Data Type Registry

Data set descriptions for automation

Stephen M Richard, USGIN

RDA Plenary 7, BoF Tokyo, Japan 日本東京、

Use cases

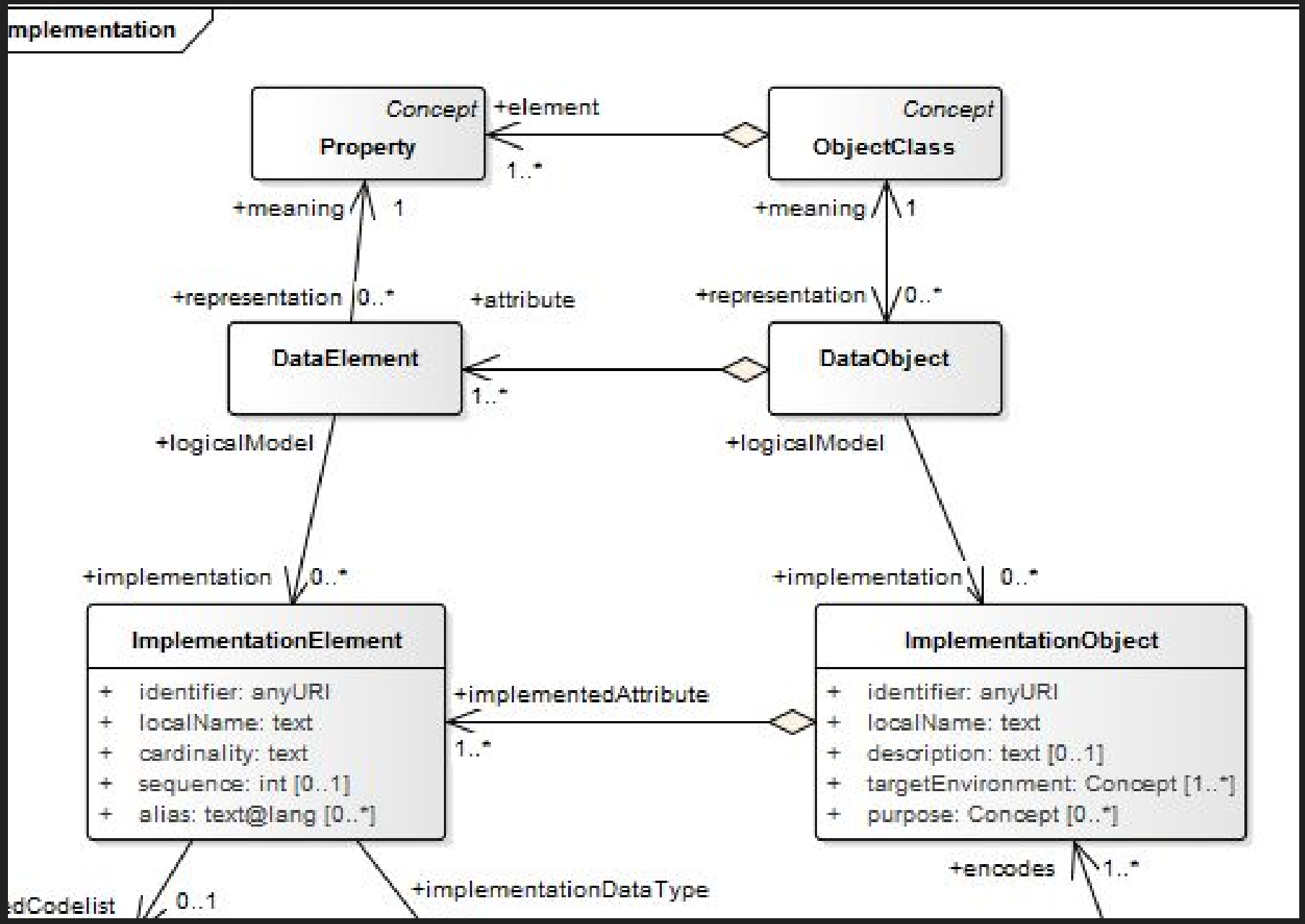
1. Reference for communities to document the meaning of entities and attributes in data that they share.
2. Discover existing data type and attribute definitions for use in constructing data models, to foster interoperability.
3. Discover resource containing information about a particular entity or property.
4. Machine-assisted data integration, based on identification of matching or ‘integratable’ attribute content.
5.

More use cases

1. Validation of data instances against a type definition.
2. Tools that spin up a UI for a particular data type.
3. Allow application to identify data sources that it can use seamlessly (MIME-type with structure)
4. Support file introspection to assist with deep data registration

Representation levels

- **Conceptual** – Entities and properties based on domain concepts, *use for semantic integration and query processing*: e.g. building, geologic fault, road, river, grain size, density
- **Logical** – the information representation framework; *use for broad data compatibility assessment*– e.g. relational, object, graph, integer, real number, term, text, generic field or attribute names
- **Physical** – detailed implementation specific, *use for low-level data integration*. E.g. xs:decimal, char(255), varchar, long integer, actual field names in a table design…


mplementation
Concept
Property
+element
1..*
Concept
ObjectClass
+meaning
1
+meaning
1
+representation
0..*
+attribute
+representation
0..*
DataElement
1..*
DataObject
+logicalModel
+logicalModel
+implementation
0..*
+implementation
0..*
ImplementationElement
+ identifier: anyURI
+ localName: text
+ cardinality: text
+ sequence: int [0..1]
+ alias: text@lang [0..*]
+implementedAttribute
1..*
ImplementationObject
+ identifier: anyURI
+ localName: text
+ description: text [0..1]
+ targetEnvironment: Concept [1..*]
+ purpose: Concept [0..*]
+encodes
1..*
edCodelist
0..1
+implementationDataType

Approach from metadata world

ISO 19110

ISO 11179

ISO 19115

US FGDC CSDGM

Common Data Model (NetCDF)

A more comprehensive model is necessary

- Integrate/harmonize various existing schemes:
  - ISO19110, ISO11179, ISO19115, FGDC CSDGM, RDA Data Type Registry 2015, NetCDF CDM, Deep Carbon Observatory, VOID…
- Proposal: **https://github.com/usgin/usginspecs "Draft UML model for information registry"**