The Loc-I Spatial Knowledge Graph

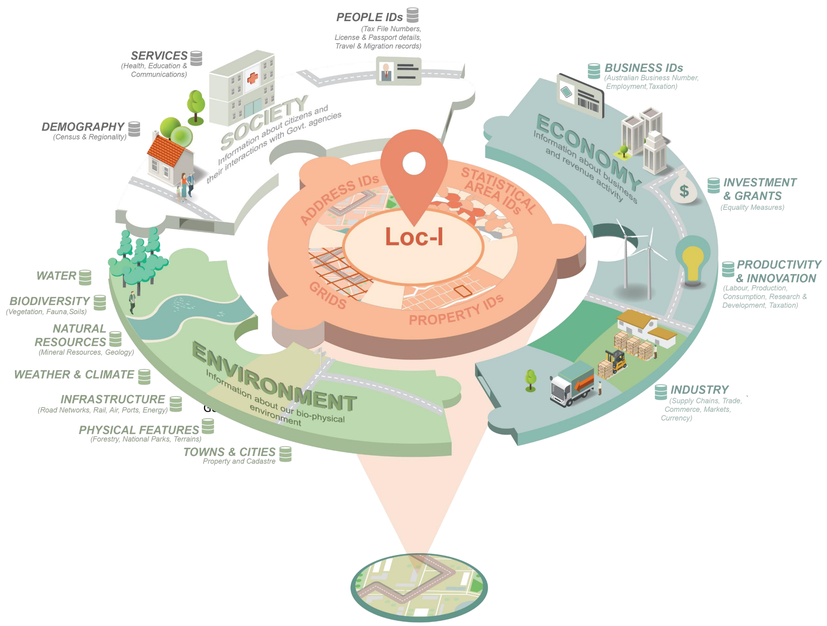
Enabling Data Integration of Location-based Data via Linked Data Approaches

Jonathan Yu, David Lemon, Paul Box, Ashley Sommer, Shane Seaton, Benjamin Leighton, Simon Cox, and Ashlin Lee (CSIRO Australia)

https://loci.cat

Introduction

The Location Index (Loc-I) project (http://locationindex.org) aims to enable government agencies to geospatially-integrate and analyse data reliably, effectively and efficiently across portfolios and information domains. Loc-I is part of the Data Integration Partnership for Australia (DIPA) initiative, which seeks to maximise government data to improve policy advice. Linked Data approaches provide solutions for data integration of location-based data.



Goals of Loc-I

1. Publish spatial data using reliable and consistent (web-)identifiers
2. Link spatial features across reporting geographies
3. Enable data integration of datasets that embed location references
4. Provide access to Loc-I data via the web - open, federated and distributed
5. Provide access to human- and machine- readable views and formats of Loc-I data

Technical Approach

Linked Data approaches and Semantic Web technologies are used to publish spatial feature identifiers and ***linksets*** (i.e. links between geographies) as richly-described web resources. Identifiers are minted using the linked.data.gov.au namespace managed by the Australian Government Linked Data Working Group. Each geography and its features are identified unambiguously via Linked Data URIs, and described using RDF according to the relevant ontologies. Content negotiation provides both human- and machine-readable multiple views and formats for a given Loc-I web identifier (refer to the LHS of ***Figure 1***). To date, we have integrated these geographies: Aus. Bureau of Statistics' Australian Standard Geography Standard (ASGS); Aus. Bureau of Meteorology's Geofabric; PSMA G-NAF Address dataset.

*A **Hybrid Spatial Knowledge Graph*** is implemented**,** which caches the spatial features and linksets, and provides powerful querying capability. It combines a RDF Graph Store (GraphDB), GIS database (PostGIS), and a text-based search engine (Elasticsearch) to provide optimised querying capability for Linked Data, Spatial and Text-based queries. Linked Data descriptions and geometric data of Loc-I spatial features and linksets are included, specifically, RDF descriptions are indexed in the RDF graph cache, GIS data in the Geo-database, and text labels and identifiers in the Text search engine component (refer to the RHS of ***Figure 1***). SPARQL and REST APIs are provided allowing users and applications to query and traverse the semantics of Loc-I features depending on the application. Loc-I applications are presented in ***Figures 2-4***).

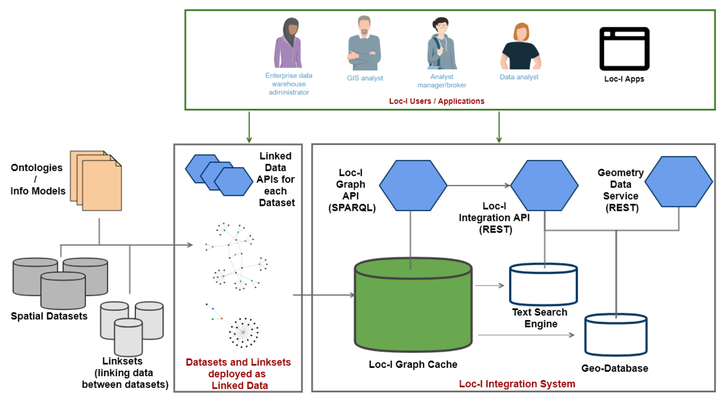
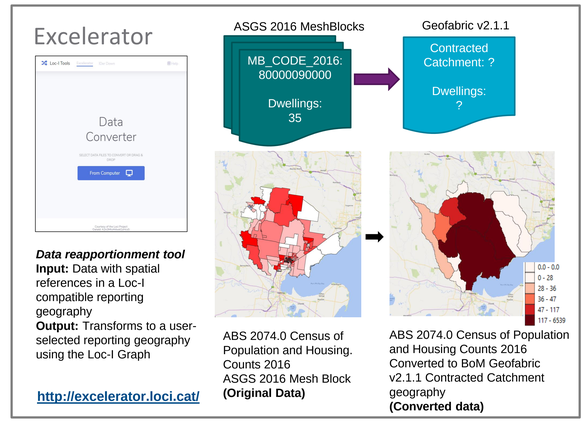


Figure 1: Loc-I spatial knowledge graph system overview

Geometry Data Service

Geometries as treated as separable first-class objects and accessed via a ***Geometry Data Service (GDS).*** The Loc-I GDS provides access to multiple views of spatial geometries for the relevant reporting geographies included in Loc-I as Linked Data. Geometry data is provided by GDS in different forms (e.g. geometry as-is, centroid, metadata-only) and formats/serialisations (e.g. WKT, GeoJSON, SHP, RDF/GeoSPARQL). The GDS achieves this via Content Negotiation on spatial features stored in the Loc-I Geo-database, which hosts the geometries. This allows spatial features in Loc-I to embed URI-references to the geometry instead of embedding a geometry literal, optimising the RDF graph store for semantic queries rather than geospatial. More information on the GDS can be found on the Github repo link: https://git.io/JepYD

Loc-I Applications



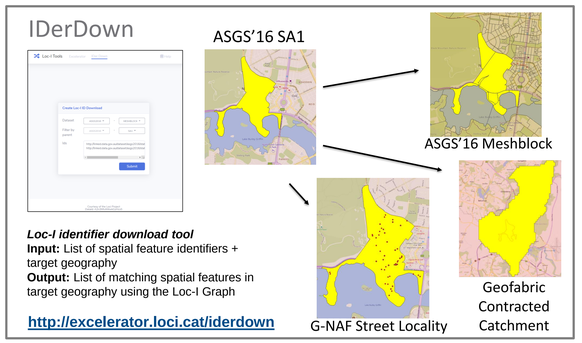
Data reapportionment tool
Input: Data with spatial references in a Loc-I compatible reporting geography
Output: Transforms to a user-selected reporting geography using the Loc-I Graph

http://excelerator.loci.cat/

ABS 2074.0 Census of Population and Housing. Counts 2016
ASGS 2016 Mesh Block
(Original Data)

ABS 2074.0 Census of Population and Housing Counts 2016 Converted to BoM Geofabric v2.1.1 Contracted Catchment geography
(Converted data)

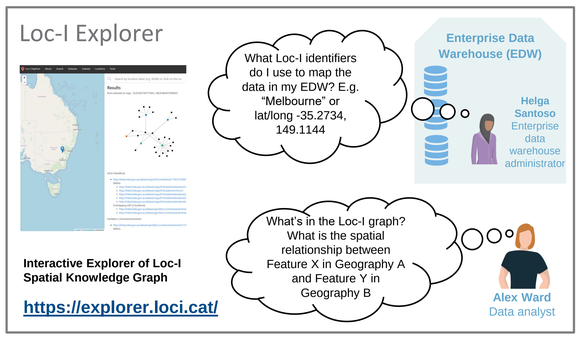
Figure 2: Loc-I Application #1 - Excelerator



Loc-I identifier download tool
Input: List of spatial feature identifiers + target geography
Output: List of matching spatial features in target geography using the Loc-I Graph

http://excelerator.loci.cat/iderdown

Figure 3: Loc-I Application #2 – IderDown



Interactive Explorer of Loc-I Spatial Knowledge Graph

https://explorer.loci.cat/

Figure 4: Loc-I Application #3 – Loc-I Explorer

As Australia's national science agency and innovation catalyst, CSIRO is solving the greatest challenges through innovative science and technology.

CSIRO. Unlocking a better future for everyone.

FOR FURTHER INFORMATION
Jonathan Yu
CSIRO Environmental Informatics Group
jonathan.yu@csiro.au
csiro.au/landandwater

ACKNOWLEDGEMENTS
We acknowledge the contributions to the initial designs of the Loc-I technical architecture and implementations from Nicholas Car (now at Surround Australia). We also acknowledge the Australian Linked Data Working Group for the governance of and minting of authoritative identifier namespaces for the Loc-I project. The Loc-I project is funded by the Data Integration Partnership for Australia (DIPA) initiative.

Australian Bureau of Statistics | Australian Government Geoscience Australia | Australian Government Department of the Environment and Energy | Australian Government Department of Agriculture and Water Resources | CSIRO