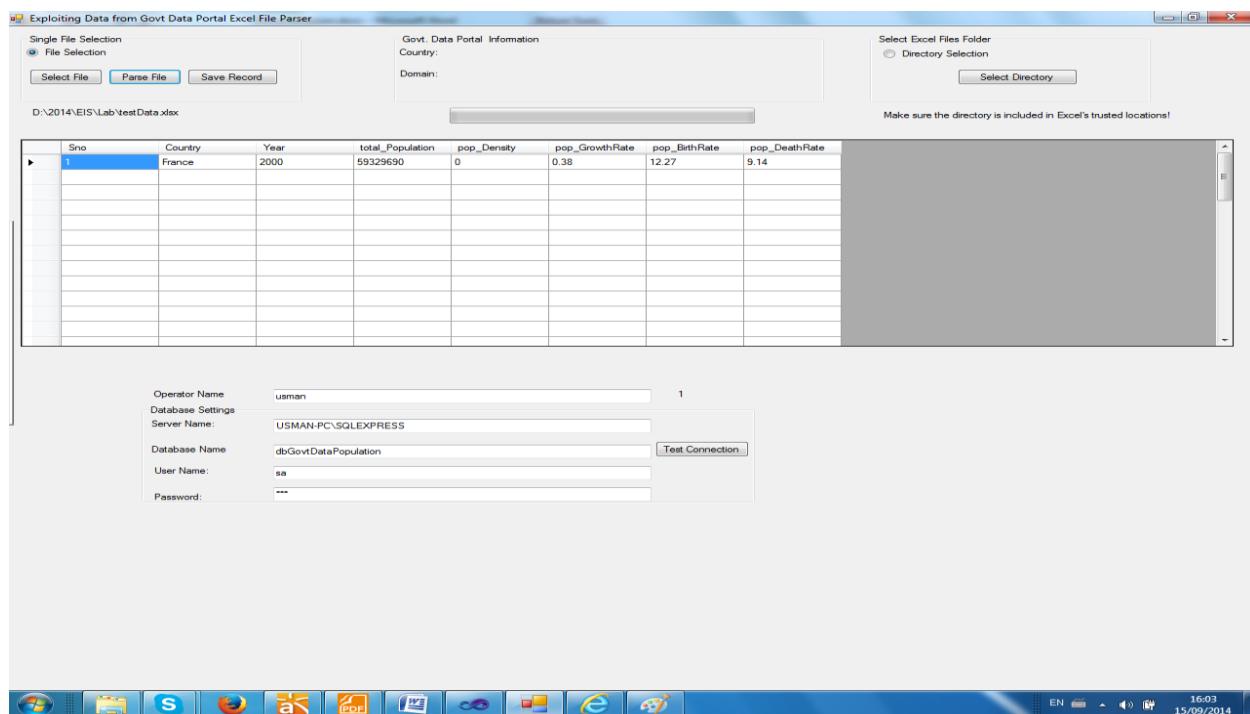


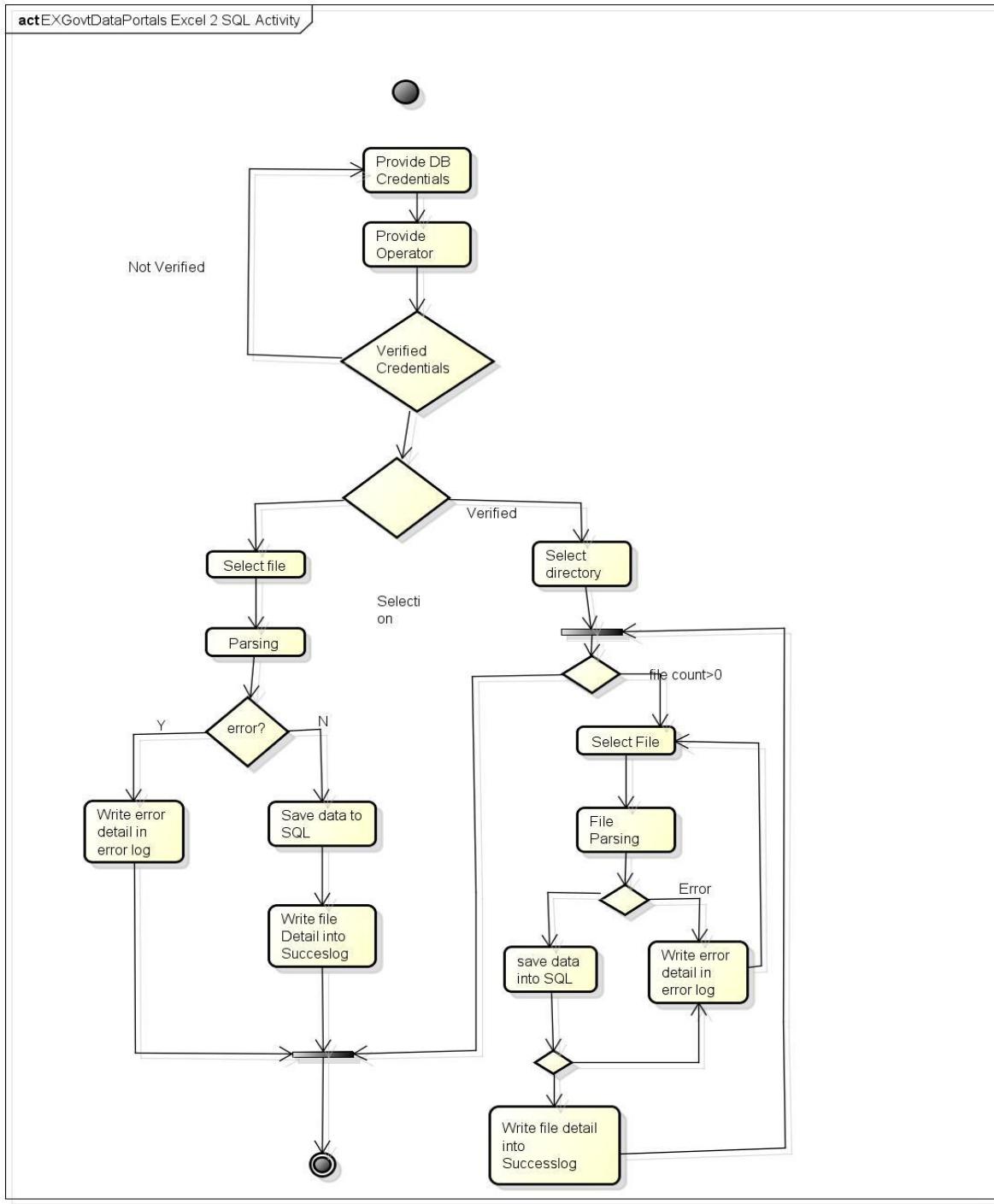
Any web server e.g. Apache needs to be set up to route the information to the correct location [61]. The only downside of this approach is that both CKAN and Drupal have to share the same theme (or at least main elements). Nonetheless, theme-sharing is generally very straightforward and is far less work than re-implementing all the features of both systems. Themes based on bootstrap version 2+ is recommended [55] as CKAN already supports this and it gives a great way of sharing style and layouts. When the two components are calling each other's JSON APIs, we also recommend using normal web API requests using e.g. drupal\_http\_request call or the python requests library[56]. This is slightly lower level than using a specific client library/Drupal module. However, it will definitely give access to the whole API and will therefore not become out of date. Also, because it is fairly simple to extend the APIs in both systems for your own custom needs, there will never be a client library/module that will support these extensions. There is a proliferation of these integration libraries/modules but any of them is not recommended as all integration cases are somewhat unique. Version 3 of the CKAN API is recommended which gives various extra features and is easy to customize. Its standard way of integration of CKAN

## 9 ExGovtDataPortalExcel2SQL Data Import Tool

This tool has been designed to import data after downloading from govt. data portals to Microsoft SQL Server automatically. It has been developed using c#, .net framework 4.5, Visual Studio 2010 and SQL Server 2008 as a windows apps. It works for both single file and a directory. This take a single file on the specified format , parse this file and show in data grid to check whether file is in proper format or not and the save the record to MS SQL . Furthermore, to select a directory or folder that may have one to unlimited files, it parse all files automatically one by one and save the record to the MS SQL database. During this process if it found some formatting error or exception or any other error during parsing, it create an error log and write all errors in the error log and move to the next file. And in case of successful importing of data to MS SQL server it create a successlog.txt and write the file name and success message in to successlog.txt and move to next file until all files in a particular directory are not processed and after processing all files user can check the success log and error log and fix the error .

Furthermore, it also process move the all success files to another success folder and error folder so that after processing user can analyze the files in detail. Its interface and architecture is as under:





Exporting Data to EIS- Microsoft SQL Architecture

# 10 Data Analysis and Visualization Tool

This tool is developed to analyze and visualize exploited data dynamically both graphically as well as in tabular form. It has been developed in C#, using Asp.net, with MS SQL server 2008 and dot Net Framework 4.5 and visual studio 2010 has been used as development tool. It visualizes dynamically and helps to analyze Population, Population Density, Growth Rate, Birth Rate, and death rate comparison of United Kingdom, Germany and France and well as it shows all these factors of individual countries for the past decades.

## 10.1 Home

### 10.1.1 Population

The total number of persons inhabiting a country, city, or any district or area

### 10.1.2 Population Density

Number of people per square kilometer or per unit Area

### 10.1.3 Growth Rate %

The average annual percent change in the population, resulting from a surplus (or deficit) of births over deaths and the balance of migrants entering and leaving a country. The rate may be positive or negative. The growth rate is a factor in determining how great a burden would be imposed on a country by the changing needs of its people for infrastructure (e.g., schools, hospitals, housing, roads), resources (e.g., food, water, electricity), and jobs. Rapid population growth can be seen as threatening by neighboring countries.

### 10.1.4 Birth Rate (Birth/1000 Population)

This entry gives the average annual number of births during a year per 1,000 persons in the population at midyear; also known as crude birth rate. The birth rate is usually the dominant factor in determining the rate of population growth. It depends on both the level of fertility and the age structure of the population.

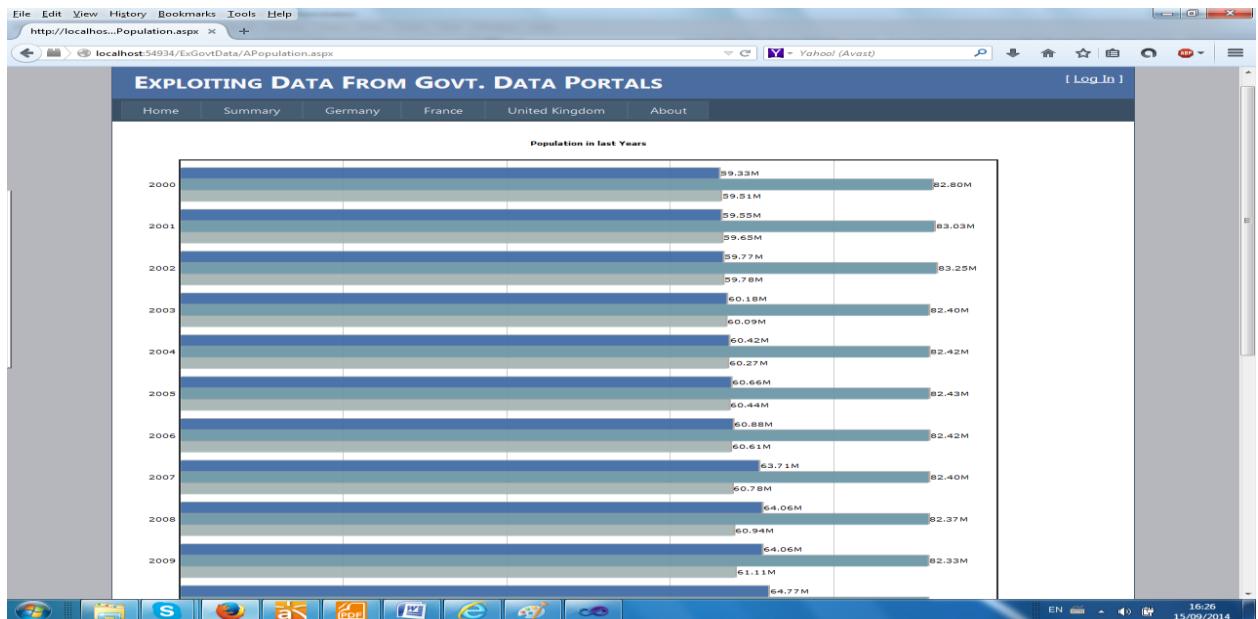
### 10.1.5 Death Rate (Death/1000 Population)

This entry gives the average annual number of deaths during a year per 1,000 populations at midyear; also known as crude death rate. The death rate, while only a rough indicator of the mortality situation in a country, accurately indicates the current mortality impact on population growth. This indicator is significantly affected by age distribution, and most countries will eventually show a rise in the overall death rate, in

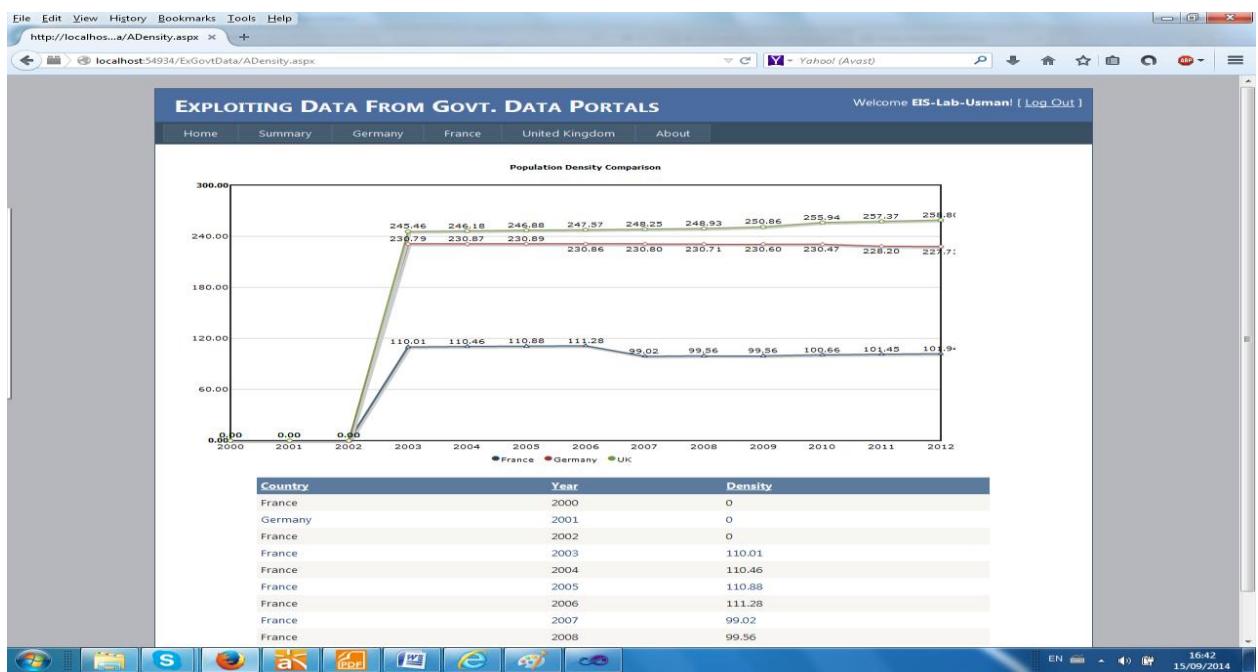
spite of continued decline in mortality at all ages, as declining fertility results in an aging population.

## 10.2 Summary

### 10.2.1 Population Germany, France and UK since 2000



### 10.2.2 Population Density Germany, France and UK



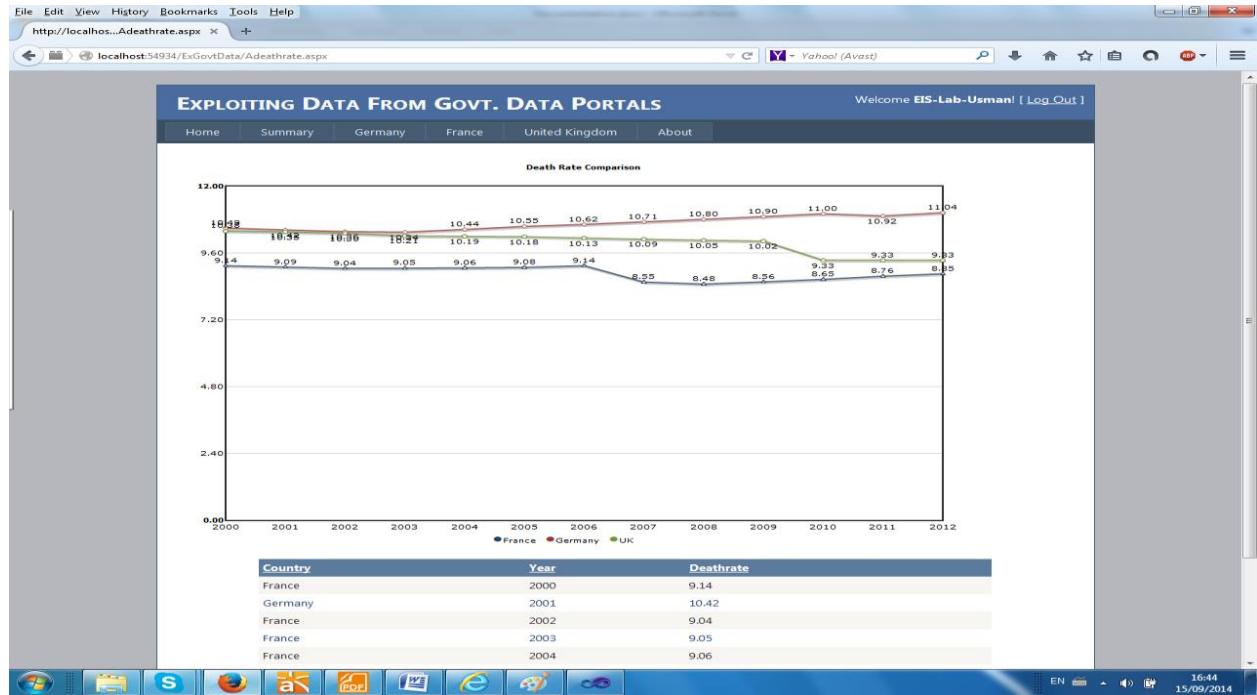
### 10.2.3 Population Growth of Germany, France and UK



### 10.2.4 Birthrate of Germany, France and UK

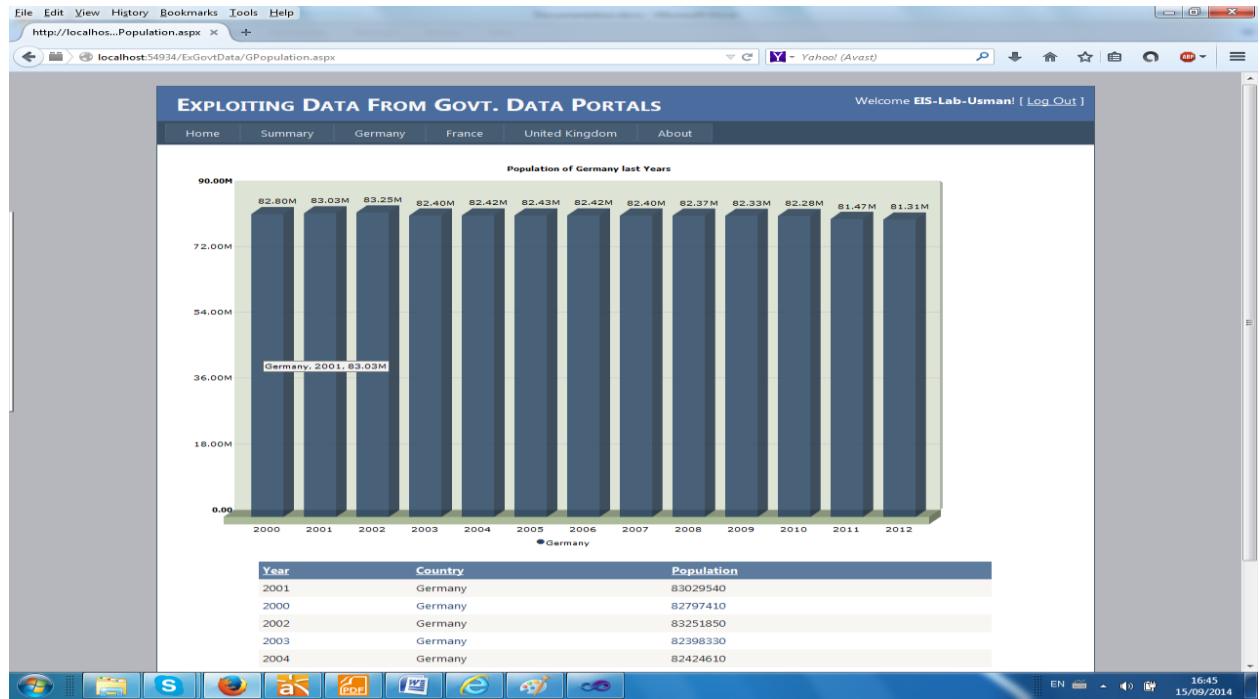


### 10.2.5 Death Rate of Germany, France and UK

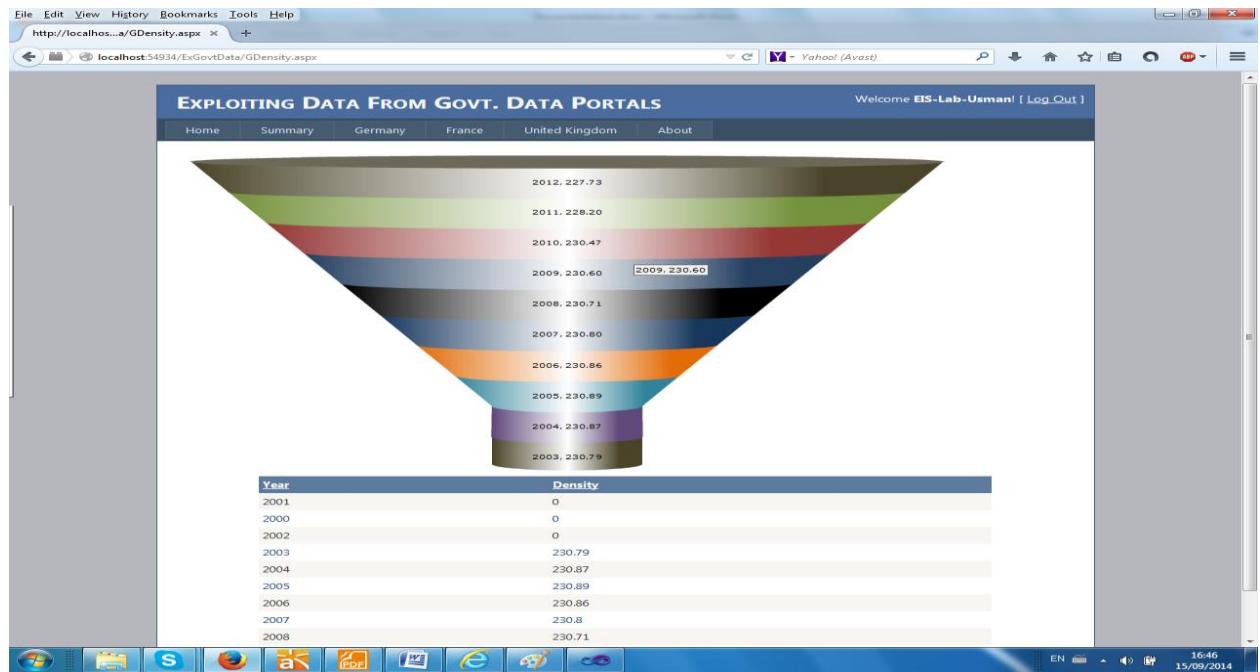


## 10.3 Data Analysis and Visualization of Germany

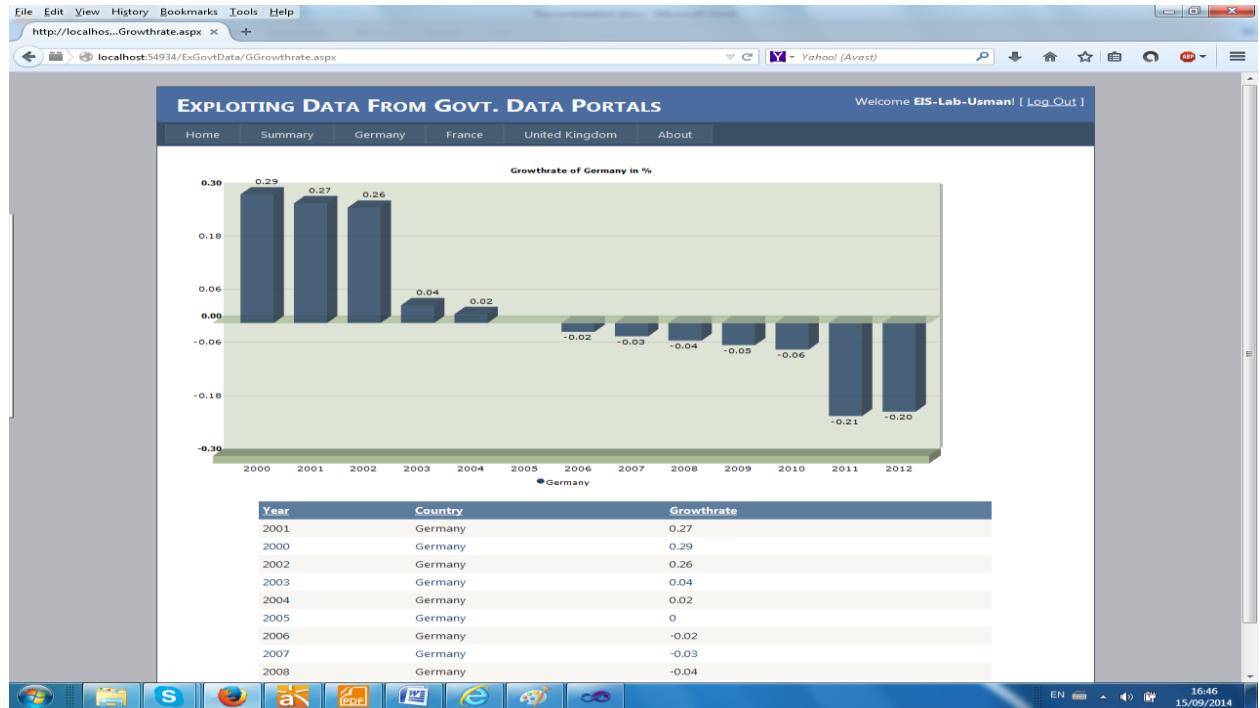
### 10.3.1 Population of Germany



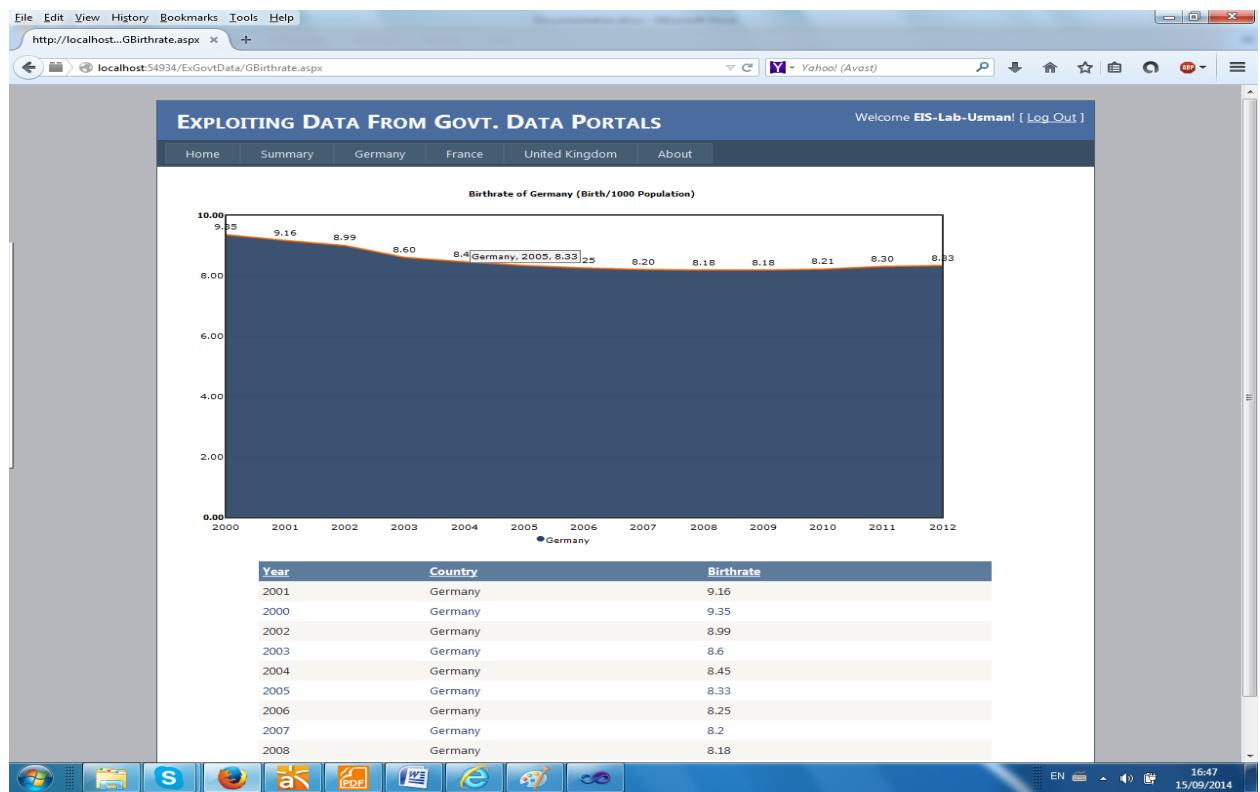
### 10.3.2 Population Density of Germany



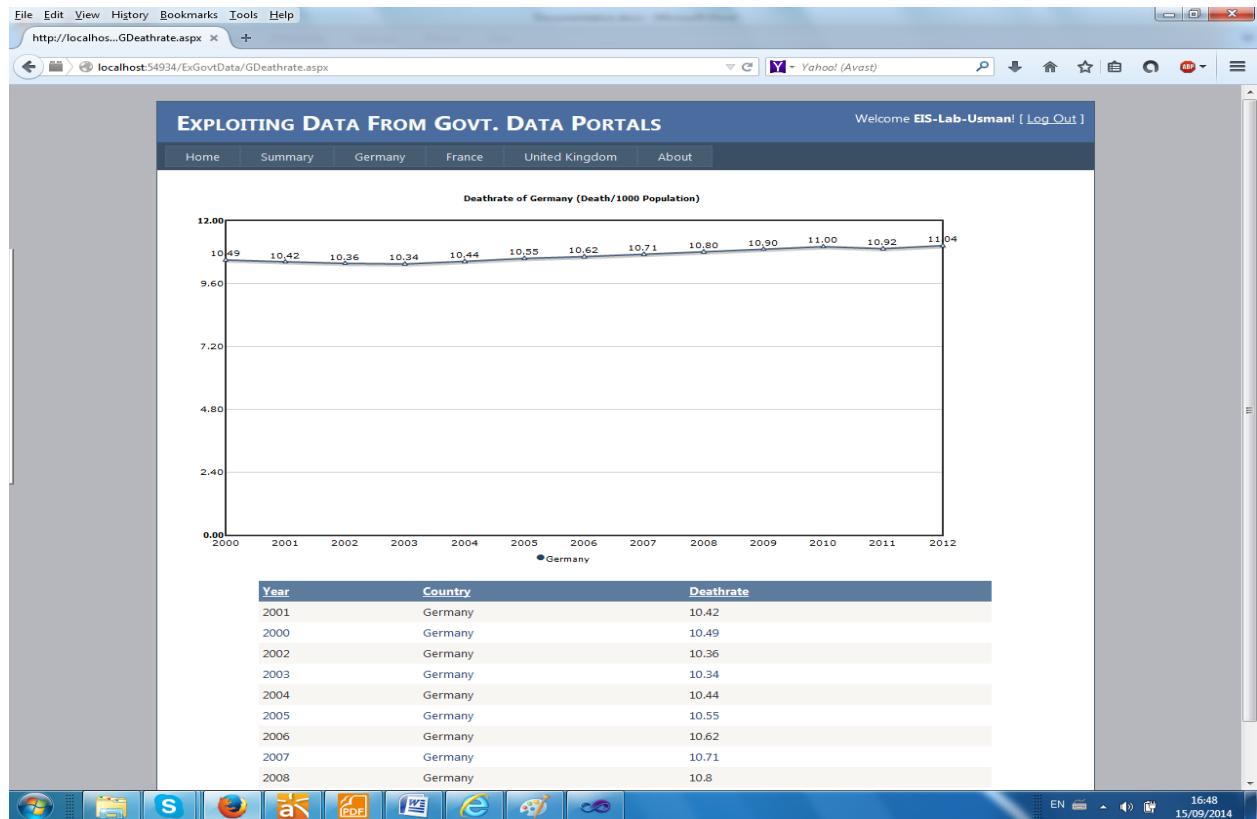
### 10.3.3 Growth Rate in Germany



### 10.3.4 Birth rate in Germany

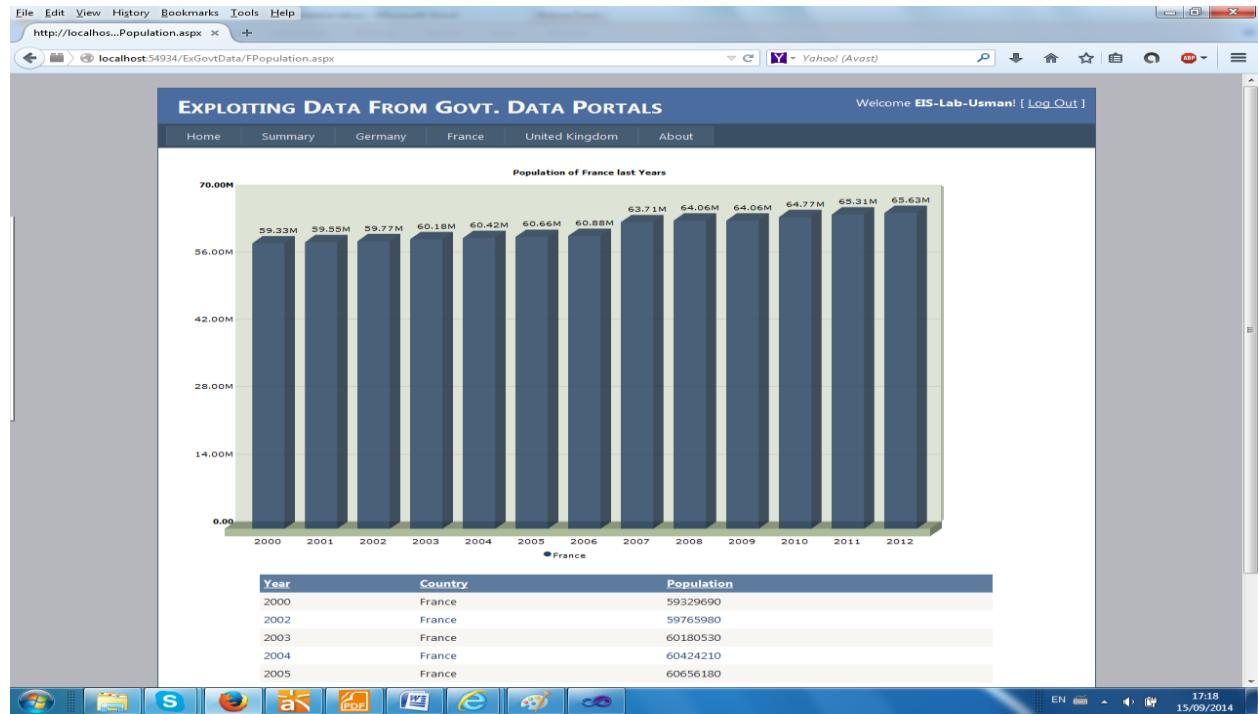


### 10.3.5 Death rate in Germany

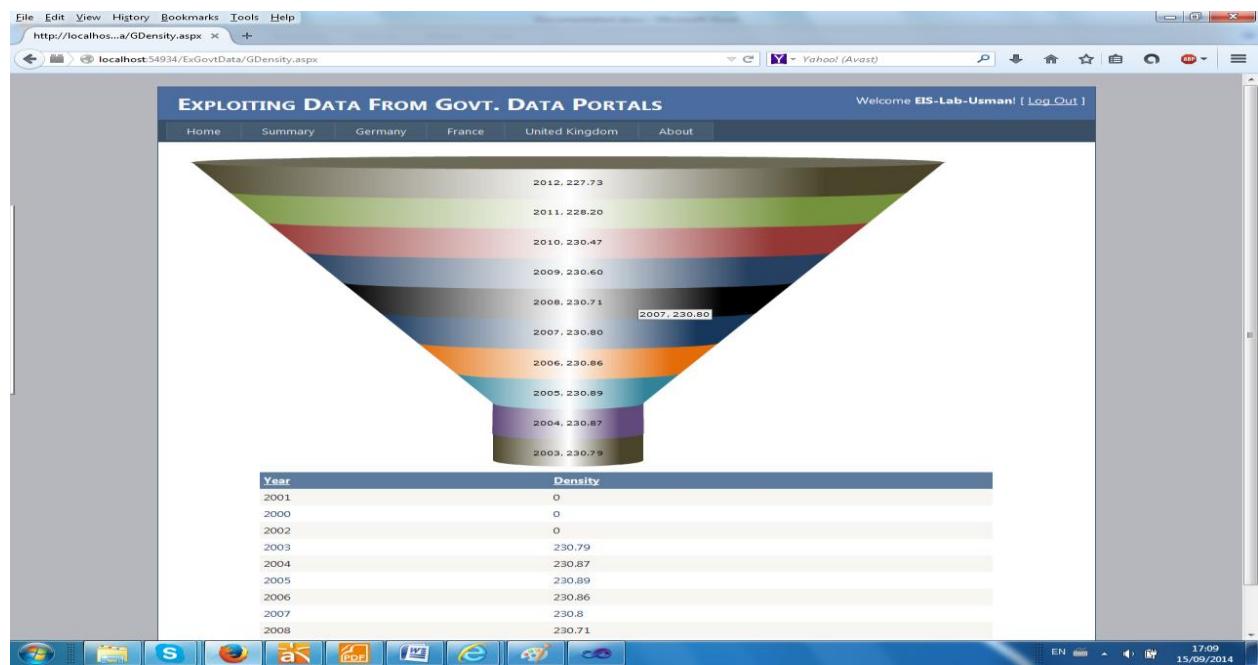


## 10.4 France Data Analysis and Visualization

### 10.4.1 Population of France



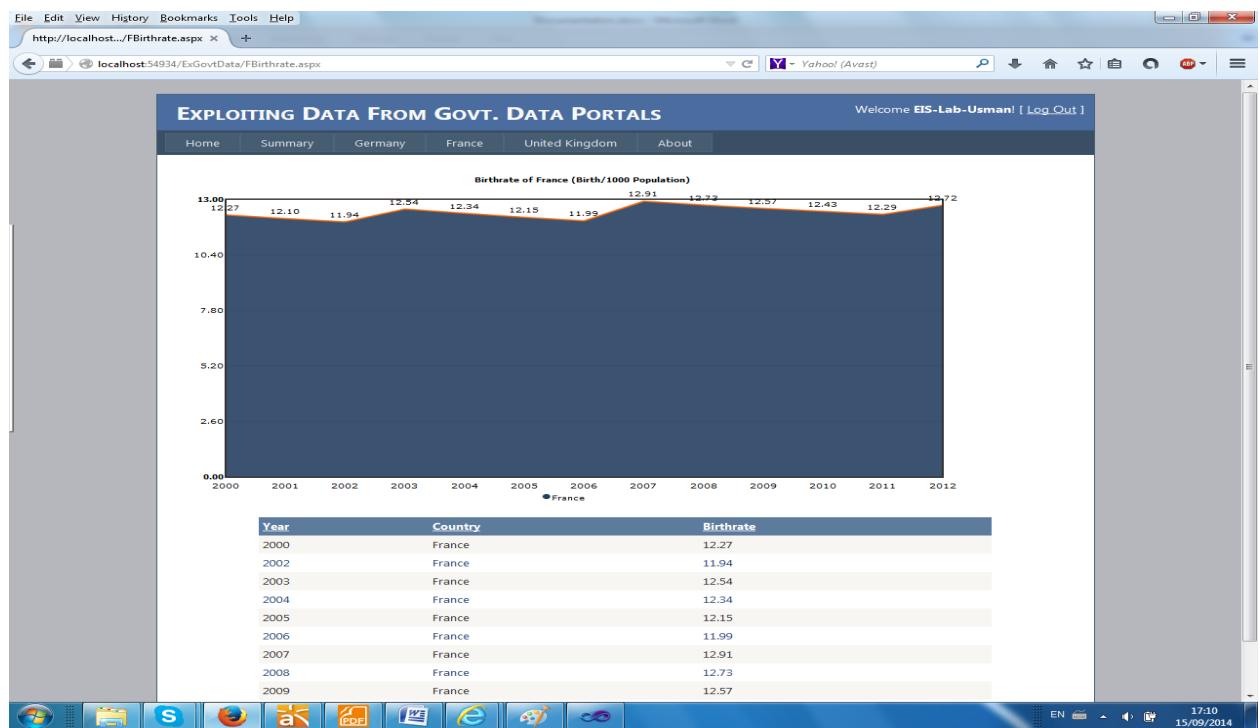
### 10.4.2 Population Density of France



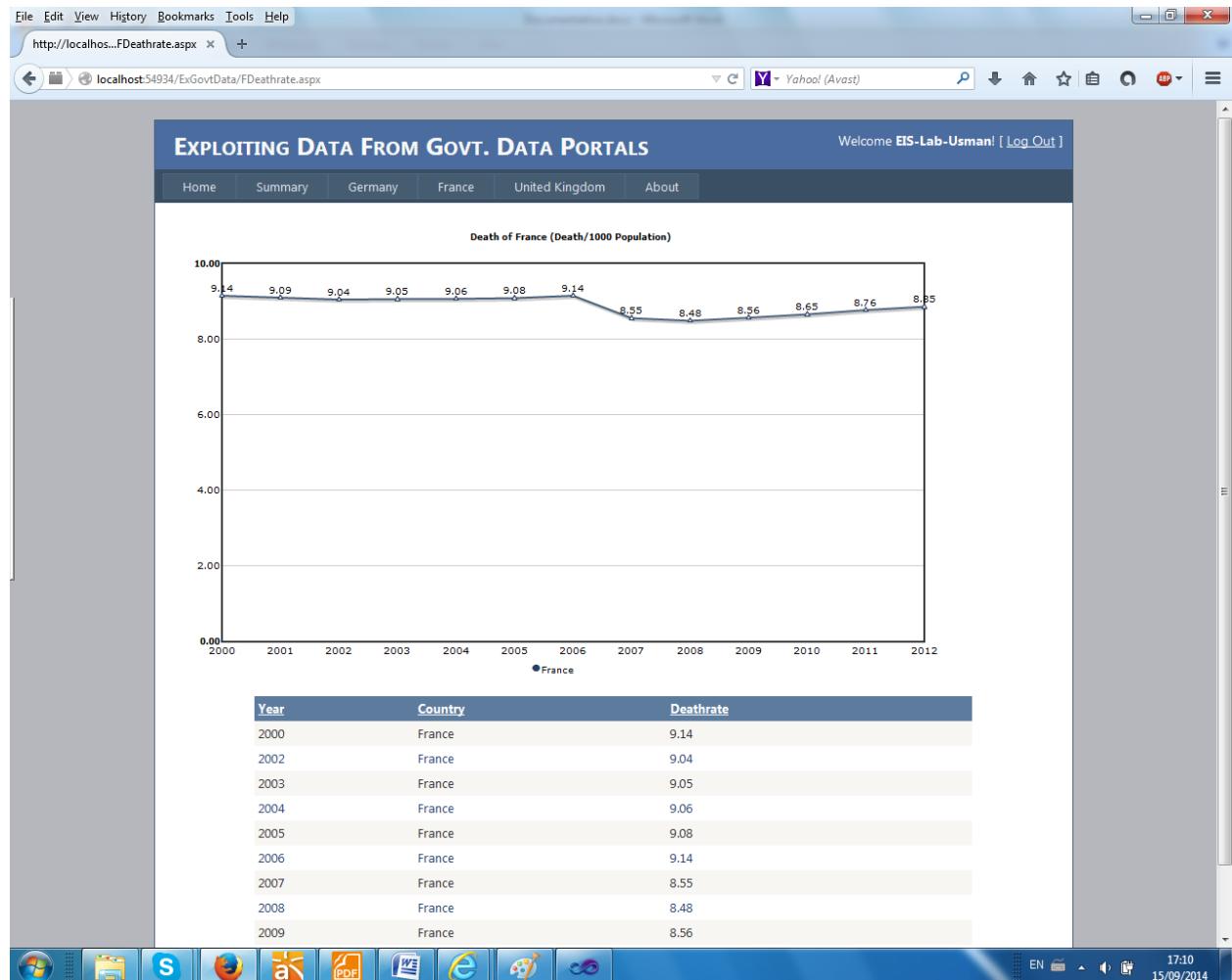
### 10.4.3 Growth Rate of France



### 10.4.4 Birth Rate of France

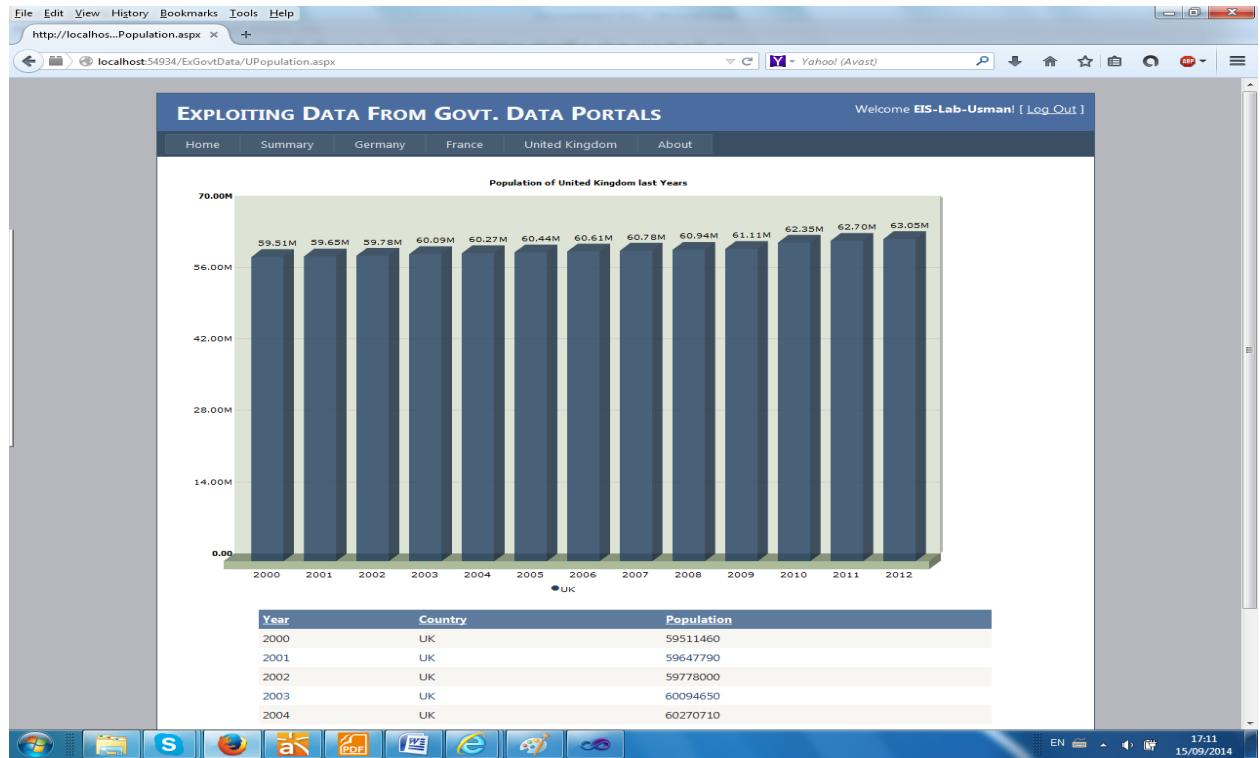


### 10.4.5 Death Rate of France

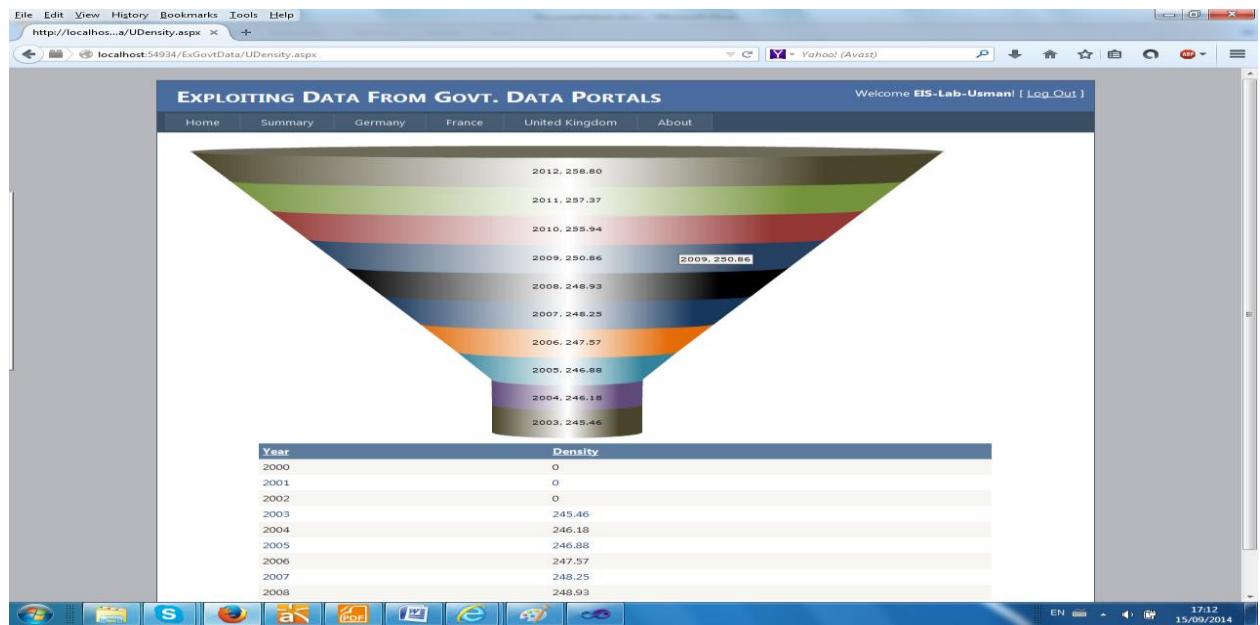


## 10.5 UK Data Analysis and Visualization

### 10.5.1 Population of UK



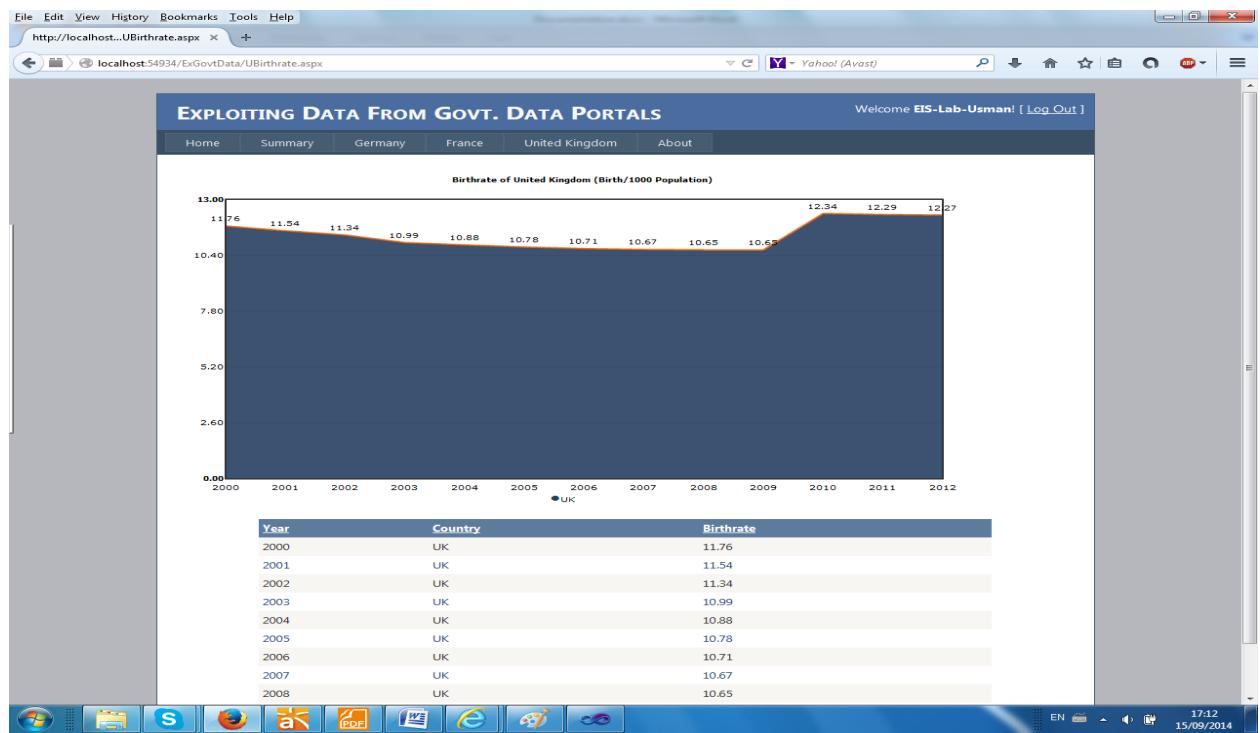
### 10.5.2 Population Density of UK



### 10.5.3 Growth Rate in UK



### 10.5.4 Birth Rate in UK



### 10.5.5 Death Rate in UK



## 11 Recommendation and future work

The current Govt. Data Portals are mostly in one language mostly native language of the portal country therefore Multi-linguistic Data Portals are recommended along with multi-lingual data sets for international researchers and public and scientists and compare policies and procedures among Governments.

At present there are too many steps to download data and it's not easy for Novice user to download and access data from data portals as data portals redirect to particular website to download data which also increase many steps to download data. It is recommended to provide all data at data portal level and there should be single data access point for information and data sets for the ease of public and to use portals more productively

Moreover, available data formats are either in human readable data format or in digital format. It is therefore recommended to use the RDF data format, JSON and Five-star (linked open) Data: Tim Berners-Lee 5-star open data format scheme for faster, easy and social access to data and more productive and usage and re-use of data.

Furthermore, there is heterogeneity of API's for Data Access among data portals although all follow Open Govt. data and metadata formats. It is therefore recommended to launch a single and one API to access all data portals at a single point. This will not only help to compare Govt. Policies and procedures but also will be so much ease for public, researcher, technical users and scientists to use and reuse this public data more productively.

Presently, due to limitation of available data formats on Govt. Data Portals “The Exploiting Govt. Data MS Excel to EIS (SQL)” export data to EIS only from MS Excel format. If in future others open format like JSON, XML, RDF etc are available then this tool may also be extended to those data format to export data to EIS as a multi data export tool.

The Data analysis and visualization web based dynamic system developed to visualize and analyze data graphically as well as in tabular form it may be further extended to represent data in linked form and also support multi domain and it may be domain independent for further research, scientific and academic usage.

# 12 References

- 1- <http://data.gov.uk/blog/integrating-ckan-and-drupal>
- 2- [http://en.wikipedia.org/wiki/Open\\_Data\\_in\\_the\\_United\\_Kingdom](http://en.wikipedia.org/wiki/Open_Data_in_the_United_Kingdom)
- 3- <https://www.gov.uk/government/publications/open-data-charter>
- 4- <https://www.gov.uk/government/how-government-works>
- 5- <https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex>
- 6- <http://neighbourhood.statistics.gov.uk/dissemination/datasetList.do?JSAllowed=true&Function=&%24ph=60&CurrentPageId=60&step=1&CurrentTreeIndex=-1&searchString=&datasetFamilyId=789&Next.x=9&Next.y=16>
- 7- [http://neighbourhood.statistics.gov.uk/dissemination/Download1.do?\\$ph=61\\_64](http://neighbourhood.statistics.gov.uk/dissemination/Download1.do?$ph=61_64)
- 8- <http://data.gov.uk/blog/datagovuk-to-go>
- 9- [wales.gov.uk/](http://wales.gov.uk/)
- 10-<http://docs.ckan.org/en/latest/api/index.html>
- 11-<http://docs.ckan.org/en/latest/index.html>
- 12-[CKAN v1.8 API](#)
- 13-[Daily updated JSON dump of the entire catalog \(data, documents, apps\).](#)
- 14-<http://www.govdata.de/ckan/api/rest/dataset/govdata-metadata-catalog>
- 15-<https://github.com/fraunhoferfokus/GovData>
- 16-[www.it-lanungsrat.de/DE/Projekte/Steuerungsprojekte/Steuerungsprojekte\\_NEGS/Open\\_Government/opengovernment.html?nn=1308406](#)
- 17-<http://www.fokus.fraunhofer.de/de/fokus/index.html>
- 18-<http://www.liferay.com/>
- 19-<https://github.com/fraunhoferfokus/ogd-metadata>
- 20-<http://numis.niedersachsen.de/>
- 21-<http://daten.berlin.de/>
- 22-<http://daten.bremen.de/sixcms/detail.php?gsid=bremen02.c.730.de>
- 23-<http://opendata.service-bw.de/Seiten/default.aspx>
- 24-<http://www.opendata.bayern.de/>
- 25-<http://daten.hamburg.de/>
- 26-<http://www.daten.rlp.de/>
- 27-<https://www-skm.destatis.de/webskm/online?nsc=true&https=1&site=>
- 28-<http://www.portalu.de/>
- 29-<http://www.geoportal.de/DE/Geoportal/geoportal.html?lang=de>
- 30-<https://www-genesis.destatis.de/genesis/online>
- 31-<http://publicdata.eu/>

- 32-<http://open-data.europa.eu/en/data/>
- 33-<http://opendefinition.org/>
- 34-[www.Sckan.org](http://www.Sckan.org)
- 35-[www.Govdata.de](http://www.Govdata.de)
- 36-[https://github.com/fraunhoferfokus/ogd-metadata/blob/master/OGPD\\_JSON\\_Schema.json](https://github.com/fraunhoferfokus/ogd-metadata/blob/master/OGPD_JSON_Schema.json)
- 37-[www.data.gouv.fr/](http://www.data.gouv.fr/)
- 38-<https://github.com/datagovuk>
- 39-<https://okfn.org/>
- 40-<https://github.com/datagovuk/ckan/tree/release-v2.2-dgu>
- 41-<https://github.com/datagovuk/ckanext-dgu/tree/stable>
- 42-<https://github.com/datagovuk/ckanext-archiver>
- 43-<https://github.com/datagovuk/ckanext-datapreview>
- 44-<https://github.com/datagovuk/ckanext-ga-report>
- 45-<https://github.com/datagovuk/ckanext-harvest/tree/2.0>
- 46-<https://github.com/datagovuk/ckanext-hierarchy>
- 47-<https://github.com/okfn/ckanext-importlib>
- 48-<https://github.com/datagovuk/ckanext-os>
- 49-<https://github.com/datagovuk/ckanext-qa/tree/2.0>
- 50-<https://github.com/datagovuk/ckanext-spatial/tree/dgu>
- 51-[http://en.wikipedia.org/wiki/Crown\\_copyright#United\\_Kingdom](http://en.wikipedia.org/wiki/Crown_copyright#United_Kingdom)
- 52-<http://misoportal.com/portal/api/request-licenced-data/432>
- 53-<http://www.w3.org/TR/gov-data/>
- 54-<http://data.gov.uk/terms-and-conditions/>
- 55-<http://twitter.github.com/bootstrap/>
- 56-<http://docs.python-requests.org/en/latest/index.html>
- 57-<http://en.wikipedia.org/wiki/G8>
- 58-[https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/207772/Open\\_Data\\_Charter.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/207772/Open_Data_Charter.pdf)
- 59-<http://fr.wikipedia.org/wiki/Etalab>
- 60-<https://github.com/etalab>
- 61-<https://gist.github.com/2782349>