

Requirements on Linked Data Consumption Platform

Jakub Klímeck, Petr Škoda, Martin Nečaský

Charles University in Prague
Faculty of Mathematics and Physics



Motivation: The 4th star problem

```
<https://data.cssz.cz/resource/observation/prehled-o-celkovem-poctu-osvc-podle-kraju/2010-03-31/VC.19/dobrovolne-dp/> a qb:Observation ;
  cssz-dimension:datum <https://data.cssz.cz/resource/reference.data.gov.uk/id/gregorian-day/2010-03-31> ;
  cssz-dimension:kraj <https://data.cssz.cz/resource/ruian/vusc/19> ;
  qb:measureType cssz-measure:dobrovolne-dp ;
  cssz-measure:dobrovolne-dp 59 ;
  qb:dataSet cssz-dataset:prehled-o-celkovem-poctu-osvc-podle-kraju .
```

```
<https://data.cssz.cz/resource/observation/prehled-o-celkovem-poctu-osvc-podle-kraju/2010-03-31/VC.19/dobrovolne-np/> a qb:Observation ;
  cssz-dimension:datum <https://data.cssz.cz/resource/reference.data.gov.uk/id/gregorian-day/2010-03-31> ;
  cssz-dimension:kraj <https://data.cssz.cz/resource/ruian/vusc/19> ;
  qb:measureType cssz-measure:dobrovolne-np ;
  cssz-measure:dobrovolne-np 13940 ;
  qb:dataSet cssz-dataset:prehled-o-celkovem-poctu-osvc-podle-kraju .
```

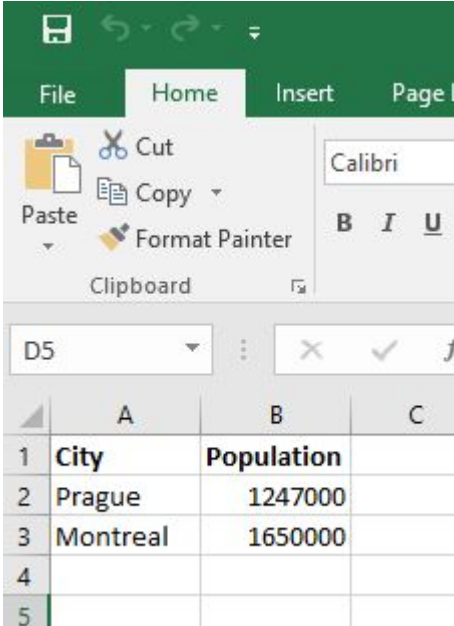
Motivation: The 4th star problem



Motivation: The 4th star problem

A bit unfair, something like showing the insides of an Excel file

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<worksheet xmlns="http://schemas.openxmlformats.
org/spreadsheetml/2006/main" xmlns:r="http://schemas.openxmlformats.
org/officeDocument/2006/relationships" xmlns:mc="http://schemas.
openxmlformats.org/markup-compatibility/2006" mc:Ignorable="x14ac"
xmlns:x14ac="http://schemas.microsoft.
com/office/spreadsheetml/2009/9/ac"><dimension ref="A1:B3"
/><sheetViews><sheetView tabSelected="1" workbookViewId="0"
/></sheetViews><sheetFormatPr defaultRowHeight="15" x14ac:dyDescent="
0.25"/><cols><col min="1" max="2" width="10.7109375" bestFit="1"
customWidth="1"/></cols><sheetData><row r="1" spans="1:2" s="1"
customFormat="1" x14ac:dyDescent="0.25"><c r="A1" s="1" t="s"
><v>1</v></c><c r="B1" s="1" t="s"><v>0</v></c></row><row r="2" spans="
1:2" x14ac:dyDescent="0.25"><c r="A2" t="s"><v>2</v></c><c r="B2"
><v>1247000</v></c></row><row r="3" spans="1:2" x14ac:dyDescent="0.25"
><c r="A3" t="s"><v>3</v></c><c r="B3"
><v>1650000</v></c></row></sheetData><pageMargins left="0.7" right="0.7"
top="0.75" bottom="0.75" header="0.3" footer="0.3"/><pageSetup
paperSize="9" orientation="portrait" horizontalDpi="4294967294"
verticalDpi="0" r:id="rId1"/></worksheet>
```



	A	B	C
1	City	Population	
2	Prague	1247000	
3	Montreal	1650000	
4			
5			

Motivation: Unmet expectations

- 4* are better than 3*. But for whom?
 - Grandma can open an Excel file. Can she open RDF?
 - Try uploading RDF files to Google drive
- Is this how it is supposed to be?
 - No!
 - There are standards
 - Discovery: DCAT, DCAT-AP
 - Syntax: RDF and serializations
 - Access: HTTP, SPARQL
 - Modelling: SKOS, DCV, ...
 - Missing: Tools
- What do the tools need to do to
 - Facilitate LOD consumption
 - Demonstrate the LOD benefits to consumers
- => 40 Requirements on Linked Data Consumption Platform (LDCCP)

3 requirements: Dataset discovery

1. Catalog support
 - CKAN API, DCAT-AP
2. Advanced discovery
 - Dataset indexing, such as Sindice, but possibly more advanced
3. Context-aware discovery
 - Recommendation of other relevant datasets based on the ones already selected for work

6 requirements: Data input

4. IRI dereferencing
 - Basic principle of Linked Data
5. RDF dump load
 - Turtle, RDF/XML, N-Triples, N-Quads, TriG, JSON-LD, RDFa
6. SPARQL querying
 - SELECT, CONSTRUCT, DESCRIBE, ASK
7. Linked Data Platform input
 - LDP Containers
8. Non-RDF data input
 - CSV, XML, JSON
9. Monitoring of input changes
 - Notifications, pipelines triggering

6 requirements: Dataset preview

10. Preview - [W3C vocabularies](#)
 - SKOS, ORG, DCV
11. Preview - [LOV vocabularies](#)
 - DCTERMS, GoodRelations, Schema.org, FOAF, vCard
12. Preview metadata
 - DCAT, DCAT-AP, VoID descriptions of datasets
13. Preview data
 - Statistics, description of datasets based on the actual data
14. Preview schema
 - Can be extracted using SPARQL queries
15. Quality indicators
 - Help users to decide whether to use a dataset or not
 - E.g. schema coverage, temporal coverage, geographical coverage, ...

2 requirements: Analysis of semantic relationships

16. Semantic relationship analysis

- Datasets interlinked?
- Shared resources?
- Temporal/geographic coverage overlapping?
- ...

17. Semantic relationship deduction

- Link discovery - SILK
- Ontology matching
- ...

7 requirements: Data manipulation

18. Vocabulary-based transformations
 - E.g. means of translating from FOAF to Schema.org, from WGS84_pos to Schema.org etc.
19. Vocabulary alignment
 - Possible semantic overlaps, suggest a transformation - ontology alignment
20. Inference
 - Inference rules, RDFS, OWL
21. Resource fusion
 - owl:sameAs, conflict resolution
22. Assisted selection and projection
 - SPARQL SELECT and FILTER or other means, graphically assisted
23. Custom transformations
 - Typical SPARQL
24. Automated data manipulation
 - Automatic transformation pipeline discovery based on some requirements

2 requirements: Provenance and license management

25. Provenance

- Record and provide provenance data (PROV-O)

26. License management

-

License Compatibility : Publishing Remixes

The combining use case covers scenarios where 2 (or more) datasets are being combined together. In this case there are at least two licenses that may influence the ability of the re-publisher to license the derived work. The table shows a 2 license scenario, the table cells show the "minimum" license that applies to the dataset.													
First License	Second License												
	CC0	CC-PDM	CC-BY-ND	CC-BY-NC-ND	CC-BY	CC-BY-SA	CC-BY-NC	CC-BY-NC-SA	ODC-PDDL	ODC-BY	ODC-ODbL	OGL 2.0	OS OpenData
CC0	No restrictions	No restrictions	-	-	CC-BY	CC-BY-SA	CC-BY-NC	CC-BY-NC-SA	No restrictions	ODC-BY	ODC-ODbL	OGL 2.0	OS OpenData
CC-PDM	No restrictions	No restrictions	-	-	CC-BY	CC-BY-SA	CC-BY-NC	CC-BY-NC-SA	No restrictions	ODC-BY	ODC-ODbL	OGL 2.0	OS OpenData
CC-BY-ND	-	-	-	-	-	-	-	-	-	-	-	-	-
CC-BY-NC-ND	-	-	-	-	-	-	-	-	-	-	-	-	-
CC-BY	CC-BY	CC-BY	-	-	CC-BY	CC-BY-SA	CC-BY-NC	CC-BY-NC-SA	CC-BY	CC-BY	ODC-ODbL	CC-BY	OS OpenData
CC-BY-SA	CC-BY-SA	CC-BY-SA	-	-	CC-BY-SA	CC-BY-SA	-	-	CC-BY-SA	CC-BY-SA	ODC-ODbL	CC-BY-SA	OS OpenData
CC-BY-NC	CC-BY-NC	CC-BY-NC	-	-	CC-BY-NC	-	CC-BY-NC	CC-BY-NC-SA	CC-BY-NC	CC-BY-NC	-	CC-BY-NC	OS OpenData
CC-BY-NC-SA	CC-BY-NC-SA	CC-BY-NC-SA	-	-	CC-BY-NC-SA	-	CC-BY-NC-SA	CC-BY-NC-SA	CC-BY-NC-SA	CC-BY-NC-SA	-	CC-BY-NC-SA	OS OpenData
ODC-PDDL	No restrictions	No restrictions	-	-	CC-BY	CC-BY-SA	CC-BY-NC	CC-BY-NC-SA	No restrictions	ODC-BY	ODC-ODbL	OGL 2.0	OS OpenData
ODC-BY	ODC-BY	ODC-BY	-	-	ODC-BY	CC-BY-SA	CC-BY-NC	CC-BY-NC-SA	ODC-BY	ODC-BY	ODC-ODbL	ODC-ODbL	OS OpenData
ODC-ODbL	ODC-ODbL	ODC-ODbL	-	-	ODC-ODbL	ODC-ODbL	-	ODC-ODbL	ODC-ODbL	ODC-ODbL	ODC-ODbL	ODC-ODbL	OS OpenData
OGL 2.0	OGL 2.0	OGL 2.0	-	-	CC-BY	CC-BY-SA	CC-BY-NC	CC-BY-NC-SA	OGL 2.0	ODC-BY	ODC-ODbL	ODC-ODbL	OS OpenData
OS OpenData	OS OpenData	OS OpenData	-	-	OS OpenData	OS OpenData	?	?	OS OpenData	OS OpenData	?	OS OpenData	OS OpenData
Partly derived from http://wiki.creativecommons.org/Wiki/cc_license_compatibility													
Note: there is a possible additional use case, which is aggregating data, unchanged. This would allow use of ND style licenses. But only if the source data remains unchanged, see: http://wiki.creativecommons.org/Frequently_Asked_Questions#If_I_create_a_collective_work_that_includes_a_work_offered_under_a_CC_license.2C_which_license.28s.29_may_I_c													

<https://github.com/theodi/open-data-licensing/blob/master/guides/licence-compatibility.md>

9 requirements: Data output and visualization

27. Manual visualization
 - User specifies, what should be in the data
28. Vocabulary-based visualization
 - Data is analyzed, visualization offered based on vocabularies
29. RDF dump output
 - Turtle, RDF/XML, N-Triples, N-Quads, TriG, JSON-LD, RDFa
30. SPARQL Update output
 - INSERT DATA
31. SPARQL Graph Store HTTP Protocol output
 - HTTP GET, HTTP PUT, HTTP DELETE, HTTP POST
32. Linked Data Platform output
 - LDP Containers
33. Tabular data output
 - SPARQL SELECT + CSV on the Web JSON-LD metadata
34. Tree-like data output
 - RDF/XML, JSON-LD or better support of mapping
35. Graph data output
 - Gephi, for images of graphs and linkage

5 requirements: Developer and community support

36. API

- APIs used by LDGP should be well-documented, standardized (REST) and usable by everyone

37. RDF configuration

- Need for configuration generation, best using one language - SPARQL

38. Repositories for sharing

- Sharing of plugins (Eclipse, ...)

39. Project reuse

- Sharing of reusable parts of consumption projects (GitHub)

40. Deployment of services

- When output is data, enable getting it/refreshing it through API

Our related efforts @ Charles University in Prague

- [LinkedPipes ETL](#)
 - Preparation and publication of RDF
 - Successor to UnifiedViews
- [LinkedPipes Visualization](#)
 - Vocabulary-based discovery of visualization pipelines
 - Successor to Payola and LDVMi
- Both going to be presented @ ESWC 2016 Demo Track, Crete, Greece

Thank you for your attention

klimek@opendata.cz

@jakub_klimek