# Linked Location Data as a Service

Stijn Goedertier[a], Nikolaos Loutas[a], Vassilios Peristeras[b]

[a] PwC EU Services, Technology Consulting
[b] European Commission, DG DIGIT

**Geographic information is provided in heterogeneous formats, creating technical and semantic barriers that hinder data consumers to combine data from various sources. Linked Data design principles can help alleviate these barriers and realise three generic use cases for data consumers to consume location data as linked data. We have demonstrated the technical feasibility for this in a Linked Data pilot that integrates address data from five different public sector organisations in Belgium. The pilot demonstrates that a Linked Data layer can be built on top of an existing geospatial implementation with a minimum of effort. It also shows that URI sets for INSPIRE spatial objects and spatial things can accommodate both XML (GML) and RDF representations.**

Visit the pilot at: http://location.testproject.eu/BEL/

## 1. INTRODUCTION: THREE GENERIC USE CASES

European public administrations maintain base registers with *authentic* location data. Examples of such base registers are the British National Street Gazetteer [NSG], the Dutch Building and Address register [BAG], or the Flemish Central Reference Database for Addresses [CRAB]. According to the European Interoperability Framework [EIF], base registers are the cornerstone of public services. The EIF recommends public administrations to develop interfaces to authentic sources and align them at semantic and technical level.

In the geospatial domain, Linked Data Design principles [TBL] have recently been proposed as a means to improve interoperability of spatial data services. Annex H of the INSPIRE Generic Conceptual model, for example, recommends the use of persistent HTTP URIs to identify spatial things and spatial objects. It mentions the following benefits:

- Enables the deployment of federated governance procedures
- Provides access to individual objects, while it remains possible to download whole datasets
- URI dereferencing and access to alternate formats (e.g. GML, RDF, JSON, CSV,...)
- Enables efficient management of shared data within and into business systems

The use of Linked Data design principles enables the following generic use cases for data consumers, which are illustrated in Figure 1 for address data:

- **Use Case 1: Disambiguate** a location by attributing an identifier (a URI) to it. A consumer queries a spatial data service to determine which location object identifier(s) (URIs) corresponds to a particular notation.
- **Use Case 2: Look up** (de-reference or resolve) a location object identifier in different formats. A user looks-up further information about a spatial object by resolving its HTTP URI. The system returns machine-readable information about the address in the format requested (e.g. GML, RDF-XML, JSON-LD notation)..
- **Use Case 3: Link** data sets via unique address identifiers. This use case is a consequence of using *common* address identifiers (preferably HTTP URIs) to denote addresses.

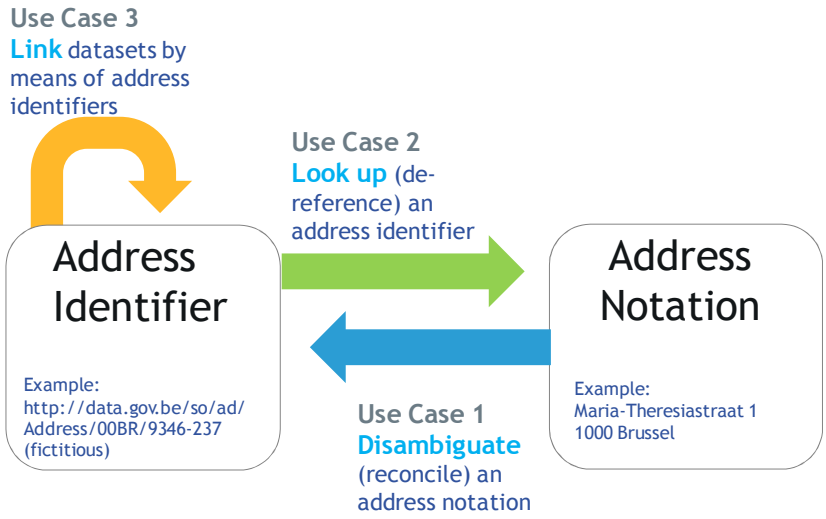


**Figure 1 Three uses cases for data consumers**

## 2. PILOT: INTERCONNECTING BELGIAN ADDRESS DATA

We have demonstrated the technical feasibility for this in a Linked Data pilot that integrates address data from five different public sector organisations in Belgium. In Belgium, the use of address data is impeded by the following obstacles:

- **Address data fragmentation**. The address data at Belgian federal level and at the three regions is housed in isolated registries maintained by the **National civil register**, the **National companies register**, and three regional registers with address data.
- **Heterogeneous address data formats**. Address data is provided using different specifications.
- **Lack of common identifiers**. Addresses, administrative units, roads, buildings, and cadastral parcels are not identified by well-formed identifiers thus making it hard to reconcile data about the same entity coming from different sources.

This situation is depicted in Figure 2. Due to these obstacles, *consumers* of address data such as national, regional, and local public administrations, businesses, and citizens make limited use of the aforementioned registers.

In the period November 2012 – February 2013, we have carried out a pilot to demonstrate that the **Core Location Vocabulary** and related **INSPIRE data specifications on addresses** can be applied to aggregate address data from various sources and contribute to overcoming the aforementioned obstacles. In particular, the pilot entails the following steps:

1. Develop (provisional) **URI sets** enabling Belgian addresses to be uniquely identified and looked up on the Web by well-formed HTTP URIs;
2. Represent existing address data from the federal and regional road and address registers using the **Core Location vocabulary** and experimental INSPIRE RDF vocabularies;
3. Put in place a **linked data infrastructure** that allows querying harmonised Belgian addresses from a SPARQL endpoint (see Figure 2).
4. Demonstrate the value of the linked data infrastructure to **disambiguate**, **lookup**, and **link** address data using simple Web-based standards such as HTTP, XML, and RDF.

The pilot demonstrates that:

- The **Core Location Vocabulary** can be used as a foundational RDF Vocabulary to *harmonise address data* that originates from disparate organisations and systems.
- The Core Location vocabulary can be ***flexibly* extended** with experimental INSPIRE RDF vocabularies (i.e. transport networks and administrative units).
- **HTTP URI sets** can be derived from INSPIRE thematic and external object identifiers for address data, allowing to create harmonised Web identifiers for, respectively, spatial things and spatial objects such as addresses.
- A **linked data infrastructure** can provide access to *harmonised, linked*, and *enriched* address data using standard Web-based interfaces (such as HTTP and SPARQL) and representation formalism (such as XHTML, RDF+XML, and GML), on top of either:
    - existing **relational/spatial database systems**, by applying conversions from a relational database to RDF**;**
    - existing **INSPIRE XML data**, by applying XSLTs to automatically generate RDF, starting from XML-encoded INSPIRE-compliant meta/data;
- The use of standard Web interfaces (such as HTTP(S) and SPARQL) can **simplify the use of address data** for humans and machines.
- URI sets for INSPIRE spatial objects and spatial things can accommodate both the **XML (GML) and RDF** world

Visit the pilot at: http://location.testproject.eu/BEL/

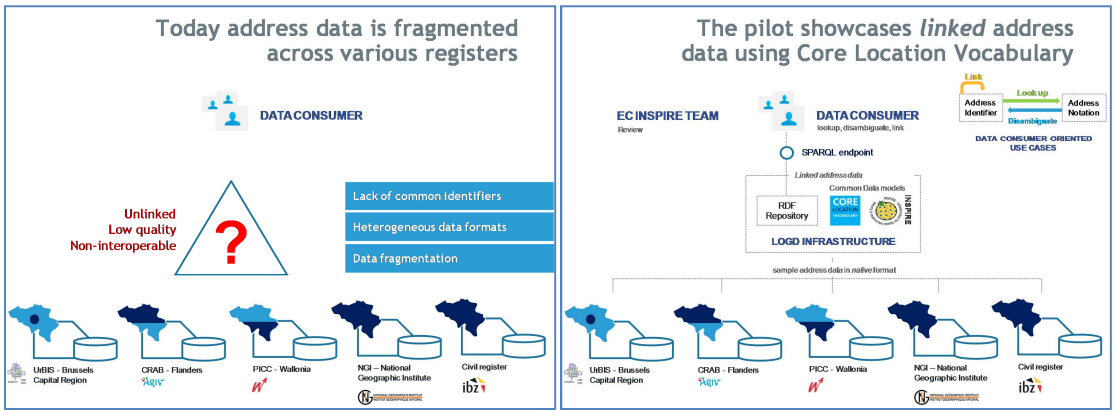


**Figure 2 INSPIRE and the Core Location Vocabulary can be used to harmonise address data**

## UC1: Disambiguate (query) address notations

- SPARQL query on the triple store
- The query is converted into SQL and hits the relational tables of several data providers

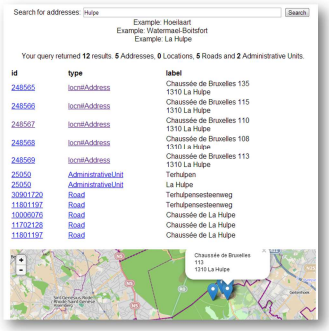


**Figure 3 Use case 1: disambiguate**

## UC2: Resolve Web identifiers

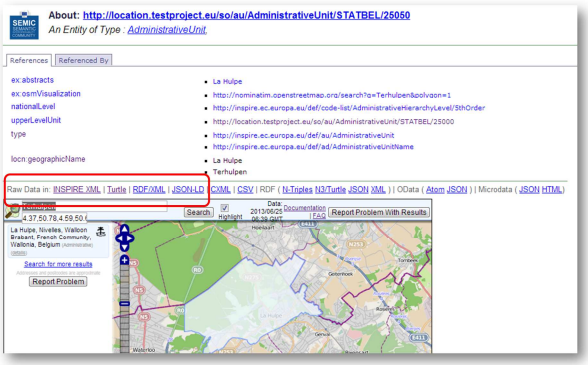


**Figure 4 Use case 2: resolve**

## UC3: Link address data


**Figure 5 Use case 3: link**

## REFERENCES


[BAG] Basisregistratie addressen en gebouwen
http://bag.vrom.nl/

[CRAB] Centraal referentieaddressenbestand
http://www.agiv.be/gis/organisatie/?artid=980

[EIF] European Interoperability Framework
http://ec.europa.eu/isa/documents/isa_annex_ii_eif_en.pdf

[INSPIRE] INSPIRE Data Specifications
http://inspire.ec.europa.eu/index.cfm/pageid/2

[ISA 1.1] Improving semantic interoperability in European eGovernment systems
http://ec.europa.eu/isa/actions/01-trusted-information-exchange/1-1action_en.htm

[ISA.2.13] Establishment of a European Union Location Framework (EULF)
http://ec.europa.eu/isa/actions/02-interoperability-architecture/2-13action_en.htm

[ISO-19160] ISO 19160 Addressing
http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=61710

[LinkedData] Tim Berners-Lee, Linked Data Design Issues, 2006.
http://www.w3.org/DesignIssues/LinkedData.html

[Location] Core Location Vocabulary
http://joinup.ec.europa.eu/asset/core_location/description

[LocationURI] Designing URI sets for Location
http://location.defra.gov.uk/2011/09/location-data-uri-sets-design-and-guidance/

[NSG] National Street Gazetteer
http://www.thensg.org.uk/iansg/link.htm?nwid=82

[PURI] D7.1.3 - Study on persistent URIs, with identification of best practices and recommendations on the topic for the MSs and the EC
https://joinup.ec.europa.eu/community/semic/document/10-rules-persistent-uris