

Evaluating automatic interlinking of authoritative geospatial data with Linked Open Data (LOD)

Co-Developed with:



- Automated Linked Data interlinking using state-of-art tools
- Facilitating extension of Ordnance Survey Ireland Linked Data with information from Linked Open Data cloud

Study Aim

- Evaluate use of state-of-art machine learning based automated Link Discovery Frameworks with real-world datasets
- Link Ordnance Survey Ireland (OSi) instances in county and townland authoritative datasets to corresponding instances in the Linked Open Data (LOD) cloud, in particular DBpedia¹ and LinkedGeoData²
- Generate and verify at least 50 links

Results (Manually Verified)

26 links from OSi county dataset to DBpedia

5 links from OSi county dataset to LinkedGeoData

25 links from OSi townland dataset to DBpedia

26 links from OSi townland dataset to LinkedGeoData

Challenges Faced

- Trial and error to decide metric specification and thresholds for machine learning algorithms
- Manual verification of generated links
- No class for a county or townland in LinkedGeoData
- No class for a townland in DBpedia
- Several townland instances in DBpedia without the property for the country being Republic of Ireland
- Several townland instances in DBpedia without a geometrical specification

Insights

- Non-uniform naming convention for labels across different datasets, presenting problems for fully automatic solutions
- Multiple naming conventions for labels of townlands from different counties within DBpedia
- Challenging to link townlands because of several townlands that have the same names and that have very close geometries
- Additional properties that could have facilitated better interlinking (like population or area of a place) unavailable directly in the OSi dataset, but as shown in our other work linkage to Central Statistics Office data (CSO) will facilitate this

Peru Bhardwaj, Dr. Christophe Debruyne, Prof. Declan O’Sullivan, ADAPT @ Trinity College Dublin, Ireland.

¹<http://wiki.dbpedia.org/>

²<http://linkedgeo.org/About>

³<http://aksw.org/Projects/LIMES.html>

⁴<http://silkframework.org/>

Experiment Methodology

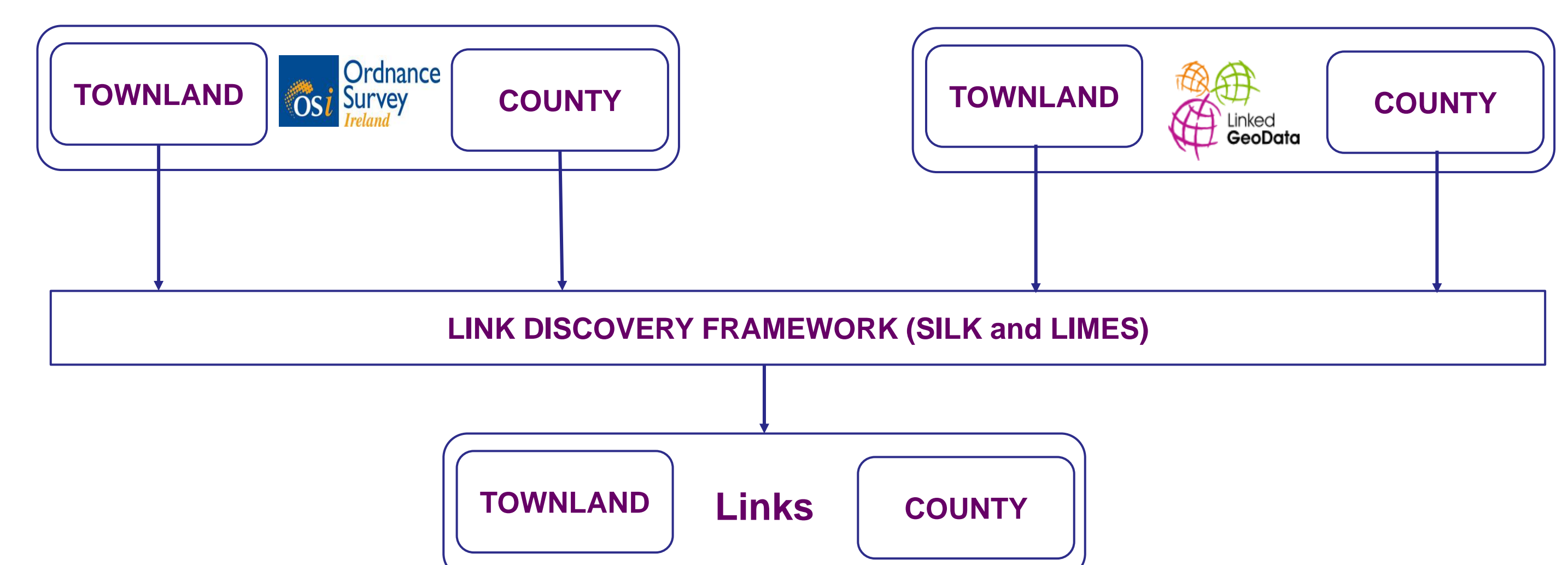
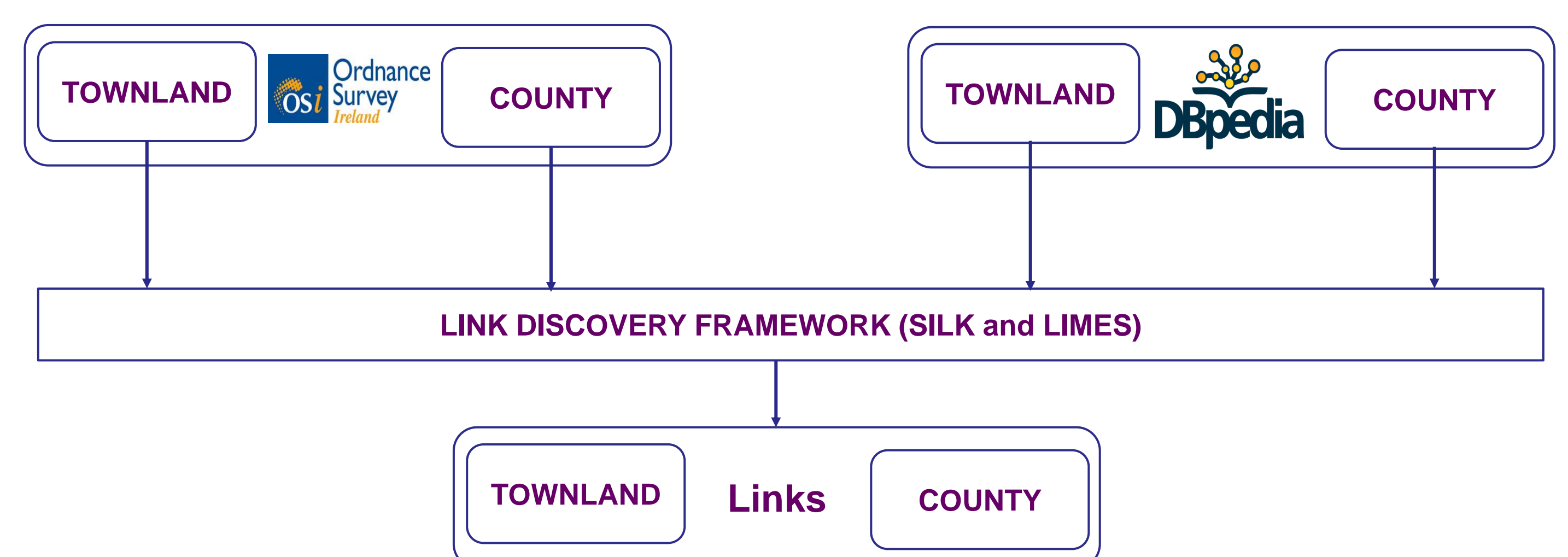
- Target datasets chosen :



AND



- SPARQL endpoints were queried and resulting triples stored in a local file for use
- Link Discovery Frameworks Tools evaluated: LIMES³ and SILK⁴
 - Similarity threshold used : 0.95
 - String matching using jaccard similarity measure
 - Geometry matching using geo-frechet (for LIMES) and wgs84 (for SILK) similarity measure



To learn more about innovative ADAPT technologies, contact: collaboration@adaptcentre.ie