FOSS4G BOF Meeting on Geodata Discovery and Metadata Models

Stefan F. Keller
GISpunkt / Institute for Software
University of Applied Sciences Rapperswil (HSR)
CH-8640 Rapperswil, Switzerland
www.gis.hsr.ch / www.ifs.hsr.ch

Short Statements (all): experiences, requirements?

Approaches:
- owscat (Tom Kralidis)
- OAI approach (Stefan F. Keller)
- A Reference Model (Rob Atkinson)
- Postponed: Google scraping (Paul Ramsey) and Client programming (Jody Garnett)
- GeoNetwork (Jeroen Ticheler)

Models for Metadata (all): Management vs. Exchange

Discussion (all):
- Pros/ Cons, Timelines, Way forward
Results...

See

- OSGeo geodata mailing list
  https://geodata.osgeo.org/servlets/BrowseList?listName=geodata

- OSGeo Wiki

Meeting too short (probably the shortest...)

Thanks to all for attending; see you again!
Vision for agile metadata

User, vision and implementation first!
Let’s unchain geodata!
Let’s make metadata sexy!

Contribution by Stefan F. Keller
Metadata is uncool (some think...)

Because its...

- for later
- for others
- for documentation

... so it’s like teeth pulling?

But wait:

- Did you ever wanted to make/have a contract?
- Ever searched a file on your disk?
- Did you ever wished to find and use data from others about your area?
Vision - Models and Protocols for Metadata Management and Exchange

Intranet

- Intranet Portal

Users

- NSDI Portal
- Search Services

Internet

- Google
- Metadata Exchange Protocol (like OAI-PMH)

Search Services

- oai_geometa
- oai_dc

Catalog Services / (Meta) Data Provider

- Catalog / Editor (e.g. GMDB)
- Catalog / Editor (e.g. Geocat)

Data Access Services / Data Warehouse, Archive

- FTP/WMS/WFS
- Data Access

and

Webservices

- Filter / Value Adding Services
- Coord.Transformer, Visualizer, Label Placer, etc.
Misunderstandings about metadata

- **ISO 19115 Core/ Comprehensive are complete?**
  - No, services lacking, Addresses are (always) complicated
  - Still serves as blueprint, but don’t expect that all attributes are implemented before hand

- **ISO 19115 is for exchange?**
  - Yes, but only for full metadata management, not for moving metadata around (lightweight ‘harvesting’)
  - Good encoding lacking?

- **Catalogs (Catalog Services) are search services?**
  - Yes, but not search engines like Google or other ones
  - Rather primarily metadata management tools

- **WMS are services on their own?**
  - No, (e.g. Coord.Transform or Viz. are). Are data access services, a metadata attribute type belonging to geodata
OAI-PMH 2.0

See www.gis.hsr.ch/wiki/OAI-PMH
Basic idea behind OAI-PMH

- Provide a protocol for aggregation (harvesting) of metadata about any kind of resource
- Origin in Open Access Initiative: Digital Libs
- Dublin Core (ns: oai_dc) is mandatory
  - Additional metadata models can be added
- OAI-PMH
  - is a protocol, needs to be implemented
  - OAI-PMH 2.0 current version, upward-compatible
- Some implementations exist... Google Sitemaps!
Architectural Overview: (Meta-)Data Providers and Search Providers

Harvester
Harvester
Harvester
Harvester

Catalog / Repository
Harvesters issue OAI-PMH requests for metadata via HTTP.
Harvesters issue OAI-PMH requests for metadata via HTTP. A Repository processes the OAI-PMH requests and has to implement the protocol.
Architectural Overview: (Meta-)Data Providers and Search Providers

Harvester

Catalog / Repository

Harvester

Request

Harvester

Index

Request

Index

Index

Harvester

Index

Harvester
<metadata>
  <oai_dc:dc
    xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/
    xmlns:dc="http://purl.org/dc/elements/1.1/"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
    http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
    <dc:title>Using Structural Metadata to Localize Experience of Digital Content</dc:title>
    <dc:creator>Dushay, Naomi</dc:creator>
    <dc:subject>Digital Libraries</dc:subject>
    <dc:description>With the increasing [..bla..] to particular communities of users. </dc:description>
    <dc:date>2001-12-14</dc:date>
    <dc:type>e-print</dc:type>
  </oai_dc:dc>
</metadata>
OAI-PMH: Sets

- Items can be organized into sets.
- Sets can either be organized flat or hierarchically.
- Sets can point to 'friends'

<table>
<thead>
<tr>
<th>setName</th>
<th>setSpec</th>
</tr>
</thead>
<tbody>
<tr>
<td>Institutions</td>
<td>institution</td>
</tr>
<tr>
<td>Oceanside University of Nebraska</td>
<td>institution:nebraska</td>
</tr>
<tr>
<td>Valley View University of Florida</td>
<td>institution:florida</td>
</tr>
<tr>
<td>Subject</td>
<td>subject</td>
</tr>
<tr>
<td>Existential Kenesiology</td>
<td>subject:kenesiology</td>
</tr>
<tr>
<td>Quantum Psychology</td>
<td>subject:quantum</td>
</tr>
</tbody>
</table>
OAI-PMH advanced: Selective Harvesting and Resumption

- **Query:**
  Harvesters can specify some constraints on which items they are interested in
  - Regarding datestamps: only items that were created, modified or deleted (optional) in a certain time period
  - Regarding sets: only items that belong to a specific set (or any of its subsets)

- **Resumption:**
  Token to deliver chunks only
OAI-PMH: Command Overview (I)

- **Identify**: retrieve information about a repository (name, protocol version, supports deletion, ...)

- **ListMetadataFormats**: lists the available metadata formats of a repository

- **ListRecords**: get either all records or a subset, must specify metadata prefix
OAI-PMH Command Overview (II)

- **ListSets**: returns the set structure of a repository

- **GetRecord**: get a specific record, must specify item’s URI and metadata prefix

- **ListIdentifiers**: like ListRecords, but retrieves only headers
Implementation Example GMDB
Implementation Example OAI-PMH Data Provider Service

GISpunkt HSR OAI-PMH 2.0 Data Provider

This is an implementation for an OAI-PMH 2.0 Data Provider, written in PHP.

This implementation (phpoai2) completely complies to OAI-PMH 2.0, including the support of on-the-fly output compression.

Links:
- Geo-Metadatenbank (GMDB) am GISpunkt HSR
- geometa.info - Die Suchmaschine für Geodienste, Geodaten und Online-Karten

Documentation:
- GMDB-Dokumentation, Support und Anregungen
- About OAI-PMH on GISpunkt HSR Wiki
- phpoai2 README and phpoai2 Changes

Query and check this Data Provider:
- Identify
- ListMetadataFormats
- ListSets
- ListIdentifiers
- ListRecords
- GetRecord

For detailed tests use the Recosrtor Explored

Any comments or questions are welcome:

Heinrich Stamer/johanns
Institute for Science Networking
stamerm@tu-ka.de

http://www.gis.hsr.ch/gmdb/
Models for Metadata Management and Exchange
Tele.../Jo's... Metadata Model

- Originator: name, address (dc:creator)
- Organization: name, address (dc:?)
- Title: dc:title
- Description: dc:description
- Publication Date: dc:date (of metadata)
- Timespan: ? (dc?)
- License terms: (dc:rights)
- Extents: georss:box? (dc:coverage)
- Spatial reference: CRS or place name?
- Spatial reference: Raster, Vector (dc:type?)
- Data source: URL => torrent, WMS, WFS, shape (dc:type?)
- Projection: (dc?)
Proposal of a Geometa Data Model

- **Relation**: URL (multivalued) - Reference to other metadata records
- **Type**: URL (preferred) or enum (multivalued) - filter services (WPS) or Protocol for data access services (DAS). Well known DAS: WMS, WFS (WxS) else WSDL
- **Identifier**: string - Unique id to identify a metadata record (URI).
- **Title**: string - Title
- **Coverage.box**: Rectangular box (mandatory) in WGS84 (e.g. georss:box)
- **Coverage.name**: String - (optional) a geographic name
- **Description**: string (multivalued) - Some free text
- **Subject**: enum - Classification from ISO 19115 as enum type
- **Language**: enum - ISO Code
- **Format**: enum (preferred) or string - File type or name of originating source system
- **Source**: URL (preferred) or string - Lineage information
- **Date**: date - Publication date or date of last change of metadata record
- **Creator**: structure - Civic Address of data owner (xAL?)
- **Publisher**: structure - Civic Address (see Creator)
- **Contributor**: structure - Address (see Creator) Leave unused?
- **Rights**: URL (preferred) or string - License information about the data
- **Audience**: string - Not used or 'GIS' as a constant