

A Lightweight Approach to Explore, Enrich and Use Data with a Geospatial Dimension using Semantic Web Technologies

Christophe Debruyne¹ ♦ Kris McGlinn¹ ♦ Lorraine McNerney² ♦ Declan O’Sullivan¹

¹ ADAPT, Trinity College Dublin, Dublin 2, Ireland

² Ordnance Survey Ireland, Phoenix Park, Dublin 8, Ireland

Context

- The Open Data Engagement Fund is an initiative by the Irish public administration to improve the availability and usage of datasets on Ireland's open data portal data.gov.ie.

Problem

- The vast majority of datasets on this portal have an explicit or implicit spatial dimension, which we would like to link with Ordnance Survey Ireland's (OSi – Ireland's national mapping agency) authoritative geospatial datasets published on data.geohive.ie.
- This would allow one to align, analyze and engage with the data with a geospatial dimension.
- But transforming, enriching, analyzing, etc. data often requires a bespoke set of tools that may be difficult to install, configure and to use for merely enriching datasets.

Research Question

- How can we leverage the process of enriching and engaging with existing datasets and authoritative geospatial data on the data.gov.ie portal?

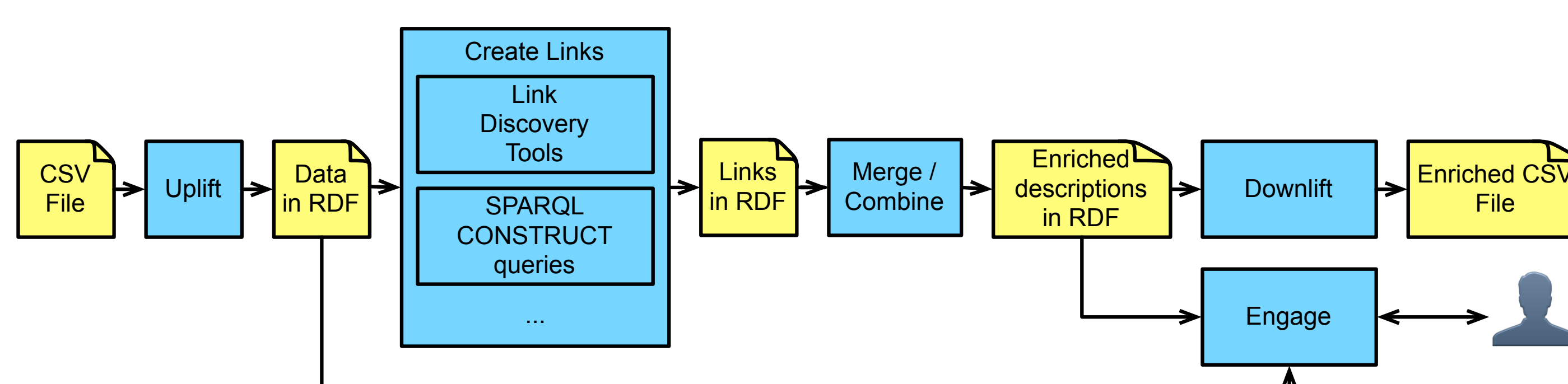
Goal

- To propose a lightweight method and tools for transforming CSV into RDF (uplift); interlink, enrich, and engage with the data; and transform the enriched RDF back into a CSV containing additional information (downlift).

Potential Impact

- Increase usage and availability of datasets by referring to authoritative geospatial datasets
- RDF representations of datasets, also improving availability
- Go beyond 3★ open data by providing a basis for Linked Data

Approach



- **Generating RDF from non-RDF**, a process called uplift, by generating an declarative mapping that reflects the CSV files original structure. The mapping language adopted is R2RML, a W3C standard for relating relational databases to RDF datasets. We thus treat CSV files as relational tables.
- **Create links and enrich** the RDF with SPARQL CONSTRUCT queries or link discovery tools.
- **Transform the enriched RDF back to CSV**, a process called downlift, by taking into account the original CSV file's structure captured in the mapping.
- **Engaging** with the data using a Triple Pattern Fragments (Verborgh et al. 2016) allowing us to formulate federated queries. A GeoSPARQL extension for the TPF client was developed to avail of spatial functions (Debruyne et al. 2016).

```
@prefix odef: <http://adaptcentre.ie/ont/odef#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix csv: <file:///...#> .
<?TripleMap>

rr:logicalTable [
  rr:sqlQuery ""SELECT rownum() AS ROW_NUM, *
    FROM fcweatherstationsp201108292221;""
];
rr:subjectMap [
  rr:template "http://www.example.org/record/{ROW_NUM}" ;
  rr:class odef:Record ;
];
rr:predicateObjectMap [
  rr:predicate csv:LONG ;
  odef:label "LONG" ;
  odef:order "5";
  rr:objectMap [ rr:column "LONG" ] ;
];
# Other predicate-object maps omitted for brevity
rr:predicateObjectMap [
  rr:predicate csv:ROW_NUM ;
  rr:column [ rr:column "ROW_NUM" ] ;
]
```

[illegible]

Results and Discussion

- Demonstrated the viability of our approach
- Method and tool available with accessible licenses
- How do we deal with loss of semantics? Provenance can be a solution (cfr. future work)
- Organization of seminars to inform public and stakeholders

Future Work

- Apply approach to more and different datasets (ongoing)
- Integrate approach with a project and provenance-centric mapping governance framework, which would ...
- Facilitate reuse (of vocabularies), traceability, and transparency of both uplift and enrichment processes.

References

- C. Debruyne, É. Clinton, D. O'Sullivan: Client-side Processing of GeoSPARQL Functions with Triple Pattern Fragments. LDOW@WWW 2017
- R. Verborgh, M. Vander Sande, O. Hartig, J. Van Herwegen, L. De Vocht, B. De Meester, G. Haesendonck, and P. Colpaert. 2016. Triple Pattern Fragments: A low-cost knowledge graph interface for the Web. J. Web Sem. 37-38 (2016), 184–206