Reducing Consumer Uncertainty

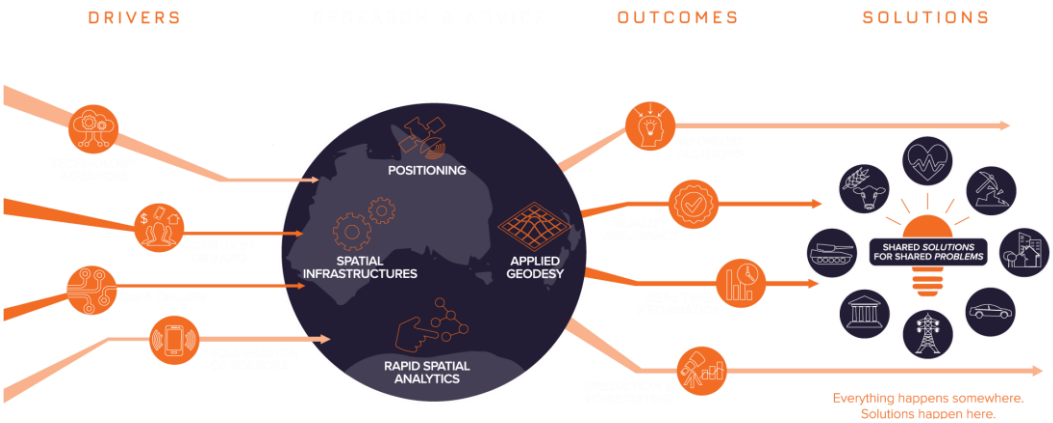
Towards a Vocabulary for User-Centric Geospatial Metadata

Dr Ivana Ivánová
Curtin University

FRONTIERSI

FRONTIERSI.COM.AU

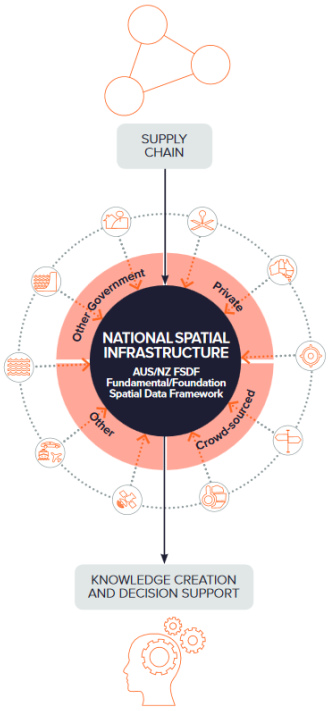
FrontierSI (former CRCSI)



FrontierSI (former CRCSI)

Spatial Infrastructures program (ANZSI) aims at building an infrastructure, which will, for example:

- assist people as they select a place to live,
- inform planners as they design new suburbs,
- aid emergency services personnel as they respond to life threatening situations and
- safely guide autonomous vehicles as they navigate our streets.



Presenter: Dr Ivana Ivánová

Research and Teaching

Affiliation:

Current:

- Senior Lecturer at Curtin University, Perth, Australia
- Research fellow at FrontierSI

Past:

- Visiting/Collaborating Professor at UNESP, São Paulo, Brazil
- Lecturer/Researcher at University of Twente, The Netherlands,
- Lecturer/Researcher at Slovak University of Technology in Bratislava, Slovakia

Research background and interest:

- spatial data quality, fitness for use, quality management
- spatial infrastructures
- provenance

Community work

Standardization:

Current:

- OGC Data Quality DWG (co-chair)
- Standards Australia IT-004 (member)

Past:

- CEN/TC 287 Outreach Group (national rep)
- Slovak Standards institute (member)
- Eurogeographics QKEN (observer)

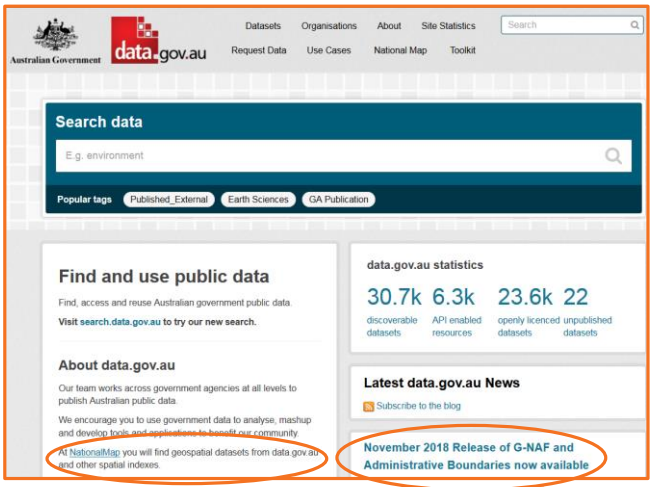
Other:

- RDA Provenance Patterns (member)
- ISPRS ICWG IV/III on Global Mapping: Updating, Verification and Interoperability (co-chair)
- UN OpenGIS Spiral 2: PostGIS course (coordinator)

Project P3.16: Reducing Consumer Uncertainty

Motivation

- The (automatic) use of spatial data is likely to increase.
- Communicating users' requirements for spatial resources is difficult.
- Producers continue to struggle to describe their products in a language meaningful to their users.
- Need to ensure interoperability between metadata from spatial and non-spatial domains via public open data interfaces.



The project

Aim:

To demonstrate through a semantic framework that linking publisher dataset attributes (mainly technical) to a consumer **vocabulary** will **make it easier for traditional and non-traditional spatial consumers to discover and access data that meets their needs**.

Specific objectives:

1. **Profile consumers** to understand how they identify 'quality' within spatial datasets to build a vocabulary based on consumer terminology;
2. **Discover and link the consumer vocabulary to the publisher's vocabulary** under a semantic framework;
3. **Track the processes that link the publisher's datasets to the consumer vocabulary** so it may be utilised by the project partners.
4. **Evaluate the consumer vocabulary in maximising the fitness-for-purpose** of spatial data queries by implementing the vocabulary with a project partner.

Duration:

12 + 3 + 3 months

Project team

Tom Spencer (PSMA) – project leader

Kylie Armstrong (FrontierSI)

Ivana Ivánová(FrontierSI)

Hasti Ziaimatin (QUT) – lead researcher

Alistair Barros (QUT)

Alireza Nili (QUT)

Greg Byrom (LINZ)

John Gallagher (DELWP)

Keith Moss (Landgate)

Wayne Patterson (DFSI – Spatial Services)

Project deliverables

1. Conference abstracts and presentations, literature review, interview templates, questionnaires;
2. Qualitative and quantitative analysis of engagements and identification of themes and priorities of data consumers;
3. Ontology representing consumer-centric data quality vocabulary;
4. Ontology specification to support implementation of the vocabulary;

Engagement: aims and outcomes

Aims and Background:

- Aim was to elicit producer and user views on geospatial data quality and fitness for use.

Survey:

- semi-structured interviews and online survey
- **23 participants:**
  - Users of geo-resources
  - Producers of geo-resources
  - Both (80%), users and producers, of geo-resources
  - Mostly from working with geo-resources in the domain of forestry, agriculture, fishing and other services.

Outcomes:

- **93%** of use data from external data providers.
- **46%** have worked with geospatial data for 2-9 years.
- **93%** make data source selection decisions based on their prior knowledge and experience with data sources.
- **53%** of the participants find selecting datasets that fit their needs a challenging task.
- **80%** of the participants consider metadata records or other supporting information when selecting datasets for use.
- **53%** of all participants believe that up to 25% of manual effort is involved in understanding the fitness for use of data sources

Engagement: more outcomes

- Identified themes and priorities of data consumers for **communicating quality of geo-resources***:
  - Spatial (5), thematic (4) and temporal accuracy (4)
  - Completeness (3)
  - Logical consistency (4)
  - Lineage (5)
  - Relevancy (5)
  - Currency (4)
  - Reliability (5)
  - Cost (3)

* *The number in () means the frequency of explicit mentioning of the term in interview*

Engagement: more outcomes

- Means for **assessing fitness-for use***:
  - Producer profile (4)
  - Dataset citation (4)
  - Data dictionary (5)
  - Quantitative quality information (5), incl. detail beyond ISO 19157 spatial/temporal resolution, scale, error estimates, uncertainty
  - Soft knowledge (6), incl. comments, known problems, potential additional use, online forums
  - Compliance with standards (3)
  - User ratings and feedback (5)
  - Community recommendation and advice (5)
  - Independent expert review (4)
  - Ease of Access
  - Licencing

* *The number in () means the frequency of explicit mentioning of the term in interview*

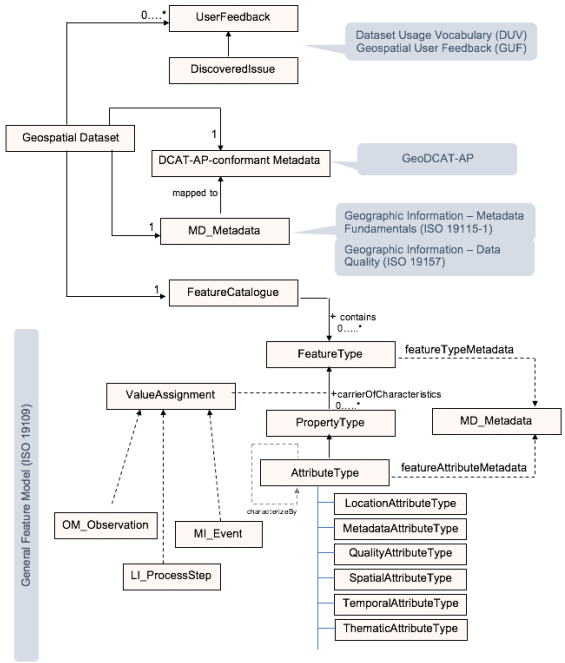
GUCM - The Ontology

Three main components:

1. Dataset Schema
2. Interoperable Metadata
3. User Feedback.

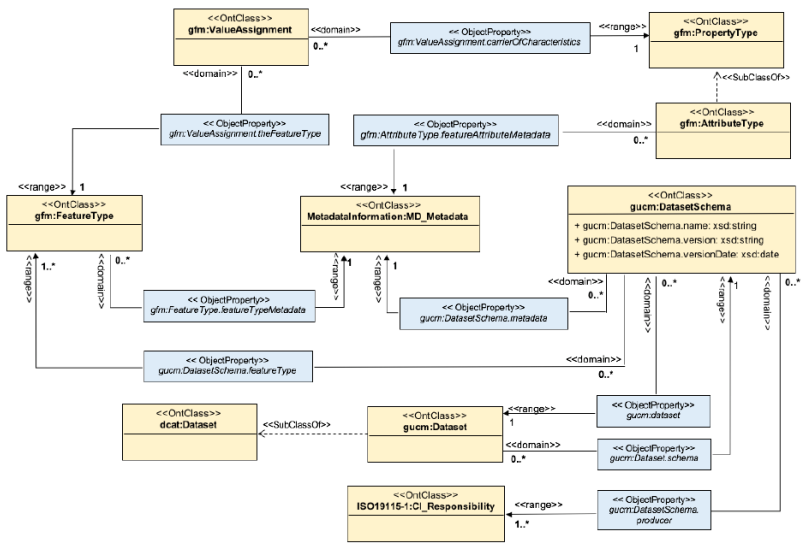
Standards used:

- ISO 19100 series
- OGC's GUF and GeoSPARQL
- W3C's DUV, DCAT, PROV
- INSPIRE's GeoDCAT-AP
- Existing ontologies: schema.org, FOAF, VCARD, SKOS, LOCN, RDFs



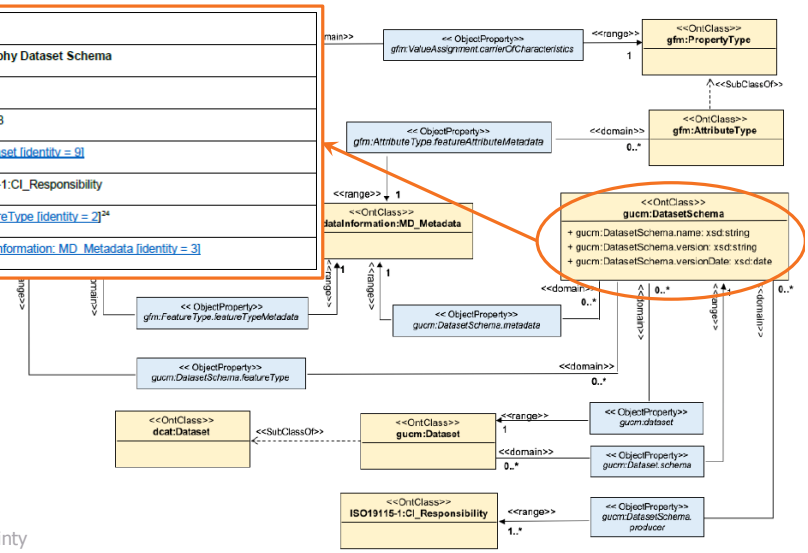
GUCM: Dataset Schema

Full description of the contents and structure of a geographic dataset in compliance with ISO 19109



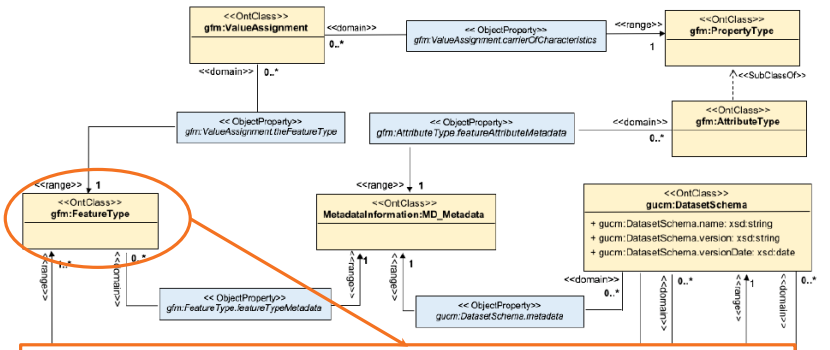
GUCM: Dataset Schema

gucm:DatasetSchema [instance identity = 1]	
gucm:DatasetSchema.name	**Hydrography Dataset Schema**
gucm:DatasetSchema.version	1.0
gucm:DatasetSchema.versionDate	23-06-2018
gucm:dataset	gucm:Dataset [identity = 9]
gucm:DatasetSchema.producer	ISO19115-1:CI_Responsibility
gucm:DatasetSchema.featureType	gfm:FeatureType [identity = 2][24]
gucm:DatasetSchema.metadata	MetadataInformation: MD_Metadata [identity = 3]



GUCM: Dataset Schema

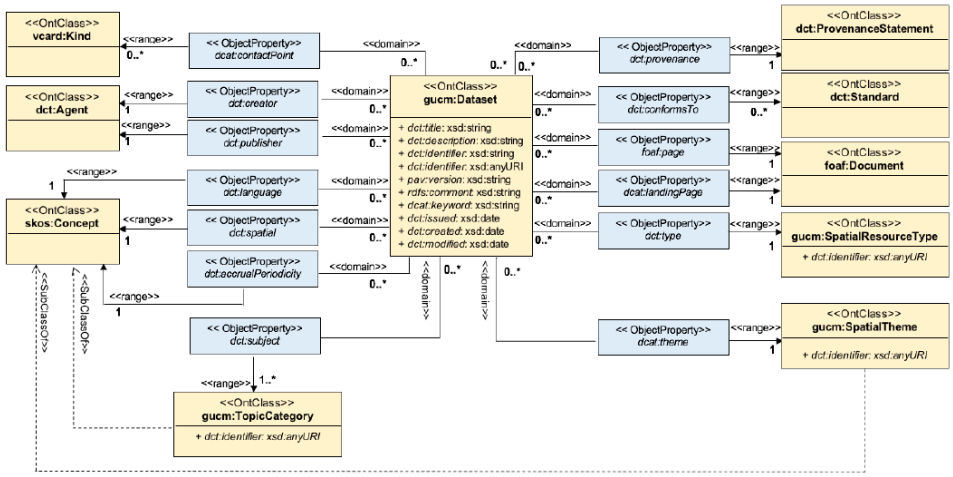
Full description of the contents and structure of a geographic dataset in compliance with ISO 19109



gfm:FeatureType [identity = 2]	
gfm:IdentifiedType.name	**Berthing Structure**
gfm:IdentifiedType.definition	A fixed (not afloat) artificial structure attached to a shore and normally used for berthing and protection of vessels
gfm:IdentifiedType.description	ISO 19109:LanguageString
gfm:IdentifiedType.designation	ISO 19109:LanguageString
gfm:FeatureType.isAbstract	False
gfm:FeatureType.theValueAssignment	gfm:ValueAssignment [identity = 4]
gfm:FeatureType.featureTypeMetadata	MetadataInformation:MD_Metadata [identity = 5]

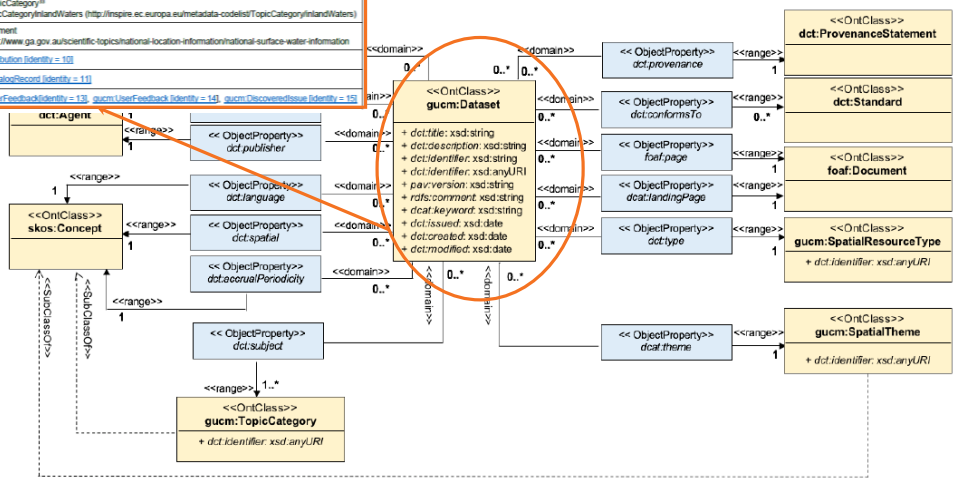
GUCM: Interoperable Metadata, p.1

Represents metadata captured by the Dataset Schema component, using concepts from domain-independent and widely-adopted ontologies, based on the GeoDCAT-AP version 1.0.1 specification

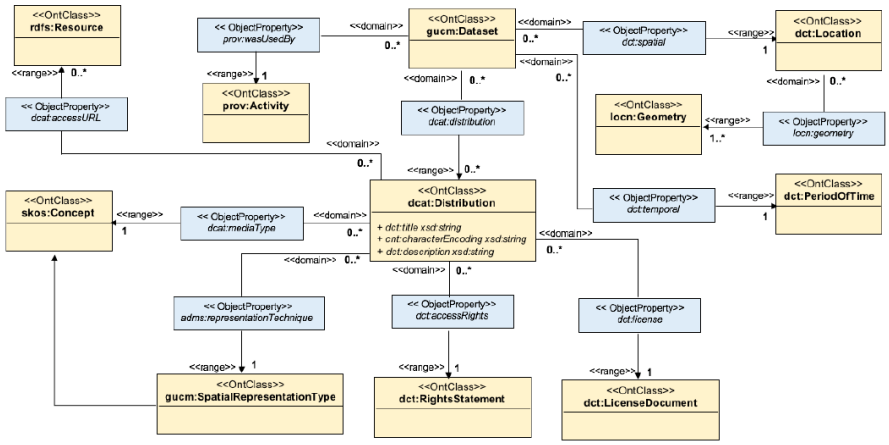


gucm:Dataset [identity = 9]	
dct:title	Hydrography
dct:description	The National Hydrography Dataset (NHD) and Watershed Boundary Dataset (WBD) are digital geospatial datasets that map the surface water of the United States and are a part of The National Map.
dct:identifier	"http://some.site/resource/12345" or "12345"
pav:version	1.0
rdfs:comment[23]	Spatial resolution (equivalent scale): 1:10000
dcat:keyword[30]	Hydrography, Ports and Harbours, Transportation Networks
dct:issued	21-03-2018
dct:created	18-03-2018
dct:modified	25-06-2018
dcat:theme	gucm:SpatialTheme[31] e.g., gucm:ThemeHydrography (http://inspire.ec.europa.eu/theme/hy)
dct:type	gucm:SpatialResourceType[32] e.g., ResourceTypeDataset (http://inspire.ec.europa.eu/metadata-codelist/ResourceType/dataset)
dct:subject	gucm:TopicCategory[33] e.g., TopicCategoryInlandWaters (http://inspire.ec.europa.eu/metadata-codelist/TopicCategory/inlandWaters)
dcat:landingPage	foaf:Document e.g., http://www.ga.gov.au/scientific-topics/national-location-information/national-surface-water-information
dcat:distribution	dcat:Distribution [identity = 10]
foaf:isPrimaryTopicOf	gucm:CatalogRecord [identity = 11]
duv:hasFeedback	gucm:UserFeedback[identity = 13], gucm:UserFeedback [identity = 14], gucm:DiscoveredIssue [identity = 15]

able Metadata, p.1 – example

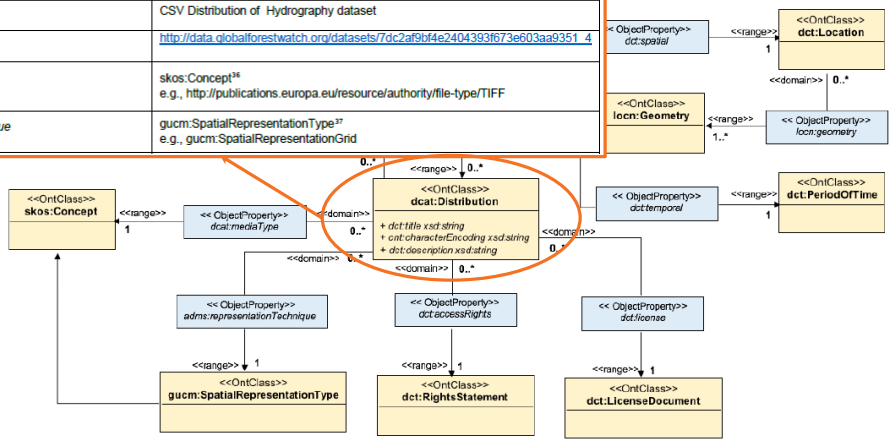


GUCM: Interoperable Metadata, p.2

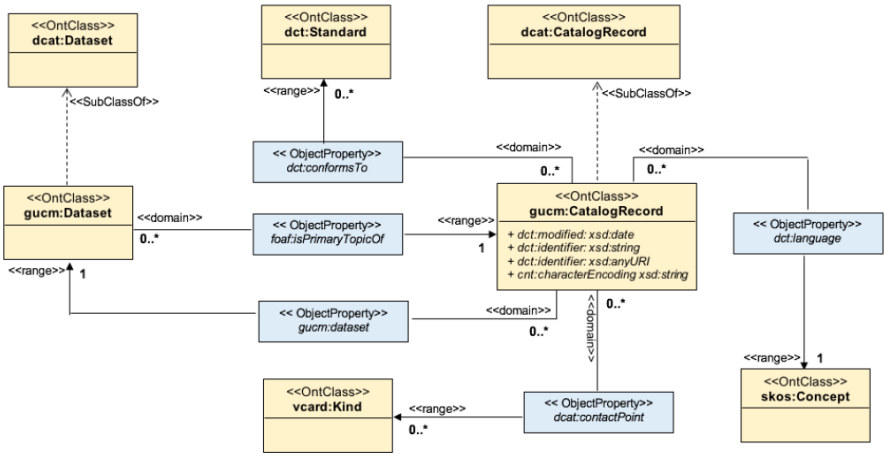


GUCM: Interoperable Metadata, p.2 – example

dcat:Distribution [identity = 10]	
dct:title	CSV Distribution — Hydrography
cnt:CharacterEncoding[34]	UTF-8
dct:description	CSV Distribution of Hydrography dataset
dcat:accessURL	http://data.globalforestwatch.org/datasets/7dc2af9bf4e2404393f673e603aa9351_4
dcat:mediaType[35]	skos:Concept[36] e.g., http://publications.europa.eu/resource/authority/file-type/TIFF
adms:representationTechnique	gucm:SpatialRepresentationType[37] e.g., gucm:SpatialRepresentationGrid

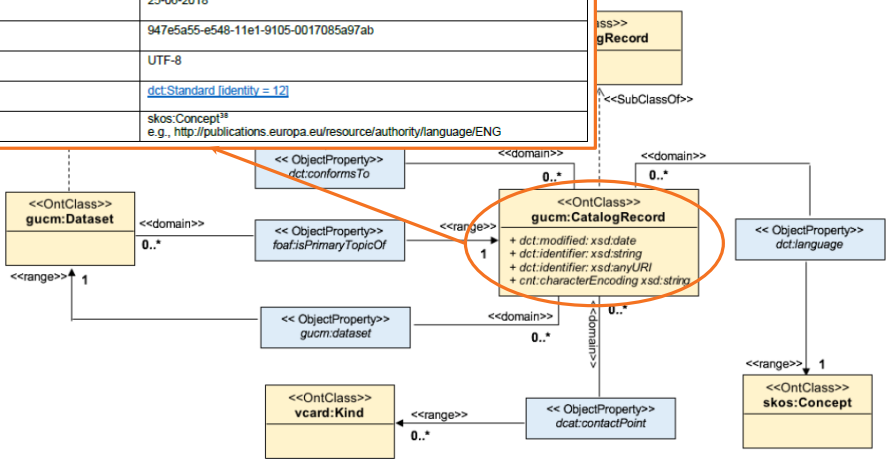


GUCM: Interoperable Metadata, p.3



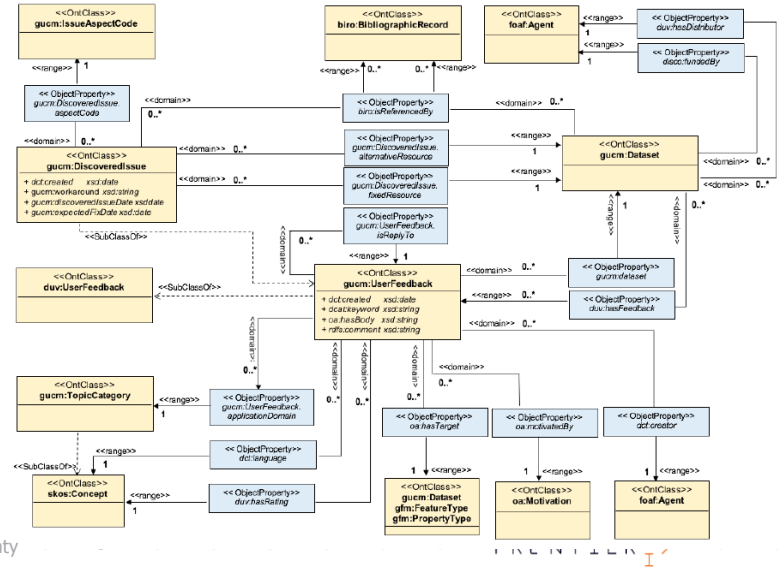
GUCM: Interoperable Metadata, p.3 – example

gucm:CatalogRecord [identity = 11]	
dct:modified	25-06-2018
dct:identifier	947e5a55-e548-11e1-9105-0017085a97ab
cnt:characterEncoding	UTF-8
dct:conformsTo	dct:Standard [identity = 12]
dct:language	skos:Concept[18] e.g., http://publications.europa.eu/resource/authority/language/ENG



GUCM: User Feedback

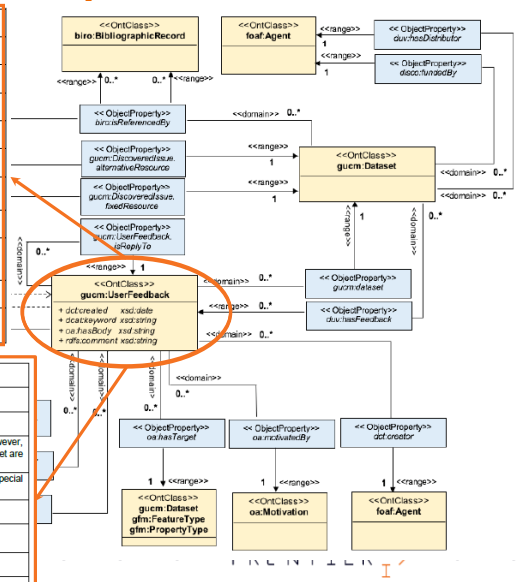
Describes users' implicit knowledge (incl. feedback, experiences, comments, questions and answers, description of encountered problems, proposed solutions and publications describing those problems, dataset ratings) using W3C's DUV and OGC's GUF.



GUCM: User Feedback – example

gucm:UserFeedback [identity = 13]	
dct:created	26-06-2018
dcat:keyword	Inland waters, definition source, boundaries
rdfs:comment	We require a definition source for this dataset, as the dataset structures, such as Mines, Rivers and Roads need to be mapped to the boundaries that we work with.
oa:hasBody	"Does this dataset have a definition source?"
gucm:dataset	gucm:Dataset [identity = 9]
oa:motivatedBy	oa:Motivation[39] e.g., oa:questioning
oa:hasTarget	gucm:Dataset [identity = 9]
gucm:UserFeedback.applicationDomain	gucm:TopicCategory[40] e.g., TopicCategoryBoundaries (http://inspire.ec.europa.eu/metadata-codelist/TopicCategory/boundaries)
dct:creator	foaf:Agent; e.g., user in the boundaries domain
dct:language	skos:Concept[41] e.g., http://publications.europa.eu/resource/authority/language/ENG

gucm:UserFeedback [identity = 14]	
dct:created	28-06-2018
dcat:keyword	Inland waters, definition source, boundaries
rdfs:comment	Feature types contain some information about the boundaries in which they have been defined; however, you're right in accessing a definition source, as boundaries defined for the feature types in this dataset are neither comprehensive nor accurate.
oa:hasBody	International Hydrographic Organisation (IHO) Hydrographic Dictionary, Part I, Volume I English; Special publication No.32
gucm:dataset	gucm:Dataset [identity = 9]
oa:motivatedBy	oa:Motivation[42] e.g., oa:replying
oa:hasTarget	gucm:Dataset [identity = 9]
dct:creator	foaf:Agent; e.g., dataset publisher or expert
gucm:UserFeedback.isReplyTo	gucm:UserFeedback [identity = 13]



Summary

- 1st version of Geospatial User-Centric Ontology (GUCM) exists together with its implementation specification.
- GUCM reflects user demands on communicating geo-resources data quality and fitness for use as indicated in comprehensive interviews with both, users and producers of geospatial resources.
- GUCM is standard compliant.
- GUCM is machine readable (currently in OWL format).

Limitations:

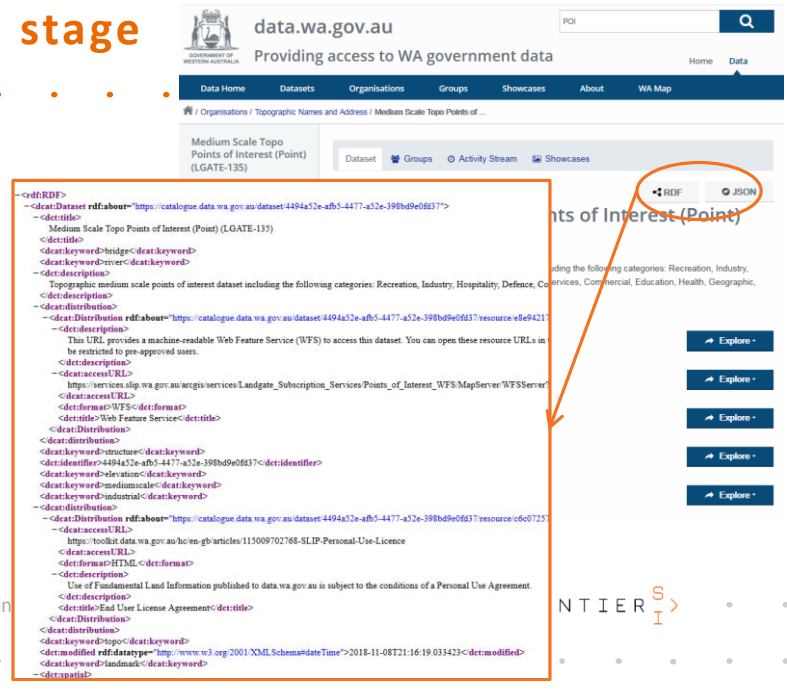
- GUCM is untested version 1 of the vocabulary (reasons: project too short, engagement phase too long)
- GUCM is unnecessarily robust – includes all relevant offline ontologies (mainly ISO 19100 series!)
- GUCM needs implementation and community feedback.

GUCM: current stage

- Project ended in July 2018

Utilization efforts under-way:

- Landgate (WA Mapping Agency) aims at enhancing their metadata exposed to the web with GUCM.
- Expected 1st results of the utilization: mid 2019



QKEN and FrontierSI collaboration

What is the way forward?

Couple of first suggestions:

- Can QKEN provide feedback on the GUCM specification (PDF available in QKEN already)?
- Any NMA's interested to implement GUCM?

Any other suggestions?

Please send any questions, suggestions, comments, or, share similar experience at:

ivana.ivanova@curtin.edu.au

Thank you!

ivana.ivanova@curtin.edu.au

FRONTIER SI

FRONTIERSI.COM.AU