# LinkedSpending: OpenSpending becomes Linked Open Data

Konrad Höffner

October 5, 2013

## Open Spending Data

- government spending data available to everyone
- what amount of money spent for which purpose, when and by which department
- demanded by the public
- increases accountability
- reduces corruption
- voters can make better informed decisions $\rightarrow$ strengthens democracy
- strengthens the government itself, more likely to commit to large projects

## Source Data

- http://openspending.org
- open platform
- public finance data from governments around the world
- more than 350 datasets, more than 17 million transactions
- updated regularily (about a dataset a week)

## Source Data

```
"sub-programme": {
 "label":"Security and safeguarding
     liberties",
 "html_url":"http://openspending.org/eu-
     budget/sub-programme/security-and-
     safeguarding-liberties",
 "name":"security-and-safeguarding-liberties
     "
},
"html_url": "http://openspending.org/eu-
    budget/entries/017dfcb58d05671ef9eb5a9f77
    fef39c8b14150c",
"amount": 41.2
```

Figure : simplified excerpt from an OpenSpending *entry*

# Why Convert the Data to RDF?

Problems with source data

- source data is structured (database) but not semantic
- open, but not linked
- data silo: own format, not interlinked to other knowledge bases, hard to integrate

Benefits of RDF

- multiple ways of access for both machines and people: resolving of URIs, SPARQL, RDF dump
- use of Linked Open Vocabularies: common vocabulary $\rightarrow$ easier integration with other spending data
- semantic web infrastructure such as Question Answering

- JSON API, no bulk download
- frequent changes
- errors in the data
- source data uses specific model for statistical observations: *data cube*
- big amount of data $\rightarrow$ performane is important to not have long waiting times
- common data model but datasets have different vocabulary

| Problems | Solution |
| --- | --- |
| JSON API | custom Java program, defining JSON path expressions |
| changes | two step process: (1) download all JSON resources (2) convert JSON to RDF |
| errors | defining error rate thresholds for accept/decline of datasets |
| data cube model | use of Linked Open Vocabulary *RDF data cube* |
| performance | use of persistent caching |
| different vocabulary | not yet solved, needs more time (cooperation with experts, student thesis?) |

## Outcome: LinkedSpending

|  | **Total** | **Average** |
|---:|---|---|
| **number of datasets** | 247 | |
| **filesize (RDF/N-Triples)** | 10 GB | 41 MB |
| **triples** | 50 million | 200 000 |
| **observations** | 2.4 million | 10 000 |

Table : total and average values (approximate)

- available at `http://linkedspending.aksw.org/`[1]
- RDF dump for access of the whole dataset
- public SPARQL endpoint for queries
- OntoWiki instance for browsing

---

[1] still under development

| information need | SPARQL Query |
|---|---|
| 1 all years which have observations in the de-bund dataset from 2020 onwards | `select distinct ?date {?o a qb:Observation .`<br>`?o qb:dataSet ls:de-bund .`<br>`?o sdmxd:refPeriod ?date .`<br>`FILTER (xsd:date(?date) >= "2020-1-1"^^xsd:date) }` |
| 2 spendings of more than 100 billion € | `select * {?o ls:amount ?a. ?o dbo:currency dbpedia:Euro.`<br>`FILTER(xsd:integer(?a)>"1E11"^^xsd:integer) }` |
| 3 datasets with multiple years | `select ?d count(?y) as ?count { ?d a qb:DataSet.`<br>`?d ls:refYear ?y. } having (count(?y)>1)` |
| 4 sums of amounts for each reference year of the dataset berlin_de | `select ?y (sum(xsd:integer(?amount)) as ?sum)`<br>`{?o qb:dataSet ls:berlin_de. ?o ls:refYear ?y. ?o ls:amount ?amou` |
| 5 datasets with currencies whose inflation rate is greater than 10 % | `select distinct ?d ?c ?r {?o qb:dataSet ?d. ?o`<br>`dbo:currency ?c. ?c dbp:inflationRate ?r. filter(?r > 10)}` |
| 6 Berlin city subsectors of research and education that have had their budget reduced from 2012 to 2013 | `...` |

| dl | cl | callret-2 |
|---|---|---|
| All Expenditure | "塞拉利昂利昂"@zh | 12 |
| Non-Salary Non-interest Budget Expenditure 2013 | "塞拉利昂利昂"@zh | 12 |
| Capital Expenditure Budget 2013 | "塞拉利昂利昂"@zh | 12 |
| All Expenditure | "Leone"@ca | 12 |
| Non-Salary Non-interest Budget Expenditure 2013 | "Leone"@ca | 12 |
| Capital Expenditure Budget 2013 | "Leone"@ca | 12 |
| All Expenditure | "Sierra Leonean leone"@en | 12 |
| Non-Salary Non-interest Budget Expenditure 2013 | "Sierra Leonean leone"@en | 12 |
| Capital Expenditure Budget 2013 | "Sierra Leonean leone"@en | 12 |
| All Expenditure | "Sierra-leonischer Leone"@de | 12 |
| Non-Salary Non-interest Budget Expenditure 2013 | "Sierra-leonischer Leone"@de | 12 |
| Capital Expenditure Budget 2013 | "Sierra-leonischer Leone"@de | 12 |