Archiving Geospatial data

Bridging the gap between Archival and Geospatial standards

Gregor Završnik, Geoarh

ISO/TC 211 Standards in practice, MARIBOR, SI
Introduction

Gregor Završnik
www.geoarh.si
Overview

• Why should we archive Geospatial data?
• What should we archive? (Important elements)
• The role of Archival and Geospatial STANDARDS
• Bridging the gap
Why should we archive Geospatial data?
Archive were where data goes to die...

Capture

Usage

Archives

GEOARH MAPPING HISTORY
Data is the new oil...
Geodata reuse

- Future savings
- Legal Safety
- Context for other data
- Modeling for the future
New and existing systems

- Conversion to long term preservation format
- Copy to a longterm medium
- Secure storage
- Metadata and documentation
- Export to archival repository
What should we archive?

Which elements are important?
Basic Archiving concepts

- Availability
- Usability
- Completeness
- Authenticity
- Provenance

Usage logic
Knowledge base
Documentation / quality
Long term preservation formats
We also need context, data quality, metadata...
Geo Information products come from complex systems...

- Ministry of culture
- Defense ministry
- Environmental agency
- Spatial planning
- Cadaster
- Ministry of agriculture
The role of Archival and Geospatial STANDARDS
OAIS (Open Archival Information System)

Start here: http://en.wikipedia.org/wiki/Open_Archival_Information_System

Full Standard: http://public.ccsds.org/publications/archive/650x0b1.PDF
Specifications for Archival Information packages

Archival Metadata standards

Geospatial standards

ARCHIVAL PACKAGE

CONTENT

CS IP

E-ARK SIP

E-ARK AIP

E-ARK DIP

CITS

CITS

CITS

Common for all Information Packages
Archival metadata standards

- METS (Packet structure)
- PREMIS (Provenience, Change history)
- EAD (Content description)
Usage of archival metadata structures
The need for Geospatial standards

GIS Tools

- Symbology, geoprocessing, exports
- Organisation, Queries, Cartographic projections, transformations...

Vector data

Raster data

Database

Lists, codepages

Models
Technical documentation

- Attribute definitions
- Feature catalogue
- Logical structure
- Visualisation (cartography)
- Common queries
- Geoprocessing workflows
Archival package containing geospatial data
Geospatial standards

- ISO 19110, *Geographic information — Methodology for feature cataloguing*
- ISO 19115-1:2014, *Geographic information — Metadata*
- ISO 19139:2007 – Geo Markup Language (GML)
- ISO 19157:2013, *Geographic information — Data quality*

What about the rest ??

- Standardisation of visualization (cartography)
- Geoprocessing (workflows, data transformation)
- Long term preservation formats
Bridging the Gap
Archivists are Extending EAD

ISO 19165 is extending ISO 19115 metadata model

+ Accuracy
+ Lineage
+ Coordinate system

+ GP_Preservation metadata
+ OAIS_Archiv
+ MD_AssociatedResources
> Preservation package description
### Proposed translation between EAD and INSPIRE metadata elements

<table>
<thead>
<tr>
<th>INSPIRE el. Nr.</th>
<th>INSPIRE el. Name</th>
<th>Explanation</th>
<th>Metadata data type</th>
<th>Proposed Cardinality</th>
<th>ISDG code</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1.1</td>
<td>Xpath expression</td>
<td>An XPath expression indicating the metadata element within the ISO 19115 / ISO 19119 UML model.</td>
<td>text</td>
<td>1..1</td>
<td>/</td>
</tr>
<tr>
<td>2.2.1</td>
<td>Resource title</td>
<td>Name by which the cited resource is known</td>
<td>text</td>
<td>1..1</td>
<td>3.1.2 Title</td>
</tr>
<tr>
<td>2.2.2</td>
<td>Resource abstract</td>
<td>Brief narrative summary of the content of the resource(s)</td>
<td>text</td>
<td>0..1</td>
<td>3.3.1 Scope and content</td>
</tr>
<tr>
<td>2.2.3</td>
<td>Resource type</td>
<td>Scope to which metadata applies</td>
<td>CodeList</td>
<td>0..1</td>
<td>(?) 3.1.5 Extent and medium of the unit of description</td>
</tr>
</tbody>
</table>

URL: Bit.ly/GeoCITS
Solution in the Slovenian national archives

- Export from existing systems
- Ingest
- Data Management
- Search
- Access

**Archival tools**
- INGEST tool
- Archival repository
- Archival Catalogue (EAD)

**Geospatial tools**
- Geoserver
- QGIS

Description unit = Parent Identifier

Geonetwork (ISO 19115; INSPIRE)
Conclusions:

• Archiving geospatial data fuels future systems

• Archiving based on standards ensures long term preservation

• We still need a lot of areas to be standardized

• Solution is by using both tools and standards side by side.
Thank You!

Gregor@geoarh.si