Reducing Consumer Uncertainty
Towards an Ontology for Geospatial User-centric Metadata
Introduction

- Cooperative Research Centre for Spatial Information (CRCSI) in Australia
- Communicate fitness for use of spatial data sources to users in spatial and other domains
- Capture and represent users’ requirements and implicit knowledge of spatial data sources
- Adopt a “user-centric” approach to geospatial metadata
- Create a vocabulary to communicate fitness for use of spatial data sources in the context of various application domains
Evaluating Geospatial Data Quality

• Geospatial data sources are being used across various user groups and domains

• Users are presented with an increasing choice of data from various data portals, repositories, and clearinghouses

• Comparing the quality of different data sources and evaluating a data source’s fitness for use, is essential

• Quality and fitness for use are dependent on the use case
Communicating Fitness for Use - Challenges

• Metadata standards are mostly focused on data production rather than potential data use and application

• Datasets are used for purposes other than what they were initially created for

• Datasets used by users with different levels of expertise

• Datasets used in various applications and domains

• End products are often derived from a variety of sources
  • difficult to issue simple statements about a product’s quality
  • impossible to label a particular dataset as “the best”
Standards and Related Initiatives

• Geospatial standards
  – ISO 19115-1:2014 - Geographic info metadata fundamentals
  – ISO 19115-2:2010 - Extensions for imagery and gridded data
  – ISO 19115-3:2016 - XML schema implementation for fundamental concepts
  – ISO 19157:2013 - Geographic information – Data quality
  – ISO/TS 19158:2012 - Geographic information - QA of data supply
  – ISO 19150-1:2012 - Geographic Information – Ontology Framework
Standards and Related Initiatives

- Geospatial Standards - continued
  - ISO 19109:2015 Geographic information — Rules for application schema
  - 19110:2016 Geographic information — Methodology for feature cataloguing

- Dublin Core Metadata Initiative (DCMI)
  - an open organization supporting innovation in metadata design and best practices across the metadata ecology

- DCAT
  - Vocabulary designed to facilitate interoperability between data catalogues published on the Web
Standards and Related Initiatives

- **Schema.org**
  - In 2011, the major search engines Google, Bing, Yahoo, and (later) Yandex created Schema.org
  - Single vocabulary (ontology) across a wide range of topics
  - Includes classes/properties for people, places, events, products, offers

- **ANZLIC Metadata Profile**

- **Foundation Spatial Data Framework (FSDF)**
  - an ANZLIC initiative that aims to deliver national coverage of standardized and quality controlled foundation spatial data for Australia and New Zealand
Standards and Related Initiatives

• Linked Open Data
  – URI: identifies an abstract or physical resource
  – RDF, SPARQL and OWL Standards
  – Data Quality Vocabulary: an extension to the DCAT vocabulary to cover the quality of the data, its update frequency, user corrections, persistence commitments, etc.
  – PROV-O: introduces a set of concepts to represent provenance information in a variety of application domains
Standards and Related Initiatives

Linked Open Data - continued

– Simple Knowledge Organization System (SKOS), a common data model for sharing and linking knowledge organization systems via the Web

– Open Annotation ontology: RDF-based specification for annotating digital resources
GeoViQua

- QualityML vocabulary
  - Extends UncertML: provides a list of alternative metrics used to quantify quality beyond uncertainty
  - Provides a set of rules for documenting quality measure parameters defined in ISO19157
- GEO Label: Graphical representation of the metadata
- Quality Model
  - Based on ISO 19115-1:2014 and ISO 19157:2013
  - Producer Quality Model
  - User Feedback Model
OGC Geospatial User Feedback Standard

• Data model for encoding user feedback about geospatial datasets or metadata records describing datasets
• Metadata that is produced by geospatial data consumers
• Complements producer supplied metadata
• Reuses and extends the ISO 19115-1:2014 data model
• User comments, questions/answers, user reports of dataset problems and proposed solutions to those problems, ratings, usage reports, citations of related datasets or publications describing usage, quality reports, relevant additional provenance information, significant events related to the use or interpretation of a dataset
Increasing Choice of Metadata Standards

• Which standards should be used?
• How much metadata should be provided?
• What quality information should be provided?
• How should quality and fitness for use be communicated in a consistent and standardized way?
• How to enable geospatial data users to determine fitness for use of spatial data sources in the context of their application and domain?
Geospatial User-centric Metadata Vocabulary

- Enable users to identify fitness for use of geospatial data sources
- Complement producer supplied metadata with user-defined fitness for use descriptions
- Contextualize user-centric metadata
  - application and domain within which data sources are used
  - profile of users that describe the user-centric metadata
- Enable producers to incorporate user-defined metadata into objective quality measures for products
- Identify various users and use cases of a dataset
Geospatial User-centric Metadata Ontology

- Transform the vocabulary into a more dynamic and well-grounded formalism; i.e., an ontology
- Describe fitness for use in machine-readable format
- Use concepts from widely-adopted ontologies and standards developed by standardization bodies; e.g., ISO/TC 211, OGC, DCMI, FSDF, W3C
- Facilitate the application of reasoning techniques developed by the Semantic Web community
Geospatial User-centric Metadata Ontology

• Describe fitness for use at various levels of granularity; e.g., dataset, feature, attribute, using the hierarchical structure of ontologies

• Publish and integrate user-centric metadata as structured data on the Web
  – where they can be linked to their corresponding data sources
  – leading to a seamless aggregation of spatial data, producer-supplied and user-centric metadata

• Enable dataset search and discovery based on quality and fitness for use criteria
Geospatial User-centric Metadata Conceptual Model

ISO 19110:2016 Geographic Information — Metadata
ISO 19115-1:2014 Metadata-Fundamentals
ISO 19157-1:2013 Data Quality

Producer Quality Lens

User Quality Lens

+ application
+ domain
+ userInformation
+ usageDescription
+ descriptionAspect (usage, fitness for purpose, limitation)
+ impact
+ rating
+ scope
  ++ Dataset
    ++ Feature/Attribute
      Application A -> Feature Types (M, N)
      Application B -> Feature Types (K, L)
+ qualityDescription
  + subjective quality/user feedback
  + MD_Metadata
  + DQ_DataQuality
  + ......................
  + ......................
Geospatial User-centric Metadata Ontology

• Enables structured and dynamic description of dataset quality
• Model can be easily extended with additional quality standards
• Enables users with various levels of expertise to describe the quality of spatial datasets, in the context of their application and domain
• Enables both producers and users of geospatial data to describe the quality of spatial datasets, using the same model
• Structured quality metadata can be queried, in order to facilitate search and discovery based on quality criteria
Geospatial User-centric Metadata Ontology

Design and Validation

• Iterative design and practical implementation
  – iterative design methodology
  – online Web portal (population of the model)
  – semi-formal feedback and recommendations from project participants and partners

• Design validation
  – formal testing and analysis of the design prototype
  – a usability study as an initial “proof-of-concept”-level validation

• Prototype implementation
  – apply and assess the model’s usefulness in the context of a real industry setting; i.e., Landgate’s open data Web portal (CKAN)
  – implement for manual use by PSMA
Future Work

- OGC Geosemantics DWG (Spatial Data on the Web Working Group)
- OGC Innovation Program
- ISO/TC 211
- AGLDWG
- data.gov.au