Spatially Enabling
Australia and New Zealand

Project 3.01 | Automatic Federation of Spatial Data Semantically

Project Leader Prof Geoff West, Dept of Spatial Sciences, Curtin University, gwest@crcsi.com.au

Research Team Mr Jeremy Siao Him Fa (Curtin University), Dr David McMeekin (Curtin University, Dr Simon Moncrieff (Curtin University), Dr Simon Cox (CSIRO)

Project Participants Curtin University, University of Canterbury
NGIS, Amristar
Landgate WA, DEPI Victoria, DNRM QLD, CSIRO, PSMA

Objectives
- To provide an automated method for seamless communication between database systems across states and territories
- To provide a method for users to query databases at all governmental levels

Outcomes
- The project will aim to deliver a functional scoped-down implementation of the federated system
- The federated system will be developed using open source programs and adapted to be applied at a large area scale

"If you can't explain it simply, you don't understand it well enough." – Albert Einstein

Introduction

Authoritative spatial data is managed by Australia's jurisdictions (states and territories). Many applications require them to be joined, either because activities occur at borders, or a uniform view of data is required. The syntactic and semantic harmonization required can be done in various locations in an information architecture. Optimally data providers may conform to a community model or standard (e.g. INSPIRE model), but there is usually little incentive for agencies, as data publication for out-of-state users is not of upmost importance. Traditionally it was left to the user to merge data from multiple sources (if they could find it). Another option is for systems or services to be deployed at an intermediate point, in a 'broker' architecture. For a subset of foundation datasets the broker role is currently played in Australia by PSMA, but their processes and distribution channels depend on bulk file transfer and significant manual intervention in the process. In this project we are exploring more automated methods for federation, using semantic technologies.

The Problem: Disparate Datasets

The picture below depicts some problems regarding the querying of differing datasets. When querying the bounding box, data needs to be fetched from various heterogeneous data sources, and hence various federation issues arise.

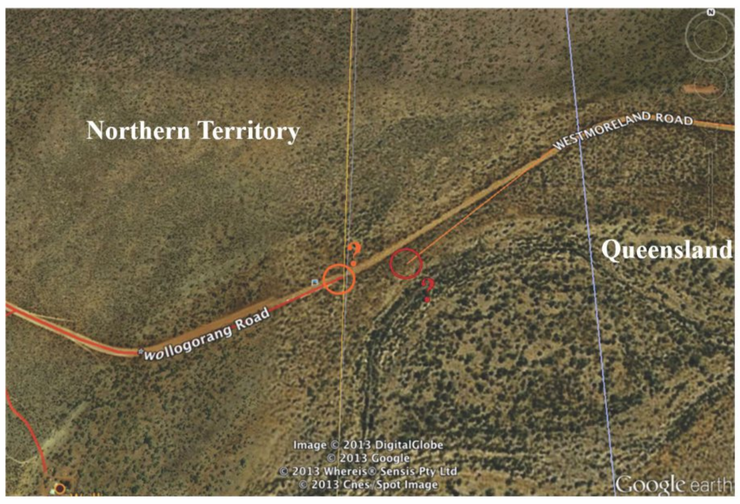

Figure 1: Disparate Datasets Problem
(Google, Northern Territory, Queensland Data)

Federation Issues

- Different road names
- Different fonts
- Different line styles and colours
- Gaps in the data
- Different locations of the state borders

Example: Differing Representation of "LandUse"

```
<?xml encoding="ISO-8859-1"?>
<!ELEMENT LandUse (LandParcel)>
<!ELEMENT LandParcel (AREA, BROAD, LU1,
  LU2, LU3, ..., JurisType, JurisName)>
<!ELEMENT AREA (#PCDATA)>
<!ELEMENT BROAD (#PCDATA)>
<!ELEMENT LU1 (#PCDATA)>
......
<!ELEMENT JurisType (#PCDATA)>
<!ELEMENT JurisName (#PCDATA)>
```

```
<LandUse>
  <LandParcel>
    <AREA>1704995.587470</AREA>
    <BROAD>A</BROAD>
    <LU1>AF</LU1>
    ......
    <JurisType>County</JurisType>
    <JurisName>EauClaire</JurisName>
  </LandParcel>
    ......
</LandUse>
```

a) Local XML data source S_1 of Eau Claire County.

```
<?xml encoding="ISO-8859-1"?>
<!ELEMENT LandUse (LandParcel)>
<!ELEMENT LandParcel (AREA, LAND_USE,
  PARCEL_ID, ..., JurisType, JurisName)>
<!ELEMENT AREA (#PCDATA)>
<!ELEMENT LAND_USE (#PCDATA)>
<!ELEMENT PARCEL_ID (#PCDATA)>
......
<!ELEMENT JurisType (#PCDATA)>
<!ELEMENT JurisName (#PCDATA)>
```

```
<LandUse>
  <LandParcel>
    <AREA>1007908.5</AREA>
    <LAND_USE>9100</LAND_USE>
    <PARCEL_ID>24710</PARCEL_ID>
    ......
    <JurisType>County</JurisType>
    <JurisName>Madison</JurisName>
  </LandParcel>
    ......
</LandUse>
```

b) Local XML data source S_2 of the City of Madison.

Differences:
- XML tags
- Names
- Information details
- Letters vs Numbers

Figure 2: LandUse Representation (Isabel Cruz – Ontology Alignment for the Semantic Integration of Heterogeneous Geospatial Data Set)

Possible Federated System

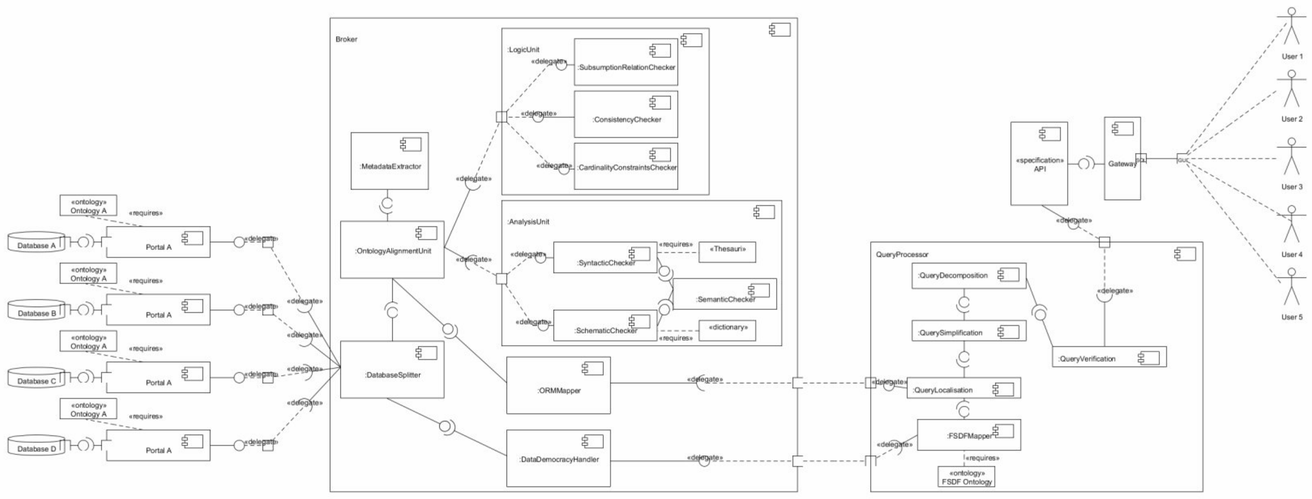

Figure 3: Sample Federated System

Landgate FESA-019 Graph

Figure 4 was produced in OWL based on Landgate's FESA-019 type and visualized in Protégé. The XML schema was retrieved via Web Feature Service.

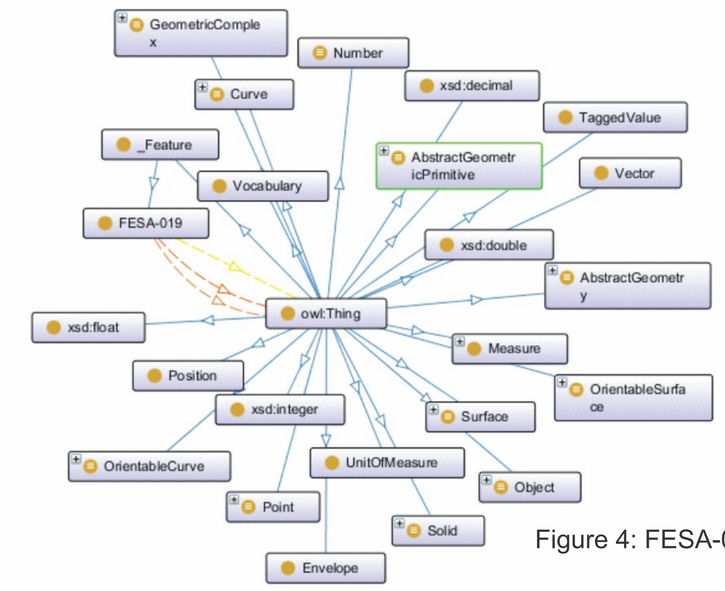

Figure 4: FESA-019 Graph

Summary: Due to the jurisdictions having different spatial data schemas and formats, database interoperability is an issue. By finding ways to federate Australia's spatial data automatically using semantic web techniques, it would allow the unification of all the disparate datasets to be done for the user. Hence, this would lead to easier access to nation wide spatial data, while lifting semantic burdens on the user.

Curtin University · 43pl · Landgate · State Government Victoria Department of Environment and Primary Industries · Australian Government Department of Industry · Business Cooperative Research Centres Program