

# Visualizing Open Data

Version of 19 May 2011

Group 4 - Rene Kaiser, Christoph Schinko, Peter Wetz and Pan Xueming

## Abstract

The driving force behind the rather recent development of the concept of *Open Data* is the wish to make data available for free, without any restrictions. This concept is not entirely new - it has its roots in the Semantic Web research community where the idea to make data from heterogenous sources readable to humans as well as machines was born. Initially, researchers were facing an imbalance of theoretical work in contrast to practical examples demonstrating the possibilities of these concepts. In recent years, this has changed due to the fact that there was not only a technical interest emerging as driving force, but *real world* issues regarding, for example, prevention of misuse of collected personal data by governmental institutions.

This survey addresses the concepts of Open Data as well as state-of-the-art formats, sources, visualizations and tools. Basic concepts, ideas and history of Open Data are the key aspects in this survey. A cross section of international as well as Austrian Open Data sources and visualizations deals with Open Governmental Data as well as statistical data. Visualization tools and libraries are presented regarding their purposes, highlights as well as potential drawbacks.



# Contents

<b>1</b>	<b>Introduction, Motivation and History</b>	<b>3</b>
1.1	Linking Open Data . . . . .	4
1.2	How do people currently access Open Data? . . . . .	6
1.3	Publishing Open Data . . . . .	6
<b>2</b>	<b>Open Data - Formats</b>	<b>9</b>
<b>3</b>	<b>Open Data Sources</b>	<b>11</b>
3.1	International Open Data Sources . . . . .	11
3.1.1	data.gov.uk . . . . .	11
3.1.2	data.gov . . . . .	12
3.1.3	Eurostat . . . . .	13
3.1.4	OECD.Stat Extracts . . . . .	14
3.1.5	WHO Data and Statistics . . . . .	14
3.2	Austrian Open Data Sources . . . . .	15
3.2.1	Open Government Data Austria . . . . .	15
3.2.2	Katalog Offene Daten Österreich . . . . .	15
3.2.3	open3.at project ubahnaufzug.at . . . . .	16
3.3	Common Properties . . . . .	16
<b>4</b>	<b>Open Data Visualizations</b>	<b>19</b>
4.1	International Open Data Visualizations . . . . .	19
4.1.1	SFpark . . . . .	19
4.1.2	Live train map for the London Underground . . . . .	20
4.1.3	Crimespotting . . . . .	21
4.1.4	OECD Regional Statistics . . . . .	21
4.2	Austrian Open Data Visualizations . . . . .	22
4.2.1	Wien Budget 2009 . . . . .	22
4.2.2	Woergl Haushalt 1949-2009 . . . . .	23
4.2.3	DataMaps.eu . . . . .	24
4.3	Common properties . . . . .	25
<b>5</b>	<b>Visualization Tools and Libraries</b>	<b>27</b>
5.1	Data Wrangler . . . . .	27
5.2	Many Eyes . . . . .	28
5.3	Exhibit . . . . .	29
5.4	Google Chart Tools . . . . .	30
5.5	JavaScript InfoVis Toolkit . . . . .	31
5.6	Protovis . . . . .	32

**6 Concluding Remarks**



# Chapter 1

## Introduction, Motivation and History

Open data is a philosophy and practice requiring that certain data be freely available to everyone, without restrictions from copyright, patents or other mechanisms of control (Wikipedia [2011]). Even though the concept is not entirely new, the *Open Data* (cf. Heath and Bizer [2011]) activities as we know them today are a rather recent development that, from the technical perspective, has its roots mainly in the Semantic Web research community. That research has focused on processing data from heterogeneous sources to make it accessible/readable for both humans and machines. While machines can access data via a defined interface in a standardized format, humans access through Web browsers in a well readable way.

Combining data from heterogeneous sources is not a straightforward task. While humans are quite well able to detect concepts with implicit relations (no explicit information, but may be inferred), machines need an *understanding* of the data to detect such facts. That understanding is, though with limitations, achieved by utilizing *semantic technologies*. The humans' advantage ends with large amounts of data, or even Web scale data. Researchers started to develop algorithms, formats and protocols for efficient processing. The key is to have a standardized format for the data that is generic enough not to exclude any of the legacy data repositories from being made available. The RDF/OWL format<sup>1</sup> was an obvious candidate and became interesting, and tools could be built based on existing work from other areas.

When aggregating data from different sources, one doesn't want to re-invent the wheel every time again, e.g. define specific and maybe even lossy data mappings. Data in RDF graphs may be distributed over the Web and furthermore allows to connect existing data sources on an entity level simply by adding further statements to (any of) the knowledgebases, declaring the relation between concepts or even declaring one concept *owl:sameAs* the other.

Researchers initially were facing the chicken and egg sort of problem that real data was needed for the development and evaluation of the algorithms, concepts and protocols, and on the other hand, without efficient technology being available, industry was slow to believe in the propagated economic opportunities and took a rather conservative approach in providing useful/valuable open data.

Due to that, Semantic Web research was facing an imbalance of theoretical work in contrast to practical examples demonstrating the value added of the ideas. For years, the community was somewhat working on an utopian assumption, namely that once the Semantic Web idea were to be realized, those concepts could finally be applied and evaluated. Ideally, every entity in the universe would be assigned an URI and businesses would start to publish all their data in RDF format referring to a set of well-known base ontologies that would make the combination (interlinking) of data sources much easier. By combining these data sources, new added value would be created with which those businesses could economically operate.

However, little of that ever became reality. Research tended to focus on *semantic technology* and disregarded the *online* and the *open* aspects. Local solutions using a single data source or intranet solutions not interacting with the rest of the world were not what people originally had in mind, though. Real, publicly available data being sparse became more and more of a problem. This situation was one of the main drivers for

---

<sup>1</sup>For details on data formats please refer to Chapter 2.

Open Data activities in the technical community.

So, what is the main motivation behind Open Data projects? Financial success is so far not the biggest driving force, it seems, and the situation in Austria is not too different from the rest of the world. Mainly, it is technical research interest on one side and the (in some cases politically motivated) request for openness of data within public institutions on the other. In the realm of *Open Government Data*, many examples focus on transparency of expenses of governmental institutions, or on insight into personal data which public institutions are collecting for administrative purposes — people want to know which data is collected about them in order to prevent misuse and unwarranted privacy intrusion (“big brother”). Further, there is work “for the good cause”. For example, some NGOs (non governmental organizations) and individuals enforce for example the concept of *participative democracy* by making data publicly available. Some might be inspired by historic examples like the famous cases of Florence Nightingale and Dr John Snow. Ultimately, there is business interest of course, but so far few success stories are known of companies working profitably with Open Data (beyond research and consulting). But already, there are signs that this is more than just a research playground for geeky scientists. Learning from the past, the Open Data movement will change in terms of ideas, purpose and technology. It remains to be seen how the current activities evolve and to what extent the current roadmaps will be realized and adapted.

The added value doesn’t come from publicly accessible data itself, it is generated by processing such data, by analyzing it, combining multiple data sources, linking the knowledge between multiple sources. As of today, the publication of data source as Open Data itself is viewed as a success story regardless if there is any direct or indirect use for it. However, an open RDF triple store with billions of entries being open has little benefit per se. Visualizations of such data sources could very well be a step into the right direction, as they would allow people to interactively explore and understand the data, and furthermore realize its value.

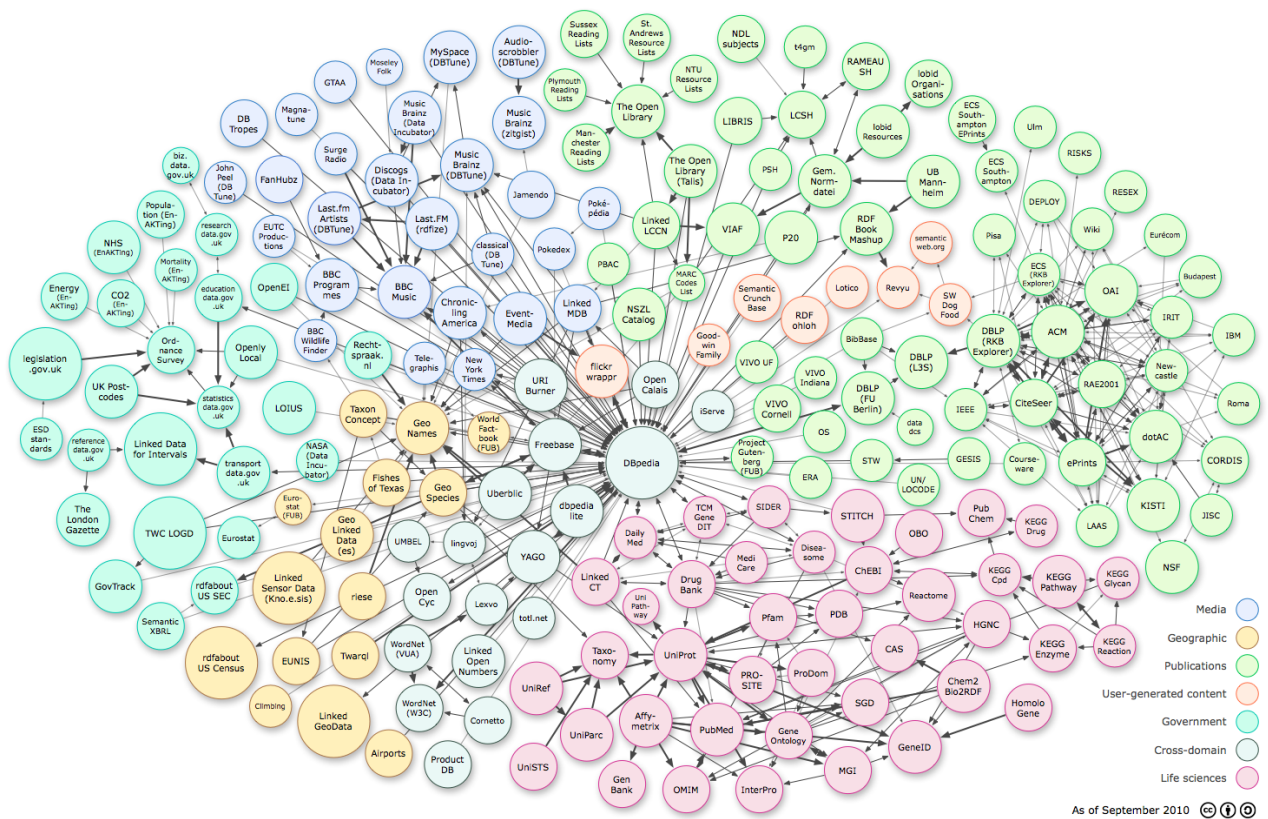
The state-of-the-art as of today, again focusing on the technical side of matters, is that a number of technical challenges need to be tackled. Open Data is usually stemming from legacy systems in all kinds of formats. Unifying the data formats in accordance with a standard — RDFizing — is a central challenge which the RDB2RDF Working Group of the W3C (W3C [2011c]) is taking care of. As one benefit of Open Data is based on the combination of different sources even beyond nice well-known mashups e.g. using geo information for visualizing data on maps, *interlinking* is another key challenge. Ontology engineering is still more of a science than an engineering field. Going beyond proof-of-concept research prototypes, practical problems appear. Most work involves only read-only access and a common mechanism for updating and extending distant Open Data repositories is missing. Such a mechanism would imply all sorts of issues such as spam and misuse which then in turn have to be dealt with themselves.

The immense data amounts themselves are of course a challenge for various processing aspects. Webscale data processing is an emerging research domain. Due to the Web based nature of Open Data, querying data sources may temporarily not be possible in a case a data source is offline, requiring caching mechanisms (or similar) for real-time applications and visualizations. In general that Open Data available is very static, few examples offer real-time updates. Semantic reasoning is required for different purposes but is a research challenge even on smaller scale data. When a reasoning mechanism inferred implicit knowledge within parts of such a huge distributed knowledgebase and some of basic statements change with high frequency, does the reasoning be done over and over again? How efficient is the access to Open Data, is it time to re-think the basic principles of the approach?

Further chapters of this document will discuss the state-of-the-art of accessing and visualizing Open Data in detail. But first, the Linking Open Data activities, current state-of-the-art of human access to Open Data besides visualizations and the typical process of publishing Open Data will be discussed.

## 1.1 Linking Open Data

Linking Open Data (Heath and Bizer [2011]; Berners-Lee [2006]; Nathan Hawks [2011]) is a key movement in the realm of Open Data. The activities are occasionally also referred to as the *Web of Data*. The well-known and ever-growing LOD cloud diagram (Richard Cyganiak and Anja Jentzsch [2010], see Figure 1.1) is a



**Figure 1.1:** The Linking Open Data cloud diagram (Richard Cyganiak and Anja Jentzsch [2010]).

visualization of data sources contributing to the Open Linked Data graph. DBpedia (Ted Thibodeau Jr [2011]) and FOAF (Friend of a Friend, Dan Brickley and Libby Miller [2011]) are very well-known examples.

According to Bizer et al [2009],

“(…) in recent years the Web has evolved from a global information space of linked documents to one where both documents and data are linked. Underpinning this evolution is a set of best practices for publishing and connecting structured data on the Web known as Linked Data. The adoption of the Linked Data best practices has led to the extension of the Web with a global data space connecting data from diverse domains such as people, companies, books, scientific publications, films, music, television and radio programmes, genes, proteins, drugs and clinical trials, online communities, statistical and scientific data, and reviews. This Web of Data enables new types of applications. There are generic Linked Data browsers which allow users to start browsing in one data source and then navigate along links into related data sources. There are Linked Data search engines that crawl the Web of Data by following links between data sources and provide expressive query capabilities over aggregated data, similar to how a local database is queried today. The Web of Data also opens up new possibilities for domain-specific applications. Unlike Web 2.0 mashups which work against a fixed set of data sources, Linked Data applications operate on top of an unbound, global data space. This enables them to deliver more complete answers as new data sources appear on the Web.”

From a technical research perspective, one focus was on the links between heterogeneous datasets, which was regarded as the key to added value of published data.

“Linked Data provides a publishing paradigm in which not only documents, but also data, can be a first class citizen of the Web, thereby enabling the extension of the Web with a global data space based on open standards - the Web of Data (Heath and Bizer [2011]).”



“The Open Data Movement aims at making data freely available to everyone. There are already various interesting open data sets available on the Web. Examples include Wikipedia, Wikibooks, Geonames, MusicBrainz, WordNet, the DBLP bibliography and many more which are published under Creative Commons or Talis licenses. The goal of the W3C SWEO Linking Open Data community project is to extend the Web with a data commons by publishing various open data sets as RDF on the Web and by setting RDF links between data items from different data sources. RDF links enable you to navigate from a data item within one data source to related data items within other sources using a Semantic Web browser. RDF links can also be followed by the crawlers of Semantic Web search engines, which may provide sophisticated search and query capabilities over crawled data. As query results are structured data and not just links to HTML pages, they can be used within other applications (W3C [2011b]).”

The movement has a famous advocate in Tim Berners-Lee who spread word in a pair of TED talks (TED Conferences, LLC [2011a,b]) and promoted W3C activities. Behind Lee, a range of researchers have shaped the research area — most with a Semantic Web research background like James Hendler, Tom Heath, Chris Bizer to just name a few. In Galway, Ireland, the DERI Enterprise Research Institute (DERI Galway [2011b]) led by Stefan Decker, is a regular contributor and also the world’s largest Semantic Web research organization. In Austria, a considerable amount of research papers was contributed by institutions like JOANNEUM RESEARCH, Salzburg Research, STI Innsbruck, Technical University Vienna, Technical University Graz and the Know Center Graz.

## 1.2 How do people currently access Open Data?

Overall, good visualization examples of Open Data are scarce, still. But, a number of projects has recently started to tackle the issue, both such using existing libraries to visualize Open Data and such creating new toolkits for Open Data visualization — examples will be discussed in Chapter 4. Without nice visualization, people use standard data browsers to access Open Data, but that has obvious limitations.


“The current state of the art with respect to the consumption of Linked Open Data for end users is RDF browsers (Berners-Lee et al; Kobilarov and Dickinson [2008]). Some tools such as Tabulator (Berners-Lee et al), Disco<sup>5</sup>, Zitgist data viewer<sup>6</sup>, Marbles<sup>7</sup>, Object Viewer<sup>8</sup> and Open link RDF Browser<sup>9</sup> can explore the Semantic Web directly. All these tools have implemented a similar exploration strategy, allowing the user to visualize an RDF sub-graph in a tabular fashion. The sub-graph is obtained by dereferencing (Berrueta and Phipps [2008]; W3C [2011a]) an URI and each tool uses a distinct approach for this purpose. These tools provide useful navigational interfaces for the end users, but due to the abundance of data about a concept and lack of filtering mechanisms, navigation becomes laborious and bothersome. In these applications, it is a tough task for a user to sort out important pieces of information without having the knowledge of underlying ontologies and basic RDF facts (Latif et al [2009]).”


The DATA.gov Catalog (DERI Galway [2011a]) is one example, a faceted browser for the data.gov catalogue, built using SIMILE Exhibit. An example is depicted in Figure 1.2.

“Faceted Wikipedia Search allows users to ask complex queries against Wikipedia. The answers to these queries are not generated using key word matching as the answers of search engines like Google or Yahoo, but are generated based on structured information that has been extracted from many different Wikipedia articles (Neofonie GmbH [2011]).”

## 1.3 Publishing Open Data

Typically, there are three basic process steps for a data source in order to be considered an Open Data source: RDFizing, interlinking and publishing. The Austria Semantic Web company (Semantic Web Company GmbH



search powered by 

[About Neofonie](#) | [About DBpedia](#) | [Imprint](#) | [Help](#)

Search

[First](#) | [Previous](#) | [Next](#) | [Last](#)

**▼ item type**

Person (3)

Scientist (3)

---

**▼ born in**

Graz (3)

---

**▼ nationality**

Austria (3)

---

**▼ died in year**

1809 (1)

1932 (1)

1878 (1)

---

**▼ born in year year**

1885 (1)

1817 (1)

1722 (1)


[Fewer](#) | [More Facets](#)


**Your Filters** Reset Filters ✕

Results 1 to 3 of 3

item type **Scientist** ✕    nationality **Austria** ✕    born in **Graz** ✕


---



**Leopold Auenbrugger** 


Josef Leopold Auenbrugger or Leopold von Auenbrugg (b. November 19, 1722, Graz, Austria; died May 17, 1809), Austrian physician who invented percussion as a diagnostic technique. On the strength of this discovery, he is considered one of the founders of modern medicine. Auenbrugger was a native of Graz in Styria, an Austrian province.

---

**Eugen Freiherr von Gorup-Besanez** 

Eugen Freiherr von Gorup-Besanez (January 15, 1817 – November 24, 1878) was an Austrian-German chemist.

---

**Ludwig Haberlandt** 

Ludwig Haberlandt (1885-1932) is known as a father of the hormonal contraception. In 1921 carried out experiments on rabbits and he demonstrated a temporary hormonal contraception in a female by transplanting ovaries from a second, pregnant, animal.

**Figure 1.2:** A DBpedia faceted browsing example: <http://dbpedia.neofonie.de/browse/rdf-type:Scientist/nationality:Austria/birthPlace:Graz>

[2011b]) has published guidelines for how to publish Open Data, from the view of Government Data (see Semantic Web Company GmbH [2011a]). In that realm, also the *the 5 stars of open linked data*, suggested by Tim Berners-Lee, have been published (cf. DERI Galway [2011c]):

- make your stuff available on the web (whatever format)
- make it available as structured data (e.g. excel instead of image scan of a table)
- non-proprietary format (e.g. csv instead of excel)
- use URLs to identify things, so that people can point at your stuff
- link your data to other people's data to provide context

The Linked Data principles (based on Berners-Lee [2006]) have been interpreted by Richard Cyganiak and Anja Jentzsch [2010] as follows:

- There must be resolvable `http://` (or `https://`) URIs.
- They must resolve, with or without content negotiation, to RDF data in one of the popular RDF formats (RDFa, RDF/XML, Turtle, N-Triples).
- The dataset must contain at least 1000 triples. (Hence, your FOAF file most likely does not qualify.)
- The dataset must be connected via RDF links to a dataset that is already in the diagram. This means, either your dataset must use URIs from the other dataset, or vice versa. We arbitrarily require at least 50 links.
- Access of the entire dataset must be possible via RDF crawling, via an RDF dump, or via a SPARQL endpoint.

Despite such guidelines, it has to be accepted that some of early examples (and some are still to be found in the LOD cloud) have serious issues regarding Open Data quality, consistency, accuracy and completeness. Ironically, the Linked Data out there is not linked to a very high degree. As apparent from the graph diagram itself, DBpedia is a positive example that a lot of other sources are referring to, but that's not the case for every linked data source.

The rest of the document is structured as follows. In Chapter 2 representations and formats of Open Data are discussed, focusing on RDF. Chapters 3 and 4 will then offer a selection of examples of both Open Data sources and visualizations. Event though some of the examples are of the *geo viz* rather than the *info viz* domain, we decided to include them as they are good representatives of the state-of-the-art. Finally, we offer concluding remarks in Chapter 5.

## Chapter 2

# Open Data - Formats

In brief, Open Data available on the Web can be found in manifold formats. Examples are CSV, XLS (proprietary, version dependent) and JSON representation. From the technical research perspective of this report, the RDF/OWL sources are the most interesting. In general, Open Data projects from the Linking Open Data initiative aim to publish in RDF. The basic idea and the most important characteristic behind that data format is to have compatibility between different sources without the need of e.g. an explicit case-dependent mapping. Entities are identified via URIs. The The LOD Cloud Tom Heath [2011] depicts many examples and two blog posts from Lee Feigenbaum explain the topic in detail (Lee Feigenbaum [2007a] and Lee Feigenbaum [2007b]). An RDF/XML example is shown in Figure 2.1.

“RDF is a standard model for data interchange on the Web. RDF has features that facilitate data merging even if the underlying schemas differ, and it specifically supports the evolution of schemas over time without requiring all the data consumers to be changed. RDF extends the linking structure of the Web to use URIs to name the relationship between things as well as the two ends of the link (this is usually referred to as a “triple”). Using this simple model, it allows structured and semi-structured data to be mixed, exposed, and shared across different applications. This linking structure forms a directed, labeled graph, where the edges represent the named link between two resources, represented by the graph nodes. This graph view is the easiest possible mental model for RDF and is often used in easy-to-understand visual explanations (W3C [2011d]).”

The RDB2RDF Working Group (W3C [2011c]) of the W3C is working on the topic of RDFizing legacy repositories. The SPARQL query language (W3C [2011e]) is already established as the common query format, comparable to the status of SQL for relational databases. The research community is further working on efficient *triple Stores* that allow fast access to massive data amounts. Still, RDF may not be the final answer, the best possible solution to that task. Some (XML) serialisations imply considerable data overheads.

```

- <rdf:RDF xml:base="http://ec.europa.eu/eurostat/ramon/rdfdata/countries/" xml:lang="en">
- <rdf:Description rdf:about="">
  <rdfs:label>Eurostat country codes</rdfs:label>
  <dcterms:title>Eurostat country codes</dcterms:title>
- <dcterms:source>
  Eurostat - http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Glossary:Country_codes
  </dcterms:source>
- <dcterms:subject>
  Country codes, Candidate countries, European Free Trade Association (EFTA), European Union (EU)
  </dcterms:subject>
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/3.0/lu"/>
  <cc:morePermissions rdf:resource="http://ec.europa.eu/geninfo/copyright_en.htm"/>
</rdf:Description>
- <Country rdf:about="BE">
  <owl:sameAs rdf:resource="http://ec.europa.eu/eurostat/ramon/rdfdata/nuts2008/BE"/>
  <owl:sameAs rdf:resource="http://sws.geonames.org/2802361"/>
  <owl:sameAs rdf:resource="http://rdfdata.eionet.europa.eu/eea/countries/BE"/>
  <code>BE</code>
  <name xml:lang="en">Belgium</name>
  <rdfs:label xml:lang="en">Belgium</rdfs:label>
  <protocolOrder rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">1</protocolOrder>
</Country>

```

**Figure 2.1:** An RDF example for Eurostat country codes: <http://ec.europa.eu/eurostat/ramon/rdfdata/countries.rdf>.

## Chapter 3

# Open Data Sources

The idea behind the term *open* has been thoroughly discussed in the introduction chapter. All kinds of sources can serve as input for the various formats Open Data can be stored in. In the following we discuss some examples of what kinds of input are available freely in the form of Open Data sources: scientific data, medical data, governmental data, environmental data. The various different Open Data formats, discussed in the previous chapter, find their applications in the form of Open Data sources.

### 3.1 International Open Data Sources

It is noticeable that there is an extensive amount of governmental data available as Open Data. This is, to a certain extent, not surprising at all. Every person on the planet lives in some sort of society being governed by one or more individuals and therefore is potentially interested in governmental data. Also it is one of the fields Semantic Web scientists viewed as a key enabler for the Open Data initiatives. A number of international websites are starting points for Open Government Data and statistical data. In the following sections a selection of prominent examples is presented.

#### 3.1.1 data.gov.uk

Non-personal data acquired by the UK government for official purposes is made available on the website *data.gov.uk* (United Kingdom Public Sector Transparency Board [2011]). The project was launched as a beta version in January 2010 offering around 2500 datasets. As of May 2011 it has yet not reached final status, but the amount of datasets has more than doubled and now counts 6600. Popular tags are:

- *health* with 1157 datasets
- *care* with 819 datasets
- *transparency* with 773 datasets
- *communities* with 659 datasets
- *child* with 608 datasets
- *health-and-social-care* with 607 datasets
- *children* with 570 datasets
- *local-government* with 568 datasets

All data is available under a worldwide, royalty-free, perpetual, non-exclusive license in/via one or more of the following formats/APIs: *CSV, XLS, XML, RDF, JSON, PDF, HTML, SPARQL*. Figure 3.1 shows the search

The screenshot shows the data.gov.uk website interface. At the top, there is a navigation bar with the HM Government logo and a 'Log in or sign up' link. Below this, the site's name 'data.gov.uk' and the tagline 'Opening up government' are displayed. A search bar with the placeholder text 'What are you looking for?' and a 'Search' button is present. A horizontal menu contains links for 'Data', 'Apps', 'Ideas', 'Forum', 'Wiki', 'Blogs', 'Linked Data', 'Resources', and 'About'. The main content area is titled 'Data' and includes a search input field with a 'Search' button. Below this, a text block states: 'Looking for something specific, or just want to know more about how Government spends your money? You'll find over 6,800 datasets to help you get answers.' There are two main browsing sections: 'Browse by publisher' and 'Browse by nation'. The 'Browse by publisher' section lists various government departments and their dataset counts, such as 'Cabinet Office (44)', 'Department for Business, Innovation and Skills (159)', 'Department for Communities and Local Government (783)', 'Department for Environment, Food and Rural Affairs (407)', 'Department for Transport (151)', 'Department of Health (1059)', and 'UK Statistics Authority (715)'. The 'Browse by nation' section includes a map of the UK and lists 'England', 'Northern Ireland', 'Scotland', and 'Wales'. On the right side, there is an 'EDITOR'S PICK' section featuring 'COINS – HM Treasury's database of public spending'. Below that is a 'Popular tags' section with a 'View all tags' link and several tags with their respective counts: 'health (2,314)', 'care (1,638)', 'transparency (1,546)', 'communities (1,318)', 'child (1,216)', 'health-and-social-care (1,214)', 'children (1,140)', and 'local-government (1,136)'. At the bottom right, there is a 'Developers' section with a note: 'You can also access this catalogue via the API or download it as a CSV/JSON dump.'

**Figure 3.1:** The search and browse interface of the website *data.gov.uk* offers an easy to use interface and various criterions to narrow down the search/browse space.

and browse interface which offers aids to narrow down the search/browse space as well as ordering of search results. Although this website offers a huge amount of data, it is not limited to provide raw numbers. There are around 100 apps (visualizations, widgets, gadgets) available to provide a meaningful view on the sources. The website builds up on CKAN (CKAN [2011]).

### 3.1.2 data.gov

Public data which the US government collects and which is not private or restricted for national security reasons is made available on the website *data.gov* (United States Government [2011]). The website was launched in late May 2009 starting with around 47 datasets. As of May 2011 it offers 379932 raw and geo-spatial datasets. Categories include:

- Agriculture with 112 results
- Arts, Recreation, and Travel with 9 results
- Banking, Finance, and Insurance with 54 results
- Births, Deaths, Marriages, and Divorces with 55 results
- Business Enterprise with 57 results
- Energy and Utilities with 117 results
- and many more...

By accessing datasets one automatically agrees to the data policy found on the website, no further licensing information is available. Datasets are available in/via one or more of the following formats/APIs: *CSV*, *JSON*, *PDF*, *RDF*, *RSS*, *XLS*, *XLSX*, *XML*, *SPARQL*, *SODA*. The search/browse interface is shown in Figure 3.2. It allows narrowing down the search/browse space as well as ordering of search results according to various criteria. Additionally, there are around 1000 apps (visualizations, widgets, gadgets, RSS feeds) available that

The screenshot shows the data.gov website interface. At the top, there is a navigation bar with links for HOME, NEXT GENERATION, RAW DATA, APPS, GEODATA, COMMUNITY, METRICS, OPEN DATA SITES, GALLERY, and WHAT'S NEW. Below the navigation bar, there is a main banner area with a large image and text. The main content area is divided into several sections, including a search bar, a list of datasets, and a sidebar with filters. The search bar is located on the left side of the main content area. The list of datasets is located in the center, and the sidebar is on the right. The datasets listed include 'Survey of Patients' Hospital Experiences (HCAHPS)', 'Hospital Outcome Of Care Measures', 'Hospital General Information', 'Nursing Home Compare - Survey Dates', 'Hospital Medicare Payment And Volume Measures', 'Nursing Home Compare - Survey Deficiency', 'Nursing Home Compare - Provider Ratings', 'Home Health Care Facilities', and 'California DRG Payments by Condition'. Each dataset entry includes a title, a brief description, and the number of views.

**Figure 3.2:** The extensive search and browse interface of the website *data.gov* offers ordering of search results and various criterions to narrow down the search/browse space (e.g. maps or charts only).

build up on the raw data to make it easily understandable. No information could be found on the software powering the website.

### 3.1.3 Eurostat

Eurostat is the statistical office of the European Union (EU) and it is situated in Luxembourg (Eurostat [2011]). The office was founded in the year 1953 and its task is to provide statistics and to enable comparisons between European countries and regions. Since 2004 all statistical data except microdata is available free-of-charge for research purposes. As of May 2011 Eurostat offers around 4000 datasets. They are available in/via one or more of the following formats/APIs: *XLS*, *TXT*, *HTML*, *PC-AXIS*, *SPSS*, *DFT*, *SDMX*. The main areas of statistical activities include:

- EU Policy Indicators
- General and regional statistics
- Economy and finance
- Population and social conditions
- Industry, trade and services
- Agriculture and fisheries
- and more...

The search interface offers various filters like recently updated or recently updated maximum number of items to narrow down the search base. Additionally browsing through the data, or downloading a snapshot of the whole database, is possible. There is no information on structured data available, however there are



unrelated efforts to convert Eurostat datasets to RDF format on CKAN (CKAN [2011]). Downloading and reproduction of data and documents for personal, non-commercial or commercial use is authorized as long as proper acknowledgment is given.

### 3.1.4 OECD.Stat Extracts

The OECD publishes statistics, among other materials, on a wide number of subjects. All statistical data is available online since 1998 via the website *stats.oecd.org* (OECD [2011d]). It offers a rich search/browse interface for its database. No additional refinement operations are available for the search interface but there is an option to store queries for later use. There are no concrete numbers on how many datasets are available. Categories include:

- Cross-Cutting and Horizontal Products
- Macroeconomics and Growth
- Finance and Investment
- Education
- Social Issues, Migration and Health
- Employment and Labour Markets
- and more...

Data is available in/via one or more of the following formats/APIs: *XLS*, *TXT*, *PC-AXIS*, *SDMX*. In addition to raw data there are charts, graphs and maps available via the website. Downloading and reproduction of data and documents for personal, non-commercial or commercial use is authorized as long as proper acknowledgment is given.

### 3.1.5 WHO Data and Statistics

The WHO publishes a number of datasets as well as visualizations via the website *www.who.int/research/en/* (WHO [2011]). It provides access to over 50 datasets on health topics including:

- Mortality and Burden of Diseases
- Child Nutrition and Child Health
- Maternal and Reproductive Health
- Epidemic-prone Diseases
- Health Systems
- and more...

Additionally there are statistical reports and indicators available. The search interface offers advanced settings like searching for specific file types or domains to narrow down the search space. Browsing the data is possible via a tree-view of categories. Data and metadata is available in/via one or more of the following formats/APIs: *XLS*, *XML*, *CSV*, *HTML*. To make the raw data more accessible there are a number of visualizations, primarily maps, since most of the information is of geospatial nature. To quote the website (WHO [2011]) on the licensing issue:

“The information presented on the website is protected under the Berne Convention for the Protection of Literature and Artistic works, under other international conventions and under national laws on copyright and neighboring rights.”

## 3.2 Austrian Open Data Sources

There is considerable activity concerning open data in Austria. The open3 network (open3 network [2011]), for example, supports a number of projects in the field of open society, open government and open data. The open3 network was founded early 2010 and positioned itself to enable knowledge transfer between the involved groups (politics, administration, population and economy). Projects like *datamaps.eu*, which provides a visualization framework, or budgetary visualizations for the years 2011 to 2014 are available on *open3.at*. An interesting project in the context of the open3 network is *ubahnaufzug.at* (*ubahnaufzug.at* [2011]) which deals with the specific problem of geospatially mapping mobility data for disabled people in Vienna. There is also a lot of publicity and movement around governmental data. Especially two websites are involved in providing and collecting information about governmental data, namely Open Government Data Austria (*gov.opendata.at*) and Katalog Offene Daten Österreich (*offener.datenkatalog.at*).

### 3.2.1 Open Government Data Austria

The goal of the Open Knowledge Forum (Verein Open Knowledge Forum Österreich [2011]) is to provide access to non-personal data. This should be achieved via a central repository providing the data as linked open data in human- and machine-readable form. In Austria, first discussions of this topic started amongst semantic web scientists who found that it would be an ideal *playground* to make use of the recently developed technologies. Also there were discussions during the Linked Data Camp (30.11.-01.12. 2009, Vienna) that it would be of great benefit if the government would open access to databases like geographic information, timetables, traffic or statistical data. After a kick-off meeting between organizations and companies in April 2010 the initiative Open Government Data Austria was formed and the website *gov.opendata.at* created.

### 3.2.2 Katalog Offene Daten Österreich

The screenshot shows the search interface of the website *offener.datenkatalog.at*. At the top, there is a search bar and navigation links: Start, Search, Add a package, Tags, Gruppen, Über. The main search area includes a search bar and filter options: Filter by  packages with open licenses  packages with downloads. Below the search bar, there are several search results:

- Geokoordinaten der U-Bahn-Stationen in Wien** (CSV, OPEN DATA). Ursprüngliche Version von Max Kossatz (www.wissenbelastet.com), überarbeitet von Robert Harm (www.harm.co.at).
- Ergebnisse der Konsultation bezgl. der EU Public Service Information - Richtlinie 2010** (PDF, XLS, XLSX, PDF, OPEN DATA). Ergebnisse der Online-Konsultation bezgl. der PSI Directive die vom 9 September 2010 bis 30 November 2010 durchgeführt wurde.
- Wien Budget 2009 Ausgaben** (JSON, OPEN DATA). Basis für die Visualisierung auf open3.at Daten wurden noch nicht vollständig übertragen, d.h. nicht alle Details bis in die 3. Ebene werden dargestellt.
- Erledigungen der Asylbehörden 1981 - 2009** (HTML, OPEN DATA). Ein von der Asylkoordination zusammengestellter Datensatz aus den Asylberichten des österreichischen Innenministeriums. Die weitere (http://www.asyl.at/fakten\_8/stat\_2010\_03.htm...
- Facebook Nutzer/innen in Österreich 2010** (CSV, OPEN DATA). Facebooknutzung nach Alter, Geschlecht und Region.

On the right side, there is a yellow box titled "Add a package" with the text: "Do you know of a dataset that should be added to Kataloge Offene Daten Österreich? Register it now."

**Figure 3.3:** The search interface of the website *offener.datenkatalog.at* offers basic options to narrow down the search/browse space (e.g. open licenses, downloads only).

A collection of Open Data that goes beyond governmental data can be found on the website *offener.datenkatalog.at*.

The goal of the initiative Katalog Offene Daten (Open Knowledge Foundation [2011]) is to make data of all kinds of different fields like politics, libraries and science freely and easily available. It is a Verein Open Knowledge Forum Österreich project and builds up on CKAN (CKAN [2011]). As of May 2011 the site seems to be offline, but there were 11 datasets of fields like government, social sciences and public transport in/via one or more of the following formats/APIs: PDF, XLS, CSV, RDF, HTML, JSON, SPARQL available. The search interface shown in Figure 3.3 offers basic options to narrow down the search space. All data is available under the Creative Commons Attribution 3.0 license.

### 3.2.3 open3.at project ubahnaufzug.at

**Alle gemeldeten Störungen**

Sollte eine gemeldete Störung bereits behoben sein, hinterlassen Sie dies bitte als Kommentar zum jeweiligen Eintrag.  
Das Administrationsteam wird den Status anschließend so rasch wie möglich ändern.

Seite 1 von 9 | 10 Störungen von 90

U-Bahn-Station	Kategorie	Status	Zusatzinfos	Gemeldet am
U4 Heiligenstadt	Aufzug	außer Betrieb	Richtung Hütteldorf	13.04.2011 13:31 Uhr
U1 Karlsplatz	Rolltreppe	außer Betrieb	Rolltreppe beim momentanen Ausgang zur Station von Badnerbahn & Strassenbahnlinie 2 ('Oper Bstg. C') ausser Betrieb 'aus Sicherheitsgrunden'.	13.04.2011 08:25 Uhr
U2 Schottentor	Rolltreppe	außer Betrieb		07.04.2011 20:15 Uhr
U2 Volkstheater	Rolltreppe	außer Betrieb	Richtung Volkstheater	30.03.2011 09:08 Uhr
U2 Aspernstraße	Aufzug	außer Betrieb	1/2	29.03.2011 21:30 Uhr
U2 Museumsquartier	Aufzug	außer Betrieb	Richtung Karlsplatz	24.03.2011 09:16 Uhr
U3 Kardinal-Nagl-Platz	Aufzug	außer Betrieb		23.03.2011 08:44 Uhr
U4 Kettenbrückengasse	Aufzug	außer Betrieb		22.03.2011 09:09 Uhr
U2 Aspernstraße	Aufzug	außer Betrieb	Richtung Erzherzog-Karl Straße	24.02.2011 12:14 Uhr
U6 Westbahnhof	Rolltreppe	außer Betrieb	Eine	23.02.2011 16:28 Uhr

<< Vorherige Seite | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 Nächste Seite >>

**Figure 3.4:** The website *ubahnaufzug.at* offers geospatially mapped mobility data for passengers of the Wiener Linien. Several options are available to ease the use of the HTML view of the data.

The project *ubahnaufzug.at* is tailored mainly for movement impaired passengers of the Wiener Linien. It offers geospatial data concerning malfunctioning elevators and escalators. The data is available in the following formats: *JSON*, *HTML*. As of May 2011 there are 90 georeferenced datasets available as can be seen in Figure 3.4. For visualization purposes it is possible to view the dataset as an augmented reality layer on the Smartphone as well as a layer in Google maps. Information concerning the license model of the data could not be found.

## 3.3 Common Properties

This section aims to find properties shared by all previously mentioned data sources. Structured data in the form of spreadsheets or HTML files is available on all inspected websites. However, open formats are not that common yet. The three big websites dealing with statistics (Eurostat, OECD.Stat Extracts, WHO Data and Statistics) offer a major part of their datasets as XLS files only, thus relying on a proprietary format. On the other hand there are the two big governmental websites (*data.gov* and *data.gov.uk*) offering a large amount of data as RDF files. This is a necessary step towards the creation of linked datasets and there is already some effort in doing so. The same goes for the Katalog Offene Daten. A very open approach regarding accessing and visualizing geo-spatial data is pursued by the *ubahnaufzug.at* project.

There are some differences regarding licensing issues of the data made available. Some websites offer their data without any details on licensing, but the majority of the websites deals with this issue. Eurostat, OECD.Stat Extracts, WHO Data and Statistics for example offer their data for free for commercial and non-commercial use but insist on proper acknowledgment. The two governmental websites offer their data free-of-charge but do have a data policy or a license model. To get a quick overview, the set of common properties is illustrated in Table 3.1.

website	structured data available	open format	linked data
<i>data.gov</i>	yes (XLS, XML, RDF, ...)	not all datasets <sup>1</sup>	288/379932 records only
<i>data.gov.uk</i>	yes (XLS, XML, RDF, ...)	not all datasets <sup>1</sup>	some records only
<i>Eurostat</i>	yes (XLS, PC-AXIS, ...)	not all datasets	none found
<i>OECD.Stat Extracts</i>	yes (XLS, PC-AXIS, ...)	not all datasets <sup>1</sup>	none found
<i>WHO Data and Statistics</i>	yes (XLS, ...)	not all datasets <sup>1</sup>	none found
<i>offener.datenkatalog.at</i>	yes (XLS, RDF, ...)	not all datasets <sup>1</sup>	very few
<i>ubahnaufzug.at</i>	yes (HTML, ...)	yes	none found

**Table 3.1:** This table shows a set of common properties of all data sources discussed in this chapter. It shows that there is structured data available in an open format on every website. However it is often not Linked Data, since formats like RDF are not common and hardly available on, for example, websites providing statistical data.

<sup>1</sup>Some data is available only in a closed format like XLS.



## Chapter 4

# Open Data Visualizations

This chapter will present some interesting and notable examples of information visualizations which are making use of Open Data. The concepts and ideas behind Open Data have already been explained in previous chapters. One has many possibilities of doing something with data gained via an Open Data repository. An attractive and useful way is to present the data in form of a visualization. There are many options of how to visually present data as can be seen in Keith Andrews [2011]. The following examples will be divided into two parts, one of them comprising international examples of visualizations and the other consisting of Austrian examples of the same.

To each shown example there is a bullet list which presents the visualization's key characteristics. This list includes the typological categorization according to Keith Andrews [2011] (if possible), the source of the visualized data and how the visualization is enriched.

### 4.1 International Open Data Visualizations

#### 4.1.1 SFpark

In April 2011 the SFpark project launched the SFpark Project (SFpark [2011a]). This project aims to provide real-time parking availability and pricing information as an Web-Application, iPhone App or Open Data Stream (see Figure 4.1). The visualization itself - incorporating a customized Google Map - is interactive. The location can be changed dynamically. Availability and pricing can be seen immediately. Generally the project aims to reduce traffic by helping drivers to quickly find open parking spaces. All in all drivers, public transit riders, bicyclists, pedestrians and business owners are helped to enjoy a less congested and less polluted city.

For this survey it is especially interesting how the data is gathered. First there are currently 8300 sensors (cf. SFpark [2011d]) to detect when and where spaces are available. These wireless sensors are able to report availability space-by-space and minute-by-minute. The project also supplies developers with an open API (SFpark [2011b], SFpark [2011c]) to provide a free public data feed including SFpark availability and pricing information.

- url: <http://sfpark.org/>
- type: interactive geographical map
- data source: [http://sfpark.org/wp-content/uploads/2011/04/SFpark\\_API\\_v01c.pdf](http://sfpark.org/wp-content/uploads/2011/04/SFpark_API_v01c.pdf)
- enrichment: custom Google Maps markers, map highlightings, dropdown menu
- interpretability: The main information (parking availability and pricing) can be drawn out from the visualization really quick. Since there are no further possibilities to look at the given data more interpretation potential can not be expected.
- creator: SFMTA — Municipal Transportation Agency (SFMTA [2011])

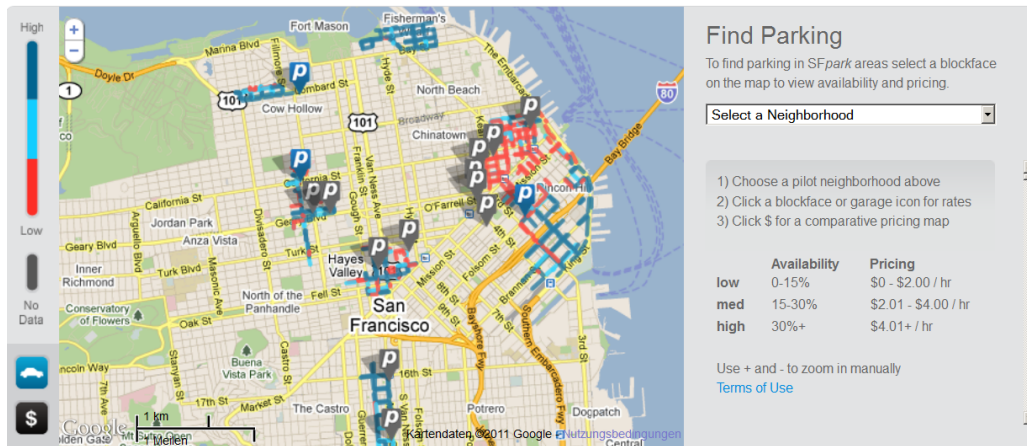


Figure 4.1: SFpark interactive visualization (screenshot taken from <http://sfpark.org/>)

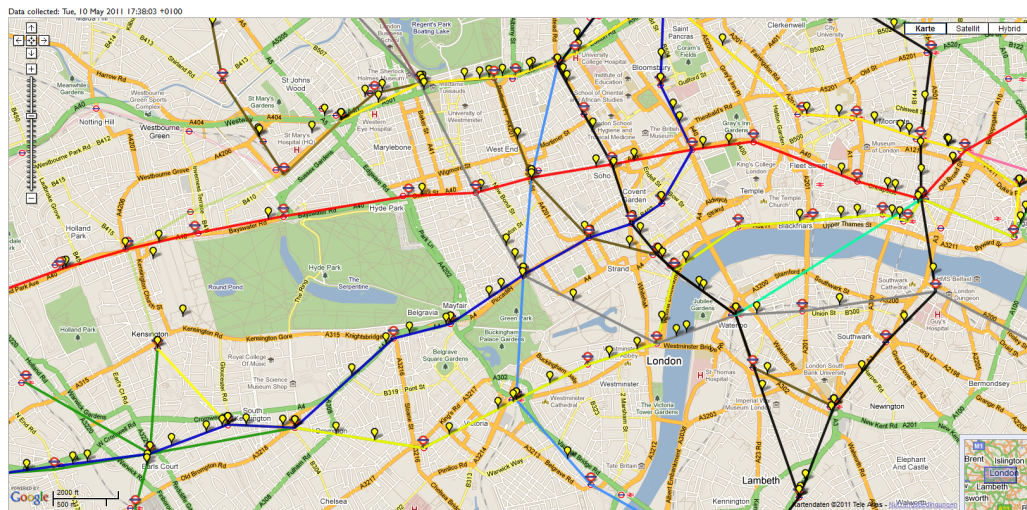


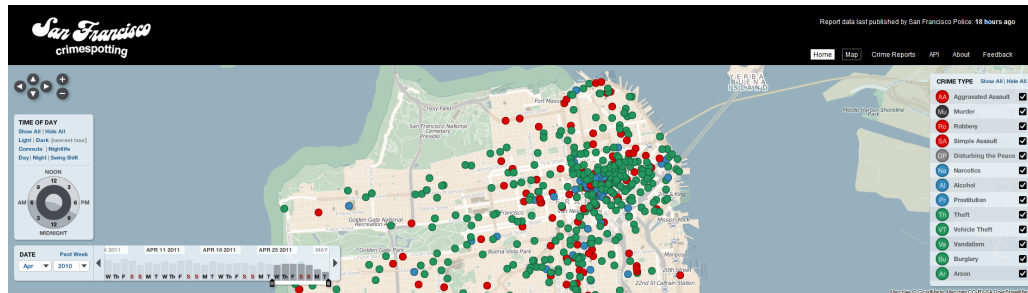
Figure 4.2: real-time London Underground train map (screenshot taken from <http://traintimes.org.uk/map/tube/>)

### 4.1.2 Live train map for the London Underground

Matthew Somerville created a real-time visualization of London's Underground trains (see Figure 4.2). He makes use of the TfL (Transport for London) API (Transport for London [2011]) to fetch the train's departure times and then applies some math to this data. Basically there is not much left to say when seeing the interactive real-time visualization (see Figure 4.2). A Google Map is used, yellow markers visualize the trains and the differently coloured lines are the train's routes.

- url: <http://traintimes.org.uk/map/tube/>
- type: interactive geographical map
- data source: <http://www.tfl.gov.uk/businessandpartners/syndication/default.aspx>
- enrichment: custom Google Maps markers representing the trains, differently coloured lines representing the routes, mouseover effect providing more information for each train
- interpretability: The important information can be pulled out of this visualization on first sight. Basically the map has a bit of an experimental character to show what is possible with the provided data feeds. Mainly it can play a role finding out anomalies in Londons underground traffic like for example congestions.





**Figure 4.3:** San Francisco Crimespotting visualization (screenshot taken from <http://sanfrancisco.crimespotting.org/>)

- creator: Matthew Somerville (Matthew Somerville [2011])

### 4.1.3 Crimespotting

Crimespotting (Crimespotting [2011]) aims at visualizing near real-time data about crimes in certain areas (see Figure 4.3). The viewer can filter and explore different types of crimes which are shown as markers on a map and define the timeframe and date of the visualized crimes. There is also a bar chart showing the distribution of crimes over days. There is also a possibility to share links directly to a particular view of the map which makes the important feature of sharing and publishing information possible. To get an idea of important ideas and goals of the application's creators there is the following quotation taken from the site (Crimespotting [2011]):

“As citizens we have a right to public information. A clear understanding of our environment is essential to an informed citizenry. [...] We believe that civic data should be exposed to the public in a more open way. With these maps, we hope to inspire local governments to use this data visualization model for the public release of many different kinds of data: tree plantings, new schools, applications for liquor licenses, and any other information that matters to people who live in neighborhoods.”

- url: <http://sanfrancisco.crimespotting.org/>
- type: interactive geographical map
- data source: <http://datasf.org/story.php?title=crime-incident-data>
- enrichment: OpenStreetMap.org customized map, custom markers, filtering of certain crime types, mouseover effects, bar chart, time frame definition, pan and zoom of the map
- interpretability: The interpretability of this map is very high. Following questions can be answered as stated on their homepage: *Is there more crime this week than last week? More this month than last? Do robberies tend to happen close to murders? We're interested in everything from complex questions of patterns and trends, to the most local of concerns on a block-by-block basis* (Crimespotting [2011]).
- creator: Stamen Design (Stamen Design [2011])

### 4.1.4 OECD Regional Statistics

At OECD (OECD [2011c]) the *Organisation for Economic Co-operation and Development* offers a highly interactive visualization of lots of different data (see Figure 4.4). Generally the data is shown on a map, as a scatter plot, table lens or as a raw data grid. The user can choose from 9 different *stories* each visualizing lots of different data again. Once the visualization is generated it still can be optimized by tweaking the legend, colour indicators and scales for example. There is a lot to explore and one can also import own data to further enrich



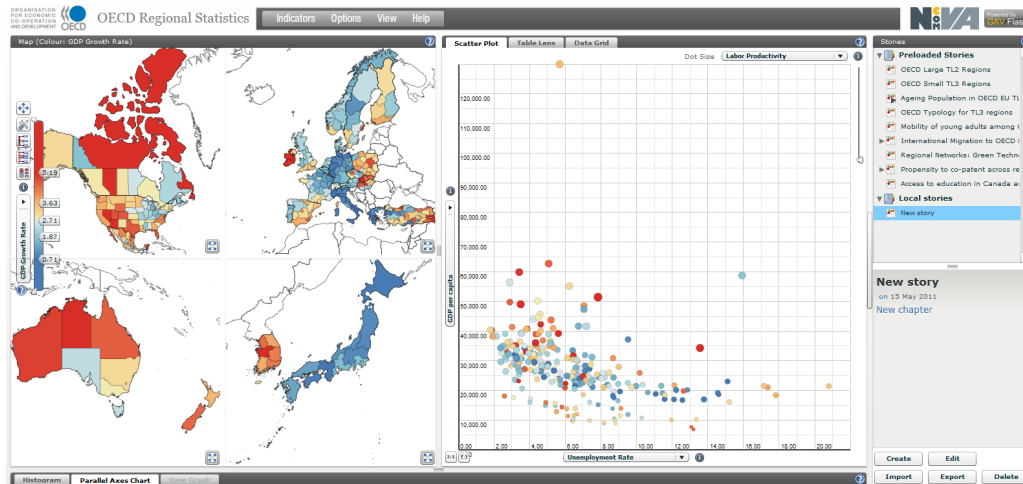


Figure 4.4: OECD Regional Statistics(screenshot from <http://stats.oecd.org/OECDregionalstatistics/>)

the pre-defined visualizations. To import your own data it has to be in a pre-defined format which is explained on the page (OECD [2011a]).

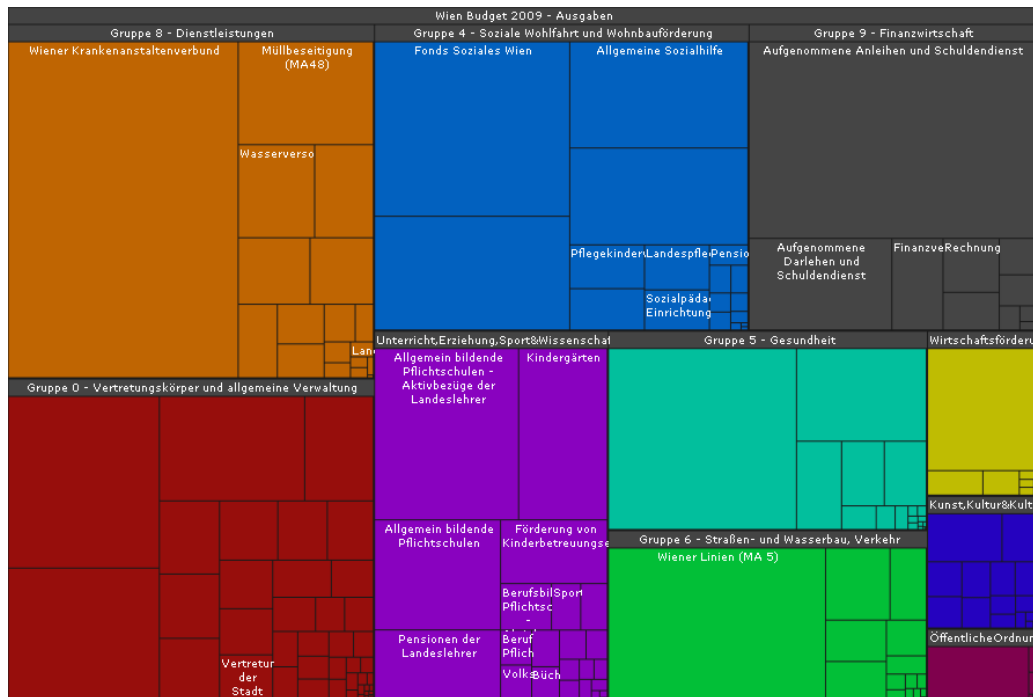
- url: <http://stats.oecd.org/OECDregionalstatistics/>
- type: interactive geographical map
- data source: <http://stats.oecd.org/Index.aspx>
- enrichment: The whole visualization basically is an eXplorer v3.5 application (Linköping University [2011b]) providing custom maps, scatter plot view, table lens view, raw data view, data import/export, and much more features.
- interpretability: The interpretability of this map is very high. Since the created maps are highly interactive a lot of different views can be generated, allowing many different conclusions to be drawn from the maps.
- creator: NCVA, Linköping University (Linköping University [2011a]) in cooperation with the OECD (OECD [2011b]).

## 4.2 Austrian Open Data Visualizations

### 4.2.1 Wien Budget 2009

At <http://www.open3.at/projekte/wien-budget-2009-visualisierung> the budget of Austria's capital Vienna has been visualized (see Figure 4.5). Ten budgetary items with subcategories up to three layers can be inspected via three different views: Squares, Stripes, Slice and Dice. By clicking on the rectangles a more detailed view gets opened. Each area also is enriched with a Mouseover effect. Moving the mouse over the visualization opens a small window supplying the viewer with more information about the current item.

- url: <http://www.open3.at/projekte/wien-budget-2009-visualisierung>
- type: Including space-filling tree browsers (tree map, market map)
- data source: <http://www.wien.gv.at/finanzen/budget/ra09/>
- enrichment: additional information with Mouseover effects



**Figure 4.5:** Vienna's budget 2009 (screenshot taken from <http://www.open3.at/projekte/wien-budget-2009-visualisierung>)

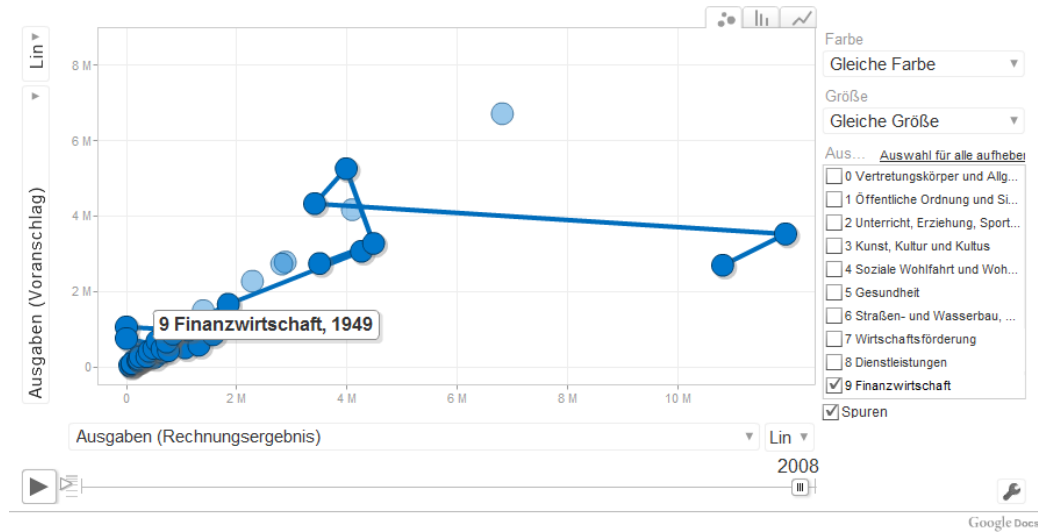
- interpretability: The size of the areas of the rectangles make it very easy to draw some trivial conclusions when looking at the visualization. For any more detailed interpretation further data should be available.
- creator: Robert Harm (Robert Harm [2011])

Further examples of Robert Harm's visualizations include the *Bundesfinanzrahmen 2011-2014 inkl. Konsolidierungsmaßnahmen als interaktive Grafik* and the *Budgetkürzungen 2011-2014 als interaktive Grafik* of Austria's government both available at <http://www.open3.at/sparpaket>. To the latter can be said that the raw data was provided as a non machine-readable sheet which had to be processed to make it machine-readable. By using the JavaScript InfoVis Toolkit (Nicolas Garcia Belmonte [2011]) then a Tree Map was generated correlating the different shortings of Austria's budget plan. During four days this visualization received 10.000 unique visitors. Therefore the creators emphasize that there is a good need for transparency and accessibility of information in Austria's public (cf. Semantic Web Company [2010]).

#### 4.2.2 Woergl Haushalt 1949-2009

At <http://www.open3.at/projekte/woergl-haushalt-1949-2009-visualisierung> Arno Abler visualized the (preliminary) budget expenses and incomes of the city Woergl (see Figure 4.6). Since this visualization makes use of the Gapminder-Tool (Gapminder [2011]) it is highly interactive. The y-axis shows the preliminary expense/income, the x-axis shows the actual expense/income. Different configurations allow the highlighting of specific categories or the illustration as a bar chart or line chart. By clicking the play-Button the animation starts.

- url: <http://www.open3.at/projekte/woergl-haushalt-1949-2009-visualisierung>
- type: Interactive scatterplot
- data source: <https://spreadsheets.google.com/ccc?key=0Agx03EpA6FWrdEM5d194TUZaUmxvTWxCLen> (Spreadsheet)



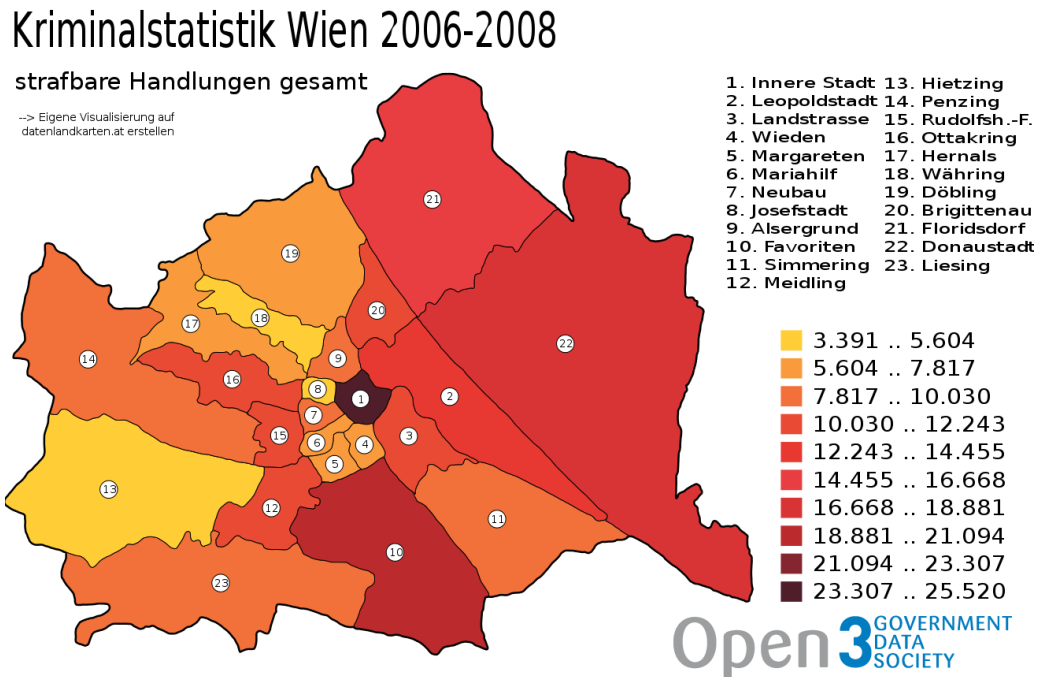
**Figure 4.6:** Budgetary expenses and income of the city Woergl from 1949 to 2009 (screenshot taken from <http://www.open3.at/projekte/woergl-haushalt-1949-2009-visualisierung>)

- enrichment: Highlighting of categories, alternate visualization as line-chart or bar-chart, variable speed of animation, adjustable colour and size of values
- interpretability: E.g. in the year 2000 a so called underspending for the category *Unterricht, Sport, Erziehung und Wissenschaft* can be seen in the graph. Furthermore in 2007 there is a huge discrepancy between preliminary and actual expenses in the *Finanzwirtschaft* sector. But since there are no further details on the numbers a more precise interpretation of these facts is not possible.
- creator: Arno Abler (Arno Abler [2011])

### 4.2.3 DataMaps.eu

The site <http://www.datamaps.eu> provides a framework to create visualizations without programming effort. They also provide raw data (datamaps.eu [2011]) which can be used to create visualizations. The page which provides the raw data lists various links. These links can be clicked to dynamically generate new visualizations via their API. As a result a graphic of Vienna's crime-rate from 2006 - 2008 has been generated, see Figure 4.7.

- url: <http://www.datamaps.eu/wp-content/gallery/kriminalitaet/kriminalitaetsstatistik2006-2008-gesamt.gif>
- type: geographical map
- data source: <http://www.wien.gv.at/statistik/daten/rtf/kriminalstatistik-jahre-bezirke.rtf> (RTF)
- enrichment: -
- interpretability: Interpretability in terms of where the crime rate is the highest is rather good (e.g. in the center of the city there is the highest crime rate). Further interpretation or conclusions are not possible.
- creator: <http://www.open3.at/>



**Figure 4.7:** crime rate in Vienna from 2006 to 2008 (screenshot taken from <http://www.datamaps.eu/wp-content/gallery/kriminalitaet/kriminalitaetsstatistik2006-2008-gesamt.gif>)

## 4.3 Common properties

To summarize the key-characteristics of the presented visualizations some common or different properties are pointed out here. The types of the visualizations are quite diverse, ranging from basic (interactive) geographical maps to interactive scatterplots and tree browsers. Some of them have certain enrichment-features in common, mostly specific mouseover-effects or alternative visualizations as bar-charts or line-charts. The source of the raw data also varies. Data is available as RDF, Spreadsheet, CSV, API (XML), etc. All in all can be said that common properties are not so easy to find since the possibilities of the availability of the data and what a visualizer can do with it is extremely high.



## Chapter 5

# Visualization Tools and Libraries

In this chapter, we discuss a couple of the most popular visualization tools and development libraries. Considering the limits in length of this report we cannot list all tools in details and we concentrate only on their most attractive features. The following tools/libraries will be observed in this section:

- Data Wrangler (Stanford University’s Visualization Group)
- Many Eyes (IBM)
- Exhibit (MIT)
- Google Chart Tools (Google)
- JavaScript InfoVis Toolkit
- Protovis (Stanford University’s Visualization Group)

In order to give readers a clear and a comparable view, we structure the discussion about each tool into 3 parts:

1. What it does? - The background, original design idea and purpose.
2. What’s cool? - The highlights, most important features, the characteristic and why user/developer should use it?
3. Is it perfect? - Drawbacks. Each tool/library has its limitations and limited feature scope. In other words, users and developers should carefully choose the most appropriate tool before they start to use it.

After understanding these checkpoints we think that the user who has to use the tools or the developer who wants to use the API to develop his own customized data visualization applications might be able to find out the most suitable tools/libraries.

### 5.1 Data Wrangler

**What it does?** The first important thing for visualizing data is collating the large amount of data from an unbending state to a designated form, so that the visualization tools can understand and transform them. Though the data analysis tools improve continuously, data analysts still spend a lot of time and effort on manipulating and assessing data quality issues, nevertheless, the results from different analysts can be very different and hardly understandable for each other. In order to save analysis time and to make the results more standard and exchangeable between different analysts and visualization tools, Sean Kandel (Kandel et al [2011]) from Stanford University’s Visualization Group have developed the Data Wrangler tool (Group [2011]), which

“allows interactive transformation of messy, real-world data into the data tables.(Group [2011])”

The output data can be used in Excel, Protovis and many other tools directly.

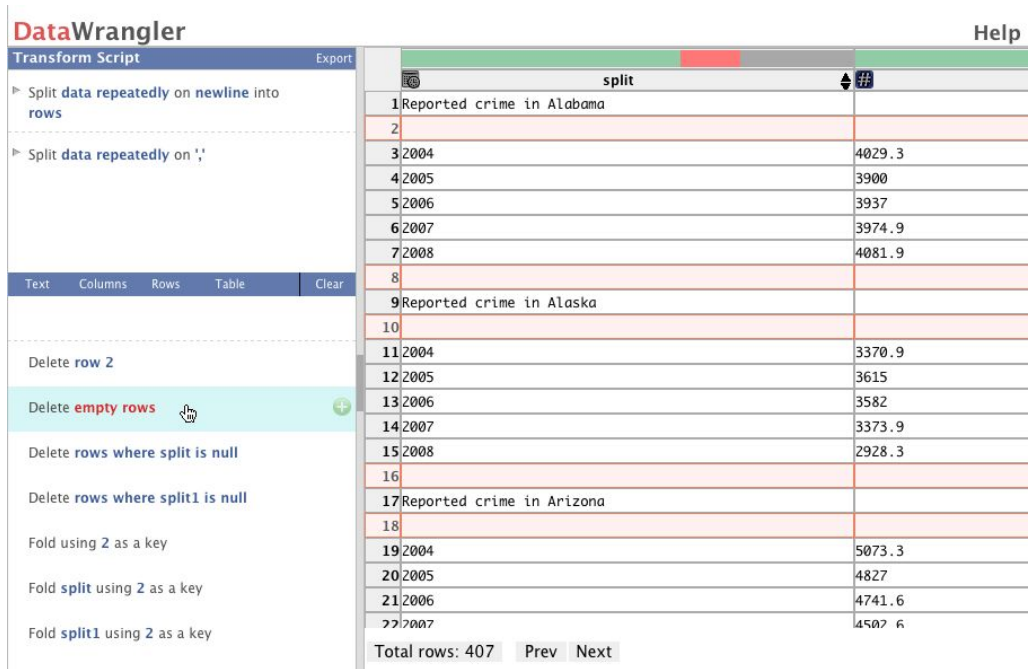


Figure 5.1: DataWrangler helps to format data into table format

**What’s cool?** Editing the text is an absolutely simple task with Data Wrangler. For example (see Figure 5.1), the empty rows are automatically detected and the hints for next operation are listed for user on the left side.

**Is it perfect?** No, there are actually couple of significant drawbacks for DataWrangler:

- Some suggestions are not useful, they could confuse the user.
- DataWrangler is a Web-based service, so it is convenient to use but it could also lead to sensitive information leaks to external sites.
- DataWrangler is still under construction, and therefore partly instable.

## 5.2 Many Eyes

**What it does?** Many Eyes project is a website that provides services to the end user to visualize data sets. It is a project of International Business Machines Corporation (IBM). Many visualization templates have been embedded into this project’s website as standard service. The purpose of Many Eyes project is mostly like this: Helping people to enrich or enhance presentations, interactive websites and sharing experiences, knowledge in an intuitive way. The Many Eyes website provides following functionalities:

- View and discuss existing visualizations and datasets. Each visualization of the Many Eyes website is attached with a forum. The viewers can comment on the visualization when they find that the view is interesting and they want to talk about the view share thoughts. Everybody can read the comments. Not only the visualizations, the viewer could also get thousands original public data sets that users uploaded to the site. Apart from the visualization, viewers can also make comments on the data set.
- Create visualizations from existing data sets. The website holds already more than 90,000 visualizations nowadays, and this number increases everyday rapidly. Viewers can choose any one from these ninety thousand visualizations and regenerate a new view for it in a more proper, reasonable and meaningful manner.

For the registered user, Many Eyes provides some couple of more attractive functionalities, such as:

- Rate the existing visualizations and data sets. Because the huge amount of the site’s content, it is very boring for the user to explore all the data sets/visualizations one by one, however, rating the visualization or data set can help people to find the best content on Many Eyes.
- Upload own data set to Many Eyes. After registration, the user can upload own data sets to the Many Eyes website, make a visualization himself and share Open Data with whole world.

### What’s cool?

- The Many Eyes project’s site has a large amount of visualizations and data sets for re-use and re-generation of visualizations, it provides already thousands of materials for further research.
- The Many Eyes site has prepared plenty of visualization formats of different catalogs (IBMResearch [2011]). Users can choose them for specific purposes, such as pie chart, line graph, network diagram and tag cloud for analyzing a text.
- The Many Eyes site is very easy to use, users are not forced to gain any technical knowledge, they can concentrate on their data set format and then get their proper visualization. Many Eyes accepts for the time being only two formats of data sets, free text for analyzing word text and a simple table form for all other visualization formats.

**Is it perfect?** No, Many Eyes is very limited on the data set format, only tables like the example in Table 5.1 can be interpreted by Many Eyes. Like Data Wrangler in last section, the Many Eyes project is also a

Fruit	Color	Calories
Apple	Red	100
Orange	Orange	110

**Table 5.1:** Simple table example

public Web-based service, so the sensitive internal information cannot be protected. So, the users who uses this service should be aware before they get start uploading the data sets.

Another limitation of Many Eyes is, although the website provides a couple of visualization formats, that customization or personalization of the visualization is impossible. So if the user has high demands on his visualization, then he has to find a specific tool or develop with another API.

## 5.3 Exhibit

**What it does?** The MIT Simile Project (open-source community [2011]) is designed to help users creating the interactive web site with advanced functionalities, such as context searching, special filtering, timeline or map visualization. Billed as a publishing framework, the JavaScript library allows easy additions of filters, searches and more other user defined criteria. An intuitive example look like following (see Figure 5.2)

Actually, the Exhibit is a relative “easy-to-use” library that we can find online for making an interactive web-based application, easy means for the professionals who has developed the Exhibit at MIT but definately not for such user whose pre-knowledge stops at Excel. Like most JavaScript libraries, Exhibit also requires more hand-coding than web services such as Many Eyes, but from the flexibility and personalization point of view, they are not comparable. On the other hand, Exhibit has clear documentation for beginners, even those who without JavaScript experience, can learn how to use it.



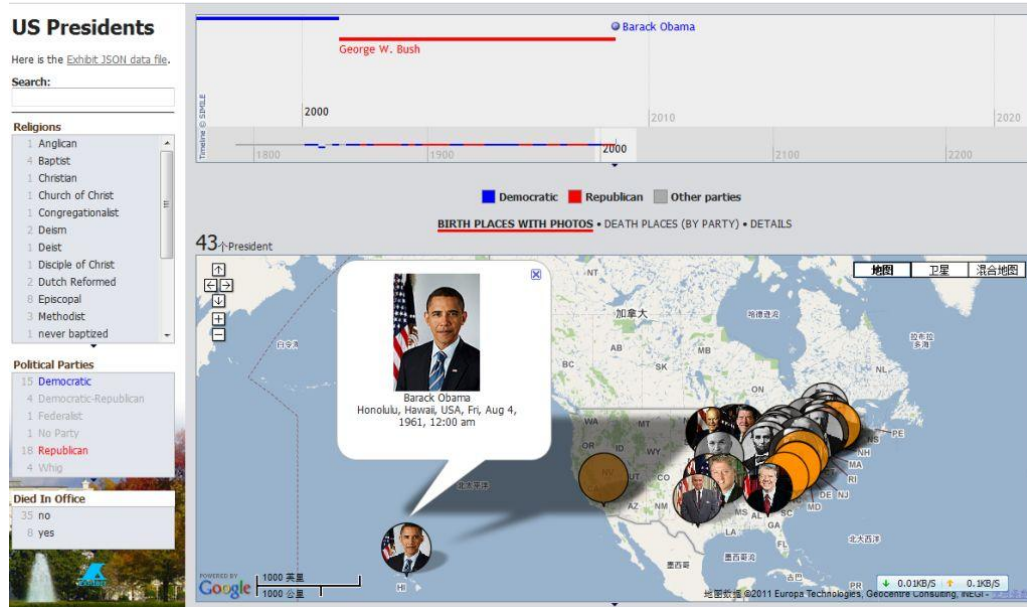


Figure 5.2: A web based interactive application example using Exhibit - U.S. Presidents

**What’s cool?**

”For those who are comfortable coding, Exhibit offers a number of views – maps, charts, timeplots, calendars and more – as well as customized lenses (ways to format an individual record) and facets (properties that can be searched or sorted). Users are much more likely to get the exact presentation they want with Exhibit than, say, Many Eyes. And their data stays local unless and until they decide to publish. (Machlis [2011])”

**Is it perfect?** The Exhibit is now a project of Google. The beginner who wants to use Exhibit must get familiar with coding of Exhibit and library syntax. It costs time, but thanks to plenty of examples and documentation, one can get start to work quickly.

**5.4 Google Chart Tools**

**What it does?** Google Chart Tools (GCT, Google [2011a]) has been developed to help users to visualize data on their own website or within Google Docs (see Figure 5.3).

There are two kinds of approaches provided by Google to create “Charts”, i) a Chart API using a “simple URL request to a Google chart server” for creating a static image. ii) A Visualization API that accesses a JavaScript library for creating interactive graphics. Google offers a comparison of data size, page load, skills needed and other factors to help user decide which option to use (Google [2011c])

For the simpler static graphics, there’s a wizard to help you create a chart from some sample formats; it goes as far as helping you input data row by row, although for any decent-size data sets – say, more than half a dozen or so entries – it makes more sense to format it in a text file.

The visualization API includes various types of charts, maps, tables and other options.

**What’s cool?** Highly customizable, with plenty of documentation. Some functions are *What You See Is What You Get* (WYSIWYG), this gives the beginners opportunity to get closer with GCT quickly and conveniently. For example, the static image chart is reasonably easy to use and features a *Live Chart Playground*, which allows user to tweak code and see the results in real time (Google [2011b]).

Start from example chart:



**Figure 5.3:** GCT offers both a wizard and API for creating Web graphics.

**Is it perfect?** To some extent, yes, but the static charts tool requires a bit more work than some of the other Web-based services, and it doesn't always offer lots of extras in return. And for the API, as with other JavaScript libraries, coding is required, making this more of a programming tool than an end-user business intelligence application. However, because of plenty examples, documentation and compatibility with all common Web browsers, it is still worth to recommend to the users and developers from beginner to expert.

## 5.5 JavaScript InfoVis Toolkit

**What it does?** InfoVis is probably not among the best known JavaScript visualization libraries, but it's definitely worth a look if the user is interested in publishing interactive data visualizations on the Web (Belmonte [2011]). Evidence - the White House is now using InfoVis for creating the Obama administration's Interactive Budget graphic (see Figure 5.4).

What sets this tool apart from many others is the highly polished graphics it creates from just basic code samples. InfoVis creator Nicolas Garcia Belmonte, senior software architect at Sencha Inc., clearly cares as much about aesthetic design as he does about the code, and it shows.

**What's cool?** The samples are gorgeous and there's no extra coding involved to get nifty fly-in effects. User can choose to download code for only the visualization types he wants to use to minimize the weight of Web pages.

**Is it perfect?** Not jet, since this is not an application but a code library, user must have coding expertise in order to use it. Therefore, this might not be a good fit for users in an organization who analyze data but don't know how to program. Also, the choice of visualization types is somewhat limited. Moreover, the data should be in JSON format.

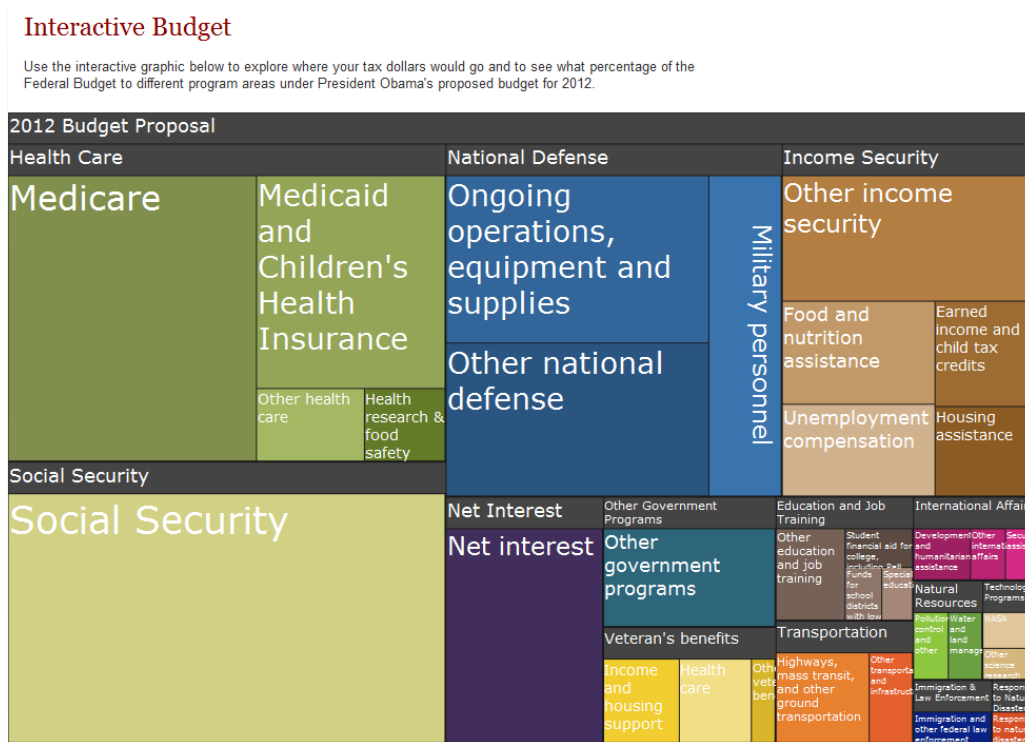


Figure 5.4: Screen shot of interactive budget graphic from White House.

## 5.6 Protovis

**What it does?** Billed as a *graphical toolkit for visualization*, the Protovis project from Stanford University’s Visualization Group (see Bostock and Heer [2011]) is one of the most popular JavaScript libraries for transforming data into visualizations. Protovis

”is designed to balance simplicity with control over the display.(Bostock and Heer [2011])”

and could be summarized in following points:

- Protovis draws custom views of data with primitive marks like bars and dots. Different with other low-level graphics libraries that quickly become tedious for visualization, Protovis defines those marks through dynamic properties that encode data, allowing inheritance, scales and layouts to simplify construction.
- It uses JavaScript and SVG for web-native visualizations.
- No plugin required (though user will need a modern web browser)!
- It is declarative and designed to be learned by example.
- It is free and open-source, provided under the BSD License.

**What’s cool?** One of the best things about Protovis is how well it’s documented, with plenty of examples featuring visualization and sample code. There are also a large number of sample visualization types available, including maps and some statistical analyses. It’s a robust tool, capable of building graphics.

**Is it perfect?** To some extent, yes. As is the case with other JavaScript libraries, it's very necessary for users to have knowledge of JavaScript. Or at least, one should know another programming language, because with helping of the documentation and samples, it's possible to copy, paste and modify code without really understanding what it's doing. Protovis is mostly declarative and designed to be learned by example. However, it is difficult to recommend that approach for nontechnical end users.



## Chapter 6

# Concluding Remarks

In this report we have presented a comprehensive summary of Open Data formats, sources and visualization examples. We have identified Open Data in various forms and for various purposes, for example scientific data, medical data, governmental data and environmental data. A fair amount of activity can be observed in the field of Open Data sources, especially concerning governmental and statistical data. Activity around Open Government Data and networks to support Open Data projects (*open3.at*) can be observed in Austria as well. A number of international as well as national websites and projects are discussed in this context and interesting results are obtained.

Structured datasets in the form of spreadsheets or HTML files are available on all inspected websites. However, open formats are not that common yet. Eurostat, OECD Statistics and WHO Data and Statistics represent websites with statistical content. They offer a major part of their datasets only as XLS files, which is a proprietary format. On the other hand *data.gov* and *data.gov.uk* offer a large amount of data as RDF files. This is necessary to create linked datasets which has already started in these projects - the same holds for the Katalog Offene Daten. Noticable are differences regarding licensing issues. Some websites offer data without details on licensing, but the majority deals with this issue. Eurostat, OECD Statistics and WHO Data and Statistics offer their data for free for commercial and non-commercial use but insist on proper acknowledgment. *Data.gov* and *data.gov.uk* offer data free-of-charge but do have a data policy or a license model.

Regarding the shown examples in section 4 can be said that there is a plethora of different examples and usages of Open Data available. It is also proven that there is already a steadily growing Open Data movement present in Austria. Just recently Vienna opened up for Open Data. It is also obvious that there is enough Open Data released. Now it is important to find useful and innovative ways to present and visualize the data. A certain degree of interactivity can be expected to be better received by the viewer than a static picture. To feature real-time visualizations, if the data service makes this possible, also can be an interesting field for Open Data usage, e.g. to inform the public about recent events regarding public security. All in all Open Data basically enables the creation of visualizations for a huge target group and for a very diverse field of applications. The difficulty is to narrow down the diversity of possibilities as good as possible to finally create an Open Data visualization in an effective and useful way.

According to the chapter 5, we know that there are plenty of tools for visualizing the datasets and meanwhile, dozens of open source libraries can be used to develop the personalized or customized visualization applications, not only as stand-alone software, but also as web-based application. Almost every tool/library has its unique features and we cannot simply decide which one is better, it's strictly dependent on the requirement of the use case. If the request is just simple enrichment of a documentation or enhance the web site, then Many-Eyes is a good choice, but if the request is a commercial level application, then one must study the different libraries very carefully and choose the most suitable one, e.g. if a better look-and-feel is required, then the JavaScript InfoVis Toolkit is the most reasonable choice. For developing an interactive application without knowledge of the JavaScript language, Exhibit is perhaps the best solution.

# References

- Arno Abler [2011]. *Arno Abler Twitter*. [https://twitter.com/arno\\_abler](https://twitter.com/arno_abler).
- Belmonte, Nicolas Garcia [2011]. *The JavaScript InfoVis Toolkit provides tools for creating Interactive Data Visualizations for the Web*. <http://thejit.org/>.
- Berners-Lee, Tim [2006]. *Linked Data - Design Issues*. <http://www.w3.org/DesignIssues/LinkedData.html>.
- Berners-Lee, Tim, Yuhsin Chen, Lydia Chilton, Dan Connolly, Ruth Dhanaraj, James Hollenbach, Adam Lerer, and David Sheets []. *Tabulator: Exploring and Analyzing linked data on the Semantic Web. Proceedings of the 3rd International Semantic Web User Interaction Workshop, 2006(i)*, page 6. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.97.950&rep=rep1&type=pdf>.
- Berrueta, D and J Phipps [2008]. *Best Practice Recipes for Publishing RDF Vocabularies. W3C Working Group Note*, (August). <http://www.w3.org/TR/swbp-vocab-pub/>.
- Bizer, Christian, Tom Heath, and Tim Berners-Lee [2009]. *Linked Data - The Story So Far. International Journal on Semantic Web and Information Systems*, 5(3), pages 1–22. <http://www.citeulike.org/user/omunoz/article/5008761>.
- Bostock, Mike and Jeff Heer [2011]. *Protovis, a graphical approach to visualization*. <http://vis.stanford.edu/protovis/>.
- CKAN [2011]. *CKAN - the Data Hub Software*. <http://ckan.org/>.
- Crimespotting [2011]. *San Francisco Crimespotting*. <http://sanfrancisco.crimespotting.org/>.
- Dan Brickley and Libby Miller [2011]. *The Friend of a Friend (FOAF) project*. <http://www.foaf-project.org/>.
- datamaps.eu [2011]. *Rohdatenverzeichnis*. <http://www.datamaps.eu/tool/list.php>.
- DERI Galway [2011a]. *DATA.gov Catalog*. <http://lab.linkeddata.deri.ie/data.gov/>.
- DERI Galway [2011b]. *DERI International*. <http://www.deri.org/>.
- DERI Galway [2011c]. *Linked Open Data star scheme by example*. <http://lab.linkeddata.deri.ie/2010/star-scheme-by-example/>.
- Eurostat [2011]. *Eurostat Home*. <http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home/>.
- Gapminder [2011]. *Gapminder: Unveiling the beauty of statistics for a fact based world view. - Gapminder.org*. <http://www.gapminder.org/>.
- Google [2011a]. *Display live data on your site*. <http://code.google.com/intl/zh-CN/apis/chart/>.

- Google [2011b]. *Live Chart Playground*. [http://code.google.com/intl/zh-CN/apis/chart/image/docs/chart\\_playground.html](http://code.google.com/intl/zh-CN/apis/chart/image/docs/chart_playground.html).
- Google [2011c]. *Which Chart Tool is Best for Me?* <http://code.google.com/intl/zh-CN/apis/chart/interactive/choosing.html>.
- Group, Stanford Visualization [2011]. *Wrangler, an interactive tool for data cleaning and transformation*. <http://vis.stanford.edu/wrangler/>.
- Heath, Tom and Christian Bizer [2011]. *Linked Data: Evolving the Web into a Global Data Space*, volume 1. Morgan & Claypool, 136 pages. <http://www.morganclaypool.com/doi/abs/10.2200/S00334ED1V01Y201102WBE001>.
- IBMResearch [2011]. *Visualization Options Available in Many Eyes*. [http://www-958.ibm.com/software/data/cognos/manyeyes/page/Visualization\\_Options.html](http://www-958.ibm.com/software/data/cognos/manyeyes/page/Visualization_Options.html).
- Kandel, Sean, Andreas Paepcke, Joseph Hellerstein, and Jeffrey Heer [2011]. *Wrangler: Interactive Visual Specification of Data Transformation Scripts*. In *ACM Human Factors in Computing Systems (CHI)*. <http://vis.stanford.edu/papers/wrangler>.
- Keith Andrews [2011]. *Information Visualisation - course notes*. <http://courses.iicm.tugraz.at/ivis/ivis.pdf>.
- Kobilarov, Georgi and Ian Dickinson [2008]. *Humboldt : Exploring Linked Data*. *Human Factors*, (April).
- Latif, Atif, Muhammad Tanvir Afzal, Anwar Us Saeed, Patrick Hoefler, and Klaus Tochtermann [2009]. *CAF-SIAL: Concept aggregation framework for structuring informational aspects of linked open data*. *2009 First International Conference on Networked Digital Technologies*, pages 100–105. <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5272079>.
- Lee Feigenbaum [2007a]. *Using RDF on the Web: A Survey*. [http://www.thefigtrees.net/lee/blog/2007/01/using\\_rdf\\_on\\_the\\_web\\_a\\_survey.html](http://www.thefigtrees.net/lee/blog/2007/01/using_rdf_on_the_web_a_survey.html).
- Lee Feigenbaum [2007b]. *Using RDF on the Web: A Vision*. [http://www.thefigtrees.net/lee/blog/2007/01/using\\_rdf\\_on\\_the\\_web\\_a\\_vision.html](http://www.thefigtrees.net/lee/blog/2007/01/using_rdf_on_the_web_a_vision.html).
- Linköping University [2011a]. *National Center vor Visual Analytics*. <http://ncva.itn.liu.se/?l=en>.
- Linköping University [2011b]. *Open Statistics eXplorer*. <http://ncva.itn.liu.se/explorer/openexp?l=en>.
- Machlis, Sharon [2011]. *22 free tools for data visualization and analysis*. [http://www.computerworld.com/s/article/9215504/22\\_free\\_tools\\_for\\_data\\_visualization\\_and\\_analysis?taxonomyId=18&pageNumber=6](http://www.computerworld.com/s/article/9215504/22_free_tools_for_data_visualization_and_analysis?taxonomyId=18&pageNumber=6).
- Matthew Somerville [2011]. *Matthew Somerville Twitter*. <http://twitter.com/dracos>.
- Nathan Hawks [2011]. *Opening Linked Data*. <http://webr3.org/blog/linked-data/opening-linked-data/>.
- Neofonie GmbH [2011]. *Faceted Wikipedia Search*. <http://dbpedia.neofonie.de/browse/>.
- Nicolas Garcia Belmonte [2011]. *JavaScript InfoVis Toolkit*. <http://thejit.org/>.
- OECD [2011a]. *Loading external data OECD Regional eXplorer*. [http://stats.oecd.org/OECDregionalstatistics/load\\_local.pdf](http://stats.oecd.org/OECDregionalstatistics/load_local.pdf).



- OECD [2011b]. *OECD*. <http://www.oecd.org/>.
- OECD [2011c]. *OECD regional statistics*. <http://stats.oecd.org/OECDregionalstatistics>.
- OECD [2011d]. *OECD Statistics*. <http://stats.oecd.org/>.
- Open Knowledge Foundation [2011]. *Katalog Offene Daten /"Osterreich*. <http://offener.datenkatalog.at/>.
- open-source community [2011]. *Exhibit creates dynamic exhibits of their collections*. <http://simile-widgets.org/wiki/Exhibit>.
- open3 network [2011]. *open3.at - Netzwerk zur F"orderung von OpenSociety, OpenGov und OpenData in /"Osterreich*. <http://www.open3.at/>.
- Richard Cyganiak and Anja Jentsch [2010]. *Linking Open Data cloud diagram*. <http://lod-cloud.net/>.
- Robert Harm [2011]. *Robert Harm Homepage*. <http://www.harm.co.at/>.
- Semantic Web Company [2010]. *Open Government Digest - November 2010*. <http://issuu.com/semwebcomp/docs/ogdd1/1>.
- Semantic Web Company GmbH [2011a]. *Open Government Data publizieren - ein Leitfaden*. <http://www.semantic-web.at/1.36.resource.315.open-government-data-publizieren-ein-leitfaden.htm>.
- Semantic Web Company GmbH [2011b]. *Semantic Web Company*. <http://www.semantic-web.at/>.
- SFMTA [2011]. *SFMTA Homepage*. <http://www.sfmta.com>.
- SFpark [2011a]. *Mayor Lee launches SFpark project*. <http://sfpark.org/2011/04/21/mayor-lee-launches-sfpark-project/>.
- SFpark [2011b]. *SFpark availability service API reference*. <http://sfpark.org/resources/sfpark-availability-service-api-reference/>.
- SFpark [2011c]. *SFpark availability service API reference*. <http://datasf.org/story.php?title=sfpark-real-time-parking-availability-and-pricing-api>.
- SFpark [2011d]. *SFpark begins sensor installation in pilot areas*. <http://sfpark.org/2010/03/05/news-story-3/>.
- Stamen Design [2011]. *Stamen Design Homepage*. <http://stamen.com/>.
- TED Conferences, LLC [2011a]. *Tim Berners-Lee on the next Web*. [http://www.ted.com/talks/tim\\_berniers\\_lee\\_on\\_the\\_next\\_web.html](http://www.ted.com/talks/tim_berniers_lee_on_the_next_web.html).
- TED Conferences, LLC [2011b]. *Tim Berners-Lee: The year open data went worldwide*. [http://www.ted.com/talks/tim\\_berniers\\_lee\\_the\\_year\\_open\\_data\\_went\\_worldwide.html](http://www.ted.com/talks/tim_berniers_lee_the_year_open_data_went_worldwide.html).
- Ted Thibodeau Jr [2011]. *DBpedia*. <http://dbpedia.org/>.
- Tom Heath [2011]. *Linked Data - Connect Distributed Data across the Web*. <http://linkeddata.org/>.
- Transport for London [2011]. *Guidelines and Support*. <http://www.tfl.gov.uk/businessandpartners/syndication/16493.aspx>.
- ubahnaufzug.at [2011]. *ubahnaufzug.at - 24h Mobilit/"at durch open data*. <http://www.ubahnaufzug.at/>.

- United Kingdom Public Sector Transparency Board [2011]. *data.gov.uk - Opening up government*. <http://data.gov.uk/>.
- United States Government [2011]. *Data.gov*. <http://www.data.gov/>.
- Verein Open Knowledge Forum Österreich [2011]. *Open Government Data - Austria*. <http://gov.opendata.at/>.
- W3C [2011a]. *Dereferencing a URI to RDF*. <http://www.w3.org/wiki/DereferenceURI>.
- W3C [2011b]. *Linking Open Data*. <http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>.
- W3C [2011c]. *RDB2RDF Working Group*. <http://www.w3.org/2001/sw/rdb2rdf/>.
- W3C [2011d]. *Resource Description Framework*. <http://www.w3.org/RDF/>.
- W3C [2011e]. *SPARQL Query Language for RDF*. <http://www.w3.org/TR/rdf-sparql-query/>.
- WHO [2011]. *WHO Data and Statistics*. <http://www.who.int/research/en/>.
- Wikipedia [2011]. *Open science data*. [http://en.wikipedia.org/wiki/Open\\_Data](http://en.wikipedia.org/wiki/Open_Data).