Clearinghouse and Metadata Concepts

Daniel Berhanu
Yaounde, Cameroon

dberhanu@uneca.org

June 16, 2004

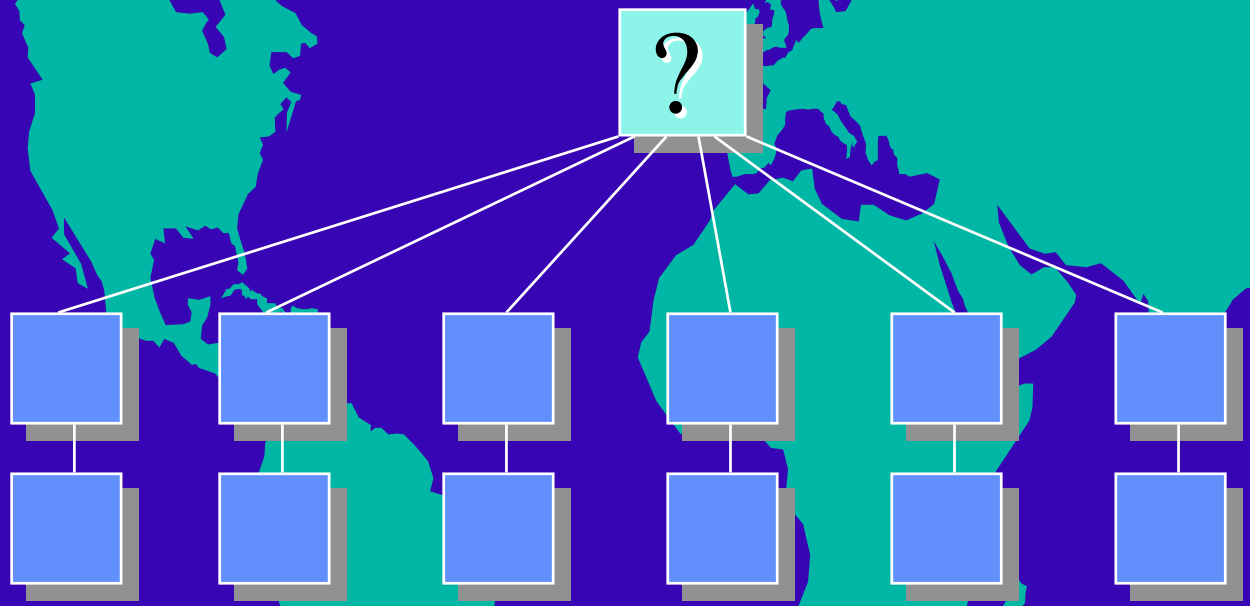
What are Metadata?

- Field level description of data and information resources, in a standard form
- Examples:
  - Identification of data resource
  - Originator and location
  - Purpose for which data set was created
  - Data collection date, processes and methods
  - Availability and access constraints
  - 'Distributor' and contact details
  - Etc.

Major Benefits of Metadata

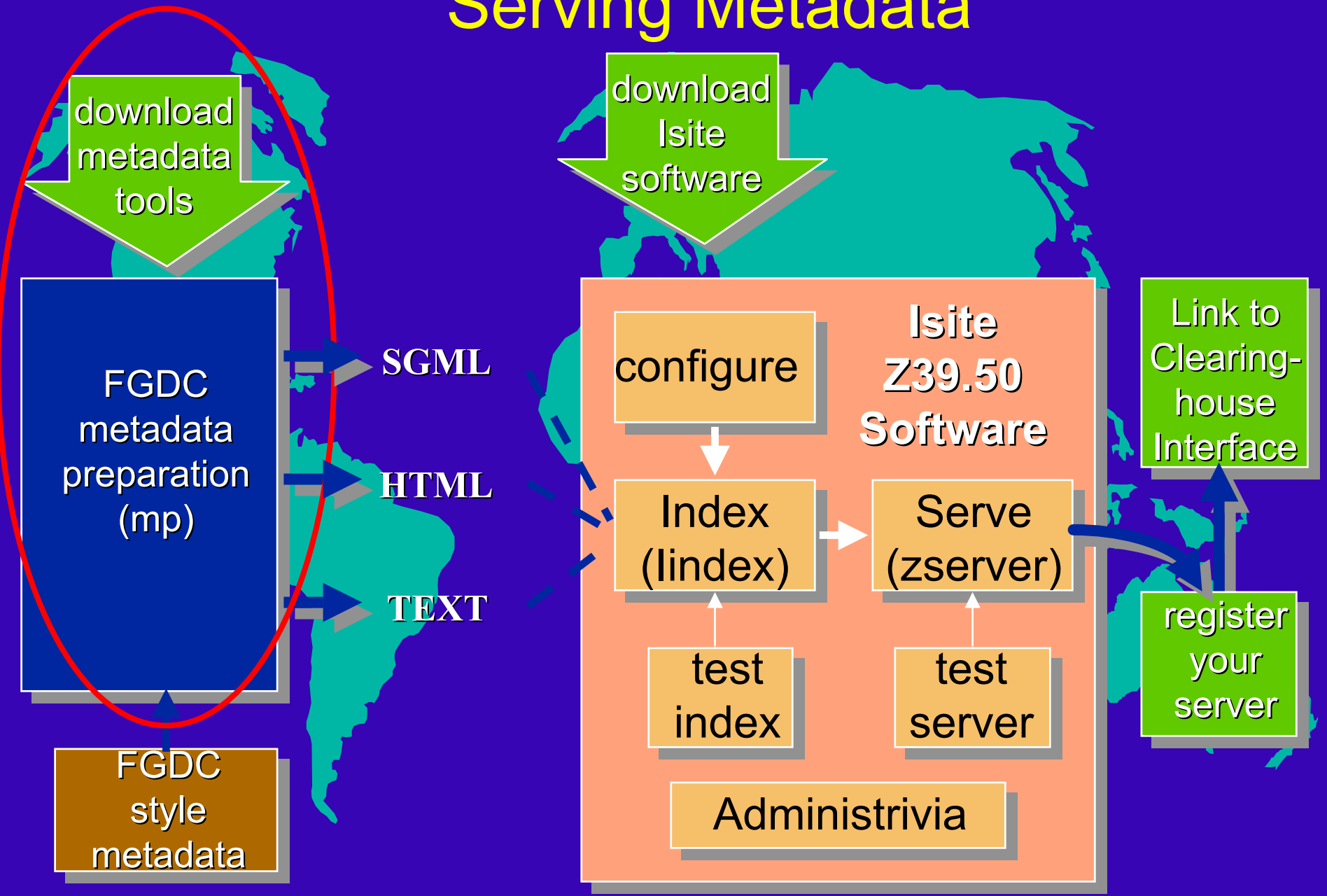
- Excellent **insight** into your data holding
- Faster access to **more suitable** data
- No loss of **knowledge** by departing staff
- Avoidance of duplicated **efforts & costs**
- Increased **client services** against lower cost
- Structured **search and comparison** of available data resources by all stakeholders

What is Clearinghouse?



- Decentralized system of searchable Internet servers containing structured metadata and data
- Each server is a “peer” in the network

Serving Metadata



Clearinghouse Provides…

- Uniform, distributed search through a single user interface to all servers worldwide
- Search for data through fields and full-text in the metadata
- A common vocabulary for metadata discovery on the Internet
- Links through to full data access, where available
- Controlled decentralization of data resources
  - Each resource is maintained by its primary producer with structured description for others to know of its existence and how to access and use it

Specify Date and Content

Specify Time Period of Content

Help...

*Specify a date or date range for desired spatial data by selecting **one** of the methods below.*

- ○ **Don't search based on time period**
- ○ **Get data whose date is** before **the date** May. 15 1998
- ◉ **Get data from** Jan. 15 1998 **through** Jun. 15 1998

Search in Full-Text (Any) or by Field

Help...

Specify search words by using one or more of the fields below.

	Search for: Transbounday	**in the field**	Any
OR / AND	Water	**in the field**	Title
OR / AND	Negotiation	**in the field**	Theme Keyword
OR / AND	Policy	**in the field**	Purpose

Specify Nodes to search

Select Data Servers to Search Help...

Specify the data sources to query and the number of records that will be retrieved from each source. Use control key combinations to make multiple selections.

- FEMA MMI Project (Q3 and HAZUS data)
- FGDC Map Service Test Node
- Forest, aquatic, and Rangeland Ecosystems in the Western USA
- Framework Data Survey
- Geography Department SDSU Clearinghouse
- Geography Network
- Geological Survey of Alabama Geospatial Data Clearinghouse Node
- Georgia GIS Data Clearinghouse
- Global Change Master Directory
- Greater Yellowstone Area Data Clearinghouse
- Guatemala - Clearinghouse /SNIG
- Guatemala Clearinghouse Universidad del Valle
- Gulf Coast NASA/UL Lafayette Regional Application Center
- Gulf of Mexico Geospatial Clearinghouse Node
- Hawaii GIS Clearinghouse Node

Search Results

Done with search!

Select the links below to view matches by database.

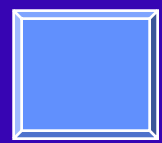
Database	Status	# Results
Africa Data Dissemination Service	Search Successful	1
FEMA MMI Project (Q3 and HAZUS data)	Search Successful	4
Forest, aquatic, and Rangeland Ecosystems in the Western USA	Search Successful	0
Framework Data Survey	Search Successful	0
Geography Network	Search Successful	18
Global Change Master Directory	Search Successful	1237

Why do Clearinghouse?

- Minimize duplication of effort in data and metadata collection and processing
- Provide means to advertise metadata collection requirements, inventory, and quality
- Support documentation of basic metadata sets for advised re-use of internal and external applications

Why not just use Web indexes?

- On-line indexes support only full-text literal search with limited support for concept search
- Indices do not support fields or other data types (date, coordinate, other numeric)
- Place names promote ambiguous geographic search and retrieval
- Server data increasingly in databases, which are not accessible to web indexes



Metadata Format Issues

- Clearinghouse requires standardized metadata fields for query, exchange, and presentation
- Example of existing standards
  - Dublin Core, GILS, FGDC, ISO
- Standardized encoding for full text and field searches
  - Standard Generalized Markup Language (SGML) being proposed as basis for metadata encoding guideline to enforce structure and help in presentation

Tag-Value Syntax

- Difficult to produce standard tag-value descriptions of data resources in XML, SGML or other encoding formats, as required for clearinghouse services

Citation:
- *Citation_Information:*
  - *Originator:* USAID / FEWS
  - *Publication_Date:* 1996
  - *Publication_Time:* 0000
  - *Title:* Mozambique - Administrative Level 3
  - *Edition:* 1
  - *Geospatial_Data_Presentation_Form:* map
  - *Series_Information:*
    - *Series_Name:* FEWS 3
    - *Issue_Identification:* 1
  - *Publication_Information:*
    - *Publication_Place:* Sioux Falls, SD, USA
    - *Publisher:* EDC/International Program
  - *Other_Citation_Details:* FEWS/EROS Data Center (EDC)
  - *Online_Linkage:* http://edcintl.cr.usgs.gov/bin/ftpadds/a=admn/b=mz

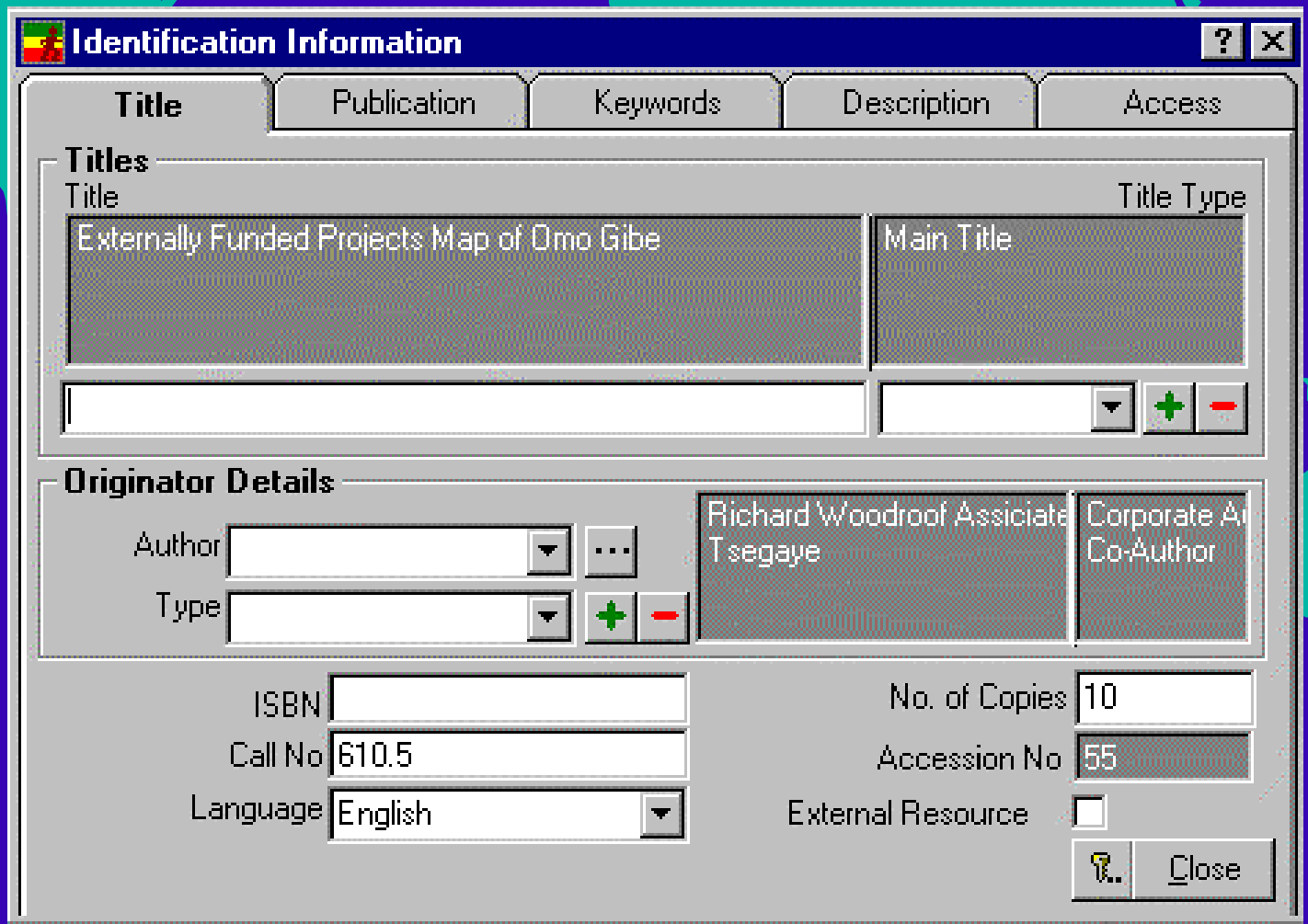
Description:
- *Abstract:*

  This ARC/INFO polygon coverage contains the third level administrative boundaries at a 1:1000000 scale.

  The coverage is part of the Famine Early Warning System (FEWS) / Associates in Rural Development

ENRAEMED: A Metadata Editor

- Software tool needed for encoding the field level descriptions into required formats



ENRAEMED

- Ethiopian Natural Resources And Environmental MetaDatabase

  -ENRAEMED means Let's Walk Together

- Objective
  - To make existing information on the nation's natural resources and the environment more accessible to potential users

Federal Partners

- Ministry of Water Resources
- Ministry of Agriculture
- Environmental Protection Authority
- Ethiopian Mapping Authority
- Ethiopian Science and Technology Commission
- Ethiopian Institute of Geological Survey
- National Meteorological Service Agency

No existing software matched our Functional Requirements

+ Current & future Geospatial MD Standards

+ 2 Bibliographical Metadata Standards

+ Extensive Defaults

+ Memo Manager

+ Freeware

+ XML import

+ Free Database

+ XML export

+ Highly scaleable

+ Element Navigator

+ Spelling Checker

+ Globally unique ID

+ Keyword Selector

+ Multi-lingual

+ Administrator

+ Map-based Bounding Coordinates entry

Maintained and supported by UNEP / FGDC

ECA Clearinghouse Project

- Assist member states to develop metadata systems and clearinghouse services
  - Point of reference for member states
  - Training at ECA
- Single point of entry to metadata collections of member states
  - "Country Spaces" for interim hosting of metadatabases
- Initial focus on geoinformation
- Registered with FGDC clearinghouse node with worldwide search access

ECA Clearinghouse …

- Will extend to other information types
- Include bibliographic metadata, providing field-level search to all data resources of ECA and partners
- Install own gateway software to provide direct access to metadata collection
  - Initially indirect access through FGDC

For More Info

- Daniel Berhanu
  - Development Information Services Division
    United Nations Economic Commission for Africa
    Phone: 251-1-443426
    dberhanu@uneca.org
- Dozie Ezigbalike
  - Development Information Services Division
    United Nations Economic Commission for Africa
    Phone: 251-1-444569
    ezigbalike.uneca@un.org