CaRSA Data Identify, Collect, and Connect: A second-generation, national GeoLD system in Australia

Nicholas J. Car[1][0000−0002−8742−7730] and
Irina Bastrakova[2]

[1] SURROUND Australia Pty Ltd., Australia &
Australian National University, Australia
nicholas.car@surroundaustralia.com
[2] Geoscience Australia
irina.bastrakova@ga.gov.au

Abstract. In 2018 – 2020, Australia built two *Linked Data* "spines" - themed collections of interoperable reference data - called LocI and LongSpine. LocI (Location Index) consists of 7 nationally-significant spatial datasets such as the Australian Statistical Geographies System. LongSpine (Longitudinal Spine of Government Functions) consists of multiple datasets of Australian government structure, such as the Australian Government Organisations Register, as well as vocabularies of government functions. Both projects interpreted existing open datasets into Linked Data form and provided online delivery of their the parts as well as infrastructure for their use as a single system.

Here described is the Climate and Resilience Services Australia's Data Identify, Collect and Connect project's next-generation implementation of the "spines" patterns and architecture and partial direct reuse of LocI. We discuss: differences with the originals, in particular LocI; original element reuse; and future-proofing steps, technical and non-technical. We also discuss the new expectations placed on this second-generation system, in particular the requirement to work with non-Linked Data spatial data systems, and how this system is pushing spatial and *Semantic Web* standards development such as DGGS and GeoSPARQL.

Keywords: Location Index · LocI · LongSpine · GeoSPARQL · DGGS · Spatial Data on the Web · Australia · data spine · national data infrastructure

1 Introduction

1.1 CaRSA Motivation

Climate and Resilience Services Australia (CaRSA) is a new Australia government cross-agency initiative[3] that will:

[3] "Australia commits to climate resilience", https://minister.awe.gov.au/ley/media-releases/australia-commits-climate-resilience

> connect and leverage the Commonwealth's extensive climate and natural disaster risk information to further prepare for and build resilience to natural disasters

Since Australia is prone to very damaging natural disasters such as bush fires, floods and droughts, this is a major government initiative allocated good resourcing and the commitment is for multiple years.

1.2 CaRSA Demonstrator Projects

Several of the demonstrator projects for CaRSA sought to test different ways of combining information from multiple government agencies relevant to natural disaster management. Traditional methods of data aggregation are being tested, such as data pooling in shared facilities, standardiseing web service-delivered information and cross-cataloging datasets, but forward-looking methdos are too. In particular, *Semantic Web* (SW) and *Linked Data* (LD) technologies[4] are being used to integrate different, but relativley similar, datasets that are published in a distributed manner and *Discrete Global Grid System* (DGGS) spatial data methods are being used to integrate spatial data from multiple sources.

This paper describes the SW/LD and DGGS approaches being implemented in CaRSA's "Data Identify, Collect, and Connect" project that we will refere to as *the project*. The project extends the approach taken by the Location Index project described in the next section.

2 LocI: The Loction Index

In 2018 - 2020, Australian spatial data and research agencies implemented a:

> national and authoritative, also federated, index for Australian spatial data using Semantic Web technologies [2]

This system, known as the Location Index (LocI) [2], aims to "better geospatially integrate and analyze data across government portfolios and information domains". The main use case addressed by LocI's is to greatly reduce the time taken by government workers in data analysis using spatial information by providing pre-integrated, authoratitive, spatial datasets that can be used in online, open data scenarios, within secure data integration environments and across the two.

Some of the interesting aspects of LocI's design include:

* federated publication of datasets via standard Linked Data APIs

[4] By "Linked Data", as opposed to "linked data" or "data linkage" etc., we mean systems and data that implement a number of *Semantic Web* technologies (RDF, OWL, SKOS, SPARQL, etc.) which are primarily defined as a series of World Wide Web Consortium (W3C) standards. The W3C's defintion of *Semantic Web* is that it is a "Web of Data", an evolved Internet able to be queried by machineswhich can draw inferences from it.

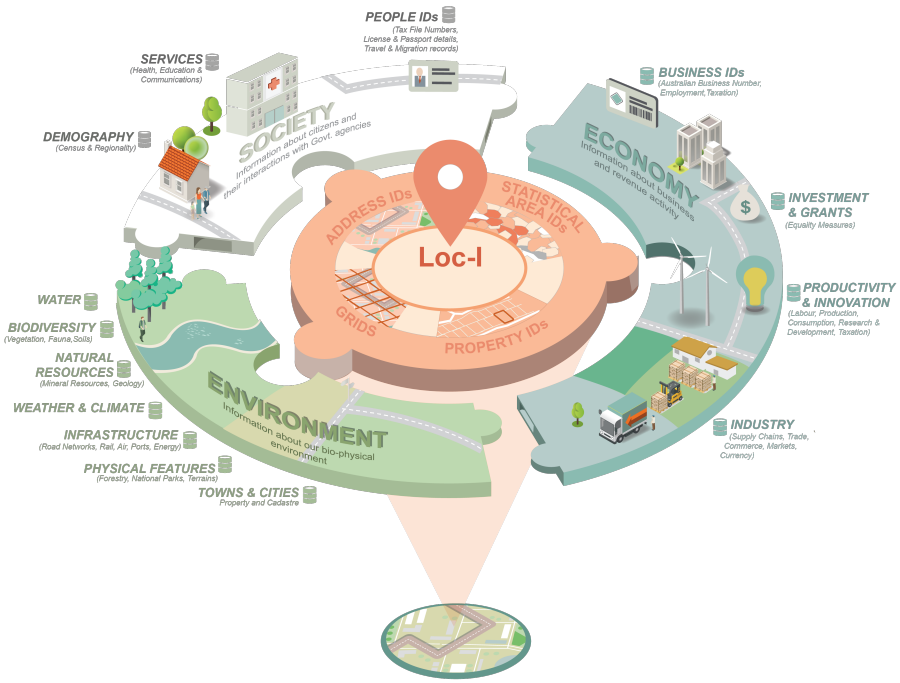

Fig. 1. A project brochure image, from [2], of LocI with respect to Australian government *Environment*, *Society* and *Economy* data

- use of VoID `Linkset` [1] instances to crosswalk datasets
  - these are independently-selectable for use meaning that a specific ccrosswalk, of potentially many, may be selected for use
- use of a *Geometry Data Service*[5] for spatial integration
  - this service extends common use of using GeoSPARQL [3] by storing `Geometry` instances seperately from the `Feature` instances they are the geometries for. This allows the geometry data to be managed in a PostGIS database [6], not a triplestore, as usually used for GeoSPARQL data.
- several different clients for different uses
  - such as *Excelerator*[7], used to upload data according to one spatial reference system and download it reapportioned according to another

LocI's initial spatial datasets are from a number of domains including environmental (the *Australian Hydrological Geospatial Fabric* [8], a collection of surface hydrology

[5] The service is online at https://gds.loci.cat/
[6] https://postgis.net/
[7] https://loci.cat/excelerator.html
[8] The original, non-RDF dataset: http://www.bom.gov.au/water/geofabric/, and the online LD version implemented by LocI: http://linked.data.gov.au/dataset/geofabric

features), human/census (the *Australian Statistical Geography Standard* spatial areas) [9], and cartographic/administrative (the *National Composite Gazetteer of Australia*)[10].

The LocI system's architecture is shown in 2 for architectural details. It shows the LocI Data Cache, which is a multi-graph triplestore, obtains its data by "pulling" RDF datasets through APIs that both interpret non-RDF data for online delivery and are also able to create static RDF versions of the datasets. All LocI datasets conform to the LocI Ontology[11] which imports the GeoSPARQL [12] and DCAT[13] ontologies. Alongside the Cache is a traditional spatial DB - PostGIS[14] used to perform fast geometry intersections.

3 CaRSA Project Changes

4 Conclusions

4.1 Future Work

* GeoSPARQL 1.1: to be completed by July, 2021
* GeoSPARQL 1.2: more adventurous extension - late 2021
* GeoSPARQL 2.0: so-far unspecified

References

1. Alexander, K., Cyganiak, R., Hausenblas, M., Zhao, J.: Describing Linked Datasets with the VoID Vocabulary. W3C Interest Group Note, World Wide Web Consortium (Mar 2011), https://www.w3.org/TR/void/
2. Car, N.J., Box, P.J., Sommer, A.: The Location Index: A Semantic Web Spatial Data Infrastructure. In: Hitzler, P., Fernández, M., Janowicz, K., Zaveri, A., Gray, A.J., Lopez, V., Haller, A., Hammar, K. (eds.) The Semantic Web. pp. 543–557. Lecture Notes in Computer Science, Springer International Publishing (2019). https://doi.org/https://doi.org/10.1007/978-3-030-21348-0_35
3. Perry, M., Herring, J.: OGC GeoSPARQL - A Geographic Query Language for RDF Data. OGC Implementation Standard, Open Geospatial Consortium (2012), http://www.opengis.net/doc/IS/geosparql/1.0

[9] Non-RDF dataset: https://geo.abs.gov.au/arcgis/services/ASGS2016/MB/MapServer/WFSServer, LD version: http://linked.data.gov.au/dataset/asgs2016
[10] LD version: https://linked.data.gov.au/dataset/placenames
[11] http://linked.data.gov.au/def/loci
[12] http://www.opengis.net/doc/IS/geosparql/1.0
[13] https://www.w3.org/TR/2014/REC-vocab-dcat-20140116/
[14] https://postgis.net/

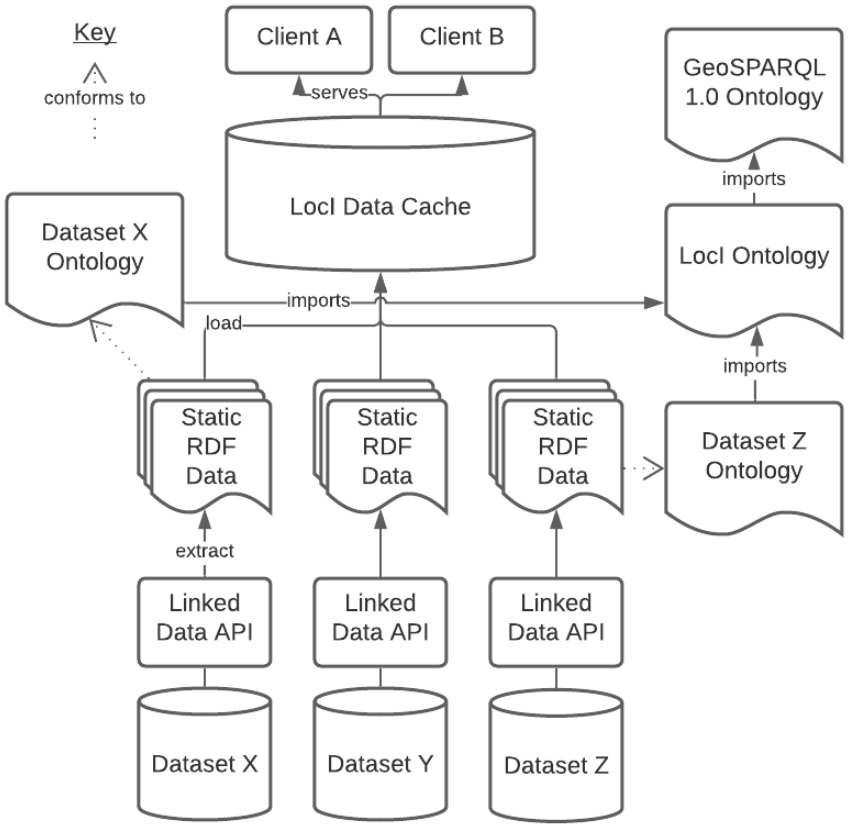

Fig. 2. An informal architecture diagram of the LocI project's *Linked Data* infrastructure.

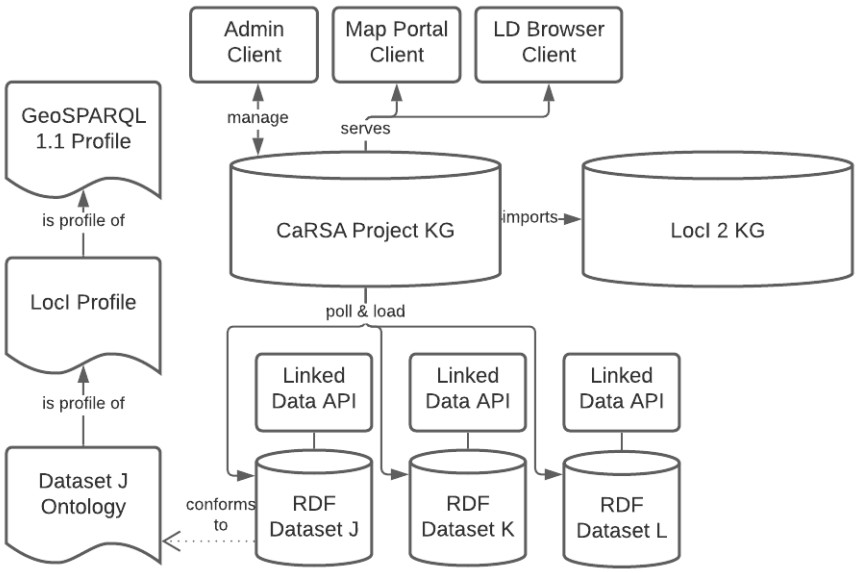

Fig. 3. An informal architecture diagram of the CaRSA project's *Linked Data* infrastructure