



LOD2 Stack and the NLP2RDF project

<http://slideshare.net/kurzum>

<http://nlp2rdf.org>

<http://lod2.eu>

Sebastian Hellmann
AKSW, Universität Leipzig



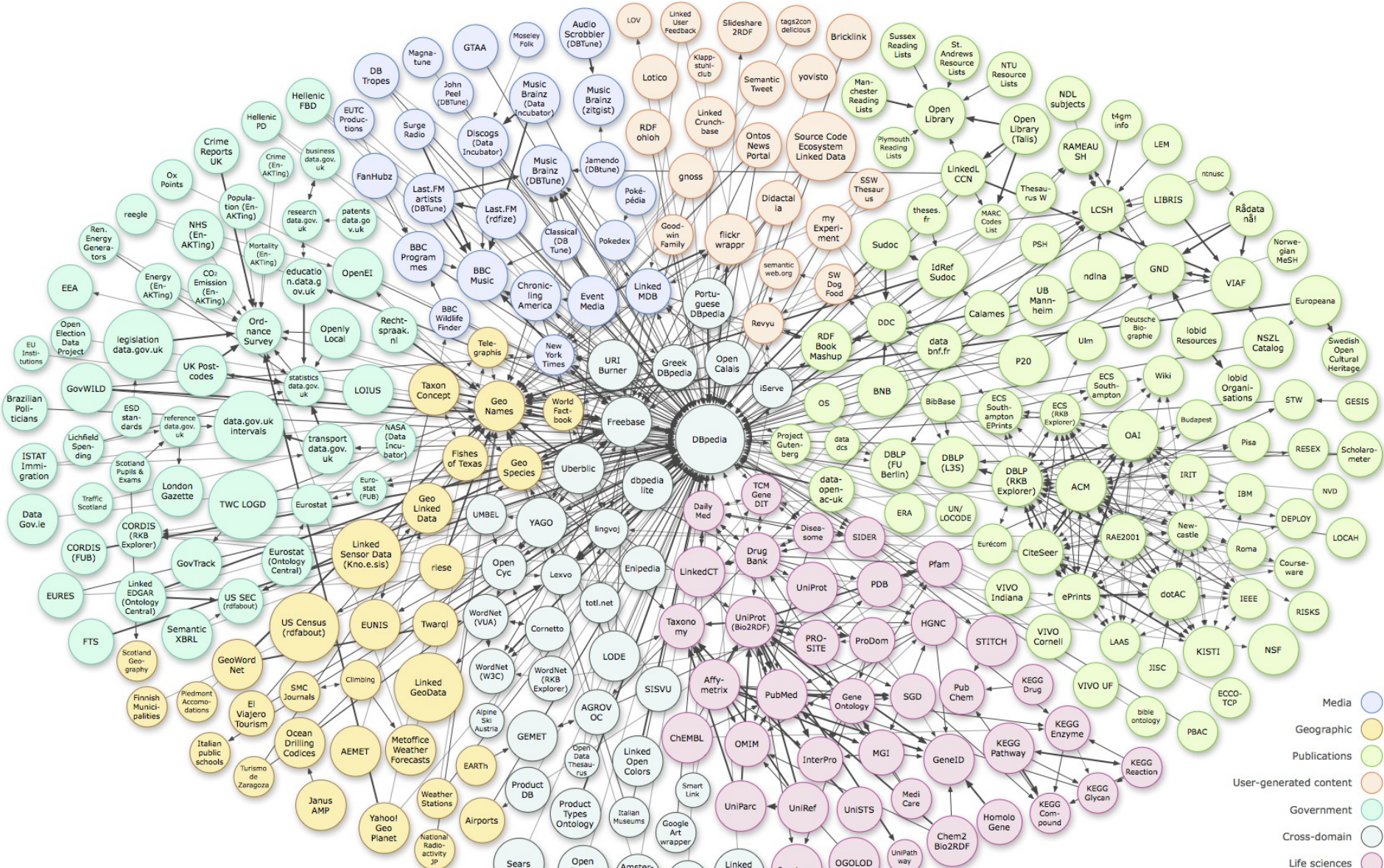
Overview

Topics:

- Introduction to LOD
- LOD2 Stack <http://stack.lod2.eu/>
- NLP Interchange Format (NIF), Version 2.0



http://lod-cloud.net



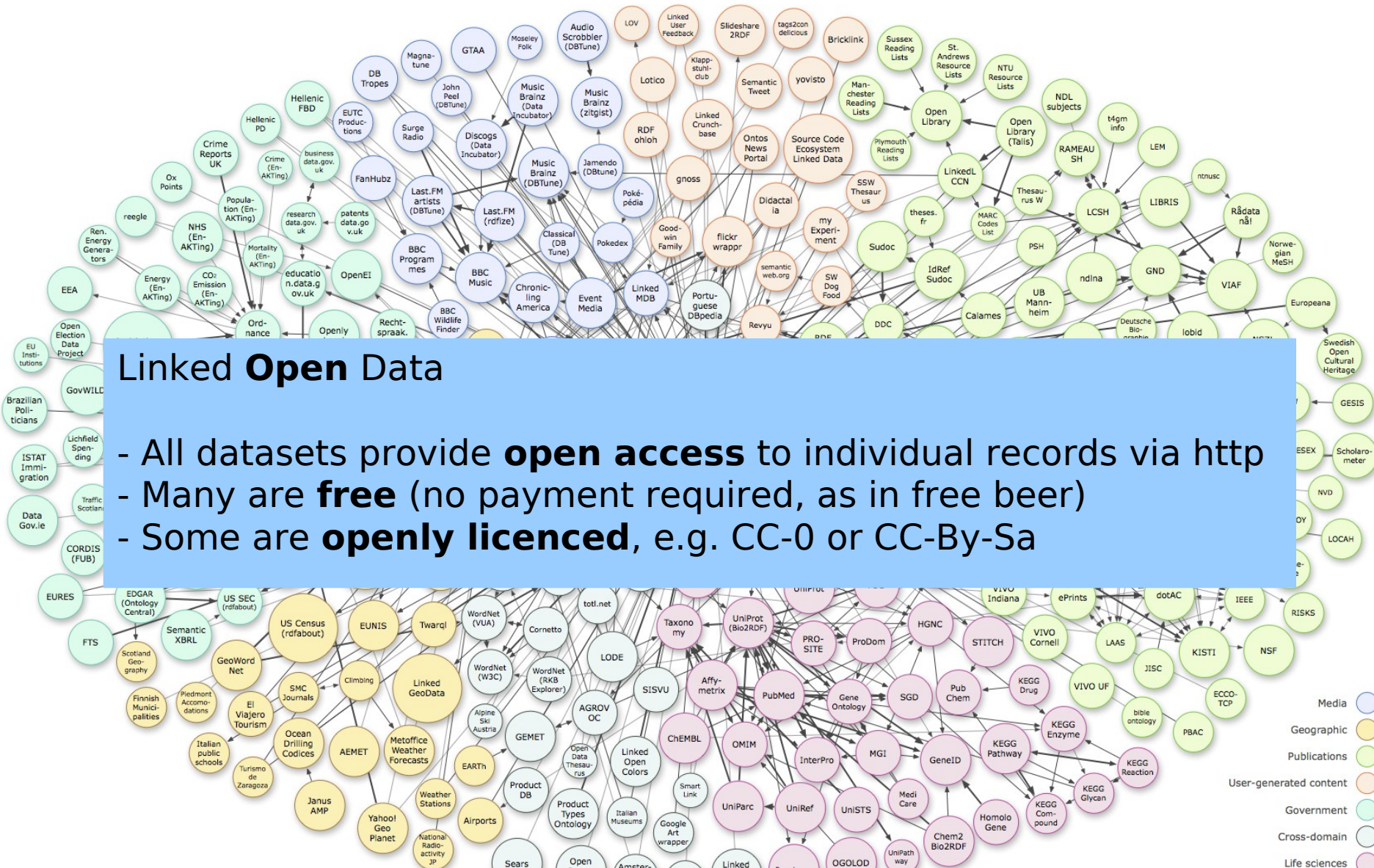
- Media
- Geographic
- Publications
- User-generated content
- Government
- Cross-domain
- Life sciences



<http://lod-cloud.net>

Linked Open Data

- All datasets provide **open access** to individual records via http
- Many are **free** (no payment required, as in free beer)
- Some are **openly licenced**, e.g. CC-0 or CC-BY-Sa



LOD2 Gang





<http://stack.lod2.eu/>

[Home](#) |
 [Stack in use](#) |
 [How to start](#) |
 [How to contribute](#) |
 [Download](#) |
 [Support](#) |
 [Online Demos](#) |
 [Press](#)

HOME ▾

Get an overview what's in.

Click yourself through the LOD2-Stack below!



The LOD2 stack comprises tools from the LOD2 partners and third parties.

The Stack

This is Version 2.0 of the LOD2 Stack, which comprises a number of tools for managing the life-cycle of Linked Data. The life-cycle comprises in particular the stages

- Extraction of RDF from text, XML and SQL
- Querying and Exploration using SPARQL
- Authoring of Linked Data using a Semantic Wiki
- Semi-automatic link discovery between Linked Data sources
- Knowledge-base Enrichment and Repair

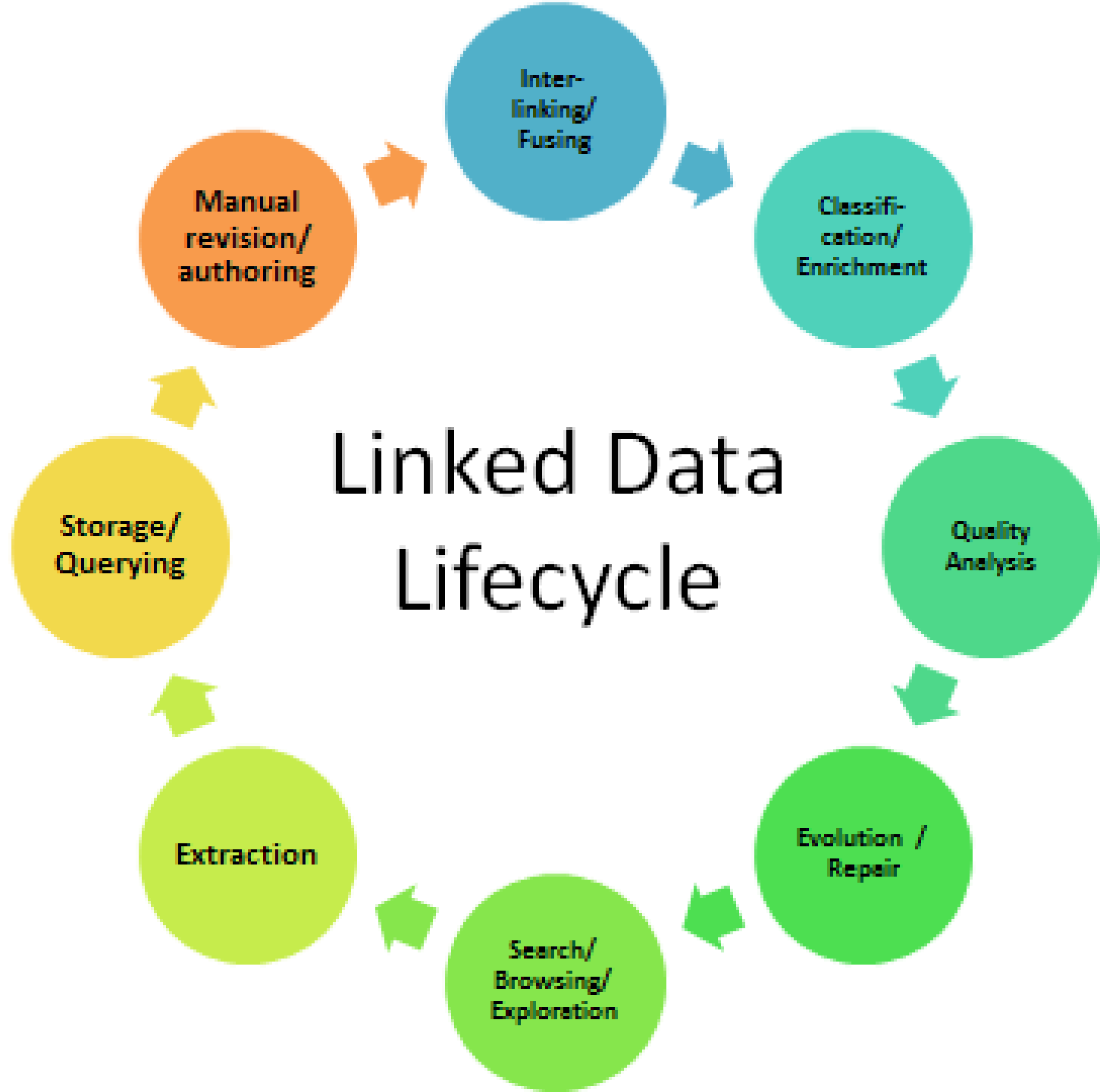
LEADING LINKED OPEN DATA TOOLS

(Blockieren)



LOD2 WEBLOG ▾

- "I like tables." - Interview with Richard Cyganiak ▾
- Added value of linking data (Polish economy data) ▾
- Semantic Data Web Lecture at





<http://stack.lod2.eu/>

#download the repository package

```
wget http://stack.lod2.eu/lod2repository\_current\_all.deb
```

#install the repository package

```
sudo dpkg -i lod2repository_current_all.deb
```

#update the repository database

```
sudo apt-get update
```

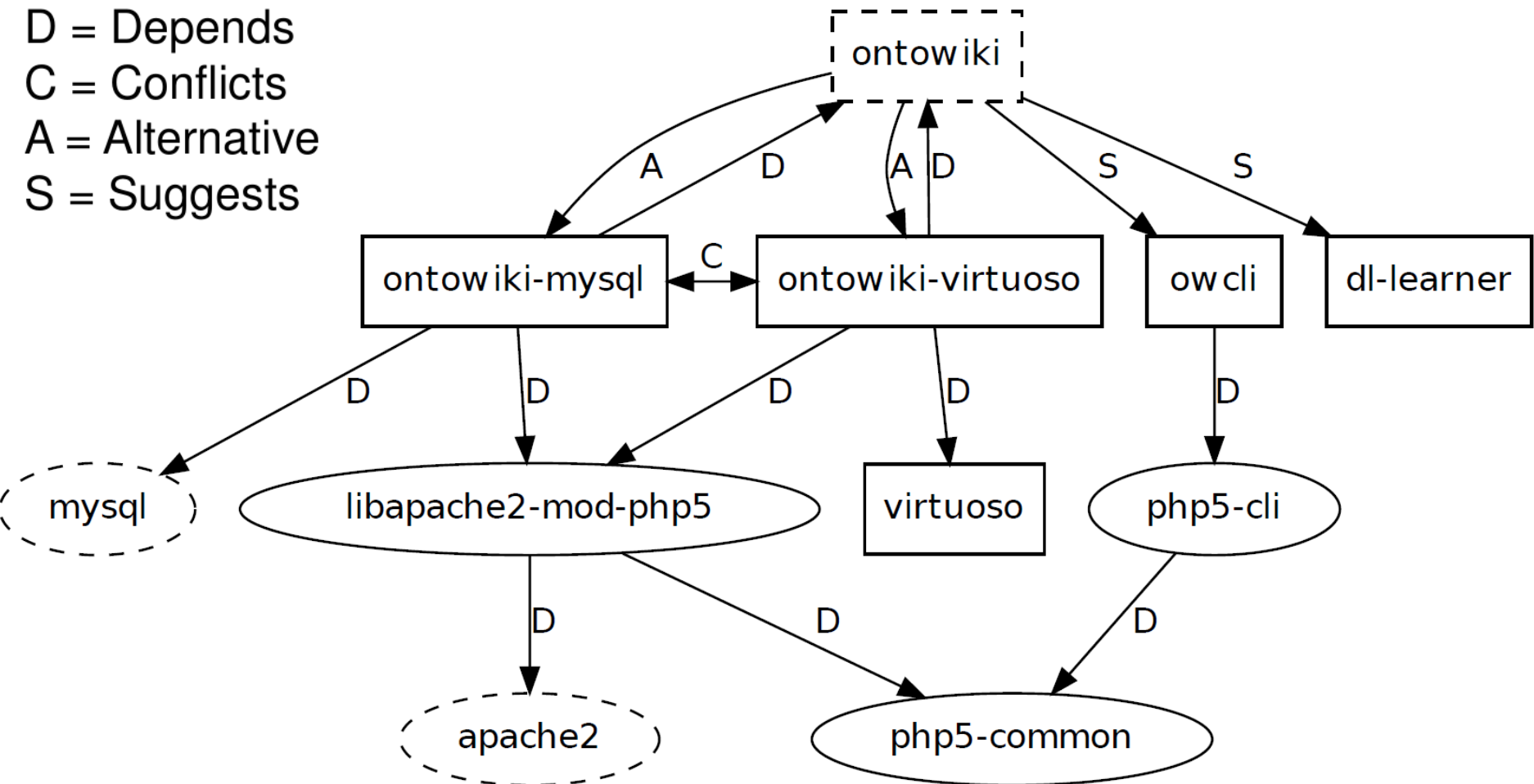
#lod2demo is a meta root package that installs all LOD2 components

```
sudo apt-get install lod2demo
```




<http://stack.lod2.eu/>

D = Depends
C = Conflicts
A = Alternative
S = Suggests





<http://stack.lod2.eu/>

Home **Stack in use** How to start How to contribute Download Support Online Demos Press

STACK
STACK IN USE ▾



- [UK Public Data](#) ▾
- [Digital Agenda Scoreboard](#) ▾
- [National Accounts Linked Data for the UK and Serbia](#) ▾
- [Increasing the Financial Transparency of European Commission Project Funding](#) ▾
- [Czech and Slovak public contracts data](#) ▾

LEADING LINKED OPEN DATA TOOLS



▾ LOD2 WEBLOG ▾

- "I like tables." - Interview with Richard Cyganiak ▾
- Added value of linking data (Polish economy data) ▾
- Semantic Data Web Lecture at SlideWiki 008 ▾



<http://stack.lod2.eu/>

Home **Stack in use** How to start How to contribute Download Support **Online Demos** Press

STACK
STACK IN USE ▾



- [UK Public Data](#) ▾
- [Digital Agenda Scoreboard](#) ▾
- [National Accounts Linked Data for the UK and Serbia](#) ▾
- [Increasing the Financial Transparency of European Commission Project Funding](#) ▾
- [Czech and Slovak public contracts data](#) ▾

LEADING LINKED OPEN DATA TOOLS



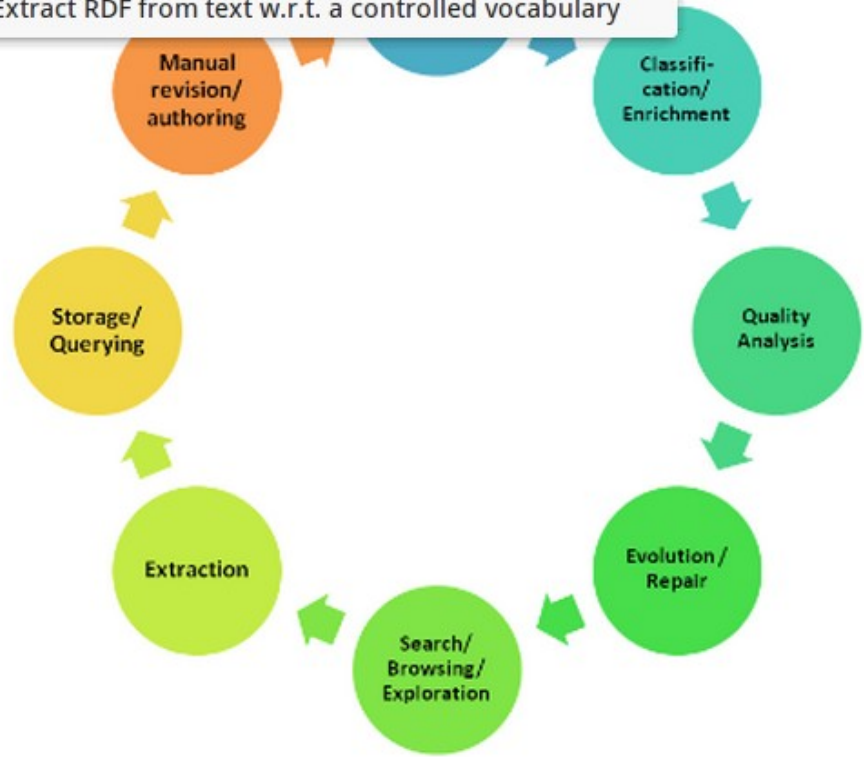
[RSS](#) ▾ **LOD2 WEBLOG** ▾

- "I like tables." - Interview with Richard Cyganiak ▾
- Added value of linking data (Polish economy data) ▾
- Semantic Data Web Lecture at SlideWiki.org ▾

- Extraction & Loading
- Querying & Exploration
- Authoring
- Linking
- Enrichment
- Online Tools and Services
- Configuration

- Upload RDF file
- Load RDF data from CKAN
- Extract RDF from XML
- Extract RDF from SQL
- Extract RDF from text w.r.t. DBpedia
- Extract RDF from text w.r.t. a controlled vocabulary

- Basic extraction
- Extended extraction



This is Version 1.0 of the LOD2 Stack, which comprises a number of tools for managing the life-cycle of Linked Data. The life-cycle comprises in particular the stages

- Extraction of RDF from text, XML and SQL
- Querying and Exploration using SPARQL
- Authoring of Linked Data using a Semantic Wiki
- Semi-automatic link discovery between Linked sources
- Knowledge-base Enrichment and Repair

You can access tools for each of these stages using the menu on top.

The LOD2 Stack is developed by the LOD2 project consortium comprising 15 research groups and companies. The LOD2 project is co-funded by the European Commission within the 7th Framework Programme (GA no. 257934).


You can find further information about the LOD2 Stack on the LOD2 project at <http://lod2.eu>.



OntoWiki (Admin)

User Extras Help Debug

Search for Resources



Knowledge Bases

Edit View

[OntoWiki System Configuration](#)

Navigation: Classes

Edit View Type

Search in Navigation

[Model](#)

[User](#)

[Usergroup](#)

[Action](#)

Properties of DefaultUserGroup

Resource

Properties History Community Source

rdf:type	Usergroup
label	DefaultUserGroup
comment	This pre-configured user group can login, register new user and edit all models except the system models.
not editable model	http://ns.ontowiki.net/SysBase/
grant access	<ul style="list-style-type: none"> <input type="checkbox"/> Register new User <input type="checkbox"/> Login <input type="checkbox"/> Rollback
editable model	<input type="checkbox"/> AnyModel
readable model	<input type="checkbox"/> AnyModel
not readable model	<ul style="list-style-type: none"> <input type="checkbox"/> OntoWiki System Configuration <input type="checkbox"/> OntoWiki System Ontology



http://stack.lod2.eu/

Silk Workbench

Workspace: movies_example Editor: movies Generate Links Learn Reference Links Population About

Undo Redo Export as Silk-LS Help

Precision = 1.00 | Recall = 0.80 | F-measure = 0.89 |

Property Paths

Source: **DBpedia**
Restriction: ?a ?p ?v .

- (custom path)
- ?a/dbpediaowl:releaseDate

Target: **linkedmdb**
Restriction: ?b ?p ?v .

- (custom path)
- ?b/linkedmdb:initial_release_date

Transformations

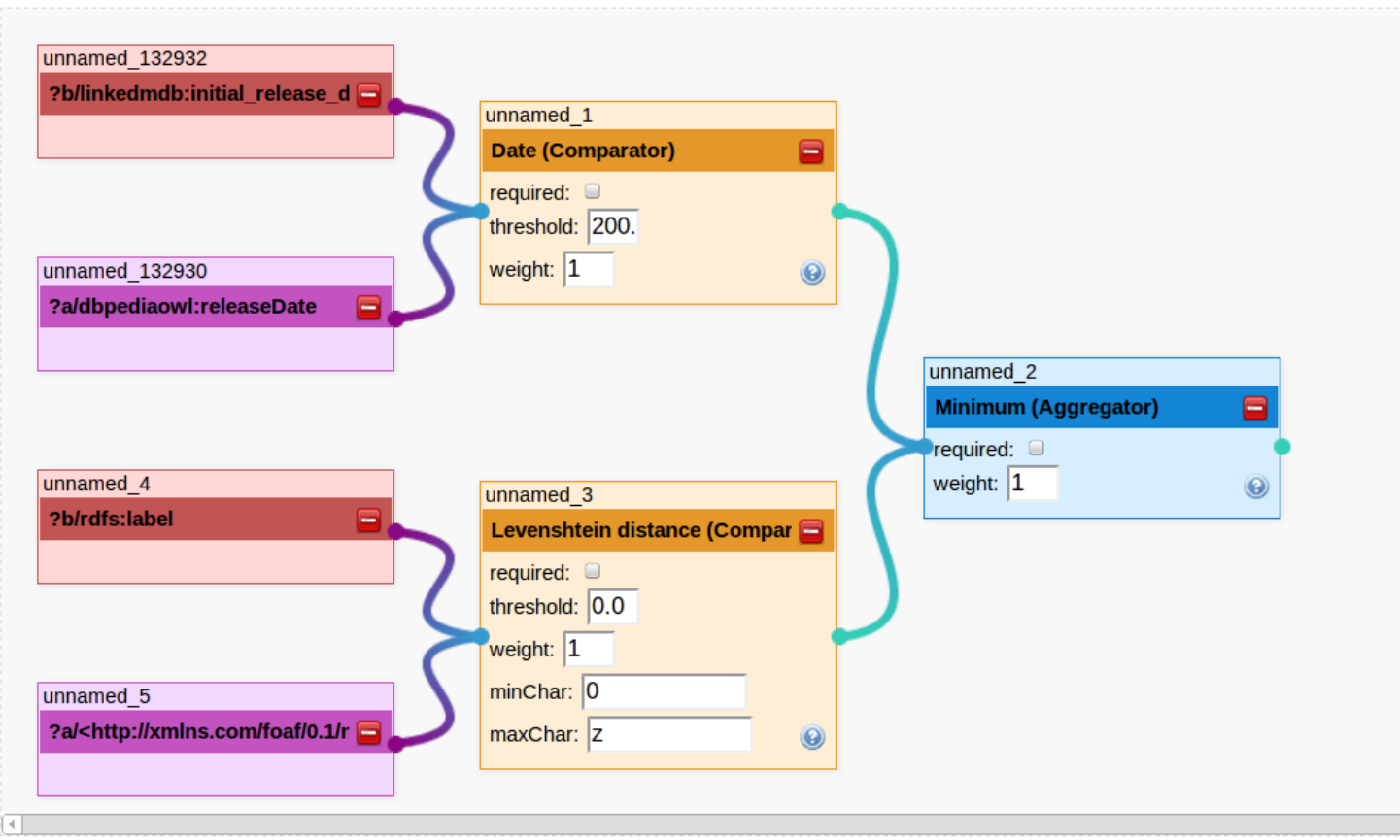
- Alpha reduce
- Capitalize

Comparators

- Date
- DateTime

Aggregators

- Average
- Euclidian distance



Link Limit: unlimited



http://stack.lod2.eu/

ORE - *Ontology Repair and Enrichment*



Knowledge Base Action

Knowledge base: SPARQL endpoint - URL: http://live.dbpedia.org/sparql

Options

Resource URI

Detect automatically

Resource type Class Object Property Data Property

Inference

Max. execution time

Max. returned axioms

Threshold

All

- DisjointObjectProperties
- EquivalentObjectProperties
- SubObjectPropertyOf
- FunctionalObjectProperty
- InverseFunctionalObjectProperty
- ObjectPropertyDomain
- ObjectPropertyRange
- SymmetricObjectProperty
- AsymmetricObjectProperty
- ReflexiveObjectProperty
- IrreflexiveObjectProperty
- TransitiveObjectProperty

Axiom types

Result

AsymmetricObjectProperty Axioms	
ACCURACY	AXIOM
<input checked="" type="checkbox"/> 99.98%	Asymmetric: author

InverseFunctionalObjectProperty Axioms	
ACCURACY	AXIOM
<input type="checkbox"/> 73.58%	InverseFunctional: author

FunctionalObjectProperty Axioms	
ACCURACY	AXIOM
<input type="checkbox"/> 88.38%	Functional: author
	Functional: http://dbpedia.org/ontology/author

ObjectPropertyRange Axioms	
ACCURACY	AXIOM
<input type="checkbox"/> 86.15%	author Range Person100007846
<input checked="" type="checkbox"/> 81.02%	author Range Person

ObjectPropertyDomain Axioms	
ACCURACY	AXIOM
<input type="checkbox"/> 99.84%	author Domain Work
<input type="checkbox"/> 92.40%	author Domain CreativeWork
<input checked="" type="checkbox"/> 84.16%	author Domain WrittenWork
<input type="checkbox"/> 80.94%	author Domain dbo:Book



http://stack.lod2.eu/



SPATIAL SEMANTIC BROWSER

Exploring: Financial Transparency System @ <http://localhost/fts/sparql>

Instances Facets Types Places

luxembourg



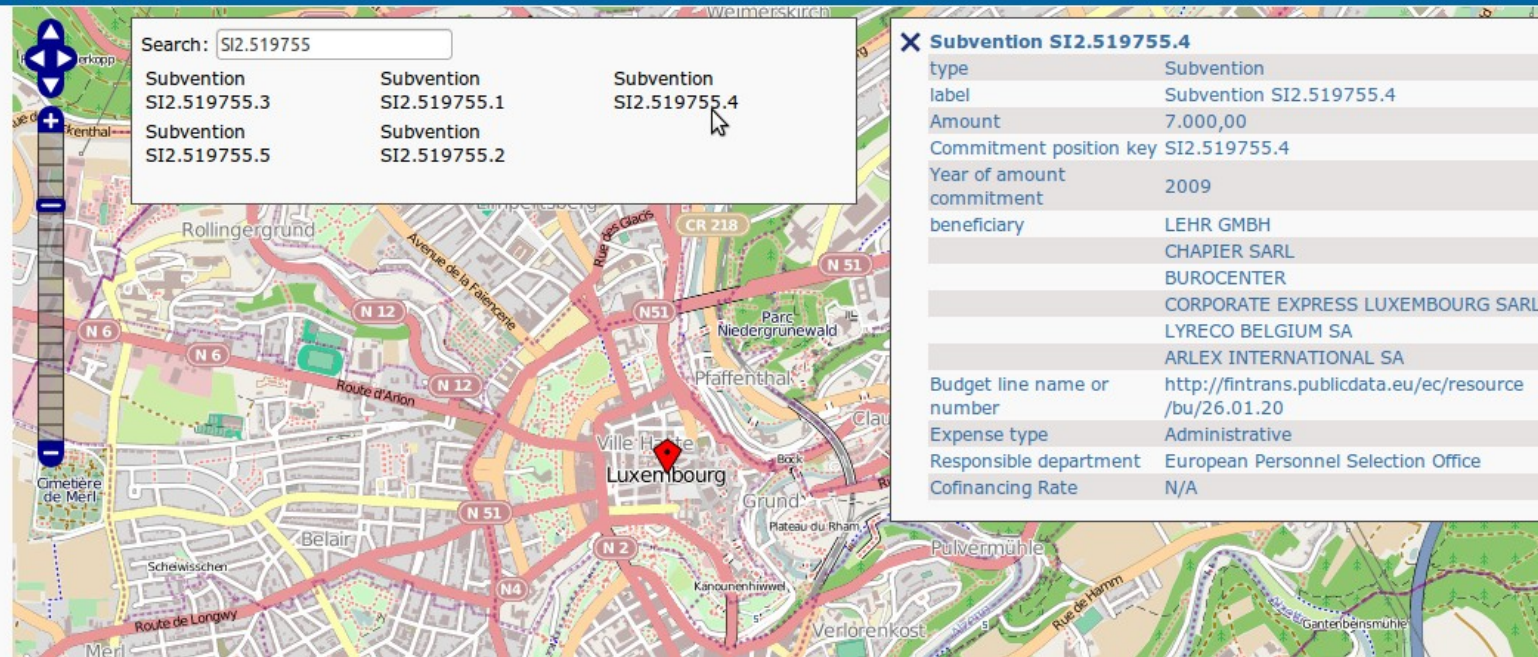
- **Luxembourg**
Europe
- **Luxembourg**
Europe
- **Luxembourg**
Canton Luxembourg
- **Luxembourg (Province de)**
French Community
- **Luxembourg**
Canton Luxembourg
- **Luxembourg**
The Legacy South
- **Luxembourg**
Don Bosco
- **Luxembourg**
Boulevard du Luxembourg

Search:

Subvention SI2.519755.3	Subvention SI2.519755.1	Subvention SI2.519755.4
Subvention SI2.519755.5	Subvention SI2.519755.2	

Subvention SI2.519755.4

type	Subvention
label	Subvention SI2.519755.4
Amount	7.000,00
Commitment position key	SI2.519755.4
Year of amount commitment	2009
beneficiary	LEHR GMBH CHAPIER SARL BUROCENTER CORPORATE EXPRESS LUXEMBOURG SARL LYRECO BELGIUM SA ARLEX INTERNATIONAL SA
Budget line name or number	http://fintrans.publicdata.eu/ec/resource/bu/26.01.20
Expense type	Administrative
Responsible department	European Personnel Selection Office
Cofinancing Rate	N/A



Making the web an exploratory place for geospatial data

http://geoknow.eu



http://stack.lod2.eu/

Silk Workbench

Workspace: movies_example

Editor: movies

Generate Links

Learn

Reference Links

Population

About

Undo

Redo

Export as Silk-LS

Help

Precision = 1.00

Recall = 0.80

F-measure = 0.89



Property Paths

Source: **DBpedia**

Restriction: ?a ?p ?v .

- (custom path)
- ?a/dbpediaowl:releaseDate

Target: **linkedmdb**

Restriction: ?b ?p ?v .

- (custom path)
- ?b/linkedmdb:initial_release_date

Transformations

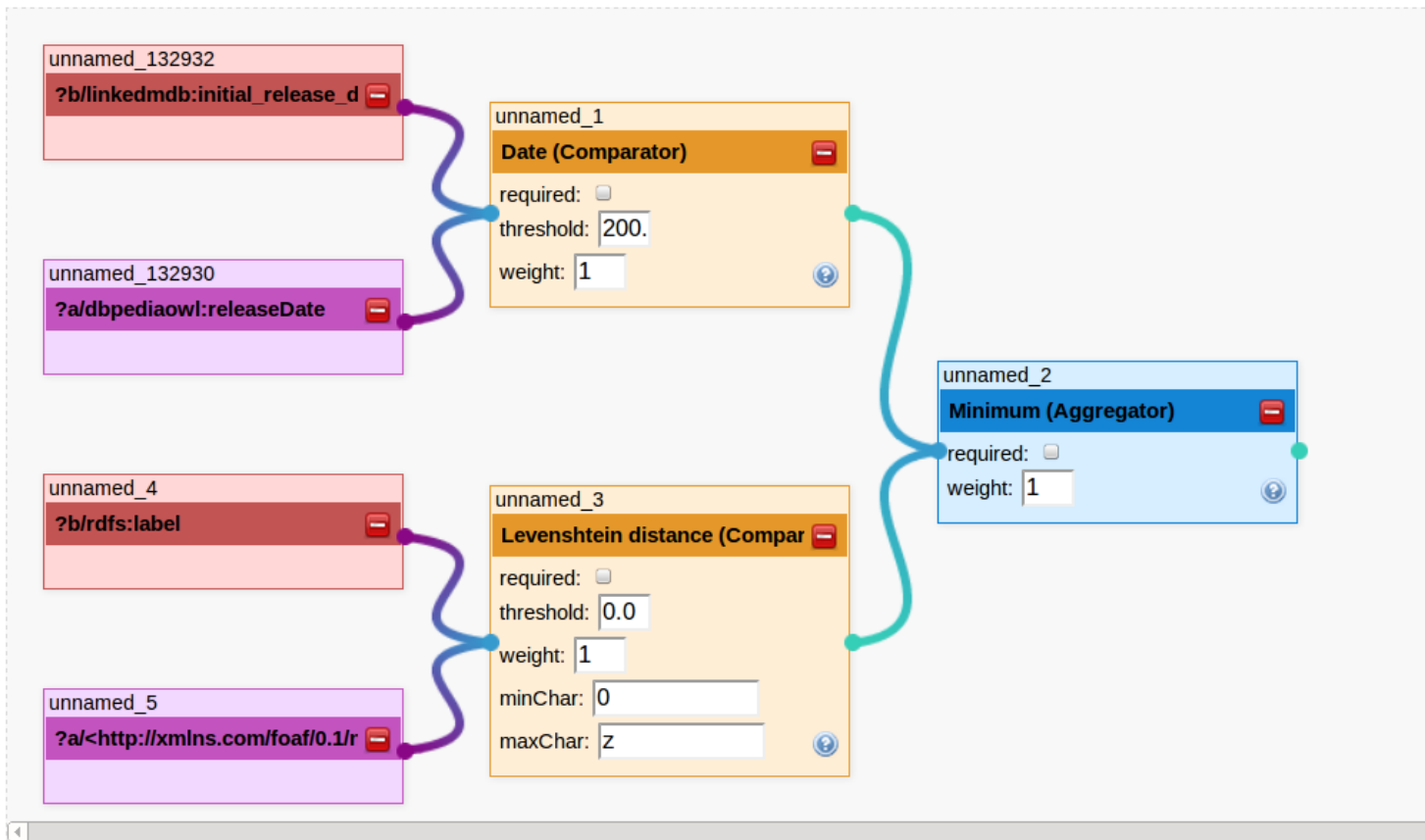
- Alpha reduce
- Capitalize

Comparators

- Date
- DateTime

Aggregators

- Average
- Euclidian distance



Link Limit: unlimited



<http://stack.lod2.eu/>

<http://demo.lod2.eu/lod2demo>

Upload RDF file
 Load RDF data from CKAN
Extract RDF from XML ▶ Basic extraction
 Extended extraction
 Extract RDF from SQL
 Extract RDF from text w.r.t. DBpedia
 Extract RDF from text w.r.t. a controlled vocabulary

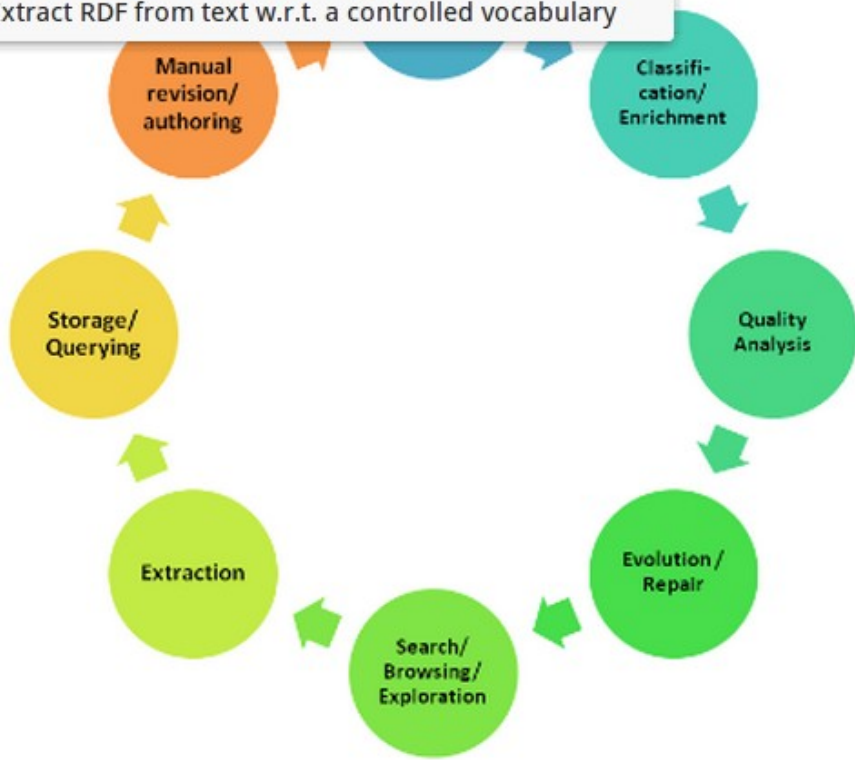
This is Version 1.0 of the LOD2 Stack, which comprises a number of tools for managing the life-cycle of Linked Data. The life-cycle comprises in particular the stages

- Extraction of RDF from text, XML and SQL
- Querying and Exploration using SPARQL
- Authoring of Linked Data using a Semantic Wiki
- Semi-automatic link discovery between Linked Data sources
- Knowledge-base Enrichment and Repair

You can access tools for each of these stages using the menu on top.

The LOD2 Stack is developed by the LOD2 project consortium comprising 15 research groups and companies. The LOD2 project is co-funded by the European Commission within the 7th Framework Programme (GA no. 257934).

You can find further information about the LOD2 Stack on the LOD2 project at <http://lod2.eu>.





Knowledge Extraction Article in Wikipedia

Article [Talk](#)

F

Knowledge extraction

From Wikipedia, the free encyclopedia

Knowledge Extraction is the creation of **knowledge** from structured (relational databases, XML) and unstructured (text, ...) in a machine-readable and machine-interpretable format and must **represent knowledge** in a manner that facilitates inference (NLP) and ETL (Data Warehouse), the main criteria is that the extraction result goes beyond the creation of structured information either the reuse of existing **formal knowledge** (reusing identifiers or ontologies) or the generation of a schema based on the data. The RDB2RDF W3C group ^[1] is currently standardizing a language for extraction of RDF from relational databases. Another goal is the mapping of Wikipedia into structured data and also the mapping to existing **knowledge** (see **DBpedia**, **Freebase** and ^[2]).

Survey of Methods / Tools

[\[edit\]](#)

Name	Data Source	Data Exposition	Data Synchronisation	Mapping Language	Vocabulary Reuse	Mapping Automat.	Req. Domain Ontology	Uses GUI
A Direct Mapping of Relational Data to RDF	Relational Data	SPARQL/ETL	dynamic	N/A	false	automatic	false	false
CSV2RDF4LOD	CSV	ETL	static	RDF	true	manual	false	false
Convert2RDF	Delimited text file	ETL	static	RDF/DAML	true	manual	false	true
D2R Server	RDB	SPARQL	bi-directional	D2R Map	true	manual	false	false
DartGrid	RDB	own query language	dynamic	Visual Tool	true	manual	false	true
DataMaster	RDB	ETL	static	proprietary	true	manual	true	true
Google Refine's RDF Extension	CSV, XML	ETL	static	none		semi-automatic	false	true
Krextor	XML	ETL	static	xslt	true	manual	true	false
MAPONTO	RDB	ETL	static	proprietary	true	manual	true	false



Knowledge Extraction from Text

LOD2 EU Project produces LOD2 Stack.

Three requirements to unlock Natural Language Processing (NLP) for the project:

1. NLP tool output is required to be in RDF
2. Scalability (less triples, focus on usefulness)
3. Common vocabulary to integrate and use NLP tools



NLP Interchange Format 2.0

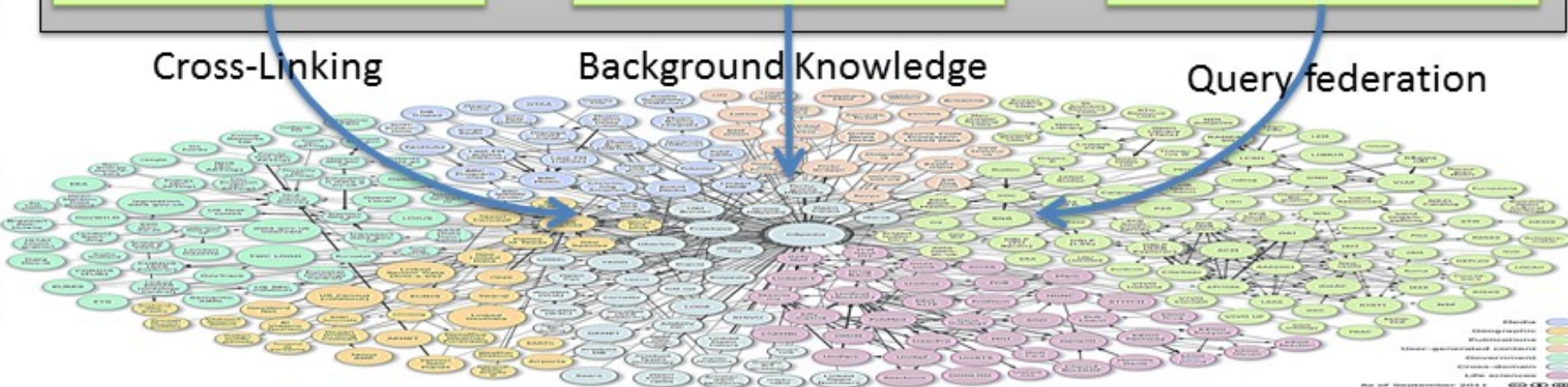
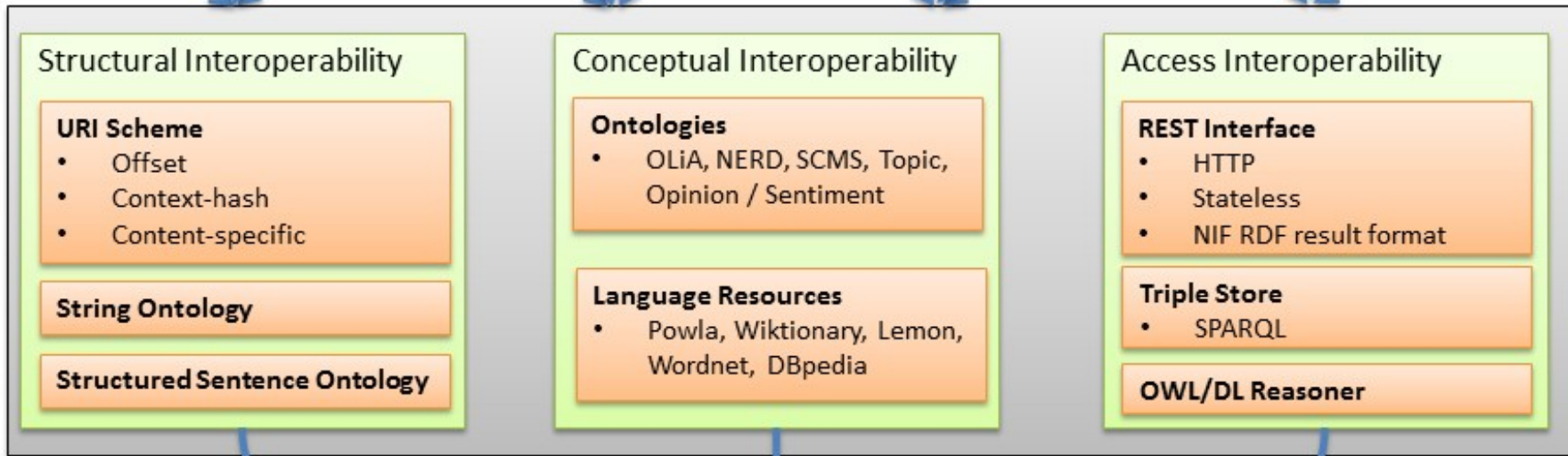
LOD2 EU Project produces LOD2 Stack.

Three requirements to unlock Natural Language Processing (NLP) for the project:

1. NLP tool output is required to be in RDF
2. Scalability (less triples, focus on usefulness)
3. Common vocabulary to integrate and use NLP tools

The **NLP Interchange Format (NIF)** is an RDF/OWL-based format that aims to achieve interoperability between Natural Language Processing (NLP) tools, language resources and annotations.

- Version 1.0 published in November 2011
- Version 2.0 is scheduled for completion within 2013





Addressing Primary Data

http://www.w3.org/DesignIssues/LinkedData.html



Google



Tim Berners-Lee

Date: 2006-07-27, last change: \$Date: 2009/06/18 18:24:33 \$

Status: personal view only. Editing status: imperfect but published.

[Up to Design Issues](#)

Linked Data

The **Semantic Web** isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other, 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, for data they links between arbitrary things described by RDF. The URIs identify any kind of object or concept. But for HTML or RDF, the same expectations apply to make the web grow:

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

Simple. In fact, though, a surprising amount of data isn't linked in 2006, because of problems with one or more of the steps. This article discusses solutions to these problems, details of implementation, and factors affecting choices about how you publish your data.

The four rules

I'll refer to the steps above as rules, but they are expectations of behavior. Breaking them does not destroy anything, but misses an opportunity to make data interconnected. This in turn limits the ways it can later be reused in unexpected ways. It is the unexpected re-use of information which is the value added by the web.





Addressing Primary Data

NIF 1.0:

http://www.w3.org/DesignIssues/LinkedData.html#offset_717_729

NIF 2.0 uses RFC 5147:

<http://www.w3.org/DesignIssues/LinkedData.html#char=717,729>

User extensions possible:

http://www.w3.org/DesignIssues/LinkedData.html#your_own_scheme

(but you have to link to documentation on how it was created)



As a Web Service

curl

```
--data-urlencode prefix="http://prefix.given.by/theClient#"
--data-urlencode input="[...]"
(--data-urlencode source="http://www.w3.org/DesignIssues/LinkedData.html"
http://nlp2rdf.lod2.eu/demo/NIFStanfordCore
```

```
http://nlp2rdf.lod2.eu/demo/NIFStanfordCore
```

```
http://www.w3.org/DesignIssues/LinkedData.html"
```

```
http://nlp2rdf.lod2.eu/demo/NIFStanfordCore
```

```
@prefix : <http://prefix.given.by/theClient#> .
@prefix str: <http://nlp2rdf.lod2.eu/schema/string/> .
:offset_717_729
  rdf:type str:OffsetBasedString ;
  str:anchorOf "Semantic Web"
  str:referenceContext :offset_0_26546 .
:offset_0_26546
  rdf:type str:Context ;
# the exact retrieval method is left underspecified
  str:occursIn <http://www.w3.org/DesignIssues/LinkedData.html> ;
# [...] are all 26546 characters as rdf:Literal
  str:isString "[...]" .
```



Vocabulary Module: OLiA

- Tibeto-Burman languages: <http://purl.org/olia/tibet.owl#VNst>
- Russian TreeTagger :
http://purl.org/olia/russ.owl#partizip_prt_sg_neut_passiv_gen_langform
- German STTS: <http://purl.org/olia/stts.owl#VAPP>
- English Penn: <http://purl.org/olia/penn.owl#VBG>

→ all map to <http://purl.org/olia/olia.owl#NonFiniteVerb>

Ontologies of Linguistic Annotation (OLiA) contain mappings for over 50 Tagsets (free and open, CC-By)



NIF 2.0 - plans

- NIF 2.0 tries to be compatible to (**Vocabulary Module**):
 - **ITS 2.0**
 - FISE used in Apache Stanbol (IKS-EU Project)
 - LAF/GrAF XML – ISO standard, recently published
 - Fragment Identifiers by IETF and W3C
 - Lemon ontology from Monnet EU Project
 - NERD ontology from EURECOM and LinkedTV EU Project
 - Xpointer/XPath URI scheme
 - Open Annotation





How you can contribute:

NIF 2.0 :

- NIF is free and open (CC-0 or CC-BY)
- All ontologies will be hosted for persistently by University Leipzig
- Sign up on the mailinglist at <http://nlp2rdf.org>
- Provide Use Cases, Requirements, Implementations at:
 - http://wiki.nlp2rdf.org/wiki/Use_cases#Use_cases
 - <http://wiki.nlp2rdf.org/wiki/Requirements#Requirements>



How you can contribute:

LOD 2 Stack

- Currently project half-time
- Most of the tools are free and open source
- Commercial rollout planned
- Many webinars available
- **You can integrate your tool via Debian package**

<http://lod2.eu>

<http://stack.lod2.eu/>



Thanks for your attention

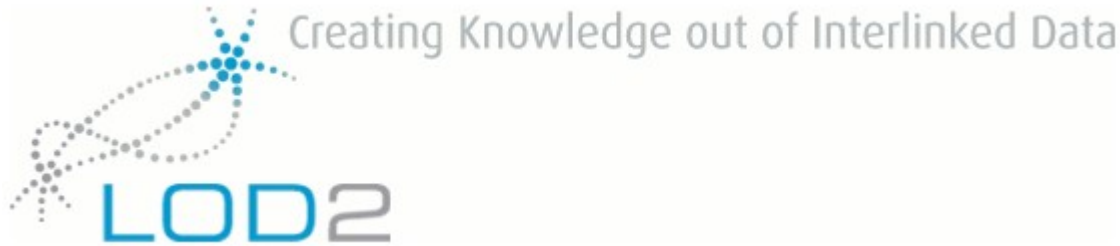
Open Community - All feedback is welcome!

<http://slideshare.net/kurzum>

Websites:

<http://nlp2rdf.org>

<http://lod2.eu>



Making the web an
exploratory place
for geospatial data

