

# 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](http://www.gis.hsr.ch) / [www.ifs.hsr.ch](http://www.ifs.hsr.ch)



# ■ ■ Agenda 14.Sept., 12:30?-13:30

- **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

[http://wiki.osgeo.org/index.php/Geodata\\_Discovery\\_Working\\_Group](http://wiki.osgeo.org/index.php/Geodata_Discovery_Working_Group)

## ■ 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...

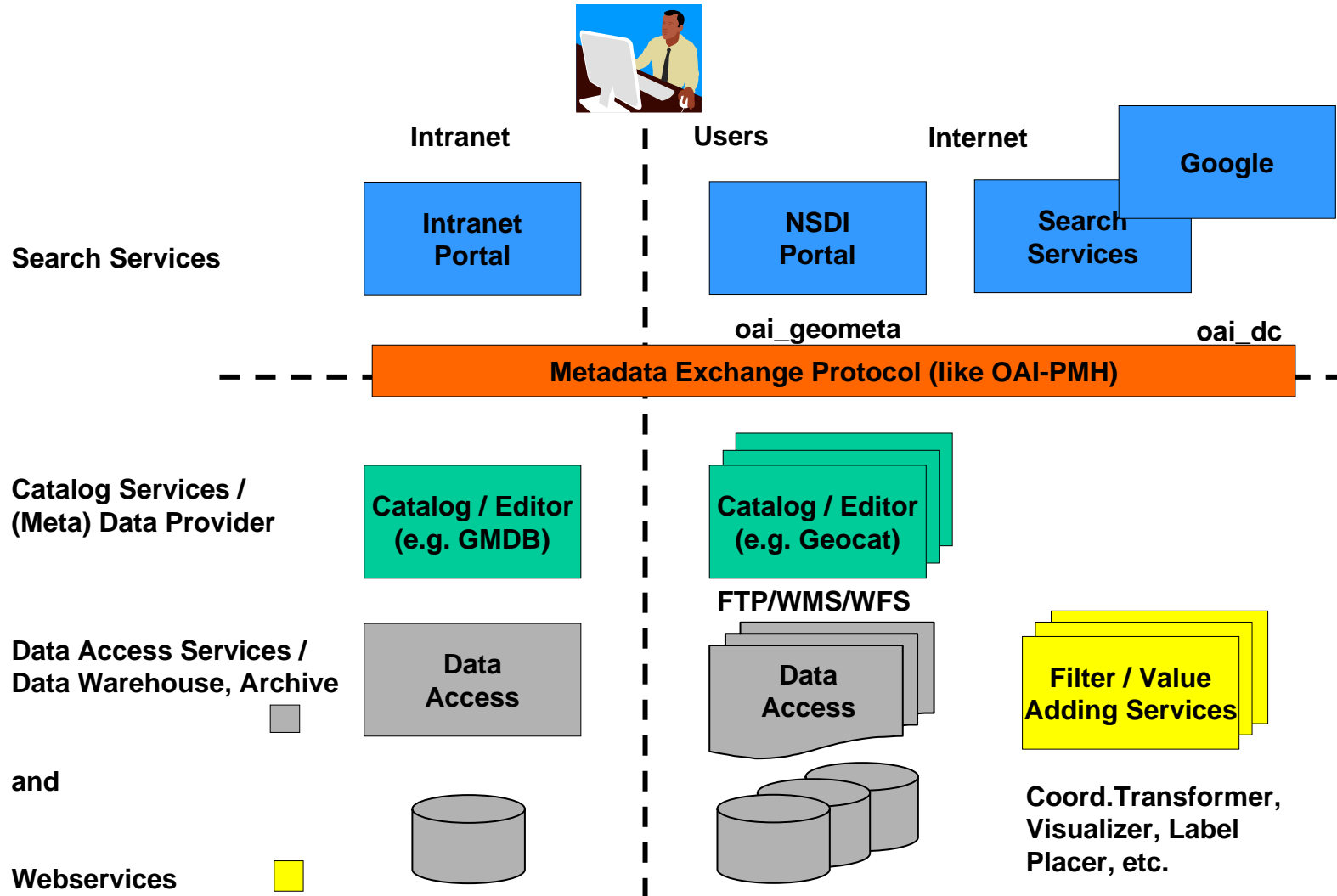
- for later
- for others
- for documentation

... 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





# Misunderstandings about metadata

- **ISO 19115 Core/Comprehensive are complete?**
  - ▶ No, services lacking, Adresses 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](http://www.gis.hsr.ch/wiki/OAI-PMH)



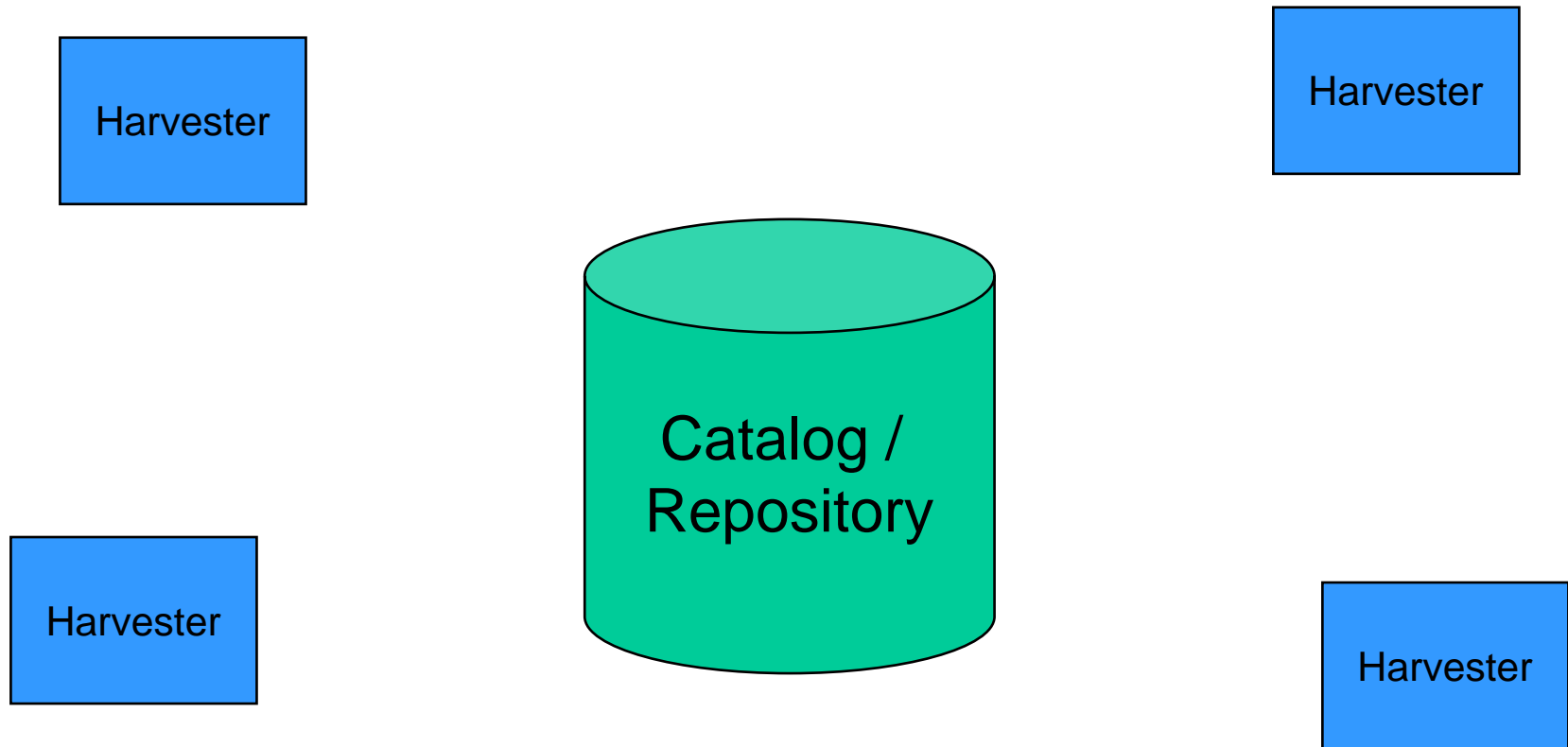


# ■ ■ Basic idea behind OAI-PMH

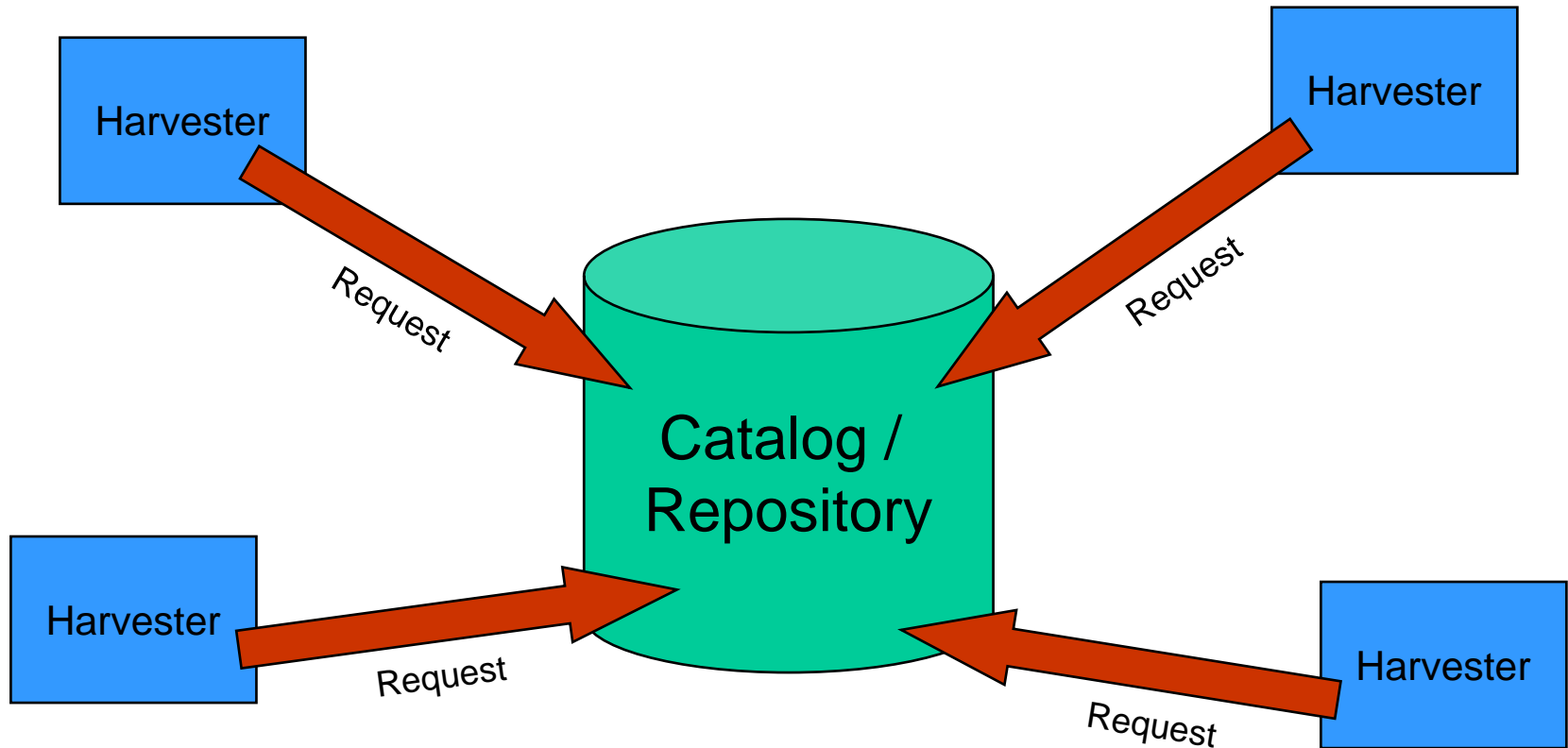
- Provide a protocol for aggregation (harvesting) of metadata about any kind of resource
- Origin in Open Access Initiative: Digital Libs
- Dublin Core (ns: oai\_dc) is mandatory
  - ▶ Additional metadata models can be added
- OAI-PMH
  - ▶ is a protocol, needs to be implemented
  - ▶ OAI-PMH 2.0 current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



# Architectural Overview: (Meta-)Data Providers and Search Providers

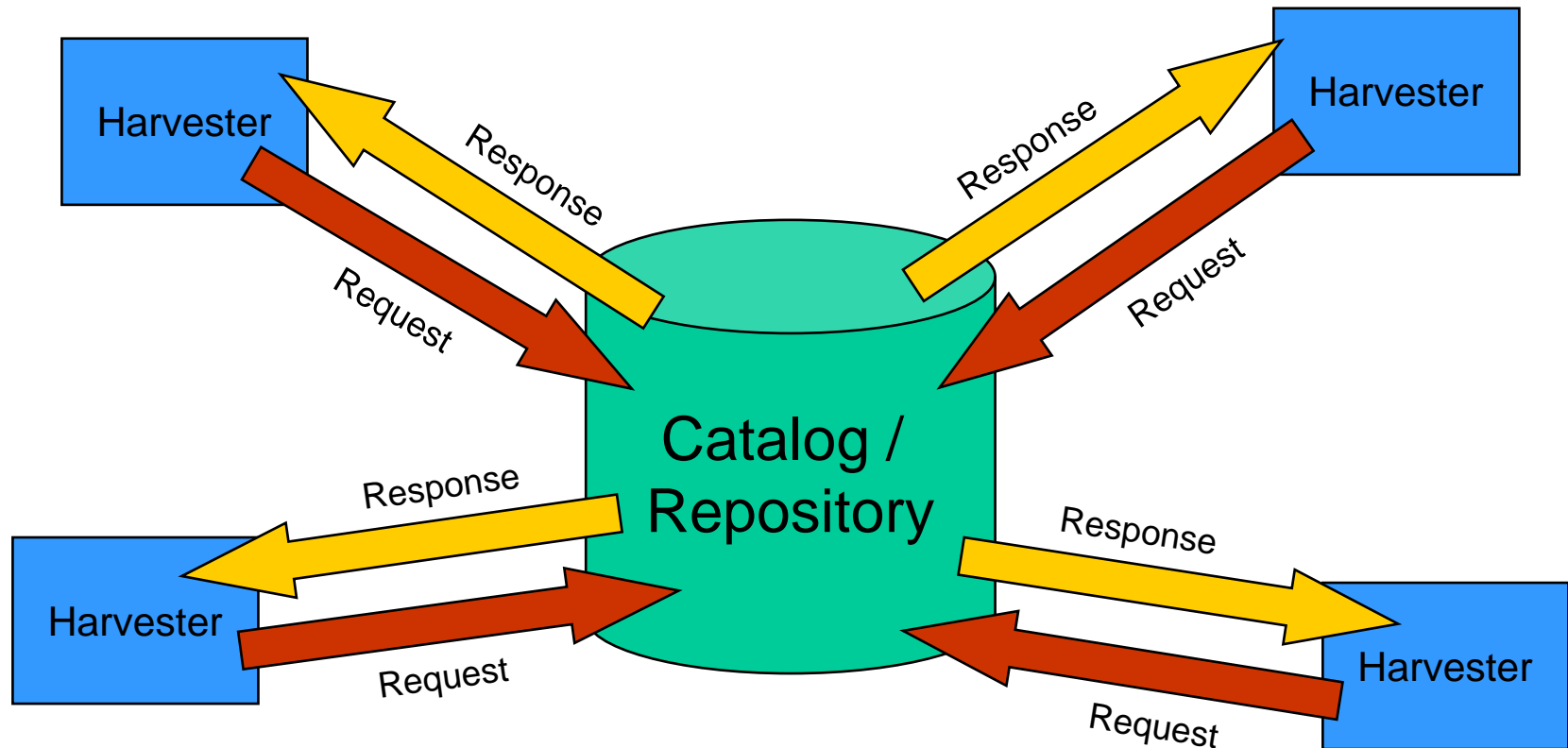


# Architectural Overview: (Meta-)Data Providers and Search Providers



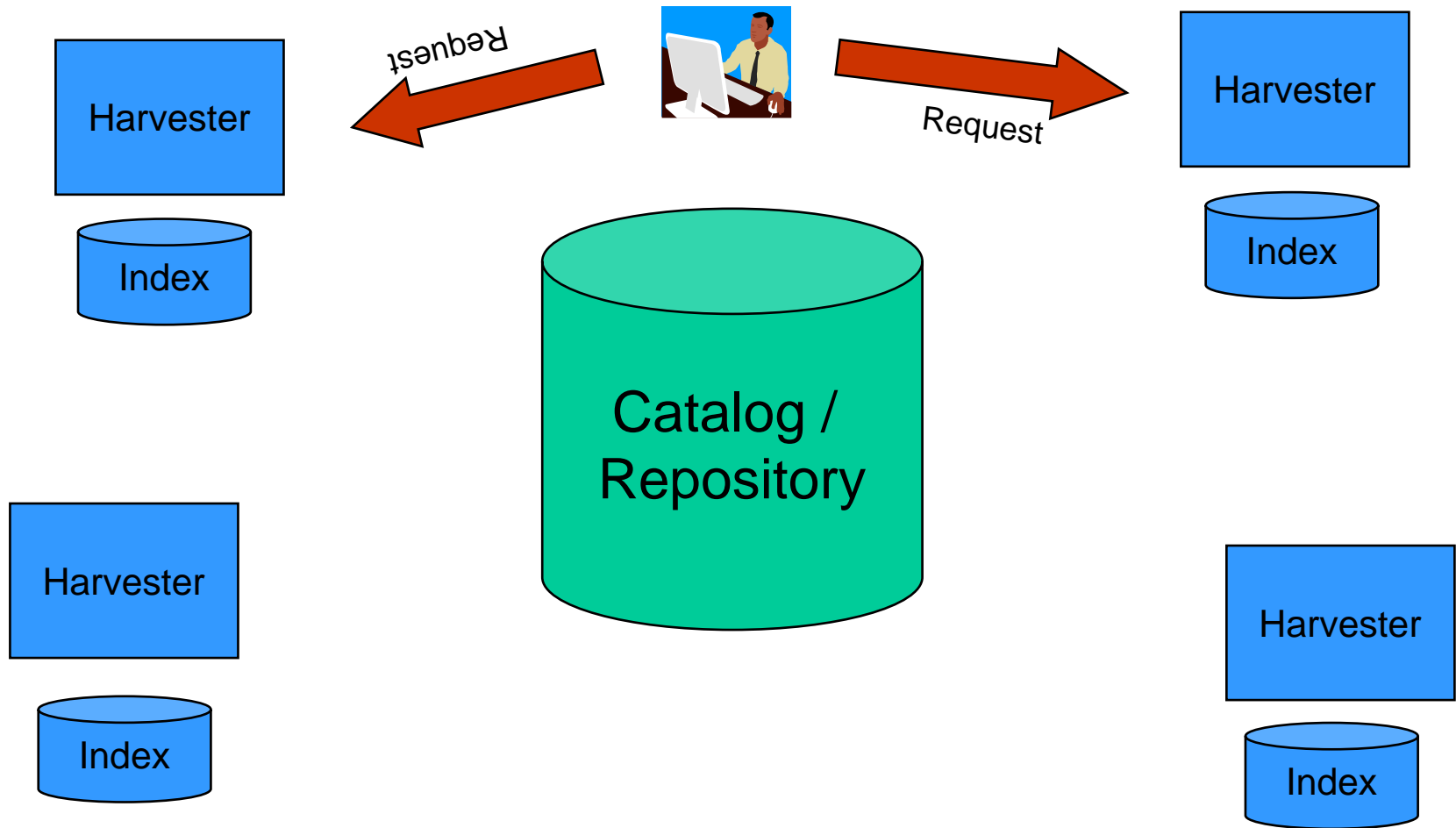
**Harvesters** issue OAI-PMH requests for metadata via HTTP.

# Architectural Overview: (Meta-)Data Providers and Search Providers



**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





# OAI-PMH: oai\_dc record format XML

```
<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>
    <dc:identifier>
      http://arXiv.org/abs/cs/0112017
    </dc:identifier>
  </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‘

setName	setSpec
Institutions	<code>institution</code>
Oceanside University of Nebraska	<code>institution:nebraska</code>
Valley View University of Florida	<code>institution:florida</code>
Subject	<code>subject</code>
Existential Kenesiology	<code>subject:kenesiology</code>
Quantum Psychology	<code>subject:quantum</code>



# 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

Geo-Metadatenbank - Titel - Mozilla Firefox

http://www.geometa.info/gmdb/metadaten\_details.php?\_key\_\_=10169

Startseite > Metadaten > Titel

### Titel

Metadaten			
Metadaten-ID	10169	Bearbeiter	Stefan Keller
Metadaten-Erstelldatum	24.02.2006	Metadaten-Änderungsdatum	09.03.2006
Typ	Plan auf Papier	Metadaten öffentlich	ja
Metadaten Online			
Untergeordnete Metadaten		Übergeordnete Metadaten	

Allgemeines			
Daten-ID/Dateiname	DatenID		
Titel	Titel		
Inhalt	Inhalt		
Vorschaugrafik	Kein Bild vorhanden		
Originalformat	SHP (.shp)	Erstelldatum	24.02.2006
Sprache	de	Änderungsdatum	24.02.2006
Status	fertiggestellt	Nachführung	keine
Bemerkungen	Bemerk		
Datenschema (Datei)	http://webgis.hsr.ch/rappinfo/		
Themen	• Grenzen		

Find: keller Find text Find Previous Highlight all Match case



# Implementation Example OAI-PMH Data Provider Service

Geo-Metadatenbank (GMDb) GISpunkt HSR OAI-PMH v2 Data Provider (phpoi2) - Mozilla Firefox

http://www.geometa.info/gmdb/oai\_pmh\_data\_provider/

## Geo-Metadatenbank

### 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 (phpoi2) 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](#)
- [phpoi2 README and phpoi2 Changes](#)

Query and check this Data Provider:

- [Identify](#)
- [ListMetadataFormats](#)
- [ListSets](#)
- [ListIdentifiers](#)
- [ListRecords](#)
- [GetRecord](#)
- For detailed tests use the [Repository Explorer](#)

Any comments or questions are welcome.

*Heinrich Stamerjohanns*  
Institute for Science Networking  
stamer#AT#uni-oldenburg.de

Find:  Find text Find Previous Highlight all Match case

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 an 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