Maintenance of INSPIRE Catalogues:

High function capabilities in support of **discovery**, **evaluation**, and **data knowledge**

Bruce Westcott
Intergraph Corporation
23 June 2010
“infrastructures for spatial information in the Member States should be designed to ensure that [...] it is easy
- to *discover* available spatial data,
- to *evaluate* their suitability for the purpose and
- to *know* the conditions applicable to their use”.

-- Directive establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)
The Importance of Metadata in SDIs

- Metadata enables discoverability;
- Discoverability facilitates collaboration;
- Collaboration is a prerequisite for interoperability;
- Interoperability makes possible a shared digital infrastructure:
  - Technical: Known structural (physical) characteristics
  - Semantic: Shared meaning
  - Business: meets informational requirements and has value(!).
Problem

- The characteristics, quality, currency, availability, and very existence of spatial data assets changes from moment to moment.

- Increasing automation in data production leads to:
  - Increase in volume of data to be described,
  - Increased ability to create **temporally-volatile datasets**
  - Increased requirement that catalogue records identify and **distinguish among similar datasets**
Challenges for Catalogue Managers

- Design a lifecycle for metadata content that is **technically** detailed, **semantically** meaningful, and has **business value**?

- Identify technologies which support the **continuous updating** of catalogue contents **at lowest effort/cost** and **greatest accuracy**.
Opinion

- Failure to address these challenges will leave us swamped in a network of Catalogues that:
  - Clog our search results (like Google with no rankings),
  - Provide for basic discovery – but:
  - Are out-of-date or abandoned, and
  - Offer little or no utility for evaluation of or knowledge about the resources named.

- Do you agree?
Catalogue Maintenance Workflows

1) Define a superset of Catalogue Elements
2) Create Profile(s) for different classes of Resources
3) Define Reuseable Elements
4) Create Templates
5) Implement Intelligent Agents
1. Elements of a Catalogue Record

- Required & Optional Elements from
  - ISO 19115 Metadata for Data
  - ISO 19119 Metadata for Services

- Definition by means of Profiling
  - **Type-1** Profile
    - Made up of ISO-standard elements
    - Example: the **INSPIRE Profile**
  - **Type-2** Profile
    - Can have additional elements
    - Defined by User Communities, National Authorities, etc.

![Figure C.1 — Metadata community profile](Image)
2. Profiles

- Different User Communities
- Different Elements to describe different Resources:
  - Dataset (vector, raster, other newer data types)
  - Dataset series
  - Service
  - Application
  - Feature type

Figure G.1 — Metadata hierarchy
## 2. Elements in a Profile (Example)

- Abstract
- Creator
- Contributor
- CouplingType
- DataQuality
- Extent
- Format; FormatVersion
- Keywords
- HierarchyLevel
- Lineage & Process Steps
- Maintenance
- MetadataCharacterSet
- MetadataLanguage
- MetadataStandardName
- MetadataStandardVersion
- Modified (Date)
- OnlineResource
- ParentIdentifier
- Publisher
- ResourceIdentifier
- ResourceCharacterSet
- ResourceLanguage
- ReferenceSystem
- RevisionDate
- ServiceIdentification
- SpatialResolution
- SpatialRepresentationType
- TopicCategory
2. Elements determined by Enterprise

- Abstract
- Creator
- Contributor
- CouplingType
- DataQuality
- Extent
- Format; FormatVersion
- Keywords
- HierarchyLevel
- Lineage & Process Steps
- Maintenance
- MetadataCharacterSet
- MetadataLanguage
- MetadataStandardName
- MetadataStandardVersion
- Modified (Date)
- OnlineResource
- ParentIdentifier
- Publisher
- ResourceIdentifier
- ResourceCharacterSet
- ResourceLanguage
- ReferenceSystem
- RevisionDate
- ServiceIdentification
- SpatialResolution
- SpatialRepresentationType
- TopicCategory
2. Elements determined by **Resource**

- Abstract
- Creator
- Contributor
- CouplingType
- DataQuality
- Extent
- Format; FormatVersion
- Keywords
- HierarchyLevel
- Lineage & Process Steps
- Maintenance
- MetadataCharacterSet
- MetadataLanguage
- MetadataStandardName
- MetadataStandardVersion
- Modified (Date)
- OnlineResource
- ParentIdentifier
- Publisher
- ResourceIdentifier
- ResourceCharacterSet
- ResourceLanguage
- ReferenceSystem
- RevisionDate
- ServiceIdentification
- SpatialResolution
- SpatialRepresentationType
- TopicCategory
2. Problematic Elements

- **Descriptive**
  - Abstract
  - Keywords
  - TopicCategory

- **Relationship with other Resources**
  - HierarchyLevel
  - ParentIdentifier
  - Lineage

- **Quality**
  - Lineage: Process Steps
  - DataQuality: Scope, Measure, Result
3. What are Reusable Elements?

- Defined entities/schema that are compounded from simple elements and:
  - Created/managed as a single instances,
  - Referenced in multiple records.
3. Reusable Elements (REs)

- Purposes:
  - Minimize data update effort and complexity,
  - Assure consistency across catalogue records.
3. Candidate REs

<<DataType>>

CI_Citation

+ title : CharacterString
+ alternateTitle [0..*] : CharacterString
+ date [1..*] : CI_Date
+ edition [0..1] : CharacterString
+ editionDate [0..1] : Date
+ identifier [0..*] : MD_Identifier
+ citedResponsibleParty [0..*] : CI_ResponsibleParty
+ presentationForm [0..*] : CI_PresentationFormCode
+ series [0..1] : CI_Series
+ otherCitationDetails [0..1] : CharacterString
+ collectiveTitle [0..1] : CharacterString
+ ISBN [0..1] : CharacterString
+ ISSN [0..1] : CharacterString

<<DataType>>

CI_ResponsibleParty

- Contact
- Address & Phone
- Online Resource
3. Candidate REs

- **MD_FeatureCatalogue**
  - Description
  - Geometry
  - Temporal (?)

- **EX_Extent**
  - **Description**
  - **Geometry**
  - Temporal (?)

- **MD_MaintenanceInfo**
  - ... And Others.
4. Templates

- Assemble templates for various classes of records, consisting of:
  - stable Enterprise-driven elements; e.g. MD_Constraints (Legal & Security),
  - Variable Enterprise elements and Resource-based elements, (and their dependencies)
- Define business rules for applying the most appropriate variables in each template, based on the resource to be described.
5. Intelligent Agents

- **Working Definition**: Automated or semi-automated processes which can derive metadata content elements from resources or processes with a minimum of operator interaction.

- **Example 1**: Extract metadata content from resources
  - Identify classes of resources as “Geospatial Data Entities” (GDEs)
  - Engineer processes to harvest and periodically refresh metadata content from GDEs based on data content and format.
5. Intelligent Agents

- **Example 2:** Extract Enterprise-based content from other systems:
  - Enterprise personnel/directory system(s) for updating `CI_ResponsibleParty` information
  - Enterprise data dictionary system(s) for managing `MD_FeatureCatalogue` instances

- **Example 3:** Implement Inheritance & Relationships
  - Engineer applications for instantiating, tracing, managing inheritance relationships (parent-child, and within hierarchies of GDEs)
  - Define content-specific business rules for inheritance
5. Intelligent Agents

- **Example 4**: Engineer “wizard” applications for generating descriptive metadata
  - **Abstract** can provide kernel of meaning for generating **Keywords**, **TopicCategory**, and other free_text. (Purpose, supplementallInformation, etc.)

- **Example 5**: Capture Lineage information
  - Engineer application for capturing flows of **LI_Processes**, referencing multiple **LI_Sources**.
  - Transform captured sequences into **LI_Lineage** fragments for inclusion in records.
Summary of Workflows

1) Define a superset of Catalogue Elements
2) Create Profile(s) for different classes of Resources
3) Define Reuseable Elements
4) Create Templates
5) Implement Intelligent Agents
Have a great INSPIRE 2010 Conference . .

. . .and thank you very much for your attention.

Visit Intergraph’s exhibit to learn more about becoming “SDI Ready”