

Publishing Linked Data Requires More than Just Using a Tool

G. Atemezing, F. Gandon, G. Kepeklian,
F. Scharffe, R. Troncy, B. Vatant, S. Villata

Datalift Data Camp – 09-10 October 2012

- 2 * 1-day camp
- Co-organized with data.gouv.fr
- 75 participants
- Goals
 - Present the Datalift platform
 - Bring your dataset ... we will lift it to 5 stars data, and interlink it with other datasets



Many working groups with a mentor

Domain	Illustrative scenario	Original dataset	Original format	Datasets for interlinking
Person	Find translations of given names to different languages.	Opendata.paris.fr	CSV	DBpedia
Tourism, Culture and Events	Lift cultural events data in the region Picardy	Yellow pages, Regional Tourism Office	XLS	EventMedia
Transport	Lift bus stops and lines transport	De Lijn bus company in Flanders (Belgium)	CSV GTFS	--
Data Catalogues	Convert data catalogues in DCAT and use thesaurus such as EUROVOC for terms classification	Opendata.gouv.fr, Opendata.montpelliernumerique.fr	XLS CSV	--
Budget of collectivities	Convert data using DataCube, use SPARQL aggregation functions	Rennes, Montpellier, Toulouse	XLS	rdf.insee.fr
Geolocation	Interlinking and publishing different shape files with temporal data	OpenstreetMap France Temporal series of agricultural data (confidential data)	SHP CSV	data.ign.fr rdf.insee.fr
Environment	Lift data of grapes of given parcel	Suez Environment INRA	CSV	DBpedia, rdf.insee.fr data.ign.fr

Datalift goals

- From raw data ...
to 5 stars interlinked data
- One click download:
 - <http://datalift.org/en/node/24>
 - https://gforge.inria.fr/scm/?group_id=2935
- Cross-platforms, triple store installed and configured
 - Sesame (default)
 - OWLIM, AllegroGraph, Virtuoso (connectors)

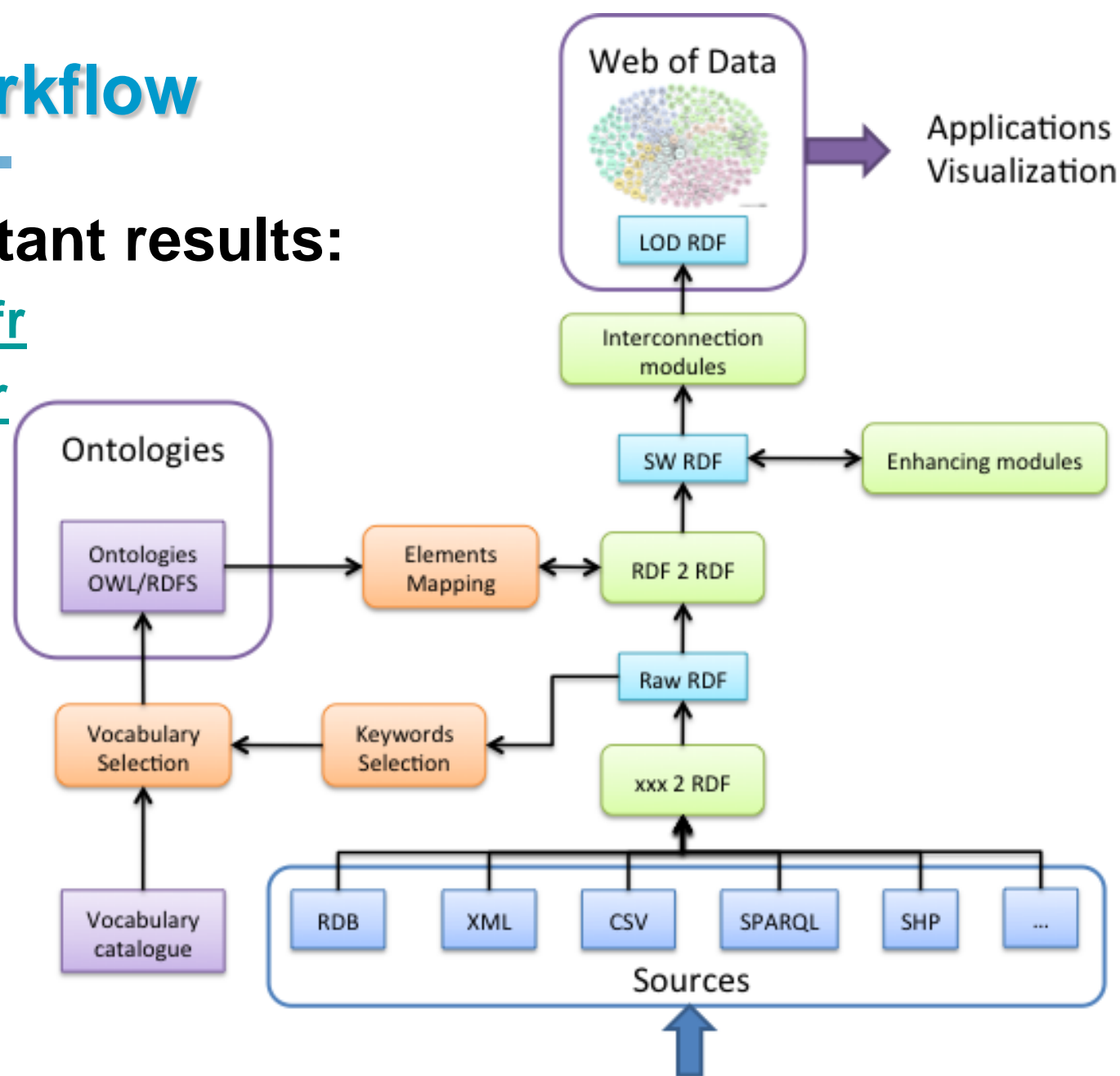


Datalift workflow

- Two important results:

- rdf.insee.fr

- data.ign.fr



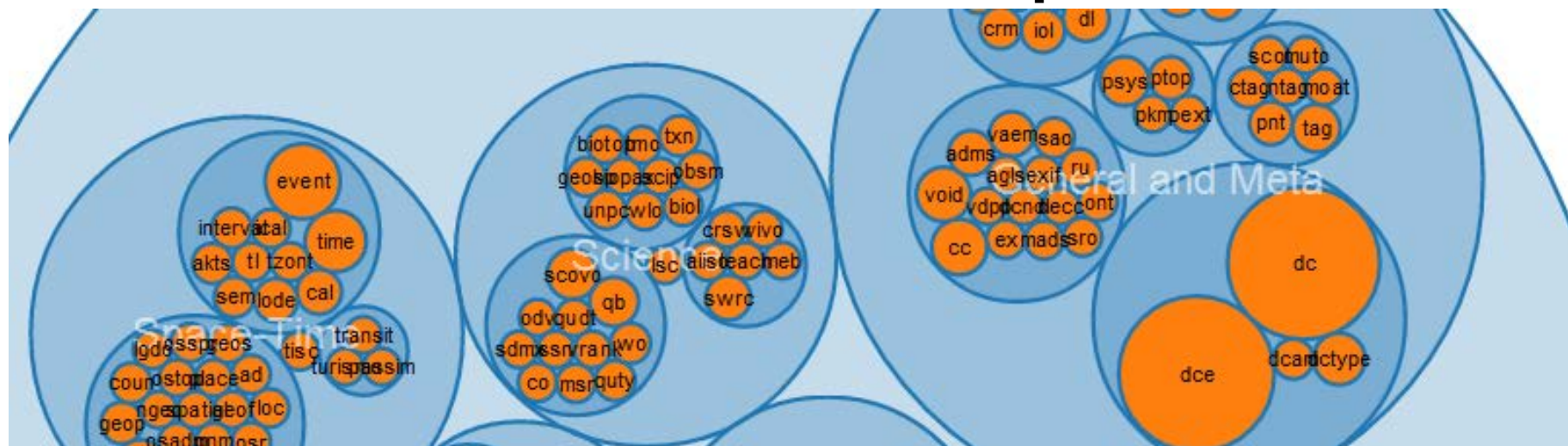
Difficulties and wishes / need support

- **Vocabulary: choose suitable vocabularies to describe the data**
 - Crash course on class/property modeling
- **RDF Conversion:**
 - Adopt a URI naming scheme to identify the data objects
 - Simple functionality proves useful: *String2URI* module
- **Automatic detection of datasets to link to**
 - Dbpedia is not the only dataset you want to link to

Linked Open Vocabularies (LOV)



- Selection of 260+ vocabularies (20k elements), <http://lov.okfn.org/>
 - LOV bot: <https://github.com/pyvandenbussche/lov>
 - LOV search: <https://github.com/pyvandenbussche/lovSearch>
 - LOV API: <https://github.com/pyvandenbussche/lovAPI>
- LOV module within the Datalift platform



Conclusion

- Multilingual vocabularies are important
- Hide the complexity of SPARQL with natural language QA systems such as QAKIS
- SHP files are important, need more tooling
- INSPIRE directive and W3C GLD vocabularies
- GTFS / DSPL and semantic web formats

“A little data lifted goes a long way”.

