Long-term preservation for INSPIRE: a metadata framework and geo-portal implementation

INSPIRE 2010, KRAKOW

Dr. Arif Shaon, Dr. Andrew Woolf
(e-Science, Science and Technology Facilities Council, UK)
Science and Technology Facilities Council

• Provide large-scale scientific facilities for UK Science, particularly in physics and astronomy

• E-Science Centre – at RAL and DL
  – Provides advanced IT development and services to the STFC Science Programme
  – Strong interest in Digital Curation of our science data; R&D Programme: DCC, CASPAR
  – Home of two key UK environmental data centres: British Atmospheric Data Centre (BADC) and NERC Earth Observational Data Centre (NEODC)
  – Active contributor in the international arena of Environmental Informatics, e.g. OGC(OWS 6), INSPIRE and ESA
Long-term Preservation of Spatial Information: Motivation for INSPIRE

- The INSPIRE directive requires global availability and uniform accessibility of heterogeneous environmental datasets across Europe through
  - common Implementation Rules (IRs)
  - Metadata standards (e.g. ISO 19115) and
  - Interoperable Data Infrastructures (e.g. INSPIRE SDI).

- Interoperability does not always guarantee sustainability over the long-term. A key question to be answered-
  - What happens to the data when a data provider ceases to exist?

- A phenomenal deluge of spatial data over the last decade
  - Triggered by the growing concerns over environmental problems, such as global climate changes
  - Ensuring sustained access to these data is becoming more difficult.

- Efficient long-term preservation is required for both current and historical spatial data exposed through INSPIRE.
Long-term Preservation of Spatial Information: Main Challenges

- Environmental data inherit the preservation challenges inherent to all digital information.
  - Existing preservation approaches and standards, such as the OAIS Reference Model should be also applicable to environmental data.

- Environmental data adds to these:
  - Highly structured and complex data models (“feature types”) that require special knowledge for accurate interpretation.
  - Static data being replaced with dynamic web services, such as the OGC web services.
  - Existing preservation approaches would need to be tailored to handle these added complexities.
  - The work presented explored the applicability of the OAIS Reference model to the preservation of environmental data.
The Current State of Play

• A relatively underexplored area until recently, when ESA announced the Long Term Digital Preservation (LTDP) initiative for their increasingly voluminous Earth Observation datasets.

• Other notable relevant initiatives:
  – The National Geospatial Digital Archive (NGDA) project funded under the National Digital Information Infrastructure and Preservation Program (NDIIPP): approach specific to US-based data
  – The Geospatial Electronic Records (GER) project: new metadata format introduced is incompatible with ISO 19115 - the metadata format required by European law and INSPIRE for describing European environmental data.
  – Some exploratory work by the Digital Preservation Coalition (DPC)
The OAIS Reference Model

- A widely adopted ISO Standard for long-term preservation of digital objects
- Defines an information model that needs to be captured for effective preservation

Facilitates data discovery, e.g. Keywords, abstract

Aids data management, e.g. Provenance history, versioning info

Information needed to render data in future, e.g. Software, hardware
Preservation Aspects of INSPIRE SDI

What already exists:
- ISO 19115 - Good for resource discovery
- Controlled vocabulary for semantic metadata validation

What is missing:
- ISO 19115 is not curation-aware
- Insufficient RI
- Data annotation is not captured

- Ad-hoc approaches to data management and storage
- Not considered in this project
A Preservation-focused INSPIRE SDI

Metadata Curation

- Catalogue 1
- Catalogue Service 1
- Catalogue <n>
- Catalogue Service <n>

Data Preservation

- Geospatial data repository 1
- Geospatial data repository <n>
- Geospatial data repository 2

User applications

RI Registry

- Security, authentication, DRM
- Direct access to data
- Search and retrieval
- Access to processed data, services

Client software

Re-render Raw Data

Distributed services

Distributed content repositories
A Preservation Profile of ISO

- Extends “MD_ApplicationSchemaInformation” used to create a particular feature view of a source geospatial dataset
- Adds information about the mapping between a source data and its application schema
- Adds information about applications/software/services required to effectively apply the mapping
- Defines additional data specific RI (e.g. data formats, storage media), mainly in the form of web-accessible resources (e.g. URL).
- Enables data providers to record RI in other formats than ISO 19115.

---

[Diagram with elements labeled: DQ_Element, OtherElement, +otherProperty, +report, RI_Lineage, RI_ProcessStep]
A Prototype Preservation-aware Geo-Portal

- Implemented a web-based portal that demonstrates the underlying functions of a preservation-aware SDI
- Based on GeoNetwork – a widely adopted open source and standards-based catalogue service, also used for the INSPIRE GeoPortal.
- Key features:
  - Recording, editing, searching and viewing metadata in the Preservation profile of ISO 19115
  - Versioning of metadata
  - Annotation of both Data and Metadata through an intuitive and user friendly wizard; captures annotation context in Xpath for metadata records
A Prototype Preservation-aware Geo-Portal (Annotation Wizard)
A Prototype Preservation-aware Geo-Portal (Annotation Wizard)
Conclusions & Future Directions

- Long-term preservation of both current and historical environmental data exposed through INSPIRE is highly important for e.g. monitoring and analysing climate change.

- Awareness is growing in Europe with the emergence of ESA LTDP, albeit not addressed in the current INSPIRE directive.

- The work presented investigates the requirements for a preservation-aware SDI for INSPIRE and presents a preservation profile of ISO 19115 that outlines the metadata requirements.

- Future work would need to focus on the implementation of efficient and interoperable data preservation solutions for the INSPIRE data repositories.
References


• National Digital Information Infrastructure and Preservation Program (NDIIPP) - http://www.digitalpreservation.gov/library/

• Geospatial Electronic Records (GER) project - http://www.ciesin.columbia.edu/ger/


Questions?