e-infrastructures austria · Deliverable

# Linked Data in a nutshell

| | |
|---|---|
| Work-Package: | Cluster I Metadata |
| Work-Package-Leader: | Susanne Blumesberger, Vienna University Library, susanne.blumesberger@univie.ac.at |
| Date: | 30.6.2016 |
| Authors and other contributors: | Susanne Blumesberger, Vienna University Library, susanne.blumesberger@univie.ac.at<br>Sandor Kopacsi, Vienna University Computer Center, sandor.kopacsi@univie.ac.at |

| | |
|---|---|
| Description: | Linked Data describes a method of publishing structured data so that it can be interlinked and become more useful through semantic queries. |
| Keywords: | Linked Data, Linked Open Data, Metadata, Semantic Web |

# Linked Data in a nutshell

compiled by
Susanne Blumesberger and Sandor Kopacsi
June 2015

# What is Linked Data? /1

**Linked Data** describes a method of publishing structured data so that it can be interlinked and become more useful through semantic queries. It builds upon standard Web technologies such as HTTP, RDF and URIs, but rather than using them to serve web pages for human readers, it extends them to share information in a way that can be read automatically by computers. This enables data from different sources to be connected and queried.

http://en.wikipedia.org/wiki/Linked_data

# What is linked data? /2

Linked Data is a method of exposing, sharing, and connecting data on the Semantic Web using URIs (Uniform Resource Identifier) and RDF (Resource Description Framework) (see http://linkeddata.org/). Metadata are the backbone of this method, making statements about data and how they relate to each other. In Linked Data these statements have to be expressed in RDF triples, which break statements in three parts:

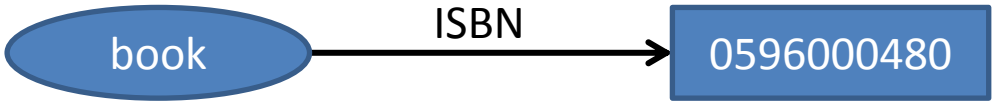
- the subject - the part that identifies the thing the statement describes,
- the predicate - the part that identifies a property or an activity or a state of the described thing.
- the object - the part that identifies the value of this property or the object of the activity when describing this thing.

In Linked Data you need URIs referencing to things by identifying, localizing and interlinking them.

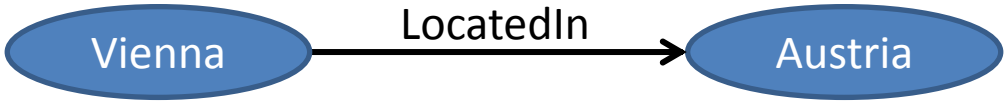
# Knowledge Representation in triplets

the knowledge can be represented in form of **triplets**, as

- **object-attribute-value**




- **subject-predicate-object**



# What is Linked Data? /3

**A metadata description in Linked Data consists of:**

- reference to a described resource, i.e. to the thing that the metadata is about, the subject of the metadata. This reference takes the form of a URL (URI) or of an unnamed placeholder that is inferred by context (e.g., that metadata embedded in a document is about the enclosing document).
- references to properties: typed relationships between the described resource and various bits of descriptive information, or between the described resource and another resource. Examples of properties, which are also known as predicates, include the fifteen elements of the Dublin Core.
- values: bits of descriptive information (string literals) or references to other entities (resources), such as people or concepts, which are related to the described resource via the properties.
- references to classes, i.e. to types, or categories, of things, such as the category books or the category people.
- references to syntax encoding schemes (RDF datatypes), i.e. specifications of how value strings map refer to things in the world
- references to vocabulary encoding schemes (VES), enumerated sets of resources of which the things referenced as values are members, the way a subject heading belongs to the VES Library of Congress Subject Headings.
- language tags indicating the language of words used in literal string values.

See more: http://wiki.dublincore.org/index.php/User_Guide

# Why Linked Data?

The Linked Data approach offers significant advantages over current practices for creating and delivering library data while providing a natural extension to the collaborative sharing models historically employed by libraries.

Linked Data and especially Linked Open Data is

- **Sharable**
- **Extensible**
- **easily re-usable**

It supports multilingual functionality for data and user services, such as the labeling of concepts identified by language-agnostic URIs. These characteristics are inherent in the Linked Data standards and are supported by the use of Web-friendly identifiers for data and concepts. Resources can be described in collaboration with other libraries and linked to data contributed by other communities or even by individuals.

Like the linking that takes place today between Web documents, Linked Data allows anyone to contribute unique expertise in a form that can be reused and recombined with the expertise of others.

More: http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/#Benefits_of_the_Linked_Data_Approach

# Principles of Linked Data /1

1. Use URIs as names for things.
2. Use HTTP URIs, so that people can look up those names.
3. Provide useful information, using the standards (RDF, SPARQL).
4. Include links to other URIs, so that they can discover more things.

# Principles of Linked Data /2

1. The first Linked Data principle advocates using URI references to identify, not just Web documents and digital content, but also real world objects and abstract concepts. These may include tangible things such as people, places and cars, or those that are more abstract, such as the relationship type of knowing somebody, the set of all green cars in the world, or the color green itself. This principle can be seen as extending the scope of the Web from online resources to encompass any object or concept in the world.

2. The HTTP protocol is the Web's universal access mechanism. In the classic Web, HTTP URIs are used to combine globally unique identification with a simple, well-understood retrieval mechanism. Thus, the second Linked Data principle advocates the use of HTTP URIs to identify objects and abstract concepts, enabling these URIs to be dereferenced (i.e., looked up) over the HTTP protocol into a description of the identified object or concept.

# Principles of Linked Data /3

3. In order to enable a wide range of different applications to process Web content, it is important to agree on standardized content formats. The agreement on HTML as a dominant document format was an important factor that made the Web scale. The third Linked Data principle therefore advocates use of a single data model for publishing structured data on the Web – the Resource Description Framework (RDF), a simple graph-based data model that has been designed for use in the context of the Web [70]. The RDF data model is explained in more detail later in this chapter.

4. The fourth Linked Data principle advocates the use of hyperlinks to connect not only Web documents, but any type of thing. For example, a hyperlink may be set between a person and a place, or between a place and a company. In contrast to the classic Web where hyperlinks are largely untyped, hyperlinks that connect things in a Linked Data context have types which describe the relationship between the things. For example, a hyperlink of the type friend of may be set between two people, or a hyperlink of the type based near may be set between a person and a place. Hyperlinks in the Linked Data context are called RDF links in order to distinguish them from hyperlinks between classic Web documents.

Linked Data: Evolving the Web into a Global Data Space http://linkeddatabook.com/editions/1.0/#htoc8 (This book provides a conceptual and technical introduction to the field of Linked Data. It is intended for anyone who cares about data – using it, managing it, sharing it, interacting with it – and is passionate about the Web.)

# Linked Data for Libraries

VIDEO: "Linked Data for Libraries" (from OCLC)
http://www.youtube.com/watch?v=fWfEYcnk8Z8x%x%

VIDEO: Linked Data for Holdings and Cataloging: The First Step Is Always the Hardest!
http://eventscribe.com/2013/ALA-Midwinter/assets/audio-flash/59289/launcher.asp?AssetID=12867

The Strongest Link: Libraries and Linked Data

This article will outline some of the benefits that linked data could have for libraries, will discuss some of the non-technical obstacles that we face in moving forward, and will finally offer suggestions for practical ways in which libraries can participate in the development of the semantic web.
http://www.dlib.org/dlib/november10/byrne/11byrne.html

# Linked Open Data/1

Linked Open Data is linked data that is open content.

The difference is most clear by the definition by Tim Berners-Lee which defines linked data by four rules, while adding a fifth (more primary) rules to define linked open data:

**“make your stuff available on the Web (whatever format) under an open license”**

http://en.wikipedia.org/wiki/Linked_open_data

More:

Linked Open Data: The Essentials. A Quick Start Guide for Decision Makers by Florian Bauer (REEEP) and Martin Kaltenböck (Semantic Web Company: http://www.semantic-web.at/LOD-TheEssentials.pdf

Mikel Egana Aranguren (2011) Publishing information in the Linked Open Data cloud. The Bioinformatics Knowledgeblog. http://bioinformatics.knowledgeblog.org/2011/07/05/publishing-information-in-the-linked-open-data-cloud/

# Linked Open Data/2

VIDEO: Linked Open Data - What is it? (from Europeana)

An introductory overview of Linked Open Data in the context of cultural institutions.

http://vimeo.com/36752317

# Some ways to Linked Data

A pattern catalogue for modelling, publishing, and consuming Linked Data von Leigh Dodds und Ian Davis under http://patterns.dataincubator.org/book/

**How to Publish Linked Data on the Web**

This document provides a general overview of the concept of Linked Data, and then describes several practical recipes for publishing information as Linked Data on the Web.

http://wifo5-03.informatik.uni-mannheim.de/bizer/pub/LinkedDataTutorial/

# Linked Data and cultural heritage

LINKED HERITAGE

Coordination of standard and technologies

for the enrichment of Europeana

Best practice report on cultural heritage linked data and metadata standards, 2011: http://www.linkedheritage.eu/getFile.php?id=229

# Linked Data Services for Theses and Dissertations

Linked Data Services for Theses and Dissertations
http://scholarsarchive.library.oregonstate.edu/xmlui/handle/1957/32977

Abstract: Linked Data presents new opportunities to expand services surrounding theses and dissertations. By creating open datasets, ETD systems can be built to interoperate with other institutional data as well as with outside metadata sources. However, much foundational work must be done before these advantages can be fully realized. This paper details work at Oregon State University to create a Linked Dataset covering the University's theses and dissertations. Using data from existing MARC and Qualified Dublin Core records, we have established a process and model for crosswalking data from existing records into a variety of Semantic Web vocabularies. Our approach is to create basic services on a dedicated thesis and dissertation interface, incrementally extending those available through our institutional repository. We describe services implemented, those in progress and plans for continued work. We also address the limitations of our existing metadata and resulting challenges in crosswalking and interoperability. While Linked Data has great promise, implementation must target specific services that can be implemented today. We plan continued work to improve our data models and to utilize new data from other linked data sources as they emerge.

# Guidelines for Linked Data

Methodological Guidelines for Publishing Linked Data

The process of publishing Linked Data has an iterative incremental life cycle model:

1) Identification and analysis of the data
2) URI design (use meaningful URIs, instead of opaque URIs)
3) Definition of the license sources
4) Reuse and leverage on data already opened up
5) Transformation of data
6) Data cleansing
7) Linking. Take the data sources selected in the specification activity and transform them to RDF according to the vocabulary created in the modelling activity
8) Dataset publication
9) Metadata publication
10) Dataset detecting
11) Register the dataset into CKAN Registry

http://delicias.dia.fi.upm.es/wiki/images/7/7a/07_MGLD.pdf

# Guidelines for URI‘s

The Resource Description Framework RDF allows users to describe both Web documents and concepts from the real world—people, organizations, topics, things—in a computer-processable way. Publishing such descriptions on the Web creates the Semantic Web. URIs (Uniform Resource Identifiers) are very important, providing both the core of the framework itself and the link between RDF and the Web. This document presents guidelines for their effective use. It discusses two strategies, called 303 URIs and hash URIs. It gives pointers to several Web sites that use these solutions, and briefly discusses why several other proposals have problems.

URIs for the Semantic Web under: http://www.w3.org/TR/cooluris/

DOIs as Linked Data: http://inkdroid.org/journal/2011/04/25/dois-as-linked-data/

# Some standards for metadata

**RDF**

The Resource Description Framework (RDF) is a framework for representing information in the Web. RDF graphs are sets of subject-predicate-object triples. RDF can be expressed in different (XML, Turtle, N-Triples, etc.) formats. Using RDF allows data to be linked and merged with other RDF datasets, resulting Linked Data.

https://wiki.dnb.de/pages/viewpage.action?pageId=68060017

http://openorg.ecs.soton.ac.uk/index.php?title=Linked_Data_Basics_for_Techies&oldid=333

http://data.bnf.fr/docs/databnf-presentation-en.pdf

**Dublin Core**

How to use DCMI Metadata as linked data:
http://wiki.dublincore.org/index.php/User_Guide/Publishing_Metadata

Presentation about Dublin Core: http://de.slideshare.net/faoaims/linking-your-resources-to-the-data-web

**MARC**

http://id.loc.gov/vocabulary/relators.html

**MODS**

Metadata Object Description Schema (MODS) is a schema for a bibliographic element set that may be used for a variety of purposes, and particularly for library applications. The standard is maintained by the Network Development and MARC Standards Office of the Library of Congress with input from users. More:

http://www.loc.gov/standards/mods/mods-overview.html

**Library Linked Data Incubator Group: Datasets, Value Vocabularies, and Metadata Element Sets**

http://www.w3.org/2005/Incubator/lld/XGR-lld-vocabdataset-20111025/

# Identifiers

ORCID provides a persistent digital identifier that distinguishes you from every other researcher and, through integration in key research workflows such as manuscript and grant submission, supports automated linkages between you and your professional activities ensuring that your work is recognized http://orcid.org/

VIAF™ (Virtual International Authority File) is an international authority file. It is a joint project of several national libraries and operated by the Online Computer Library Center (OCLC)

The aim is to link the national authority files (such as the German Name Authority File) to a single virtual authority file. In this file, identical records from the different data sets are linked together.

VIAF: http://viaf.org/

List of Author identifications:
https://repinf.pbworks.com/w/page/13779410/Author%20identification

# Open Data in Austria

http://www.opendataportal.at

Presentation by Martin Kaltenböck:

http://www.slideshare.net/MartinKaltenboeck/open-data-portal-odp-sterreich-prsentation-bei-der-opendatach-2014-in-zrich

# Glossar from DINI AG KIM

http://dini-ag-kim.github.io/glossar/glossar.html

Sustainable data storage and the provision of data for use by third parties are the central roles of science. e-Infrastructures Austria is a federally funded program for the coordinated expansion and continued development of data repositories across Austria, and is made possible by a grant from the Austrian Ministry of Science, Research and Commerce (BMWFW). This program enables the safe archival and lasting availability of electronic publications, multimedia objects and other digital data from the research and teaching fields. Concurrently, topics relating to research data management and digital archiving workflows will be addressed.

**The working area is organized in twelve Work-Package-Clusters:**

Cluster A Monitoring of Document Repositories within the Partner Network
*Patrick Danowski (IST Austria)*

Cluster B Planning and Implementation of a „National Survey" for Research Data
*Christian Gumpenberger (University of Vienna)*

Cluster C Designing a Knowledge Network: Development of a reference structure for the construction of Repositories
*Paolo Budroni (University of Vienna)*

Cluster D Infrastructure
*Raman Ganguly (Vienna University Computer Center)*

Cluster E Legal and Ethical Issues
*Seyavash Amini (Counsellor-at-law, University of Vienna)*

Cluster F Open Access
*Andreas Ferus (Academy of Fine Arts Vienna)*

Cluster G Visual Data modeling
*Martin Gasteiner (University of Vienna)*

Cluster H Life Cycle Management
*Andreas Rauber (Technical University Vienna)*

Cluster I Metadata
*Susanne Blumesberger (University of Vienna)*

Cluster J Permanent backup of the data
*Adelheid Mayer (University of Vienna)*

Cluster K Data from scientific and artistic-scientific research processes
*Bernhard Haslhofer (Austrian Institute of Technology)*

Cluster L Cross-project issues (technical and non-technical)
*Andreas Jeitler (University of Klagenfurt)*

University of Vienna Library and Archive Services, 1010 Vienna, Universitätsring 1, T: +43-1-4277-15172, office@e-infrastructures.at

e-infrastructures austria