LINKED OPEN DATA: THE ESSENTIALS

THE CLIMATE KNOWLEDGE BROKERING EDITION

Florian Bauer, Martin Kaltenböck

Linked Open Data: The Essentials

The Climate Knowledge Brokering Edition

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

2nd edition 2016

Imprint

Published by:

edition mono/monochrom
Zentagasse 31/8, 1050 Wien
fax/fon: +431/952 33 84
edition-mono@monochrom.at

Publication:

1st edition (2011), 2nd edition (2016)

ISBN:

978-3-902796-54-7

Cover design & Layout:

Susan Härtig (Semantic Web Company)

Production Editor:

Thomas Thurner (Semantic Web Company)

Proofreading:

John Tkacik (REEEP), Maria van Veldhuizen (REEEP)

Print:

DGS – Druck & Grafik, Vienna, Austria

PDF Version:

PDF Version: A PDF version of Linked Open Data: The Climate Knowledge Brokering Edition for download is available here: http://www.semantic-web.at/LOD-TheEssentials2016.pdf

Introductory Remarks

Editorial Notes by Geoff Barnard (CDKN), Martin Schöpe (BMU), Martin Hiller (REEEP) and Martin Kaltenböck (SWC)

Geoff Barnard

Knowledge Management Strategy Advisor, Climate and Development Knowledge Network (CDKN)

With the growing demand for reliable information on climate issues, and the urgency of the climate challenges facing us, efficient data and information sharing is surely a prerequisite for effective action. The problem is not the lack of reports, briefings and datasets out there; it's the overabundance of them. How do you find and access what's relevant to you? And how do you make intelligent use of it once you have?

Linked open data principles provide the key to solving this conundrum, and this book provides a much-needed practical guide on how to apply them. It will be especially valuable for the new generation of climate knowledge brokers that are emerging to provide links between producers and users of climate knowledge, and to connect people to the information and data they need.

REEEP has been championing linked open data principles in the climate and renewables sectors for over a decade. CDKN has been supporting its work since 2011, when the Climate Knowledge Brokers (CKB) Group was formed. With more and more online information sources becoming available, a trend we dubbed "portal proliferation syndrome", there was an urgent need for both organising principles and practical tools to navigate that growing ocean of climate information.

Working with partners from the CKB Group, REEEP has made real progress on both fronts, culminating in the launch of the "Climate Tagger" in 2015. Built on a smart multi-lingual thesaurus of climate- related terms, it allows content to be automatically "tagged" with relevant key words, making it far easier to find. In the process it turns "open data" into "linked open data", and unlocks a whole new range of possibilities for intelligent sharing and reuse.

Linked Open Data: The Essentials is a "must read" for anyone thinking of setting up a new climate portal or knowledge platform. If you want to make the most of your investment in climate knowledge, start here!

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.

(taken from 2012 first edition)

Martin Hiller

Director General, REEEP – Investing in clean energy markets

In theory, 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.

In practice, the challenge is how to sort through and effectively use the ever-increasing volume of information available. Linked Open Data (LOD) points to one possible solution: the LOD movement calls on organizations 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.

Linked Open Data: The Essentials was developed to give decision makers a quick overview of the LOD concept and how to engage with the process in their respective organizations. In response to the success of the first edition we at REEEP are proud to be able to present this new "Knowledge Brokering Edition".

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. KNOWLEDGE BROKERING

Linked Open Data as an important enabler for successful climate knowledge brokering

What does it mean to be a "Climate Knowledge Broker"

Human beings face, in climate change, an unprecedented challenge; one which dramatically impacts the most fundamental systems we have created for our safety and survival – such as agriculture, water and energy – as well as the ecosystems in which we live and our quality of life. Our decision making will become increasingly climate constrained; many of these decisions will pertain to areas never before considered to be climate-dependent. With this unprecedented challenge is an unprecedented need for relevant and tailored information to inform these decisions.

At one time, the field of climate information might have been considered the exclusive domain of climate scientists, populated by inscrutable models and projections. Now, this domain is expanding rapidly, as enormous volumes of climate-relevant information and knowledge are produced to meet demand across different sectors of society. This information is used in ways never before imagined, by people often never before considered part of the audience.

Although awareness of climate change has increased steadily over recent years, the field of climate information in many ways lags behind, and is certainly nowhere near fulfilling current, much less expected demands. The existing climate knowledge system will not be enough to supporting the climate mitigation and adaptation actions necessary for our societies to survive and thrive.

In large part, the problem revolves around the volume and availability of, and access to, climate information. Many climate information users

are deluged by the enormous number of reports, scenarios, datasets, tool-kits and other information products released each year; consequently, they are too often unable to find the information they need. Others are left out by massive informational gaps – typically related to language or regional applicability. Still others are not yet aware of the role climate will have in their analyses, planning and decision making. This is where *climate knowledge brokers* enter the picture as demand-led intermediaries, translators and advocates for climate-conscious decision making. Climate knowledge brokers work between information producers and people who use the information, screening and evaluating the flood of resources to provide tailored knowledge at crucial times.

Climate knowledge brokers come in all shapes and sizes – they can include radio moderators, teachers, leaders at farming co-ops, weather presenters, sellers of seeds, assistants to managers and politicians. They have an important role to inform climate relevant decisions based on best practice and quality data, and their relevance will only increase in the years ahead.

Leading climate knowledge brokers commit to open data

The Climate Knowledge Brokers[1] (CKB) Group is a growing alliance of leading global, regional and national knowledge brokers specializing in climate and development information. It brings together a diverse set of information players, from international organizations to research institutes, NGOs and good practice networks, and covers the full breadth of climate related themes. CKB uses the terms "climate information" and "climate knowledge" as short-hand to encompass information and knowledge that is not only about the climate, but also about systems and human activities that have impact on, or are affected by, climate change.

As an agreed standard on how to provide this best available climate knowledge to those that need it, the CKB group has adopted seven principles[2] .

One of those principles clearly states the commitment to (Linked) Open Data: *"We promote open knowledge; meaning we have an open mindset, are actively seeking to share our knowledge and want to work with others who have the same attitude. We pledge to apply open knowledge, sharing and collaboration as standards in our work".*

Open Data and its relevance for successful climate knowledge brokering

For some consumers of climate information, a synthesis of information will be sufficient to make a decision. Others will need to deepen their understanding of issues that are particularly important. They may feel the information offered only partially covers their circumstances and wish to investigate further, or perform different analyses themselves. In addition, a sophisticated information consumer will not take everything they hear or read at face value. When offered information based on someone else's interpretation or analysis, they will want to check its accuracy and relevance to their own situation. In order to follow up, they need access to the data and information behind what is presented. To be able to judge whether they trust the information they need to know where it came from. This is particularly true for the knowledge broker, who must tailor information to varied needs of different target audiences – a synthesis prepared for one group of stakeholders will not fulfil all needs of another group. The unrestricted ability to explore original, primary and timely data allows the successful knowledge broker to deliver a purposeful knowledge product designed for the precise needs of a decision maker. Examples of decision makers in need of climate information include policy makers, town planners, disaster response teams, investors, producers of agricultural goods, travel agents and many more. Of course, not all these users of climate information can be served with the same product, and to create an adaptive and

dynamic portfolio of information offers, knowledge brokers require reliable access to all relevant data.

Not all climate-related information is currently openly available. In some cases, this is for commercial reasons: the business model of the information provider may depend upon selling information. This poses a problem for users who need the information but cannot afford it. At the same time, if the company does not get paid, the information will not be collected and produced in the first place.

In other cases, where climate data and information are produced using public funds, it is harder to defend the reasoning for a denial of access.

Even when information is made available, adequate references are not always supplied, sources are not always quoted and access to raw data is often not consistently provided (for instance, we were told that national meteorological institutes often cannot or simply do not share their raw data). When raw data is available, it is often not in an open data format that would make it easy for users to perform analyses themselves. This is where CKB, and other advocates of Open Data, see tremendous potential, and it is the reason why we are dedicated to increasing and accelerating the opening-up of climate data.

Links

(1) Climate Knowledge Brokers – www.climateknowledgebrokers.net/

(2) The 7 CKB Principles – www.climateknowledgebrokers.net/what-is-ckb/the-7-ckb-principles/

Further reading

- Climate Change Knowledge Portal (CCKP), The World Bank – http://sdwebx.worldbank.org/climateportal/
- CKB Manifesto – manifesto.climateknowledgebrokers.net/

Quiz – Chapter 1

Q1: Which of these statements describe parts of the role of a climate knowledge broker? (multiple answers possible)

- [ ] To buy and sell data about climate change
- [ ] To produce the information on climate change decision-makers need
- [ ] To screen existing information on climate change and adapt it to user needs
- [ ] To make decision-makers aware of the fact that they should consider climate change

Q 2: True or false:

- [ ] Only people with special training can call themselves climate knowledge brokers
- [ ] The main reason why climate change information is not used enough in decision-making is that not enough information is produced
- [ ] In order to be a successful climate knowledge broker, collaboration and sharing are key
- [ ] If detailed interpretations and syntheses of climate data are freely available, it is still helpful to make the raw data publicly accessible as well
- [ ] Climate knowledge brokers can have many different official job titles – from gardeners to politicians

Q 3: Why must a climate knowledge broker tailor his information for different target audiences? (multiple answers possible)

- [ ] This is his business model – he needs to sell different "dossiers"

- ☐ Recipients in need of climate information can include many different types of people – out of one and the same data set, they may need very different parts for their decision making
- ☐ Once a climate knowledge broker has created a perfect synthesis out of raw data, he can use this for all purposes
- ☐ Since the climate knowledge broker acts as a "translator", his role is to present facts in a way that they make sense to a particular group of stakeholders

Q 4: In what ways (if at all) is Open Data useful for climate knowledge brokering? (multiple answers possible)

- ☐ It allows the climate knowledge broker to verify claims by checking the actual facts (raw data, source)
- ☐ It gives the climate knowledge broker a chance to combine information from different sources to present a "fresh" view on a certain topic
- ☐ In fact, it doesn't really matter if data is open or not – a good climate knowledge broker will use what he can find on the internet in whatever way he needs it
- ☐ Only access to all relevant data allows a climate knowledge broker to deepen her understanding on issues that are particularly important
- ☐ Information based on someone else's interpretation or analysis can be taken and re-interpreted by the climate knowledge broker – Open Data doesn't really come into this

Answers:

Q1: 1 (F) 2 (F) 3 (T) 4 (T)
Q2: 1 (F) 2 (F) 3 (T) 4 (T) 5 (T)
Q3: 1 (F) 2 (T) 4 (F) 4 (T)
Q4: 1 (T) 2 (T) 3 (F) 4 (T) 5 (F)

2. 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 in 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), which was 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. Since 2011 62 additional governments have joined the Open Government Partnership.

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. Its focus is only on data and information produced or commissioned by government or government-controlled entities and is not individual-related.

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 produced not by just governmental institutions and includes those from, but also by other relevant stakeholder groups such as business/industry, citizens, NPOs and NGOs, science or education.

Some of the best-known institutions currently undertaking open data activities include the World Bank[4], the United Nations[5], REEEP[6], the New York Times[7], The Guardian[8], Open Knowledge International (formerly OKFN)[9] and the Open Data Institute (ODI)[10].

In 2007, 30 Open Government advocates came together in Sebastopol, California, USA to develop a set of OGD principles[11] that underscored why OGD is essential for democracy. In 2010, the Sunlight Foundation[12] 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 for open data.

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

1. **Data Must Be 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[13] 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 Kroes, former Vice-President of the European Commission responsible for the Digital Agenda, stated a strong commitment to OGD through her an-

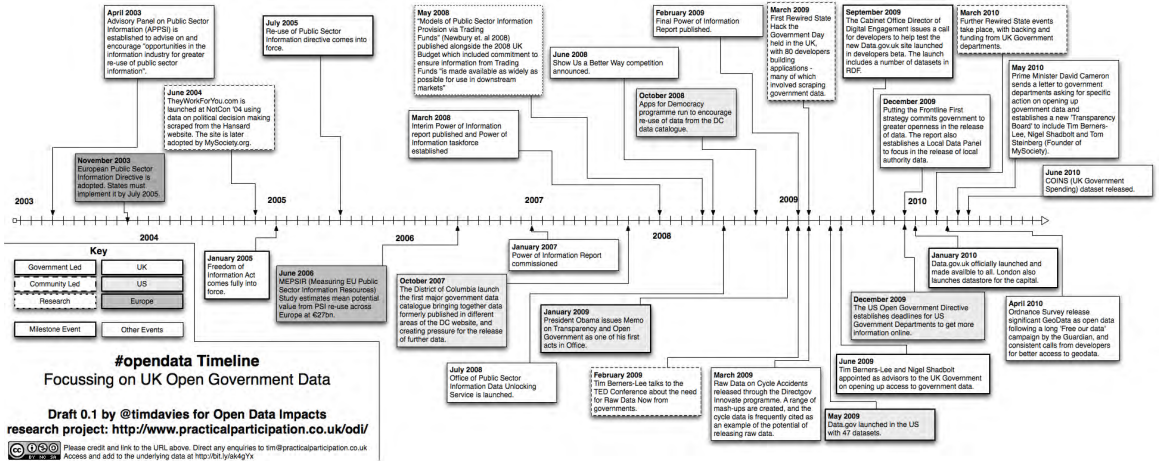
nouncements of an EC data portal by early 2012, which by 2016 provided nearly 10,000 datasets[14] and a Pan-European data portal acting as a single point of access for all European national data portals, which was launched in November 2015 and provided 600,000 datasets in 2016[15]. Open Data is an important part of both the Digital Agenda for Europe[16] and the European e-government Action Plan 2011 - 2015[17]. In December 2011 the EC furthermore announced its Open Data Strategy for Europe: Turning Government Data into Gold[18].

The earliest adopters of national open data activities and initiatives were definitely the governments of the United States of America[19], Australia[20], the Scandinavian countries and the UK government[21]. All of these countries have demonstrated strong 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: data.reeep.org

(7) Open Data New York Times: http://www.nytimes.com/column/raw-data

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

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

(10) The Open Data Institute: http://theodi.org/

(11) 8 Principles of Open Government Data: https://opengovdata.org/ or https://public.resource.org/8_principles.html

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

(13) EC Open Data Portal: https://data.europa.eu/euodp/en/data

(14) European Open Data Portal: https://www.europeandataportal.eu/

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

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

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

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

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

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

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

Further Reading

- Open Government, Wikipedia: http://en.wikipedia.org/wiki/Open_government
- Open Knowledge International, OGD website: http://opengovernmentdata.org/
- Open Data, Wikipedia: http://en.wikipedia.org/wiki/Open_data
- Open Knowledge Foundation Blog: http://blog.okfn.org/
- The Open Data Institute: http://theodi.org/
- European Open Data Portal: https://www.europeandataportal.eu/
- The Open Data Index – tracking the state of government open data: http://index.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 consuming 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, visualize 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 optimize access to these items (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.

Many thanks to Andy Seaborne for pointing out the CSV bug, to Kerstin Forsberg for suggesting "data highlighting" in the 4 and 5 star examples, and to Vassilios Peristeras for proposing that we explain not only the "what" but also the "why."

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 organizations, products and services to create portals, platforms, internet-based services and applications.

LOD is domain-independent and penetrates various areas and domains, which gives it 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 principles for the consumption and publication of LOD to create powerful and innovative services for knowledge management, decision making and general data management. Best practice examples Climate Tagger[6] and OpenEI[7] showcase a tool and a platform utilizing LOD. Another popular example of applied Open Data in the energy and climate context is the SolarMedAtlas[8].

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://5stardata.info/en/

(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) Climate Tagger – www.climatetagger.net

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

(8) SolarmedAtlas – www.solar-med-atlas.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
- BabelNet, a semantic network of lexicographical and language data: http://babelnet.org/
- Linked Open Data -- Creating Knowledge Out of Interlinked Data, free e-book: http://link.springer.com/book/10.1007%2F978-3-319-09846-3

Quiz – Chapter 2

Q1: Open Government Data is thought to enable: Tick what applies (multiple answers possible)

- ☐ Transparency
- ☐ Democracy
- ☐ Participation
- ☐ Collaboration

Q2: Some organization are spearheading Open Data – Please name a few:

.....

Q3: What is considered "OGD" – the 8 original principles: Tick what applies (multiple answers possible)

- [ ] Data Must Be Complete
- [ ] Data Must Be Primary
- [ ] Data Must Be Timely
- [ ] Data Must Be Accessible
- [ ] Data Must Be Machine-Processable
- [ ] Access Must Be Non-Discriminatory
- [ ] Data Formats Must Be Non-Proprietary
- [ ] Data Must Be License-free
- [ ] Data must be Peer-reviewed
- [ ] Data must be Unique
- [ ] Data must be Aggregated

Q4: Under the 5 Stars Model, data that is available under an open license, structured in a non-proprietary format would be considered: (tick what applies)

- [ ] 2star data
- [ ] 3star data
- [ ] 4star data

Answers:
Q1: 1 (T) 2 (T) 3 (T) 4 (T)
Q2: WorldBank, the United Nations, REEEP, the New York Times, The Guardian and the Open Knowledge Foundation
Q3: 1 (T) 2 (T) 3 (T) 4 (T) 5 (T) 6 (T) 7 (T) 8 (T) 9 (F) 10 (F) 11 (F)
Q4: 1 (F) 2 (T) 3 (F)

3. 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 correlations between 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 to integrate information from different sources, otherwise known as mashing data, are often too time-consuming and too costly.

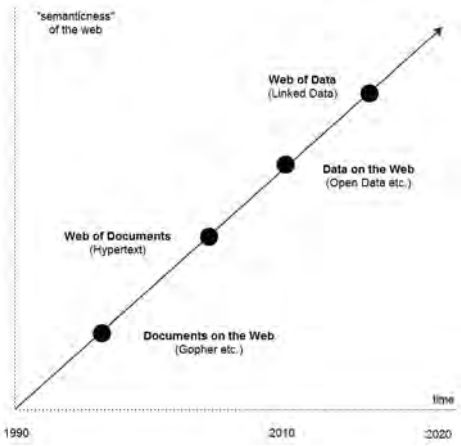
Two driving factors can cause this unpleasant situation:

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 is 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 specific application to be built on top. If we stop emphasizing 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. Open data means in its core that the data is open to any kind of application and this can be achieved if we use open standards like RDF[1] 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 recognized as mainstream (like the web we all know today), there are a lot of LOD already available. The so called LOD cloud[2] has exceeded 100 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 the most up-to-date version of the LOD Cloud diagram of 2014 below:

Why should we link data on the web and how can 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 they are and how they will make use of 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 posts: they are like emails 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 shared it on the web and now you receive lots of spam.

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 between systems, standards play the most crucial role. Either the transmitter or the receiver has to convert or map their 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: 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-harmonized internet where all of those layers are based on 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 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 of us, if not all. 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

In correspondence to the three points above, here are the steps already taken 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[3],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 developed 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, at probably the lowest cost 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, information is still being hidden away.

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 re-used and linked to another resource. Re-usage can only be encouraged by providing information about available information. In order to undertake this task in a sustainable manner, information must be recognized as an important resource that should be managed like any other resource.

Examples for LOD applications

Linked open data is widely available in several industries, such as in the following four examples:

- Linked data in libraries[4]: focusing on library data exchange and the potential for creating globally interlinked library data; exchanging and jointly utilizing 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[5]: 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[6]: re-using public sector information (along the PSI Directive[7]); improving internal administrative processes by integrating data based on linked data; and interlinking government and non-government information.
- Cultural Heritage linked data[8]: numerous EU cultural heritage institutions (GLAM sector including but not limited to: galleries, libraries, archives and museums) share and interlink metadata and data via the Europeana network.

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 GO & NGO sector, more and more media companies are jumping on the bandwagon. Their assumption is that ever more industries will perceive linked data as a cost-efficient way to integrate data.

Linking information from different sources also plays an important role in open innovation processes[9]. If data can be put in a new context, more and more valuable applications – and therefore knowledge – will be generated.

Links

(1) http://www.w3.org/RDF/

(2) http://www.lod-cloud.net

(3) http://dbpedia.org/

(4) Jan Hannemann, Jürgen Kett (German National Library): "Linked Data for Libraries" (2010) http://www.ifla.org/past-wlic/2010/149-hannemann-en.pdf

(5) http://obofoundry.org/

(6) https://www.europeandataportal.eu/

(7) https://en.wikipedia.org/wiki/Directive_on_the_re-use_of_public_sector_information

(8) http://labs.europeana.eu/api/linked-open-data-introduction

(9) https://en.wikipedia.org/wiki/Open_innovation

Quiz – Chapter 3

Q1: Bring into right order the evolution of "semanticness" of the web:

- [ ] Data on the web
- [ ] Web of documents
- [ ] Web of data
- [ ] Documents on the web

Answers:

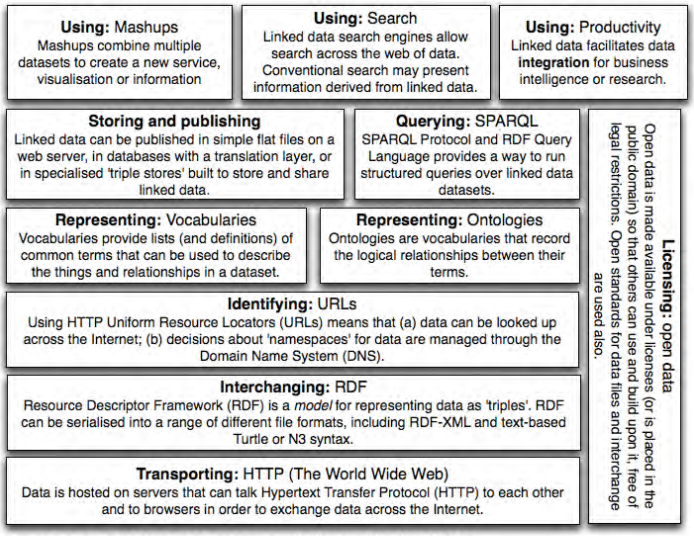
Q1: 1 (3) 2 (2) 3 (4) 4 (1)

4. 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 organization. We also provide recommendations for further reading 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 gives a technical overview of the necessary building blocks for 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

4.1. Publishing Linked Open Data

First steps for publishing your content as LOD

The ideas behind, benefits of and effort required for publishing LOD have already been discussed in the previous chapters of this publication following the 5 Star Model of OGD. Publishing LOD 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 will help your organization to become an important data hub within your domain.

Quick guide for publishing LOD

We have prepared a short guide to 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.

Analyze 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 cleaning your data and information by removing 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 changes 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 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 an often-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 officially recommended by W3C for semantic web data models. Remember to include your specified license(s) in 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 optimized data processing and integration for data (re-)use and allow for the creation of new knowledge from your data sets by putting them into new context with other data. Evaluate and carefully choose the most relevant data sets to be linked with your own data sets.

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 pro-

viding 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 publish 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] visualizes 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 whether you are using HTML or RDF, the same recommendations 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 information about the 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:

- Official UK Legislation: http://www.legislation.gov.uk/
- REEEP Climate Tagger API: http://www.climatetagger.net/climate-tagger-api/
- EU project: LATC – LOD around the clock: http://latc-project.eu/
- Global Buildings Performance Network (GBPN): http://www.gbpn.org/databases-tools/mrv-tool/about
- Statistical Data published as linked open data: http://stats.lod2.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/home/bernadette-hyland.html

(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://wifo5-03.informatik.uni-mannheim.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.cabinet-office.gov.uk/resource-library/designing-uri-sets-uk-public-sector
- Linked Data Patterns, Dodds & Davies: http://patterns.dataincu-bator.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
- Collection of Linked Open Vocabularies: https://lov.okfn.org/dataset/lov/
- Basel Register of Thesauri, Ontologies & Classifications: http://bartoc.org/

Quiz – Chapter 4.1

Q1: What is RDF … Tick what applies (multiple answers possible)

- [ ] Resource Description Framework
- [ ] Official W3C recommendation for semantic web data models
- [ ] Name of a LOD vocabulary
- [ ] Ratified Data Framework
- [ ] Powerful data model for LOD

Q2: What is true/false in regard to licenses … Tick what applies (multiple answers possible)

- [ ] Open Data doesn't need a license
- [ ] An open license such as creative commons should be part of metadata
- [ ] A license is important so your data cannot be misused
- [ ] All open licenses are the same

Q3: When publishing Open Data … Tick what applies (multiple answers possible)

- [ ] I can charge a small fee
- [ ] Data must be reviewable (contact person indicated)
- [ ] Data will be available over time (permanence)
- [ ] I must create a custom-built vocabulary

Answers:
Q1: 1 (T) 2 (T) 3 (F) 4 (F) 5 (T)
Q2: 1 (F) 2 (T) 3 (F) 4 (F)
Q3: 1 (F) 2 (T) 3 (T) 4 (F)

4.2. 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.

Organizations can benefit and reach competitive advantage through the possibility to: 1) spontaneously generate dossiers and information mash ups from distributed information sources; 2) 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 organizations usually provide high quality data and information). A very good approach for this evaluation is to use one of the globally available open data catalogues such as The Data Hub[1]. 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 other data sets.

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 to match different vocabularies of the consumed (internal and external) data sets often occurs. This is relevant to ensure smooth data integration through vocabulary alignment[2]. Another concern is the fact that LOD sources are not absolutely stable and 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 use in return.

To conclude, please consider some best practice examples for consuming LOD from these LOD players:

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

Links

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

(2) 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/
- 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
- Collection of Linked Open Vocabularies: https://lov.okfn.org/dataset/lov/
- Basel Register of Thesauri, Ontologies & Classifications: http://bartoc.org/

Linked Open Data Tool Box Collection

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

- **PoolParty Semantic Suite:** http://www.poolparty.biz
  Services and tools for LOD-based metadata Management, enterprise search, text mining and data integration
- **Unified Views:** https://unifiedviews.eu/
  Data Integration and Linked Data Warehouse Framework
- **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
- **Stardog:** http://stardog.com/
  Enterprise data unification platform built on smart graph technology

- **GraphDB:** http://graphdb.ontotext.com/graphdb/
  Enterprise ready Semantic Graph Database
- **Callimachus Project:** http://callimachusproject.org/
  A framework for data-driven applications using linked data

Quiz – Chapter 4.2

Q1: How can organizations benefit from consuming Linked Open Data? Tick what applies (multiple answers possible)

- ☐ Spontaneously generate dossiers and information mash ups from distributed information sources
- ☐ Create applications based on real time data
- ☐ Create new knowledge out of this interlinked data
- ☐ When no clear license is available, they can use it to beef up their own datasets without the need for attributions
- ☐ Put the raw data on their websites as is and draw some of the web traffic away from the original source

Q2: When re-using data, I have to … Tick what applies (multiple answers possible)

- ☐ Credit it to the sources
- ☐ Ensure all information is correct
- ☐ Log-in to the provider's database

Answers:
Q1: 1 (T) 2 (T) 3 (T) 4 (F) 5 (F)
Q2: 1 (T) 3 (F) 3 (F)

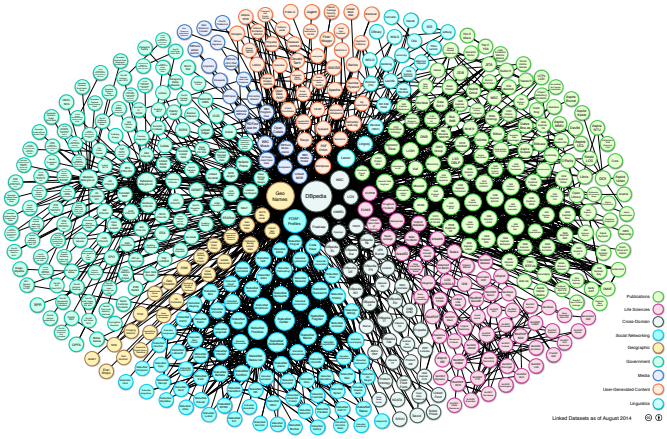
5. BEST PRACTICE AND EXAMPLES

Five 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 it now includes data sets from diverse organizations and data providers including media (e.g. BBC, NYT), governments (e.g. US, UK) as well as user-generated content. To visualize key LOD providers and their linkages, Richard Cyganiak (DERI) and Anja Jentzsch (Freie Universität Berlin) regularly compile a diagram of 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 detailed statistical breakdown was published in 2014.[4]

* 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 – August 2014"

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" [5], a community effort to extract structured information from Wikipedia and "geonames" [6], which provides RDF descriptions of millions of geographical locations worldwide.

The next chapters highlight five 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:

- Climate Tagger – a suite of software tools to help organizations in the climate and development fields streamline and catalogue their data and information resources
- openEI – a portal providing various energy data sets in a semantic wiki
- Enipedia – an active exploration into the applications of wikis and the semantic web for energy and industry issues
- GBPN's Tool for Building Energy Performance Scenarios
- Development Initiatives – Joined-up Data Standards project

All five best practices are part of the LOD cloud diagram shown above, and their data sets are interlinked with other data providers to maximize the benefits of using LOD technology. More examples can be found on in the CKAN[7] 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) State of the LOD Cloud 2014 http://linkeddatacatalog.dws.informatik.uni-mannheim.de/state/

(5) DBpedia: http://wiki.dbpedia.org/about

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

(7) 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
- Adoption of the Linked Data Best Practices in Different Topical Domains, Max Schmachtenberg, Christian Bizer, and Heiko Paulheim, ISWC2014 paper http://dws.informatik.uni-mannheim.de/fileadmin/lehrstuehle/ki/pub/SchmachtenbergBizerPaulheim-AdoptionOfLinkedDataBestPractices.pdf

5.1. Climate Tagger

streamlining and cataloguing data and information resources

climatetagger

A tool for analyzing online content based on an LOD thesaurus

REEEP's Climate Tagger[1] is a suite of tools to help knowledge-driven organizations in the climate and development arenas streamline and catalogue their data and information resources, and connect them to the wider climate knowledge community. Climate Tagger utilizes Linked Open Data, and is based on the tried-and-true reegle Tagging API (now Climate Tagger API). It was first introduced by REEEP in 2011 to help its network better catalogue and connect data, and is backed by the expansive Climate Smart Thesaurus, which was developed in collaboration with domain experts.

The extensive vocabulary organized in the Climate Smart Thesaurus was compiled and further developed over a number of years, and is still regularly extended to include an ever-growing wealth of terms relevant to the climate and clean energy sectors. From the beginning, the thesaurus was built and maintained as fully Linked Open Data, using the software PoolParty[2], and formatted as semantic standard using the Simple Knowledge Organization System (SKOS)[3]. The thesaurus provides links both to and from other relevant LOD thesauri and its entire content can be queried via a SPARQL endpoint[4].

The thesaurus is currently maintained in five different languages, and each term ("concept") has a permanent web address – "Uniform Resource Identifier" (URI) – so it can be referenced and linked to other LOD sources. Currently thesaurus concepts are linked at concept level to DBPedia[5], GEMET[6] and the FAO's AgroVoc[7] thesaurus. This means that definitions, additional translations or synonyms from such third parties are readily available in real time.

The Climate Tagger API – the machine-to-machine interface for high quality tagging

Climate Tagger API[8] utilizes the Climate Smart Thesaurus to analyze any textual information published online by comparing its content to all thesaurus concepts. An algorithm is used to attach a relevance score to extracted concepts and return this information back to the user. The extracted keywords are connected to additional information from the thesaurus, allowing Climate Tagger not only to tag content, but also to enrich a web page with suitable definitions.

Web developers can request an API key and specify the exact information they want to have returned for their content. Using the API allows freedom in terms of use cases, whether Climate Tagger is used for tagging resources, placing them into pre-defined categories or requesting translations of keywords to cross-reference multi-lingual resources is entirely up to the user.

Another advantage is the possibility of setting a timer for automated updates to make sure any web content is always tagged with Climate Tagger's latest version. In 2016, the whole section on renewable energy was overhauled in collaboration with IRENA, and a comparative analysis between REEEP, the Caribbean Community Climate Change Centre (CCCCC) and the Secretariat of the Pacific Regional Environment Programme (SPREP) has further increased the depth of thesaurus content overall.

To make integration of Climate Tagger even easier, Climate Tagger is also available for standard use case applications as a plugin for Drupal and WordPress. Using Climate Tagger for an online platform is free of charge.

Some use cases for Climate Tagger

Case 1 – To organise a previously unsearchable document database

Your organization has amassed a vast, valuable storehouse of electronic documents. Unfortunately, they are not particularly well-ordered or categorized. They also lack tagging for online searchability, so visitors to your own website have no effective way of finding and accessing the wealth of information that is already there.

Using Climate Tagger, there is fast, easy and reliable way of making your storehouse of documents searchable – and findable! Climate Tagger can tag all of your reports, articles and scientific papers instantly. Because the system relies on a thesaurus with clearly defined terms and interrelationships, it will automatically use a consistent set of keywords to describe content. This means you can place documents in useful clusters, make them all accessible to others and increase the impact of your work.

Several organizations in the clean energy field are already using Climate Tagger to improve document tagging and more are looking at the best way to integrate it into their systems at the moment.

Case 2 – To cross-link your own resources

Your web portal offers a vast amount of online resources. They're actually organized and categorized quite well, but no two documents are cross-linked with each other.

Climate Tagger will instantly index all of your resources and can help create a simple application to automatically suggest related content from your own pool of documents.

For example, OpenEI has already built a widget to suggest related articles to their readers, helping them to find what they're looking for.

Case 3 – To offer definitions, synonyms and links

You work hard to offer your users up-to-date relevant information – but at the same time you know that misunderstandings are not unusual when trying to describe a complex topic.

By integrating Climate Tagger into your publishing environment, you can offer your users a simple system that extracts particular topics from your articles and offers definitions, synonyms and links to open resources such as Wikipedia.

For instance, the Eldis development knowledge platform is using this system to offer their users a "what do we mean by" info box that includes definitions and synonyms retrieved from Climate Tagger.

Case 4 – Multiple language tagging

Your website is offered in multiple languages. To accurately tag your resources, you would need to hire multiple native-speakers with expertise in clean energy or climate change topics – a time-consuming and costly undertaking!

Climate Tagger can automatically extract quality-assured keywords from original documents in five languages: English, Spanish, French, Portuguese and German.

You can connect related resources within and across languages, and offer your users an intuitive structure for retrieving documents. Climate Tagger helps you by semantically scanning your resources and extracting the most relevant keywords and geographic location.

Climate Tagger was developed by REEEP in collaboration with the US National Renewable Energy Laboratory's Open Energy Information program, the Stockholm Environment Institute's weADAPT program and the Institute for Development Studies' Eldis program, as part of the Climate Knowledge Brokers Group.

Climate Tagger is made possible thanks to the generous financial support of a number of donors, including the Climate and Development Knowledge Network, the Federal Government of Germany, IRENA and the Climate Technology Centre and Network of the United Nations Environment Programme.

Links:

(1) Climate Tagger website: http://www.climatetagger.net/

(2) PoolParty Thesaurus & Taxonomy Management Software https://www.poolparty.biz/taxonomy-thesaurus-management/

(3) W3C – SKOS Simple Knowledge Organization System: https://www.w3.org/2004/02/skos/

(4) W3C – SPARQL endpoints https://www.w3.org/wiki/SparqlEndpoints

(5) DBpedia (LOD version of Wikipedia) http://wiki.dbpedia.org/

(6) GEMET Thesaurus https://www.eionet.europa.eu/gemet/

(7) FAO's AGROVOC Thesaurus http://aims.fao.org/vest-registry/vocabularies/agrovoc-multilingual-agricultural-thesaurus

(8) Climate Tagger API: http://www.climatetagger.net/climate-tagger-api/

Further reading:

- Wikipedia – Application Programming Interface: https://en.wikipedia.org/wiki/Application_programming_interface
- Controlled vocabularies: http://www.controlledvocabulary.com/
- W3C – Resource Description Framework (RDF): https://www.w3.org/RDF/

5.2. OpenEI Definitions

OpenEI publishing and consuming open energy information

OpenEI, short for open energy information, is a collaborative knowledge-sharing platform[1] with free and open access to energy-related data, models, tools, and information. OpenEI features over 60,000 content pages, more than 1,500 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
Open Energy Information

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 increases the utility of the information, making it more accessible and useful to others.

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 and 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 inquires 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:

- http://en.openei.org/lod/

5.3. Enipedia

Opportunities and Challenges in using Linked Data for Power Plants

Enipedia[1] is a project that uses a Semantic Wiki and Linked Data principles to help organize and navigate information about power plants all over the world. The use of Linked Data is a natural fit as questions about the power sector often span different networks of facts. As an example, Enipedia contains information on both companies and power plants related to:

- Companies can have subsidiaries which are other companies (and these companies may have subsidiaries as well)
- A power plant is owned by a company, has a generating capacity, a fuel type, coordinates, is located within a country, etc.

Given these connections, using SPARQL queries, we can generate different views such as:

- What is the total generating capacity of a company and all of its subsidiaries throughout the whole ownership chain?
- How much power is generated using different fuel types per country?
- In which countries does a company own which types of power plants?

Multiple Versions of the "Truth"

While Linked Data enables the generation of generating diverse conclusions from the same data, a major challenge in working with power sector data is that there is not one single version of the truth. Finding the definitive truth often requires combining information from multiple sources describing a plant's location, generating capacity, and emissions. The various data sources consulted do not always agree, which can occur for several reasons.

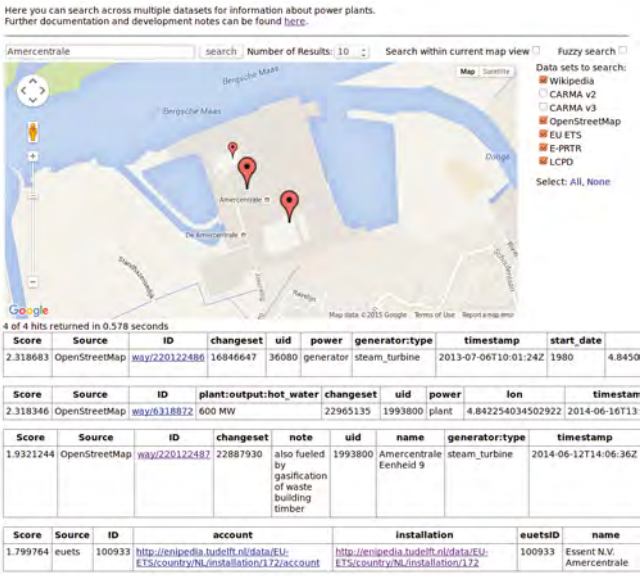
First, much of the data governments possess on the power sector is collected to comply with regulation driven reporting obligations, and is not necessarily collected with the aim to provide researchers with a broad, easy to process overview of the power sector. It is not uncommon to find that the data lacks unique identifiers and that the exact names used for installations change over time. Furthermore, different data sets may offer different levels of detail (i.e. data for individual generators vs. aggregated data for the overall power plant). The data may also be sensitive to changes over time as plants may be refurbished, expanded, or shut down.

Locating Entities Described in Multiple Heterogeneous Data Sets

To deal with this, we need a system that can facilitate matching and disambiguation. For this, we have created an API based on Elasticsearch[2] which allows users to search data that is not necessarily well structured. We use a collection of both official and crowdsourced databases, using schemas that match those used in the original sources, without attempting to align these schemas with the one in use on Enipedia. The use of both official and crowdsourced databases is important as they each have specific strengths and weaknesses. Official government sources generally offer complete information generated through data gathered "inside the fence" of the facilities. This is information that is simply not feasible to crowdsource. The crowdsourced dataset OpenStreetMap does not provide consistently detailed information on power plant properties, but it contains accurate geographic coordinates (often missing from official sources), as the facilities are traced using aerial photographs. Wikipedia also is useful for finding links to primary sources which can aid in data verification.

A key feature of Elasticsearch is that it allows us to flexibly search multiple data fields. This is important since one cannot just match power plants based on their names, due to the fact that in various databases, the name of a power plant may also include words related to its owner,

city, region or even fuel type. The option to alter the type of matching is important as well. Among other query options, Elasticsearch allows for what is called a "common terms query", which minimizes the importance of words that occur frequently in a set of text. This is useful since matches on rare terms are more informative than matches on common terms such as "power plant". The advantage of this is that since the tool only considers term frequency, it is language independent, and one does not have to have knowledge of specific language terms (i.e. kraftwerk, centrale thermique, etc.) to be able to use it effectively. Furthermore, Elasticsearch also allows for geographic queries, so that users can search for all plants matching a text query within a bounding box, or simply leave the text query empty to find all plants within a region that are mentioned in the different databases.



Score	Source	ID	changeset	uid	power	generator:type	timestamp	start_date	
2.318683	OpenStreetMap	way/220122486	16846647	36080	generator	steam_turbine	2013-07-06T10:01:24Z	1980	4.8450

Score	Source	ID	plant:output:hot_water	changeset	uid	power	lon	timestam
2.318346	OpenStreetMap	way/6318872	600 MW	22965135	1993800	plant	4.842254034502922	2014-06-16T13

Score	Source	ID	changeset	note	uid	name	generator:type	timestamp
1.9321244	OpenStreetMap	way/220122487	22887930	also fueled by gasification of waste building timber	1993800	Amercentrale Eenheid 9	steam_turbine	2014-06-12T14:06:36Z

Score	Source	ID	account	installation	euetsID	name
1.799764	euets	100933	http://enipedia.tudelft.nl/data/EU-ETS/country/NL/installation/172/account	http://enipedia.tudelft.nl/data/EU-ETS/country/NL/installation/172	100933	Essent N.V. Amercentrale

Capturing Ambiguity in Linked Data

At the moment, this tool allows us to augment existing data by hand. This data augmentation can be taken a step further using a strategy that could be termed "Loosely Coupled" Linked Data, which means that two different entities are not necessarily related by owl:sameAs, but rather via links defined by SKOS[3] such as skos:broadMatch, skos:narrowMatch and skos:relatedMatch. In other words, "Loosely Coupled" Linked Data allows us to describe supersets, subsets and entities with (potentially ambiguous) overlap. A key issue to consider is that determining the exact nature of links between entities, especially in large datasets, can be prohibitively expensive. The best way to deal with this depends on whether one views the publishing of a dataset as an end goal, or as a step in an iterative process with users continuously reviewing, criticizing, correcting and amending it. If the data is published as part of a continual iterative process, then it is useful to capture the degree of (un)certainty, as this helps to reduce the search space for people who use the data later and have relevant knowledge that can be used to improve it.

A different approach for dealing with multiple versions of the "truth" is used by the Wikidata project, which uses collections of "statements" about entities[4]. For example, the exact population of Berlin can never truly be known, but there are estimates about it with varying degrees of accuracy. Each of these estimates is part of a statement which contains information about its source. All of the statements can be examined and compared, and the one which users agree is the most accurate is marked as the "preferred" statement.

We can conclude from this discussion that Linked Data is a powerful concept in practice. However, implementing it still poses many challenges, especially when one tries to combine multiple data sources that are not compiled in a way that facilitates interoperability. To some extent, this can be dealt with by improving publishing practices of the original data, although there are tools and approaches available which can help facilitate data comparison and integration.

Links

(1) http://enipedia.org

(2) https://github.com/cbdavis/enipedia-search

(3) W3C – SKOS Simple Knowledge Organization System: https://www.w3.org/2004/02/skos/

(4) Wikidata: Claims and statements https://www.wikidata.org/wiki/Wikidata:Glossary#Claims_and_statements

Text by: Dr. Chris Davis (c.b.davis@rug.nl)

5.4. GBPN

Tool for Building Energy Performance Scenarios

The Global Buildings Performance Network (GBPN)

The Global Buildings Performance Network (GBPN)[1] is a globally organized and regionally focused organisation with the mission is to provide policy expertise and technical assistance to advance building energy performance and realize sustainable built environments for all. It was founded in 2010, and it aims to contribute to the building sector achieving its full energy savings and CO2 mitigation potential of more than 2.1 Gt by 2030. The GBPN is mandated to advance knowledge and expertise on improving the energy performance of buildings and to provide the structure to achieve this. It is coordinated by a global hub based in Paris, and represented regionally by an office in Beijing and partner organizations in Brussels, Washington D.C. and Delhi. The GBPN works to achieve the transformational changes required in the building sector to tackle climate change while promoting economic and social well-being.The network has developed extensive partnerships with international organizations, governments, industries, research institutes and NGOs. Its project partners also extend its reach to South East Asia, while its experts network includes Latin American and African expertise.

Tool for Building Energy Performance Scenarios

This online tool[2] enables users to interactively engage with modeled data for three possible energy mitigation scenarios for the building sector, globally and per region, until 2050.

To facilitate a rapid transition to sustainable energy in the building sector, GBPN supports a path of ambitious policy actions, based on an energy scenario in which state of the art for energy efficiency becomes the norm in both new and existing buildings only ten years from now. This tool shows users where the world could be in terms of building energy use by 2050 under several scenarios, depending on the ambitiousness of policy decisions and technology choices.

The tool was developed using the 3CSEP-HEB (High Efficiency Buildings) model, which takes a performance-oriented approach to building energy use analysis. Instead of component-oriented methods, the tool takes a systemic perspective: the performance of whole buildings is studied and these performance values are used as key inputs in the scenarios.

The model generates projections for the entire world. Results are provided for the following key 11 regions: North America (NAM), Western Europe (WEU), Eastern Europe (EEU), Former Soviet Union (FSU), Latin America and the Caribbean (LAC), Pacific OECD (PAO), Centrally Planned Asia (CPA), Pacific Asia (PAS), South Asia (SAS), Middle East and Africa (MEA), and Africa (AFR). The tool also provides results for the four key zones: China, Europe, India and the United Sates.

Linked Open Data: a Strong and Stable Basis of the Tool

Linked open data is used for publishing the data as well as for the analysis, in the form of (complex) querying of the data along the specified prediction model. The data was collected and the scenarios (of prediction) were calculated in a Microsoft Access Database[3] and all data was converted to RDF using the RDF Datacube Vocabulary[4] for the devel-

opment of the methodology and the prediction model[5]. The collected data concerning building types, regions and climate zones was linked to DBpedia[6] to offer more comprehensive information about these topics. The use of this RDF data – stored in a Virtuoso Universal Server (Triple Store)[7] – enables complex queries. The user of the tool selects the required filter from a set of available filters (which are aligned with the dimensions of the data cube) and is guided through the process by a very intuitive user interface. After selecting the filters, the user receives visualizations of the results that can be downloaded as images to be used e.g. in presentations, and the user can also download the underlying data of the visualized results in several formats (rdf, csv, Json). The visualization component of the tool is realized using HighCharts[8] and the overall tool is seamlessly embedded into the GBPN Knowledge Platform, which uses Drupal Open Source CMS[9].

A very special feature of the tool is that for any filtered selection managed by the user in the graphical user interface, the user is able to receive and see (or use) the underlying SPARQL query. This enables users to learn the SPARQL query language as well as to better understand the underlying data model. Finally it allows the user to easily query all of the data available via the API provided in the form of a SPARQL endpoint[10]. The data is available with an open license (Creative Commons Attribution 3.0[11]), which means all data can be used for any purposes as long as an attribution is added to the data source.

Links & further Reading

(1) GBPN, Global Buildings Performance Network: http://www.gbpn.org/

(2) Tool for Building Energy Performance Scenarios: http://www.gbpn.org/databases-tools/mrv-tool/about

(3) Microsoft Access Database: https://en.wikipedia.org/wiki/Microsoft_Access

(4) The RDF Datacube Vocabulary: https://www.w3.org/TR/vocab-data-cube/

(5) Details on the Methodology: http://www.gbpn.org/databases-tools/mrv-tool/methodology

(6) DBpedia: http://wiki.dbpedia.org/

(7) Virtuoso Universal Server: https://virtuoso.openlinksw.com/

(8) Highcharts JavaScript Library: http://www.highcharts.com/

(9) Drupal Open Source CMS: https://www.drupal.org/

(10) SPARQL query language: https://en.wikipedia.org/wiki/SPARQL

(11) Creative Commons 3.0: https://creativecommons.org/licenses/by/3.0/

5.5. DEVINIT

Joined-up Data Standards

Development Initiatives

Development Initiatives[1] mission is to ensure that decisions about the allocation of finance and resources result in an end to poverty, increase the resilience of the world's most vulnerable people, and leave no one behind. To deliver this, the organization works in three key thematic areas:

Theme 1: Poverty – Measuring the progress of people out of poverty

Data has the power to bring about a greater understanding of poverty and enables the accurate measurement of people's progress. With the right data, one can look beyond frequently used aggregate "income" or "consumption" poverty statistics and consider the reality of the lives of people in poverty.

Development Initatives' work seeks to:

- Drive commitments and much-needed investment towards improving poverty data, prioritizing disaggregated data and civil administration systems.
- Unpack the poverty data that does exist, bringing in new global and subnational data to start building a clearer and more accurate picture of poverty.

Theme 2: Resources – Investments to end poverty and build resilience

Data is essential for effectively identifying, linking and targeting resources to drive poverty eradication and sustainable development.

Development Initatives' work seeks to:

- Improve decision-making on the allocation of resources through our "all resources" analysis and tools such as our Development Data Hub.
- Increase the transparency of resource flows to improve understanding of the allocation of resources and their effectiveness, in part through our core technical role in the International Aid Transparency Initiative (IATI).

Theme 3: Data use – Improving data use for sustainable development

Development Initiatives believes that for decision-makers and advocates to better understand poverty and the resources available to address it, they need better data.

Development Initatives' work seeks to:

- Understand and remove the barriers to data use by creating and testing solutions, and by driving behavior change.
- Collaborate across the open data community and harness drivers for data use, such as the Data Revolution for Sustainable Development.

The Joined-up Data Standards Initiative for Better Information

Joined-Up Data Standards[2] is a project jointly implemented by Development Initiatives (DI) and Publish What You Fund. The project has initiated a multi-stakeholder process to develop technical and process solutions for the harmonization of global data standards at critical intersections such as time, location and organizations. In the context of an emerging data revolution for sustainable development, this work is vital for data users from the local to the global level who need better information for improved decision making.

So far, 4 projects are working on mapping and linking international standards to each other and the Joined-Up Data Standards project – using linked open data technologies AND making all data available as open data for use and re-use. The standards were imported and/or created (modeled) using the PoolParty Semantic Suite modeler (3) and are available for user view in the form of a website (a linked data frontend) as well as via API in the form of a SPARQL endpoint.

These available joined-up data standard projects are:

INDICATORS[4]: This project hosts international monitoring data standards (indicators) such as Sustainable Developmental Goals, Millennium Developmental Goals (MDGs) and World Development Indicators (WDI). The indicators are mapped in a machine- and human-readable format and linked to both each other and across the projects maintained on this thesaurus such as "sectors", " supranational regions and grouping" and "household surveys". The cross-mapping of standards is an ongoing process and the list of the data standards in this project will increase as our work progresses.

SECTORS[5]: This project links international data standards used to classify socio-economic activities. The project offers a machine-readable cross-walk (mapping) between major standards developed by the United Nations, World Bank and the Organization for Economic Co-operation and Development (OECD). The data standards in this project

are also mapped in a machine-readable way in the same way as other projects in this Thesaurus such as: indicators, surveys, supranational regions and groupings as well as Natural Resource Governance Institute (NRGI).

SUPRANATIONAL REGIONS AND GROUPINGS[6]: This project contains machine- and human-readable mappings of international classifications of countries with respect to economic, geographic, health or security groupings. The classifications include those by the United Nations, the Organisation for Economic Co-operation and Development (OECD), World Bank, World Health Organisation, and International Monetary Fund, among others. This project is cross-linked to other data standards hosted on this thesaurus.

SURVEYS[7]: This project provides a machine-readable cross-walk between standard questionnaires used in international household surveys. Currently, the project provides a mapping of the DHS Program Demographic and Health Surveys (DHS) to UNICEF Multiple Indicator Cluster Surveys (MICS). These surveys will be soon cross-mapped to the other projects in this thesaurus, such as "supranational regions and groupings", denoting where the surveys took place, "indicators", mapping the international indicators to the source question and finally "sectors", to show the alignment between national and international socio-economic sectors.

Links & further Reading

(1) Development Initiatives: http://devinit.org/

(2) Joined-up Data Standards: http://joinedupdata.org/

(3) PoolParty Semantic Suite: http://www.poolparty.biz

(4) Indicators Project: http://joinedupdata.org/Indicators.html

(5) Sectors Project: http://joinedupdata.org/Sectors.html

(6) Supranational Regions and Groupings Project: http://joinedupdata.org/geo-pol.html

(7) Surveys Project: http://joinedupdata.org/Surveys.html

6. APPENDIX

6.1. Authors

Martin Kaltenböck, CMC

Martin Kaltenböck studied communication, psychology and marketing Martin Kaltenböck studied communications, psychology and marketing at the University of Vienna. He is Co-Founder, 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 semantic information management, linked (open) data and open (government) data. He is a certified management consultant and an invited expert of OGD Austria, a governmental cooperation. He working as an invited expert at W3C, and is a member of the Steering Board of the European Data Forum, which he chaired in Athens, Greece in 2014. In 2015 he co-founded the ODI Node Vienna together with 3 Austrian university partners.

Publications

- Enterprise 2.0 - Introduction, Principles, Use Cases and Tools
- Open Government Data (OGD) White Book Austria 2011
- Linked Open Data -- Creating Knowledge Out of Interlinked Data

Contact

Phone: +43 (0)1 402 123 50
E-Mail: m.kaltenboeck@semantic-web.at
Slideshare: http://www.slideshare.net/MartinKaltenboeck
LinkedIn: http://www.linkedin.com/in/martinkaltenboeck

Mag. Florian Bauer

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 catalyze the market for renewable energy and energy efficiency, with a primary focus on emerging markets and developing countries. As Chief Operating Officer at REEEP, Florian oversees REEEP's knowledge tools such as reegle.info (a linked open data clean energy information portal) and the Climate Tagger (a tool to automatically tag unstructured content and connect it with similar content on other websites). Florian co-authored the "Climate Knowledge Brokers Manifesto" and plays a leading and consulting role in Open Data and Linked Open Data in the international development and climate change community. 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
- Climate Knowledge Brokers Manifesto

Contact

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

Mag. Andreas Blumauer

Andreas Blumauer earned a Master's degree in business, specializing 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 organizations. In 2000, he co-founded punkt. netServices, a specialized 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 semantic web.

Publications

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

Contact

Phone: +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

Denise Recheis

Denise studied media in the UK. She later discovered her interest in technology and studied engineering for renewable energy in Vienna, Austria. She joined REEEP in 2010 and developed her skills in taxonomy management, and became a keen advocate for Linked Open Data. As Product Manager of Climate Tagger, she demonstrates the benefits of using (Linked) Open Data in the fields of clean energy, climate change and sustainability. She manages the sector's most comprehensive Linked Open Data vocabulary describing clean energy and climate and is open to exploring synergies with other stakeholders in the international development sector for the purpose of collaboration and mutual learning.

Publications

- The Open Book

Contact

Phone: +43 (0) 1 260 263 495
E-Mail: denise.recheis@reeep.org

6.2. Attribution & Credits

Image credits and copyrights

The pictures used for this book were obtained from the photography libraries.

Sponsoring Partners

CDKN – Climate Development Knowledge Network

http://cdkn.org

CKB – Climate Knowledge Brokers

www.climateknowledgebrokers.net/

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

The new edition of this book was created as a capacity building manual for Climate Knowledge Brokers supported and funded by CDKN; it is based on the first edition which 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 could not have been made without the help of 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, Geoff Barnard, Marianne Osterkorn, Martin Hiller, Anna Florowski, John Sheridan, José Manuel Alonso, Bitange Ndemo, Jon Weers, Thomas Thurner, Vince Reardon, Jena Wuu, Günther Friesinger, Sigmund Kluckner, John Tkacik, Maria van Veldhuizen, Quinn Reifmesser, Chris Davis

After the success of the original "Linked Open Data- The Essentials" guide from 2012, we are proud and happy to now be able to present this new edition, titled "Linked Open Data - The Climate Knowledge Brokering Edition".

This book is a quick start guide for decision makers and climate knowledge brokers seeking to get up to speed with the concept of Linked Open Data (LOD), and are looking to make their organization part of the movement.

It provides a brief overview of LOD and its relevance for successful knowledge brokering, as well as 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 are climate knowledge brokers, and how is Open Data relevant to them?
- How can Open Data help my organization to communicate more effectively?
- What do I need to take into account when developing a LOD strategy for my organization, and what needs to be done technically in order to open up and publish data sets?
- How can I add value to my own data sets by consuming LOD from other sources, and how can that support my role as a knowledge intermediary?
- What can be learned from five case studies of best practice in LOD?
    - REEEP's Climate Tagger www.climatetagger.net
    - NREL's Open Energy Information Portal http://en.openei.org
    - Enipedia http://enipedia.tudelft.nl/wiki/Main_Page
    - GBPN's Tool for Building Energy Performance Scenarios
    - Development Initiatives' Joined-up Data Standards

REEEP® CKB CLIMATE KNOWLEDGE BROKERS SEMANTIC WEB COMPANY linking data to knowledge

edition mono monochrom ISBN: 978-3-902796-54-7