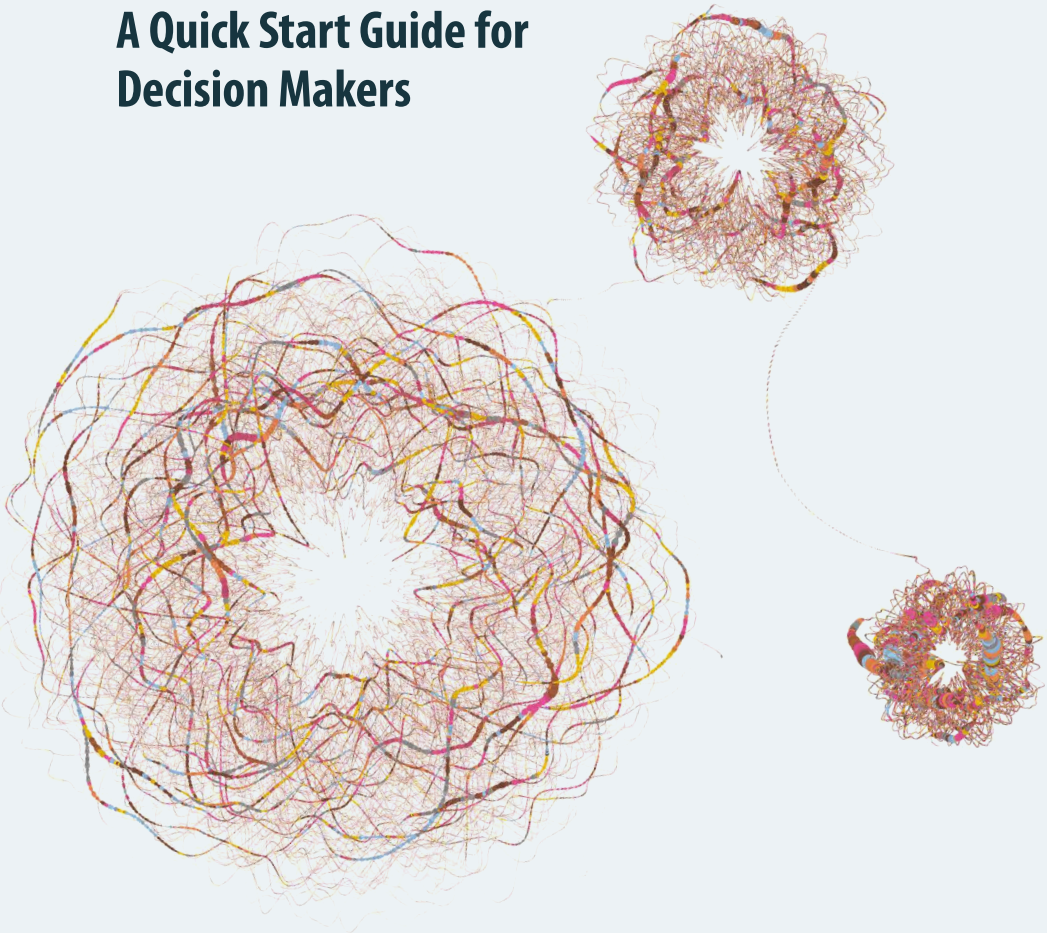
Florian Bauer, Martin Kaltenböck

Linked Open Data: The Essentials

A Quick Start Guide for Decision Makers



Linked Open Data: The Essentials

A Quick Start Guide for Decision Makers

by Florian Bauer (REEEP) and Martin Kaltenböck (Semantic Web Company)

Imprint

Published by:

edition mono/monochrom, Vienna, Austria
ISBN: 978-3-902796-05-9

Print:

DGS – Druck- u. Graphikservice GmbH

Design & Layout:

Susan Härtig (Semantic Web Company)

Production Editor:

Thomas Thurner (Semantic Web Company)

Proofreading:

Jena Wuu & Vince Reardon

PDF Version:

A PDF version of Linked Open Data: The Essentials for download is available here: http://www.semantic-web.at/LOD-TheEssentials.pdf

Introductory Remarks

by Martin Schöpe (BMU), Martin Hiller (REEEP) and Martin Kaltenböck (SWC)

RDir Dr. Martin Schöpe

Head of Division KI II 3, Federal Ministry for the Environment, Nature Conservation and Nuclear Safety, Germany

The efficient management of data and information is critical for the global economy. In particular, considering the explosive growth in the renewable energy and energy efficiency sectors, data and knowledge sharing can ensure better decision-making and disaster management, ensure more powerful project development and promote effective financing mechanisms for sustainable energy.

As one of the initiators and principal supporters of the reegle.info clean energy information portal, the German Federal Ministry for the Environment, Nature Conservation and Nuclear Safety (BMU) has supported the global exchange of clean energy information since reegle's inception in 2006.

Today, Open Government Data (OGD) is emerging as a major movement in knowledge sharing. The basic premise is to open up publicly-owned data and information from governmental institutions and to make it available in machine-readable formats for easy re-use and cross-combination by citizens, industry, media, and academia – as well as by government itself. The OGD movement has the power to fuel greater transparency, to enable collaboration between stakeholders, and last but not least to spur new economic activity.

The actual technology that drives and enables OGD is known as Linked Open Data (LOD). To accelerate knowledge sharing in the field of clean energy, the BMU is sponsoring "Linking Open Data to Accelerate Low-Carbon Development," a workshop for decision makers in clean energy organisations that will be held at the Masdar Institute in January 2012, and organised by the Renewable Energy and Energy Efficiency Partnership (REEEP).

To accompany this workshop, you have in your hands a useful publication, *Linked Open Data: The Essentials*, which provides a succinct introduction to the topic for decision makers and project developers. I hope you will find in it the inspiration for developing your own data and information management strategy.

Martin Hiller

Director General, Renewable Energy and Energy Efficiency Partnership (REEEP)

Today, the internet makes the wealth of human knowledge available to anyone, anywhere. From a clean energy perspective, this makes the internet one of the most potent capacity building tools possible. The challenge is how to sort through and effectively use the ever-increasing volume of information available.

Linked Open Data (LOD) is a growing movement for organisations to make their existing data available in a machine-readable format. This enables users to create and combine data sets and to make their own interpretations of data available in digestible formats and applications.

With the aim of accelerating this movement in the clean energy arena, the Renewable Energy and Energy Efficiency Partnership (REEEP) is sponsoring a workshop entitled „Linking Open Data to Accelerate Low-

Carbon Development" at the Masdar Institute in January 2012, with the kind support of the German Federal Ministry for the Environment, Nature Conservation and Nuclear Safety (BMU). This workshop seeks to assist key energy and development organisations in taking necessary steps to open up their data sets and to maximise the interconnection between knowledge brokers and providers.

Linked Open Data: The Essentials was developed to accompany this workshop and to give decision makers a quick overview of the LOD concept and how to accelerate the process in their respective organisations. We trust you will find the publication to be useful reading.

Martin Kaltenböck

Managing Partner & CFO, Semantic Web Company GmbH, Austria

Data management has become a crucial factor for business success and innovation. Efficient handling of Linked (Open) Data and metadata in the fields of public administration and industry is key. With a combination of social software methods and technologies, organisations can benefit and reach competitive advantage.

The publication *Linked Open Data: The Essentials* gives decision makers a good overview of Open Government, Open Government Data, Open Data and Linked Open Data (LOD). It highlights the potentials and benefits of LOD, providing a quick guide with the most important steps for LOD publishing, a consumption strategy for your organisation and four best practice examples of LOD in use.

Table of Contents

1. FROM OPEN DATA TO LINKED OPEN DATA

A brief history and factbook of Open Government, Open (Government) Data & Linked Open Data

This introductory chapter will describe the principles of linking data; define important terms such as Open Government, Open (Government) Data and Linked Open (Government) Data; and explain relevant mechanisms to ensure a solid foundation before going more in-depth. Each subsequent chapter explains a specific topic and suggests additional resources, such as books and websites, for gaining more detailed insights on a particular theme. We hope that by introducing you to the possibilities of Linked Open Data (LOD), you will be able to share our vision of the future semantic web.

Open Government & Open (Government) Data

When we talk about Open Government today, we refer to a movement that was initiated by "The Memorandum on Transparency and Open Government"[1] (The Transparency Directive), a memorandum signed by US President Barack Obama shortly after his inauguration in January 2009. The basic idea of Open Government is to establish a modern cooperation among politicians, public administration, industry and private citizens by enabling more transparency, democracy, participation and collaboration. In European countries, Open Government is often viewed as a natural companion to e-government[2].

An important excerpt of the memorandum reads: "My Administration is committed to creating an unprecedented level of openness in Government. We will work together to ensure the public trust and establish a system of transparency, public participation, and collaboration. Openness will strengthen our democracy and promote efficiency and effectiveness in Government."

The Open Government partnership[3] was launched on September 20, 2011, when the eight founding governments (Brazil, Indonesia,

Mexico, Norway, Philippines, South Africa, United Kingdom, United States) endorsed an Open Government declaration, announced their countries' action plans, and welcomed the commitment of 38 governments to join the partnership. In September 2011, there were 46 national government commitments to Open Government worldwide.

Some of the most important enablers for Open Government are free access to information and the possibility to freely use and re-use this information (e.g. data, content, etc). After all, without information it is not possible to establish a culture of collaboration and participation among the relevant stakeholders. Therefore, Open Government Data (OGD) is often seen as a crucial aspect of Open Government.

OGD is a worldwide movement to open up government/public administration data, information and content to both human and machine-readable non-proprietary formats for re-use by civil society, economy, media and academia as well as by politicians and public administrators. This would apply only to data and information produced or commissioned by government or government-controlled entities and is *not* related to data on individuals..

Being open means lowering barriers to ensure the widest possible re-use by anyone. With OGD, a new paradigm came into being for publishing government data that invites everyone to look, take and play!

The often-used term "Open Data" refers to data and information beyond just governmental institutions and includes those from other relevant stakeholder groups such as business/industry, citizens, NPOs and NGOs, science or education.

Some of the best-known institutions currently undertake Open Data activities include the World Bank[4], the United Nations[5], REEEP[6], the New York Times[7], The Guardian[8] and the Open Knowledge Foundation (OKFN)[9].

In 2007, 30 Open Government advocates came together in Sebastopol, California, USA to develop a set of OGD principles[10] that underscored why OGD is essential for democracy. In 2010, the Sunlight Foundation[11] expanded these to 10 principles. Even though these principles are neither set in stone nor legally binding, they are widely considered by the global open (government) data community as general guidelines on the topic.

Government Data shall be considered "open" if the data is made public in a way that complies with the principles below:

1. **Data must ce complete**
   All public data is made available. The term „data" refers to electronically-stored information or recordings, including but not limited to documents, databases, transcripts, and audio/visual recordings. Public data is data that is not subject to valid privacy, security or privilege limitations, as governed by other statutes.
2. **Data must be primary**
   Data is published as collected at the source, with the finest possible level of granularity, and not in aggregate or modified forms.
3. **Data must be timely**
   Data is made available as quickly as necessary to preserve the value of the data.
4. **Data must be accessible.**
   Data is available to the widest range of users for the widest range of purposes.
5. **Data must be machine-processable**
   Data is structured so that it can be processed in an automated way.
6. **Access must be non-discriminatory**
   Data is available to anyone, with no registration requirement.

7. **Data formats must be non-proprietary**
   Data is available in a format over which no entity has exclusive control.
8. **Data must be license-free**
   Data is not subject to any copyright, patent, trademark or trade secrets regulation. Reasonable privacy, security and privilege restrictions may be allowed as governed by other statutes. Compliance to these principles must be reviewable through the following means:
   - » A contact person must be designated to respond to people trying to use the data; or
   - » A contact person must be designated to respond to complaints about violations of the principles; or
   - » An administrative or judicial court must have the jurisdiction to review whether the agency has applied these principles appropriately.

The two principles added by the Sunlight Foundation are as follows:

9. **permanence**
   Permanence refers to the capability of finding information over time.
10. **usage costs**
    One of the greatest barriers to access to ostensibly publicly-available information is the cost imposed on the public for access – even when the cost is *de minimus*.

It has been acknowledged that the worldwide OGD movement originated in Australia, New Zealand, Europe and North America, but today we also see strong OGD engagement and activity in Asia, South America and Africa. For example, Kenya started Africa's first data portal[12] in July 2011.

The European Commission (EC) has also put the issue high up on its agenda and is actively pushing OGD forward in Europe. Neelie

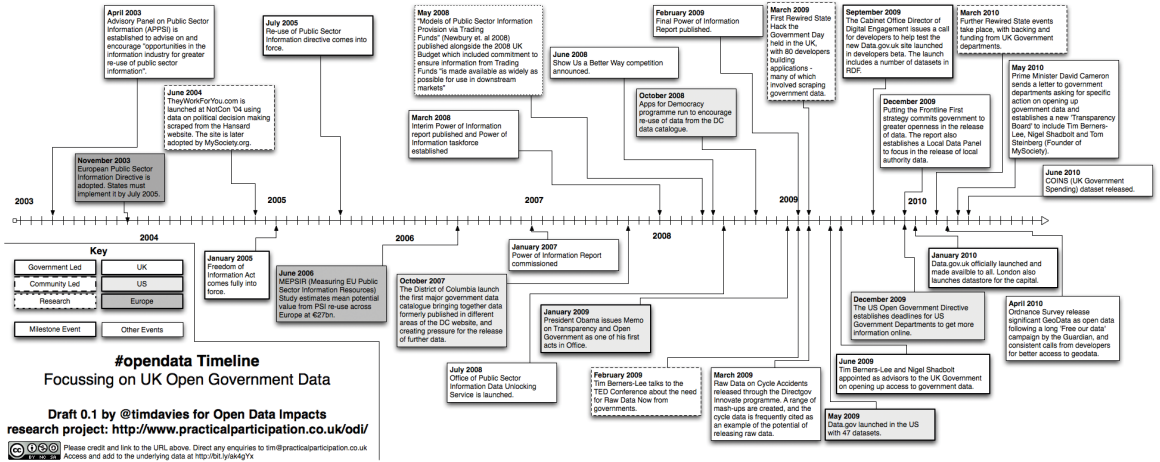
Krose, Vice-President of the European Commission responsible for the Digital Agenda, has stated strong commitment to OGD through her announcement of an EC data portal by early 2012 and for a Pan-European data portal acting as a single point of access for all European national data portals by 2013. Open Data is an important part of both the Digital Agenda for Europe[13] and the European e-government Action Plan 2011–2015[14]. In December 2011 the EC furthermore announced its Open Data Strategy for Europe: Turning Government Data into Gold[15].

The current leading countries in national Open Data activities and initiatives are definitely the governments of the United States of America[16], Australia[17], the Scandinavian countries and the UK government[18]. All of these countries have a high political commitment to both Open Data and central Open Data portals, and they all have a strong Open Data community. These innovative countries and the people behind them can be considered the pioneers of OGD.

Two very good resources about the worldwide OGD movement are:

- SWC world map of Open Data initiatives, activities and portals: http://bit.ly/open-data-map
- OKFN comprehensive list of data catalogs curated by experts from around the world: http://datacatalogs.org

For a good example of a national OGD process, please refer to the following "UK Open Government Data Timeline" by Tim Davies:



Links

(1) The Memorandum on Transparency and Open Government: http://www.whitehouse.gov/the_press_office/TransparencyandOpenGovernment

(2) e-government, Wikipedia: http://en.wikipedia.org/wiki/E-Government

(3) Open Government Partnership: http://www.opengovpartnership.org

(4) Open Data World Bank: http://data.un.org

(5) Open Data United Nations: http://data.worldbank.org

(6) Open Data REEEP: http://data.reegle.info

(7) Open Data New York Times: http://data.nytimes.com

(8) Open Data The Guardian: http://www.guardian.co.uk/world-government-data

(9) Open Knowledge Foundation: http://okfn.org

(10) 8 Principles of Open Government Data: http://www.opengovdata.org/home/8principles

(11) Sunlight Foundation: 10 principles of Open Government Data: http://sunlightfoundation.com/policy/documents/ten-open-data-principles

(12) Kenya Open Data Portal: http://opendata.go.ke

(13) Digital Agenda for Europe: http://ec.europa.eu/information_society/digital-agenda

(14) eGovernment Action Plan Europe 2011–2015: ec.europa.eu/information_society/activities/egovernment/action_plan_2011_2015

(15) Announcement: Open Data Strategy for Europe: http://bit.ly/s5FiQo

(16) Open Data Catalogue United States of America: http://data.gov

(17) Open Data Catalogue of Australia: http://data.gov.au

(18) Open Data Catalogue United Kingdom: http://data.gov.uk

Further Reading

- Open Government, Wikipedia: http://en.wikipedia.org/wiki/Open_government
- Open Knowledge Foundation, OGD website: http://opengovernmentdata.org
- Open Data, Wikipedia: http://en.wikipedia.org/wiki/Open_data
- Open Knowledge Foundation Blog: http://blog.okfn.org

Putting the L in Front: From Open Data to Linked Open Data

As mentioned above, OGD is all about opening up information and data, as well as making it possible to use and re-use it. An OGD requirements analysis was conducted in June 2011 in Austria and highlights the following eleven areas to consider when thinking about OGD:

1. Need for definitions
2. Open Government: transparency, democracy, participation and collaboration
3. Legal issues
4. Impact on society
5. Innovation and knowledge society
6. Impact on economy and industry
7. Licenses, models for exploitation, terms of use
8. Data relevant aspects
9. Data governance
10. Applications and use cases
11. Technological aspects

When considering how to fully benefit from OGD in concrete cases, it is clear that interoperability and standards are key. This is where LOD principles come into play.

To fully benefit from Open Data, it is crucial to put information and data into a context that creates new knowledge and enables powerful services and applications. As LOD facilitates innovation and knowledge creation from interlinked data, it is an important mechanism for information management and integration.

There are two equally important viewpoints to LOD: publishing and consuming. Throughout this guide, we will always address LOD from both the publishing and consumption perspectives.

The path from open (government) data to linked open (government) data was best described by Sir Tim Berners-Lee[1] when he first presented his 5 Stars Model at the Gov 2.0 Expo in Washington DC in 2010. Since then, Berners-Lee's model has been adapted and explained in several ways; the following adaptation of the 5 Stars Model[2] by Michael Hausenblas[3] explains the costs and benefits for both publishers and consumers of LOD.

★	Information is available on the Web (any format) under an open license
★★	Information is available as structured data (e.g. Excel instead of an image scan of a table)
★★★	Non-proprietary formats are used (e.g. CSV instead of Excel)
★★★★	URI identification is used so that people can point at individual data
★★★★★	Data is linked to other data to provide context

What are the costs and benefits of ★ web data?

As a consumer ...	As a publisher ...
✓ You can see it. ✓ You can print it. ✓ You can store it locally (on your hard drive or on an USB stick). ✓ You can enter the data manually into another system.	✓ It is easy to publish.

What are the costs and benefits of ★★ web data?

As a consumer, you can do everything that you could do with ★ web data, plus:	As a publisher ...
✓ You can directly process it with proprietary software to aggregate it, perform calculations, visualise it, etc. ✓ You can export it into another (structured) format.	✓ It is easy to publish.

What are the costs and benefits of ★★★ web data?

As a consumer, you can do everything that you could do with ★★ web data, plus:	As a publisher ...
✓ You do not have to pay for a format over which a single entity has exclusive control	✓ It is easy to publish.

What are the costs and benefits of ★★★★ web data?

As a consumer, you can do everything that you could do with ★★★ web data, plus:	As a publisher ...
✓ You can link to it from any other place, either on the web or locally. ✓ You can bookmark it. ✓ You can re-use parts of the data.	✓ You will need to invest some time slicing and dicing your data. ✓ You will need to assign URIs to data items and think about how to represent the data. ✓ You have fine-granular control over the data items and can optimise their access (e.g. load balancing, caching, etc.)

What are the costs and benefits of ★★★★★ web data?

As a consumer, you can do everything that you could do with ★★★★ web data, plus:	As a publisher ...
✓ You can discover new data of interest while consuming other information. ✓ You have access to the data schema.	✓ You will need to invest resources to link your data to other data on the web. ✓ You make your data discoverable. ✓ You increase the value of your data.

LOD is becoming increasingly important in the fields of state-of-the-art information and data management. It is already being used by many well-known organisations, products and services to create portals, platforms, internet-based services and applications.

LOD is domain-independent and penetrates various areas and domains, thus proving its advantage over traditional data management. For example, the project LOD2[4] Creating Knowledge Out of Interlinked Data, which is funded by the European Commission under the 7th

Framework Programme, develops powerful LOD mechanisms and tools based on three real use cases: OGD, linked enterprise data and LOD for media and publishers. For further reading on linked open (government) data, please refer to the Government Linked Data (GLD) W3C working group[5].

The following chapters discuss the benefits of LOD, as well as basic LOD consuming and publishing principles for creating powerful and innovative services for knowledge management, decision making and general data management. The best practice examples reegle.info[6] and OpenEI[7] show how LOD can have a great impact on their respective target groups. Another popular example of applied OGD is legislation.gov.uk.

Links

(1) Sir Tim Berners-Lee (Wikipedia): http://en.wikipedia.org/wiki/Tim_Berners-Lee

(2) 5 Stars Model on Open Government Data by Michael Hausenblas: http://lab.linkeddata.deri.ie/2010/star-scheme-by-example

(3) Michael Hausenblas: http://semanticweb.org/wiki/Michael_Hausenblas

(4) LOD2 – Creating Knowledge Out of Interlinked Data: http://www.lod2.eu

(5) GLD W3C Working Group: http://www.w3.org/2011/gld/charter

(6) Clean energy info portal reegle.info: http://www.reegle.info

(7) Open Energy Info (OpenEI): http://en.openei.org

Further reading

- Linked Data, Wikipedia: http://en.wikipedia.org/wiki/Linked_data
- Linked Data – Connect Distributed Data Across the Web: http://linkeddata.org

- Linked Data: Evolving the Web into a Global Data Space, Heath and Bizer: http://linkeddatabook.com
- Linking Government Data, David Wood (Editor), Springer; 2011 edition (November 12, 2011), ISBN-10: 146141766X, ISBN-13: 978-1461417668
- W3C Linking Open Data Community Project: http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData

2. THE POWER OF LINKED OPEN DATA

Understanding World Wide Web Consortium's (W3C)[1] vision of a new web of data

Imagine that the web is like a giant global database. You want to build a new application that shows the correspondence among economic growth, renewable energy consumption, mortality rates and public spending for education. You also want to improve user experience with mechanisms like faceted browsing. You can already do all of this today, but you probably won't. Today's measures for integrating information from different sources, otherwise known as mashing data, are often too time-consuming and too costly.

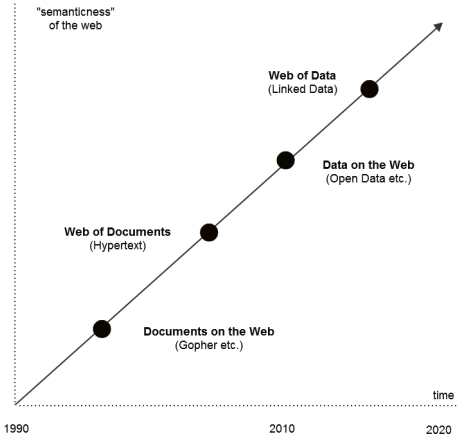
Two driving factors can cause this unpleasant situation:

First of all, databases are still seen as „silos", and people often do not want others to touch the database for which they are responsible. This way of thinking is based on some assumptions from the 1970s: that only a handful of experts are able to deal with databases and that only the IT department's inner circle is able to understand the schema and the meaning of the data. This is obsolete. In today's internet age, millions of developers are able to build valuable applications whenever they get interesting data.

Secondly, data is still locked up in certain applications. The technical problem with today's most common information architecture is that metadata and schema information are not separated well from application logics. Data cannot be re-used as easily as it should be. If someone designs a database, he or she often knows the certain application to be built on top. If we stop emphasising which applications will use our data and focus instead on a meaningful description of the data itself, we will gain more momentum in the long run. At its core, Open Data means that the data is open to any kind of application and this can be achieved if we use open standards like RDF[2] to describe metadata.

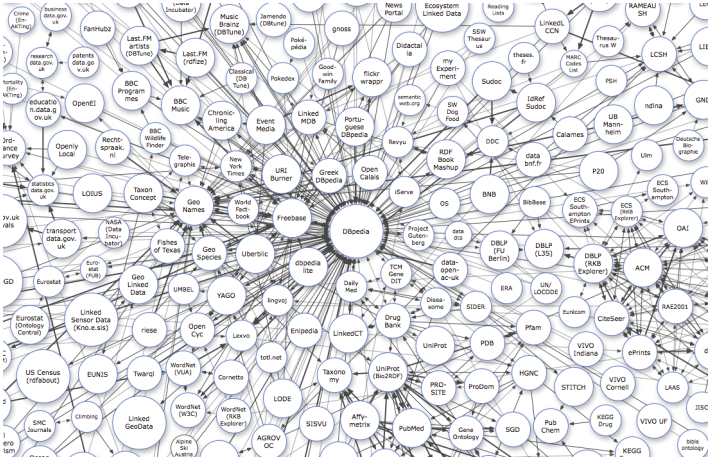
Linked Data?

Nowadays, the idea of linking web pages by using hyperlinks is obvious, but this was a groundbreaking concept 20 years ago. We are in a similar situation today since many organizations do not understand the idea of publishing data on the web, let alone why data on the web should be linked. The evolution of the web can be seen as follows:



Although the idea of Linked Open Data (LOD) has yet to be recognised as mainstream (like the web we all know today), there are a lot of LOD already available. The so called LOD cloud[3] covers more than an estimated 50 billion facts from many different domains like geography, media, biology, chemistry, economy, energy, etc. The data is of varying quality and most of it can also be re-used for commercial purposes.

Please see a current version of the LOD Cloud diagram of 2011 as follows:



Why should we link data on the web and how do we do it?

All of the different ways to publish information on the web are based on the idea that there is an audience out there that will make use of the published information, even if we are not sure who exactly it is and how they they will use it. Here are some examples:

- Think of a twitter message: not only do you not know all of your followers, but you often don't even know why they follow you and what they will do with your tweets.
- Think of your blog: it´s like an email to someone you don't know yet.
- Think of your website: new people can contact you and offer new surprising kinds of information.
- Think of your email-address: you have shared it on the web and receive lots of spam since then.

In some ways, we are all open to the web, but not all of us know how to deal with this rather new way of thinking. Most often the „digital natives" and „digital immigrants" who have learned to work and live with the social web have developed the best strategies to make use of this kind of „openness." Whereas the idea of Open Data is built on the concept of a social web, the idea of Linked Data is a descendant of the semantic web.

The basic idea of a semantic web is to provide cost-efficient ways to publish information in distributed environments. To reduce costs when it comes to transferring information among systems, standards play the most crucial role. Either the transmitter or the receiver has to convert or map its data into a structure so it can be „understood" by the receiver. This conversion or mapping must be done on at least three different levels: used syntax, schemas and vocabularies used to deliver meaningful information; it becomes even more time-consuming when information is provided by multiple systems. An ideal scenario would be a fully-harmonised internet where all of those layers are based on exactly one single standard, but the fact is that we face too many standards or „de-facto standards" today.

How can we overcome this chicken-and-egg problem? There are at least three possible answers:

- Provide valuable, agreed-upon information in a standard, open format.
- Provide mechanisms to link individual schemas and vocabularies in a way so that people can note if their ideas are "similar" and related, even if they are not exactly the same.
- Bring all this information to an environment which can be used by most, if not all of us. For example: don't let users install proprietary software or lock them in one single social network or web application!

A brief history of LOD

Corresponding to the three points above, here are the steps already done by the LOD community:

- W3C has published a stack of open standards for the semantic web built on top of the so-called „Resource Description Framework" (RDF). This widely-adopted standard for describing metadata was also used to publish the most popular encyclopedia in the world: Wikipedia now has its „semantic sister" called DBpedia[4], which became the LOD cloud's nucleus.
- W3C´s semantic web standards also foresee the possibility to link data sets. For example, one can express in a machine-readable format that a certain resource is exactly (or closely) the same as another resource, and that both resources are somewhere on the web but not necessarily on the same server or published by the same author. This is very similar to linking resources to each other using hyperlinks within a document, and is the atomic unit for the giant global database previously mentioned.
- Semantic web standards are meant to be used in the most common IT infrastructure we know today: the worldwide web (WWW). Just use your browser and use HTTP! Most of the LOD cloud's resources and the context information around them can be retrieved by using a simple browser and by typing a URL in the address bar. This also means that web applications can make use of Linked Data by standard web services.

Already reality – an example

Paste the following URL in your browser: http://dbpedia.org/resource/Renewable_Energy_and_Energy_Efficiency_Partnership and you will receive a lot of well structured facts about REEEP. Follow the fact that REEEP „is owner of" reegle (http://dbpedia.org/resource/Reegle) and so on and so forth. You can see that the giant global database is already a reality!

Complex systems and Linked Data

Most systems today deal with huge amounts of information. All information is produced either within the system boundaries (and partly published to other systems) or it is consumed "from outside," "mashed" and "digested" within the boundaries. Some of the growing complexity has been caused in a natural way due to a higher level of education and the technical improvements made by the ICT sector over the last 30 years. Simply said, humanity is now able to handle much more information than ever before with probably the lowest costs ever (think of higher bandwidths and lower costs of data storage).

However, most of the complexity we are struggling with is caused above all by structural insufficiencies due to the networked nature of our society. The specialist nature of many enterprises and experts is not yet mirrored well enough in the way we manage information and communicate. Instead of being findable and linked to other data, much information is still hidden.

With its clear focus on high-quality metadata management, Linked Data is key to overcoming this problem. The value of data increases each time it is being re-used and linked to another resource. Re-usage can only be triggered by providing information about the available information. In order to undertake this task in a sustainable manner, information must be recognised as an important resource that should be managed just like any other.

Examples for LOD applications

Linked Open Data is already widely available in several industries, including the following three:

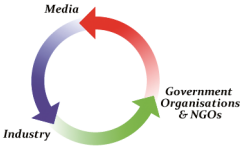
- Linked Data in libraries[5] : focusing on library data exchange and the potential for creating globally interlinked library data; exchanging and jointly utilising data with non-library institutions;

growing trust in the growing semantic web; and maintaining a global cultural graph of information that is both reliable and persistent.

- Linked Data in biomedicine[6]: establishing a set of principles for ontology/vocabulary development with the goal of creating a suite of orthogonal interoperable reference ontologies in the biomedical domain; tempering the explosive proliferation of data in the biomedical domain; creating a coordinated family of ontologies that are interoperable and logical; and incorporating accurate representations of biological reality.
- Linked government data: re-using public sector information (PSI); improving internal administrative processes by integrating data based on Linked Data; and interlinking government and non-government information.

The future of LOD

The inherent dynamics of Open Data produced and consumed by the "big three" stakeholder groups – media, industry, and government organizations/NGOs – will move forward the idea, quality and quantity of Linked Data – whether it is open or not:



Whereas most of the current momentum can be observed in the government & NGO sectors, more and more media companies are jumping on the bandwagon. Their assumption is that more and more industries will perceive Linked Data as a cost-efficient way to integrate data.

Linking information from different sources is key for further innovation. If data can be placed in a new context, more and more valuable applications – and therefore knowledge – will be generated.

Links

(1) The World Wide Web Consortium (W3C): is an international community that develops open standards to ensure the long-term growth of the Web: http://www.w3.org

(2) Resource Description Framework (RDF): http://www.w3.org/RDF

(3) Linked Open Data Cloud: http://www.lod-cloud.net

(4) DBpedia: http://dbpedia.org

(5) Jan Hannemann, Jürgen Kett (German National Library): „Linked Data for Libraries" (2010) http://www.ifla.org/files/hq/papers/ifla76/149-hannemann-en.pdf

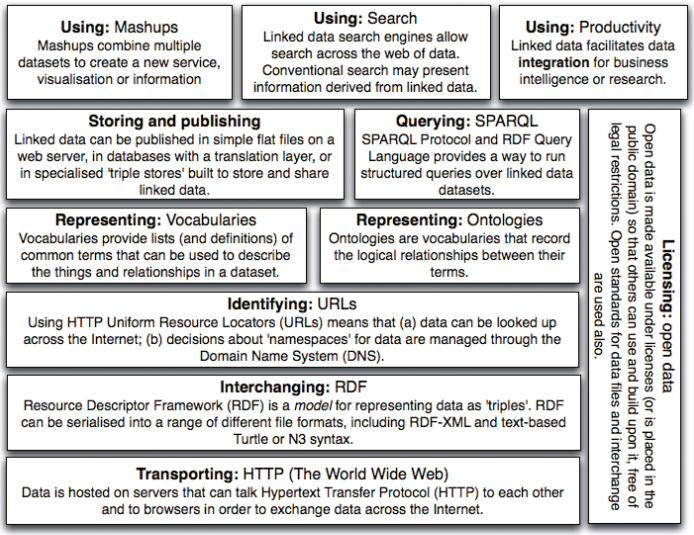
(6) OBO Foundry: http://obofoundry.org

3. LINKED OPEN DATA START GUIDE

A quick guide for your own LOD strategy and appearance

The following two sections review LOD publication and consumption and provide the essential information for establishing a powerful LOD strategy for your own organisation. We also provide further reading recommendations for anyone seeking more technical details on LOD publishing and consumption, as well as a list of the most important software tools for publishing and consuming LOD.

The following figure is a technical overview of the necessary blocks for building your strategy for LOD publishing and consumption.



Elements of the Linked Open Data Puzzle (revision 2) - 2nd May 2011. CC BY-SA-NC
Draft sketch by Tim Davies (@timdavies / tim@practicalparticipation.co.uk) for IKM Working Paper on Linked Open Data for Development. Comments welcome. Search 'linked open data puzzle' on http://www.opendataimpacts.net for latest version.

Idea based on Semantic Web Stack at http://en.wikipedia.org/wiki/Semantic_Web_Stack

3.1. Publishing Linked Open Data

First steps for publishing your content as LOD

The idea, benefit and effort of publishing LOD have already been mentioned in the previous chapters of this publication, and were discussed along the 5 Star Model of OGD. Publishing Linked Open DataLOD provides a powerful mechanism for sharing your own data and information along with your metadata and the respective data models for efficient re-use. Going LOD helps your organisation to become an important data hub within your domain.

Quick guide for publishing LOD

We have prepared a short guide for the most important issues that need to be taken into account when publishing LOD as well as a step-by-step model to get started.

Analyse your data

Before you start publishing your data, it is crucial to take a deeper look at your data models, your metadata and the data itself. Get an overview and prepare a selection of data and information that is useful for publication.

Clean your data

Data and information that comes from many distributed data sources and in several different formats (e.g. databases, XML, CSV, Geodata, etc.) require additional effort to ensure easy and efficient modelling. This includes ridding your data and information of any additional information that will not be included in your published data sets.

Model your data

Choose established vocabularies and additional models to ensure smooth data conversion to RDF. The next step is to create unified

resource identifiers (URIs)[1] as names for each of your objects. To ensure sustainability, remember to develop data models for data that change over time.

Choose appropriate vocabularies

There are lots of existing RDF vocabularies for re-use; please evaluate appropriate vocabularies for your data from existing ones. If there are no vocabularies that fit your needs, feel free to create your own.

Specify license(s)

To ensure broad and efficient re-use of your data, evaluate, specify and provide a clear license for your data to avoid its re-use in a legal vacuum. If possible, specify an existing license that people already know. This enables interoperability with other data sets in the field of licensing. For example, Creative Commons[2] is a commonly-used license for OGD.

Convert data to RDF

One of the final steps is to convert your data to RDF[3], a very powerful data model for LOD. RDF is an official W3C recommendation for semantic web data models. Remember to include your specified license(s) into your RDF files.

Link your data to other data

Before you publish, make sure that your data is linked to other data sets; links to your other data sets and to third party data sets are useful. These links ensure optimised data processing and integration for data (re-)use and allow for the creation of new knowledge from your data sets by putting them into a new context with other data. Evaluate and choose carefully the most relevant data sets to be linked with your own.

Publish and promote your LOD

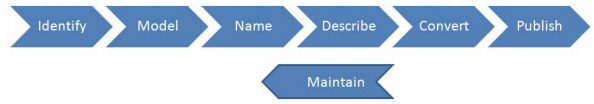
Publish your data on the web and promote your new LOD sets to ensure wide re-use – even the best LOD will not be used if people cannot find it! Alongside other ways of promotion it is a great idea to add your LOD sets into the LOD cloud[4], a visual presentation of LOD

sets by providing and updating the meta-information about your data sets on the data hub[5]. Remember to always provide human-readable descriptions of your data sets to make the data sets „self-describing" for easy and efficient re-use.

For a similar approach, we recommend the „Ingredients for high quality Linked (Open) Data" by the W3C Linked Data Cookbook[6]. The essential steps to publishing your own LOD are:

1. Model and link the data
2. Name things with URIs
3. Re-use vocabularies whenever possible
4. Publish human- and machine-readable descriptions
5. Convert data to RDF
6. Specify an appropriate license
7. Announce the new Linked Data Set(s)

The following life cycle of Linked Open (Government) Data by Bernadette Hyland[7] visualises the path for LOD publishing:



The Four Rules of Linked Data (W3C Design Issues for Linked Data[8]) are also a good place to start understanding LOD principles:

The semantic web isn't just about putting data on the web – that is the old „web of pages." It is about making links, so that a person or machine can explore the semantically connected „web of data." With Linked Data, you can find more related data.

Like the web of hypertext, the web of data is constructed with documents on the web. However, unlike the web of hypertext, where links are relationships anchors in hypertext documents written in

HTML, LOD functions through links between arbitrary things described by RDF. The URIs identify any kind of object or concept, but regardless of HTML or RDF, the same expectations apply to make the web grow:

1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the established standards (e.g. RDF, SPARQL)
4. Include links to other URIs, so that more things can be discovered

Furthermore, it is crucial to provide high quality information for developers and data workers about your data. Provide information about data provenance as well as its data collection to guarantee smooth and efficient work with your data.

To ensure widest possible re-use, provide a (web) API[9] on top of the published data sets that allows users to query your data and to fetch data and information from your data collection tailored to their needs. A web API enables web developers to easily work with your data.

Here are some best practice examples for publishing LOD:

- UK official National Open (Government) Data Portal – Linked Data Area: http://data.gov.uk/linked-data
- Official UK Legislation: http://www.legislation.gov.uk
- reegle.info LOD portal: http://data.reegle.info
- EU project: LATC – LOD around the clock: http://latc-project.eu

Links

(1) Uniform Resource Identifier, URI on Wikipedia: http://en.wikipedia.org/wiki/Uniform_resource_identifier

(2) Creative Commons: http://creativecommons.org

(3) Resource Description Framework (RDF): http://www.w3.org/RDF/ RDF on Wikipedia: http://en.wikipedia.org/wiki/Resource_Description_Framework

(4) The LOD Cloud: http://richard.cyganiak.de/2007/10/lod

(5) The Data Hub (formerly CKAN): http://thedatahub.org

(6) W3C Linked (Open) Data Cookbook: http://www.w3.org/2011/gld/wiki/Linked_Data_Cookbook

(7) Bernadette Hyland: http://3roundstones.com/about-us/leadership-team/bernadette-hyland

(8) W3C Design Issues for Linked Data: http://www.w3.org/DesignIssues/LinkedData.html

(9) Web API: http://en.wikipedia.org/wiki/Web_API
or Web Service: http://en.wikipedia.org/wiki/Web_service

Further Reading

- How to publish Linked Data on the Web, Bizer et al: http://www4.wiwiss.fu-berlin.de/bizer/pub/linkeddatatutorial
- Linked Data – Connect Distributed Data across the Web: http://linkeddata.org
- Linked Data: Evolving the Web into a Global Data Space, Heath and Bizer: http://linkeddatabook.com
- Designing URI Sets for the UK Public Sector: http://www.cabinetoffice.gov.uk/resource-library/designing-uri-sets-uk-public-sector
- Linked Data Patterns, Dodds & Davies: http://patterns.dataincubator.org/book/linked-data-patterns.pdf
- Linking Government Data, David Wood (Editor), Springer; 2011 edition (November 12, 2011), ISBN-10: 146141766X, ISBN-13: 978-1461417668

3.2. Consuming Linked Open Data

First steps for consuming content as LOD

Consuming LOD enables you to integrate and provide high quality information and data collections to mix your own data and third party information. These enriched data collections can act as single points of access for a specific domain in the form of a LOD portal and as an internal or Open Data warehouse system that enables better decision making, disaster management, knowledge management and/or market intelligence solutions.

Organisations can benefit and reach competitive advantage through the possibility to: 1) spontaneously generate dossiers and information mash ups from distributed information sources; create applications based on real time data with less replication; and 3) create new knowledge out of this interlinked data.

Quick guide for consuming LOD

Here are the most important issues and milestones to consider when consuming LOD:

Specify concrete use cases

Always specify concrete (business) use cases for your new service or application. What is the concrete problem you would like to solve? What data is available internally and what will you need from third party sources?

Evaluate relevant data sources and data sets

Based on your concrete use case(s), the next step is to evaluate relevant LOD sources for data integration. Find out what data sources are available and what quality of data these third party sources offer (data quality is often associated with the information source itself; well known organisations usually provide high quality data and information). A very good approach for this evaluation is to use a LOD

search engine like Sindice[1] or one of the globally available Open Data catalogues such as The Data Hub[2]. Also consider data set update cycles and when the data was last updated.

Check the respective licenses

Evaluate the licenses for use and re-use provided by the owners of the data. Avoid using data where no clear and understandable license is available. If in doubt, contact the respective data holders and clarify these questions. It is also important to know what license these data sets provide for mashing up data sets with others.

Create consumption patterns

Creating consumption patterns specifies in detail exactly which data is re-used from a certain data source. Not all data in a set will be relevant to the specified use case(s), in which case you can develop consumption patterns that clearly specify only the relevant data in the set.

Manage alignment, caching and updating mechanisms

When LOD is consumed, the need for matching different vocabularies of the consumed (internal and external) data sets often occurs. This is relevant to ensure smooth data integration by vocabulary alignment[3]. Another concern is the fact that LOD sources are not absolutely stable nor always available to consume data in real time. To prevent a specific data set from being unavailable at a certain time, create caching mechanisms for specific third party data and information. Another important issue is to consume up-to-date information; a feasible approach here is to implement updating mechanisms for LOD consumption. Please see the „Linked Open Data Tool Box Collection" at the end of this chapter for more information.

Create mash ups, GUIs, services and applications on top

To serve your users and to create powerful LOD applications or services on top of mashed up LOD, it is crucial to provide user-friendly graphical user interfaces (GUIs) and powerful services for end users.

Establish sustainable new partnerships

When using third party data and information, contact the data providers to build new partnerships and offer your own data for vice-versa use.

As a conclusion, please consider some best practice examples for consuming LOD from these LOD players:

- UK Organograms: http://data.gov.uk/organogram/hm-treasury
- reegle.info country profiles: http://www.reegle.info/countries
- EU project: LATC – Linked Open Data Around-The-Clock: http://latc-project.eu

Links

(1) Sindice – the semantic web index: http://sindice.com/

(2) The Data Hub: http://thedatahub.org

(3) Vocabulary / Ontology Alignment on Wikipedia: http://en.wikipedia.org/wiki/Ontology_alignment

Further Reading

- Second International Workshop on Consuming Linked Data: http://km.aifb.kit.edu/ws/cold2011
- Semantic Web for the Masses, paper by Lisa Goddard: http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/3120/2633
- Linked Data: The Future of Knowledge Organization on the Web: http://www.iskouk.org/events/linked_data_sep2010.htm
- Linked Data: Evolving the Web into a Global Data Space, Heath & Bizer: http://linkeddatabook.com

Linked Open Data Tool Box Collection

The Linked Open Data Tool Box Collection provides a list of important software tools and services for LOD publishing and consumption.

- **PoolParty Product Family**: http://www.poolparty.biz
  Services and tools for LOD-based metadata management, enterprise search, text mining and data integration
- **LOD2 Technology Stack**: http://lod2.eu/WikiArticle/TechnologyStack.html
  Technology stack for LOD by the R&D project LOD2 - Creating Knowledge Out of Interlinked Data
- **Silk**: http://www4.wiwiss.fu-berlin.de/bizer/silk
  A link discovery framework for the web of data
- **LIMES**: http://aksw.org/Projects/LIMES
  Link discovery framework for metric spaces
- **Virtuoso Universal Server**: http://virtuoso.openlinksw.com
  Universal server for Linked Data consumption, storage and retrieval
- **Callimachus Project**: http://callimachusproject.org
  A framework for data-driven applications using Linked Data
- **Sindice**: http://sindice.com
  The semantic web Index
- **RDF Alchemy / The LOD Manager**: http://www.semantic-web.at/linked-data-manager
  Management suite for scheduling alignment, caching and updating mechanisms for LOD

A good resource for LOD projects and suppliers (in the domain of eGovernment) is W3C's community directory, which was launched in late 2011: http://dir.w3.org

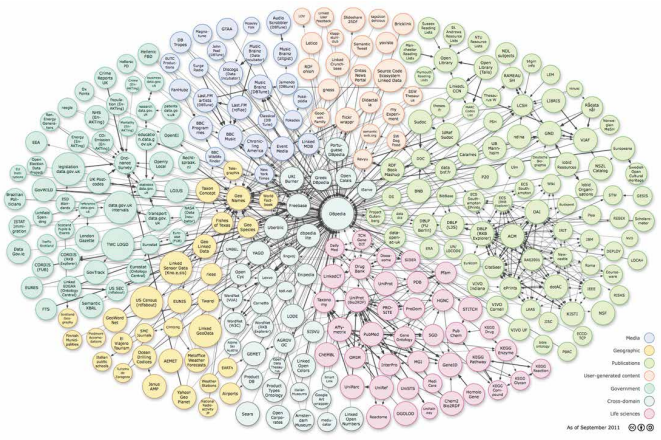
4. BEST PRACTICE AND EXAMPLES

Four best practice examples to showcase the power of LOD

LOD is still a relatively new field. One of the first projects to kickstart the web of Linked Data sets was the Linking Open Data Project[1], which identified existing data sets available under open license, converted them to RDF according to the Linked Data principles, and published them on the web. The Linking Open Data Project is a community effort founded in January 2007 and supported by the W3C Semantic Web Education and Outreach Group [2].

Since 2007, the web of data has grown enormously and now includes data sets from diverse organisations and data providers including media (e.g. BBC, NYT), governments (e.g. US, UK) as well as user-generated content. To visualise key LOD providers and their linkages, Richard Cyganiak (DERI) and Anja Jentzsch (Freie Universität Berlin) regularly compile a diagram for the linking Open Data cloud[3], in which each node represents a data set published as Linked Data and inter-node arcs illustrate connections between datasets. As of September 2011, the datasets in the LOD cloud consist of over 31 billion RDF triples* and are interlinked by around 504 million RDF links.

* A RDF triple is an expression that defines a way in which you can represent a relationship between objects in a dataset. Usually there are three parts to a triple: Subject, Predicate and Object.



"Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch. http://lod-cloud.net – September 2011"

As you can see in the cloud diagram, some of the data providers have become well-known and established as popular linking hubs in the web of data. Prominent examples are „DBpedia"[4], a community effort to extract structured information from Wikipedia and „geonames"[5] that provides RDF descriptions of millions of geographical locations worldwide.

The next chapters highlight four successful examples of pioneering websites and applications that already use LOD to enrich their own content and to publish their data sets in RDF as LOD for free re-use by external parties:

- reegle.info – a clean energy information portal focusing on high quality information on renewable energy, energy efficiency and climate-compatible, development-related topics
- openEI – a portal providing various energy data sets in a semantic wiki
- Legislation.gov.uk and Legislate Mobile – a portal with data from the official national government archive of the United Kingdom

- Solar Med Atlas: Socio-Economic Data Integration - within the broader framework of the portal solar-med-atlas.org, socio-economic information for Mediterranean countries is presented based on data pulled from LOD sources

All four best practices are part of the LOD cloud diagram shown above, and their data sets are interlinked with other data providers to maximise the benefits of using LOD technology. More examples can be found on the CKAN[6] directory „Data Hub," a registry of open knowledge data sets and projects.

Links:

(1) Linking Open Data Project: http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData

(2) W3C Semantic Web Education and Outreach Group: http://www.w3.org/2001/sw/sweo

(3) The Linked Open Data (LOD) Cloud: http://lod-cloud.net

(4) DBpedia: http://dbpedia.org/About

(5) Geonames: http://www.geonames.org/ontology

(6) The Data Hub: http://thedatahub.org

Further Reading:

- Linked Data – The Story So Far, Christian Bizer, Tom Heath, Tim Berners-Lee, International Journal on Semantic Web and Information Systems (IJSWIS) (2009): http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf
- Linking Government Data, David Wood (Editor), Springer; 2011 edition (November 12, 2011), ISBN-10: 146141766X, ISBN-13: 978-1461417668
- The Joy of Data - A Cookbook for Publishing Linked Government Data on the Web, Bernadette Hyland und David Wood: http://www.springerlink.com/content/n30nq362wr678101

4.1. reegle.info Country Profiles

reegle.info consuming LOD to provide comprehensive clean energy country profiles

This best practice example on how to produce and consume Linked Data is based on the country profiles of the clean energy information gateway reegle.info[1]. These comprehensive dossiers focus heavily on energy-related issues and the entire content is fetched from different Open Data sources in a self-maintaining way that ensures that reegle users can always access the latest high quality information in a visually appealing presentation.

Reegle has already established itself as a popular information portal in the fields of renewable energy and energy efficiency. With more than 200,000 users a month (October 2011), its services are widely used. Reegle was launched in 2006 but underwent a complete re-design in 2010, not only in style but also in content, technology and services. Reegle now offers all of its data under W3C standards, i.e. it is open and Linked Data in a non-proprietary format (Resource Description Framework – RDF).

As a consumer and provider of Open Data, reegle has collected its information from the most highly-rated sources and offers this data in 243 individual country profiles.

reegle's country energy profiles

Reegle offers comprehensive energy-related information on 243 countries and regions, and enriches retrieved external Open Data with REEEP's own policy and regulatory overview. All countries display relevant energy statistics from established sources such as the UN and the World Bank as well as all reegle stakeholder ("actors") active in the relevant country. Projects outputs are also offered wherever available and round out this unique information output.

Maps, tools and programs from NREL were also integrated, as their data portal OpenEI.org also uses Linked Data technology, offers data sets in RDF format, and provides a SPARQL endpoint for accessing their datasets.

Consuming and providing Open Data

New sources meeting reegle's quality requirements are constantly reviewed and added to improve reegle's country profiles. Right now, data is provided by UN Data, World Bank Data, DBpedia, Eurostat, OpenEI, RES-Legal and REEEP. Reegle is also merging UN and World Bank data to show the share of renewables in any given country, excluding nuclear power. Using LOD makes it possible to process

multiple data sets and provide added value by combining different data sets. Without the possibility of using machine-readable data sets, a huge amount of manpower would be needed to provide this service.

Reegle also established itself in the LOD cloud as a provider of Open Data and is listed in CKAN's directory of Open Data sets. As a provider of an ever-increasing amount of data, reegle has also taken on the challenge of building the SKOS-based „Renewable Energy and Climate Compatible Development Thesaurus" with full Linked Data capacities to structure and retrieve energy and climate data.

Benefits of using Linked Data

One of the largest benefits of using and providing Linked Open Data sets for reegle is the split of responsibilities. For a small organisation like REEEP, it would not be possible to have a huge team of people to maintain a large database with information on clean energy. Linked Open Data sets make it possible to provide reegle users with a comprehensive and up-to-date overview on various energy-related and country-specific topics. Instead of updating this information manually, it is directly linked to data providers' information, so any updates are reflected immediately.

As a provider of Linked Open Data sets[2], reegle offers developers of other applications / websites the opportunity to easily extract and use all reegle / REEEP data free of charge through the SPARQL endpoint. This ensures that its datasets are widespread and reach their target audience even if they do not retrieve it directly from reegle.

Reegle is a product of REEEP[3] and REN21[4].

Links:

(1) Clean energy info portal reegle: http://www.reegle.info

(2) reegle data portal: http://data.reegle.info

(3) REEEP – the Renewable Energy and Energy Efficiency Partnership: http://www.reeep.org

(4) REN21 – the Renewable Energy Policy Network for the 21st Cetury: http://www.ren21.net

Further reading:

- Open Data info: http://blog.reegle.info/blog/tag/open-data
- Developer Guide: http://data.reegle.info/developers/guide
- Thesaurus Guide: http://data.reegle.info/thesaurus/guide

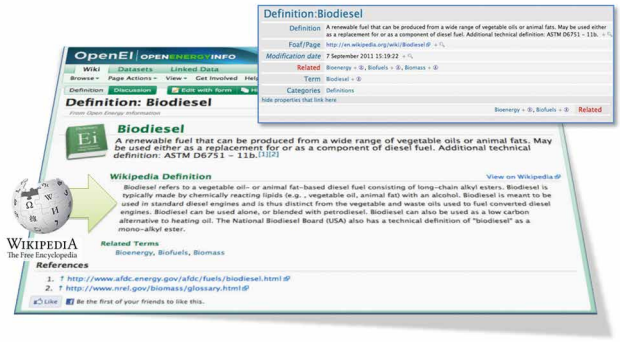
4.2. OpenEI Definitions

OpenEI publishing and consuming open energy information

OpenEI, short for open energy information[1], is a collaborative knowledge-sharing platform with free and open access to energy-related data, models, tools, and information. OpenEI features over 55,000 content pages, more than 600 downloadable data sets, regional gateways on a variety of energy-related topics, and numerous online tools. Among these tools is the „OpenEI Glossary"[2], an online glossary of energy terms, which is a model for producing and consuming Linked Data.

OpenEI's data-sharing mission

Sharing data is a key component of OpenEI's mission. Linked Data principles were adopted to ensure that the vast majority of information found on OpenEI is made available in a variety of universal formats, including restful API, RDF and SPARQL. Exposing data in these machine-readable formats makes the information instantly useful to other Linked Data-savvy enterprises.



Wherever possible, the OpenEI Glossary features related terms and definitions collected from other sources. This is made possible by the Linked Data services provided by other agencies such as DBPedia and reegle. OpenEI obtains this information through RDF or SPARQL endpoints in real time, ensuring that the information provided to the user is always current.

All talking about the same thing

Providing this additional information strengthens users' understanding of the defined terms while semantically linking the resources to a central concept. The result is a machine-readable relationship among multiple data sources. This relationship allows machines to automatically determine the term associated with each definition and is essential in collecting the appropriate definitions for a specific term.

In the case of OpenEI Definitions, this is achieved using a semantic property called *foaf:page*. Because OpenEI, DBPedia, and reegle all refer to the same Wikipedia article as a page discussing the same topic, OpenEI is able to automatically generate the SPARQL necessary to fetch definitions from each of the other sites. In short, by semantically linking the OpenEI definition to a central concept (the Wikipedia page), the developers gain easy access to all other information linked to that same concept.

Returning the favor

Linked Data on OpenEI is about more than just developing informative pages that leverage multiple sources to provide a comprehensive user experience. By properly attributing definitions with the appropriate semantic properties, OpenEI can be sure that each piece of leveraged content points back to its original source. Furthermore, the OpenEI definition for each term is also published by assigning the value of the semantic property foaf:page to the same Wikipedia URI. In the case of OpenEI, properties assigned to definitions and other semantic content are automatically published to OpenEI's SPARQL and RDF endpoints. This completes the semantic relationship so that anyone looking to explore the meaning behind a concept will find the OpenEI definition alongside DBpedia's and reegle's in the LOD cloud.

Benefits of using Linked Data

Linking multiple definitions to the same concept strengthens the understanding of that concept and enables both human and machine inquirers to confirm that they are, in fact, defining the same thing. But what about the business benefits? By linking OpenEI data to others, OpenEI can essentially outsource select content and information, allowing those portions of the site to be actively maintained by the appropriate subject matter experts. This allows the OpenEI team to focus their efforts on areas of unique expertise. Plus, being a resource for other sites, such as reegle, has led to a marked increase in web traffic on OpenEI. In October 2011, OpenEI saw an additional 1,200 unique visitors referred from sites like reegle as a result of sharing Linked Data.

Text by: Jon Weers, NREL (Jon.Weers@nrel.gov)

This work was supported by the U.S. Department of Energy under Contract No. DE-AC36-08-GO28308 with the National Renewable Energy Laboratory.

Links:

(1) OpenEI Portal: http://en.openei.org

(2) OpenEI Glossary: http://en.openei.org/wiki/Glossary

Further reading:

- LOD on Open Energy Information: http://en.openei.org/lod
- FOAF Project: http://www.foaf-project.org

4.3. The official home of UK legislation

Legislation.gov.uk as a best practice example of publishing & consuming LOD

Legislation.gov.uk[1] is an official government archive of the United Kingdom, managed by the National Archives and is a best practice example of publishing Open Government Data. The portal offers access to all published UK legislation, so that it can be shared and re-used by citizens, community groups and businesses. The available data covers about 800 years in time, beginning with 1267 and including the most recent changes in UK legislation for everyone to discover and use.

What can users find on legislation.gov.uk?

Legislation.gov.uk compiles the legislative content from the OPSI (Office of Public Sector Information)[2] website and revised legislation from the Statute Law Database[3] to provide a single legislation service that fulfills all criteria of a modern government portal. The OPSI and SLD websites will be decommissioned shortly and users will be re-directed to the new service.

Types of legislation covered include UK public and local acts, church measures and instruments, ministerial orders and acts of the Parliament of Great Britain.

The main and general objective of legislation.gov.uk is to deliver a high-quality public service for people who need to consult, cite, and use legislation as part of the traditional web of documents. Publishing the UK's Statute Book as data, for people to take, use, and re-use for whatever purpose or application they may need it for, is a service made possible through the new web of data.

Research has also shown that users of the OPSI website and UK Statute Law Database were often confused about the relevance of legislation accessed. Therefore, it is a primary goal to make legislation.gov.uk intuitive and to make it obvious whether a certain piece is current or historical. Legislation can be searched by year, as more recent data sets from the last few decades tend to be complete whereas historical data sets are often partial. A timeline is now available for advanced users to navigate through particular Acts over the centuries and makes legislation.gov.uk a worthy successor to previous portals.

A map view is also available to quickly see which legislation applies to all of the UK or to just England, Scotland, Wales or Northern Ireland.

Open Government Data as Linked Data

The UK legislative data portal was developed fully in line with W3C standards to ensure that citizens can enjoy all the benefits of consuming and re-using LOD as well as comfortably browse through the portal's webpages.

Persistent URIs were used for the legislative data, and there are currently three types of URIs for legislation data on legislation.gov.uk – identifier URIs, document URIs and representation URIs. These URIs are already being picked up and used by other semantic web pioneers – which is exactly the idea behind creating these persistent URIs.

To ensure the possibility and incentive to link legislative data with other Linked Data, important concepts and ideas are formally defined by statute. Since all concepts of legislation data have now been defined

by statute, it is possible to use these definitions to connect terms with appropriate statutes in a thesaurus or ontology. For example, the ESD toolkit[4] created a controlled vocabulary of services by different authorities and linked them to the corresponding identifier URIs of powers and duties from legislation.gov.uk. A controlled vocabulary, like a thesaurus or an ontology, drastically improves a database's efficiency when being searched. Drawing different ways of describing certain legislative concepts, like synonyms or even misspellings, under a single term is how the ESD Toolkit's controlled vocabulary vastly improves the searchability of the data and improves the whole user experience. Plus, connecting governing legislation and enacting authorities adds concrete value to public information and citizens.

Relying on LOD standards also makes it possible for legislation.gov.uk to give accurate information about when a section is repealed, by what piece of legislation, and when that repeal comes into force. Again, this is possible by connecting, or linking, the right pieces of information through persistent URIs.

Publishing UK legislation in this way has made the UK statute book an important contribution to both the web of data and the old-fashioned web of documents.

Application of legislative data

The mobile legislate app[5] provides an iPhone-friendly way to access legislation.gov.uk data whilst out and about. Through the application, it is easy to view either HTML or PDF versions of legislation data. The interface has been optimised for the smaller screen; interesting information can be bookmarked; it is possible to explore a specific legal matter in-depth. Even though data is cached for 24 hours to ensure smooth user experience, new legislation

should appear in the app for browsing within 24 hours of it being released publicly. Sharing noteworthy pieces of legislation on Facebook is also made easy through this app.

Accessibility details

Technologies used on legislation.gov.uk include XHTML + RDFa, CSS and JavaScript. It is possible to convert to PDFs and Braille copies may be obtained.

The whole legislation.gov.uk website is based on an open API which makes available the raw data behind it for other users. To see this underlying data, just append /data.xml or /data.rdf to the URL.

All data is available freely under the Open Government License[6].

Article by Denise Recheis (REEEP)

Links

(1) The Official Home of UK Legislation: http://www.legislation.gov.uk

(2) Office of Pulic Sector Information: http://en.wikipedia.org/wiki/Office_of_Public_Sector_Information

(3) Statute Law Database: http://en.wikipedia.org/wiki/UK_Statute_Law_Database

(4) Education for Sustainable Development Toolkit: http://www.esdtoolkit.org

(5) Mobile Legislate App: http://mobilelegislate.com

(6) Open Government Licence: http://www.nationalarchives.gov.uk/doc/open-government-licence

Further reading

- Linking UK Government Data (John Sheridan): http://www.slideshare.net/semwebcompany/linking-uk-government-data-john-sheridan

4.4. Solar Atlas for the Mediterranean: Socio-Economic Data Integration

Solar Atlas for the Mediterranean platform publishing and consuming LOD

Originally, the "Solar Atlas for the Mediterranean" (Solar-Med-Atlas) platform[1] was built to provide online resources for global horizontal and direct normal irradiance data for the southern and eastern Mediterranean region. The portal features a section on socio-economic country profiles for renewable energies and energy efficiency, consuming Linked Open Data that is published by the project partners. Currently, data is available for 11 countries, with topics ranging from energy consumption to energy production to average price of electricity and regulatory frameworks.

The Solar-Med-Atlas is available to all interested public in the field of solar potential in the region, however is geared towards solar energy project planners and implementers. It also serves as an information platform for policy makers and others working with solar and renewable energy. The socio-economic pages were included to provide another layer of information for these audiences, informing them not only about the solar potential, but also the country backgrounds.

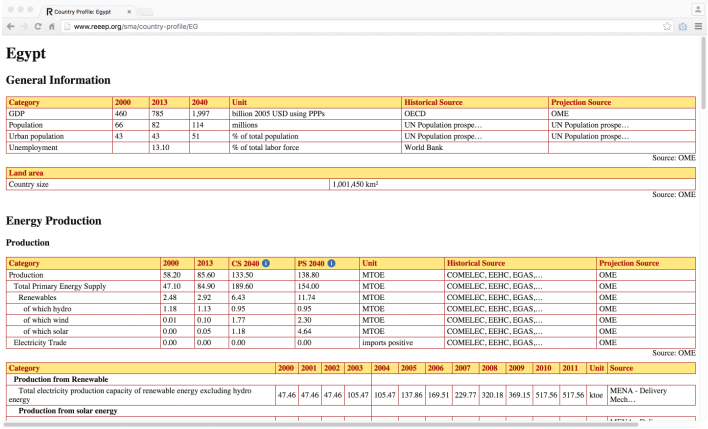
Due to technical restrictions, the data published within this project is hosted on data.reeep.org[2]. REEEPs online data store is used to host project partners' data, which is available for everyone to be reused under the Open Government Licence through the SPARQL endpoint using RDF.

The socio-economic part of the online platform then in turn consumes the LOD from data.reeep.org. The data is published as a website widget, which can also be implemented into other websites that do not have the possibility to consume LOD.

The Solar Med Atlas Project is led by the German Aerospace Center (DLR), in collaboration Geomodel, Transvalor, Armines, RCREEE, OME,

RENAC and REEEP. It is supported by the International Climate Initiative of the Germany Ministry of Environment, Nature Conservation and Nuclear Safety.

The Socio-Economic Country Profiles



Country Profile: Egypt

www.reeep.org/sma/country-profile/EG

Egypt

General Information

Category	2000	2013	2040	Unit	Historical Source	Projection Source
GDP	460	785	1,997	billion 2005 USD using PPPs	OECD	OME
Population	66	82	114	millions	UN Population prospe…	UN Population prospe…
Urban population	43	43	51	% of total population	UN Population prospe…	UN Population prospe…
Unemployment		13.10		% of total labor force	World Bank	

Source: OME

Land area	
Country size	1,001,450 km²

Source: OME

Energy Production

Production

Category	2000	2013	CS 2040	PS 2040	Unit	Historical Source	Projection Source
Production	58.20	85.60	133.50	138.80	MTOE	COMELEC, EEHC, EGAS,…	OME
Total Primary Energy Supply	47.10	84.90	189.60	154.00	MTOE	COMELEC, EEHC, EGAS,…	OME
Renewables	2.48	2.92	6.43	11.74	MTOE	COMELEC, EEHC, EGAS,…	OME
of which hydro	1.18	1.13	0.95	0.95	MTOE	COMELEC, EEHC, EGAS,…	OME
of which wind	0.01	0.10	1.77	2.30	MTOE	COMELEC, EEHC, EGAS,…	OME
of which solar	0.00	0.05	1.18	4.64	MTOE	COMELEC, EEHC, EGAS,…	OME
Electricity Trade	0.00	0.00	0.00	0.00	imports positive	COMELEC, EEHC, EGAS,…	OME

Source: OME

Category	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	Unit	Source
Production from Renewable														
Total electricity production capacity of renewable energy excluding hydro energy	47.46	47.46	47.46	105.47	105.47	137.86	169.51	229.77	320.18	369.15	517.56	517.56	ktoe	MENA - Delivery Mech…
Production from solar energy														

The data provided within the Socio-Economic data store is provided by two expert organizations, both focusing on renewable energies and energy efficiency in the Mediterranean and in the Arab region.

The data from both organizations was merged into one data store at REEEP, providing an enriched data set of historic data as well as projections and scenarios of future developments. Additionally, regulatory frameworks are included to give an overview of the countries' current situation of policies regarding renewable energy and energy efficiency efforts.

Links

(1) http://www.solar-med-atlas.org/

(2) http://data.reeep.org/

5. APPENDIX

5.1. Authors

Martin Kaltenböck, CMC

Author of Chapter 1 & 3

Martin Kaltenböck studied communication, psychology and marketing at the University of Vienna. In 2000, he co-founded punkt. netServices, an Austrian company specialising in information and knowledge management and Enterprise 2.0 solutions. He is Managing Partner and CFO at Semantic Web Company (SWC), where he is responsible for finance and operations. Furthermore, he leads numerous projects in national and international research, industry and public administration. His regular speaking engagements and publications cover the fields of Enterprise 2.0, social semantic web, Linked Open Data and Open Government Data. He is a certified management consultant, a member of the Executive Board of the Austrian Chapter of the Open Knowledge Foundation (OKFO) and an invited expert of OGD Austria, a governmental cooperation. He is also currently working as an invited expert at W3C.

Publications

- ZukunftsWebBuch 2010 - Chances and Risks of the Future Web
- Enterprise 2.0 - Introduction, Principles, Use Cases and Tools
- Open Government Data (OGD) White Book Austria 2011

Contact

Tel: +43 (0)1 402 12 35-25
E-Mail: m.kaltenboeck@semantic-web.at
Slideshare: http://www.slideshare.net/MartinKaltenboeck
LinkedIn: http://www.linkedin.com/in/martinkaltenboeck

Mag. Florian Bauer

Author of Chapter 4

Florian Bauer earned a Master's Degree in IT Management and a Bachelor's Degree in Software Engineering from the Vienna University of Technology. Since 2007, he has worked at the Renewable Energy and Energy Efficiency Partnership (REEEP), a non-profit, specialist change agent aiming to catalyse the market for renewable energy and energy efficiency, with a primary focus on emerging markets and developing countries. As Operations and IT Director at REEEP, Florian oversees the product management and strategic development of the clean energy information portal reegle.info which has emerged as a key access point for high quality energy information and data sets. Over the last few years, he has been focusing on transforming reegle into a key provider for LOD data sets in the field of clean energy. Prior to joining REEEP, Florian worked as a project manager and consultant for Siemens Austria, where he managed intercultural projects, and he also founded a web design firm in 2003.

Publications

- Environmental Software Systems. Frameworks of eEnvironment - data.reegle.info – A New Key Portal for Open Energy Data
- Information Technology Safety-Concepts in Europe

Contact

Tel: +43 (0)1 260 26 37 14
E-Mail: florian.bauer@reeep.org
Skype: reeep_f.bauer
LinkedIn: http://www.linkedin.com/in/florianbauer

Mag. Andreas Blumauer

Author of Chapter 2

Andreas Blumauer earned a Master's degree in business, specialising in business informatics, at the Vienna University of Economics and Business and at the Vienna University of Technology. He started his career in 1996 as software developer for financial services organisations. In 2000, he co-founded punkt. netServices, a specialised provider of Enterprise 2.0 solutions. Andreas is Managing Partner at Semantic Web Company (SWC), where he is responsible for various consulting projects as well as for the product management of PoolParty, a Linked Data management platform. He regularly lectures at universities throughout Austria in the fields of knowledge management systems, social software and semantic technologies. Andreas has been a pioneer in the area of the semantic web and Linked Data since 2002, and is co-editor and editor of one of the first comprehensive books in German about the semantic web.

Publications

- Social Semantic Web - Web 2.0, was nun?
- Using Linked Data in Thesaurus Management

Contact

Tel: +43 (0)1 402 12 35-27
E-Mail: a.blumauer@semantic-web.at
Slideshare: http://www.slideshare.net/ABLVienna
LinkedIn: http://at.linkedin.com/pub/andreas-blumauer/6/46/b

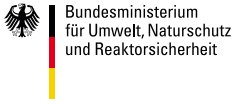
5.2. Attribution & Credits

Image credits and copyrights

The pictures used for this book were obtained from the photography libraries and individual photographers credited below.

- Ferrás – Corporate Design Wien (reegle best practice montage – p.44) – http://www.ferras.at
- Jer Thorp (flickr), Fotocredit Frontcover , CC BY – http://blprnt.com

Sponsoring Partners



Federal Ministry for the Environment, Nature Conservation and Nuclear Safety (BMU)

http://www.bmu.de

Renewable Energy & Energy Efficiency Partnership (REEEP)

http://www.reeep.org

Semantic Web Company (SWC), Vienna – Austria

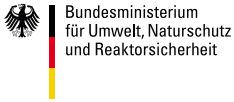
http://www.semantic-web.at

LOD2 – Creating Knowledge out of Interlinked Data

http://lod2.eu

Event Partners/Sponsors

This book was created during the preparation of the REEEP- and SWC-organised event, "Linking Open Data to Accelerate Low-Carbon Development," held in January 2012 in Abu Dhabi with the goal of providing an easy-to-understand guide on first steps steps in consuming and producing LOD. The event was kindly supported by:



Federal Ministry for the Environment, Nature Conservation and Nuclear Safety (BMU)

http://www.bmu.de

International Renewable Energy Agency (IRENA)

http://www.irena.org

National Renewable Energy Laboratory (NREL)

http://www.nrel.gov

Masdar Institute

http://www.masdar.ac.ae

Acknowledgments

This book would not have been possible without the help of all those who have contributed ideas, discussions, demos and texts. We would like to thank everyone who helped us (names listed in random order):

Martin Schöpe, Marianne Osterkorn, Martin Hiller, Anna Florowski, John Sheridan, José Manuel Alonso, Bitange Ndemo, Jon Weers, Thomas Thurner, Denise Recheis, Günther Friesinger , Vince Reardon, Jena Wu

This is a quick start guide for decision makers who need to quickly get up to speed with the Linked Open Data (LOD) concept, and who want to make their organization a part of this movement.

It gives a quick overview of all key aspects of LOD, and gives practical answers to many pertinent questions including:

- What do the terms Open Data, Open Government Data and Linked Open Data actually mean, and what are the differences between them?
- What do I need to take into account in developing a LOD strategy for my organization?
- What does my organization need to do technically in order to open up and publish its data sets?
- How can I make sure the data is accessible and digestible for others?
- How can I add value to my own data sets by consuming LOD from other sources?
- What can be learned from four case studies of best practices in LOD?
    - REEEP's clean energy information portal reegle.info
    - NREL's Open Energy Information Portal
    - The official home of UK legislation: legislation.gov.uk
    - Socio-Economic information in the Solar Atlas for the Mediterranean
- What are the potentials offered by this fundamental step-change in the way data is shared and consumed via the web?

REEEP®

SEMANTIC WEB COMPANY
linking data to knowledge

edition mono

ISBN: 978-3-902796-05-9