



# The NOAA Data Center Metadata Enterprise: *IT Solutions*

NOAA National Coastal Data Development Center  
NOAA National Oceanographic Data Center  
NOAA National Geophysical Data Center  
NOAA National Climatic Data Center

26-27 August, 2009  
Asheville, NC, USA



# Logistics

- Safety
- Restrooms
- Lunch
- Group Dinner Tonight at Tupelo Honey Café
- Tour Tomorrow



# Purpose

- To examine, from a technical standpoint, the feasibility of a joint development path with NCDDC and GeoNetwork leading toward the longer term vision as described in the suite of enterprise metadata functions and requirements
- To determine technical options for presentation to the Data Center Directors

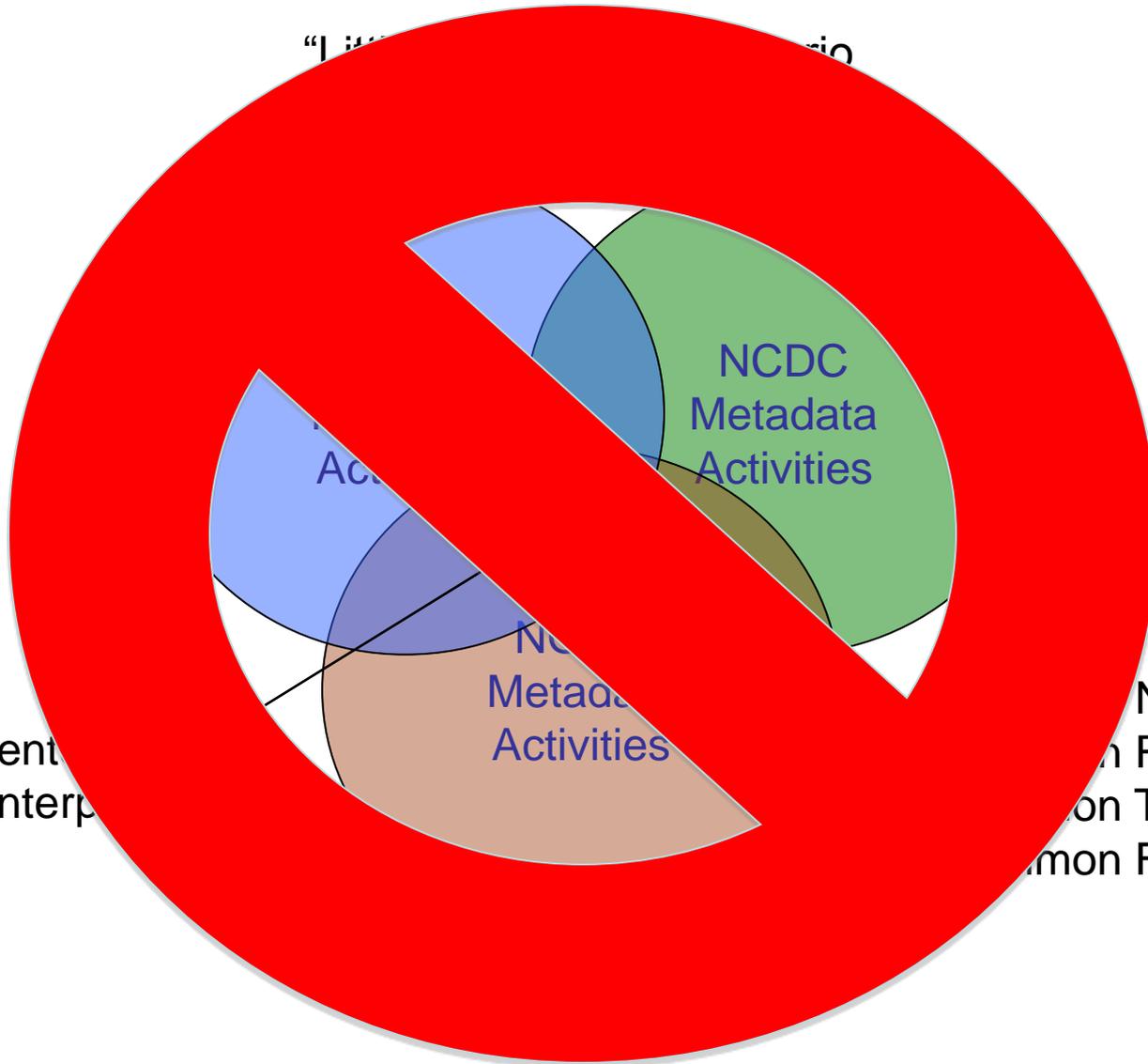


# Goal

- To provide several options for moving forward on an enterprise solution for metadata including pros/cons, timeliness, cost, resources, etc.



# The Metadata Enterprise



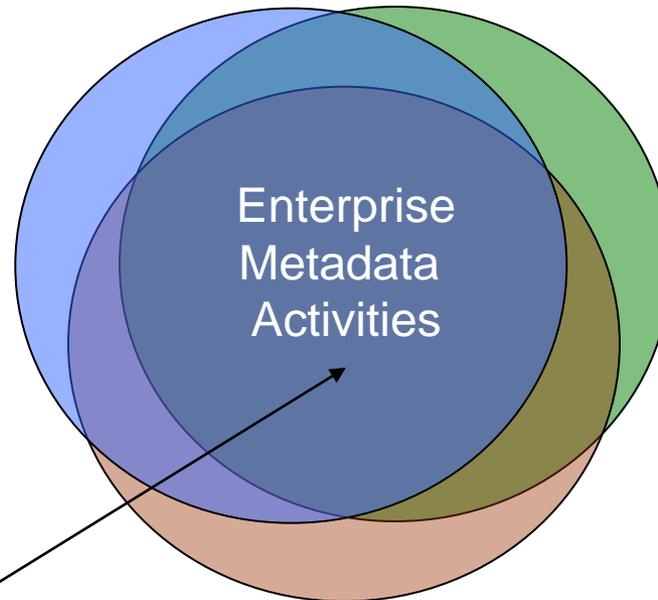
The Data Center  
Metadata Enterprise

Needs  
Requirements  
Tools  
Common Functions



# The Metadata Enterprise

“Lots in common” Scenario



The Data Center  
Metadata Enterprise

Common Needs  
Common Requirements  
Common Tools  
Common Functions



# Enterprise Functions

The Data Center Enterprise Metadata System

## Metadata Manipulation Functions

Import

Convert

Export

Validate

Publish

Edit

## Management/Admin Functions

Manage Controlled  
Vocabularies

Manage  
Components

Manage DM  
Data

Support Queries

Generate  
Reports

Control  
Access

Manage  
Workflow

Link to Archival  
Storage

Handle  
Versions

Minimize  
Duplicates

## System-Wide or Cross-Cutting Functions/Requirements

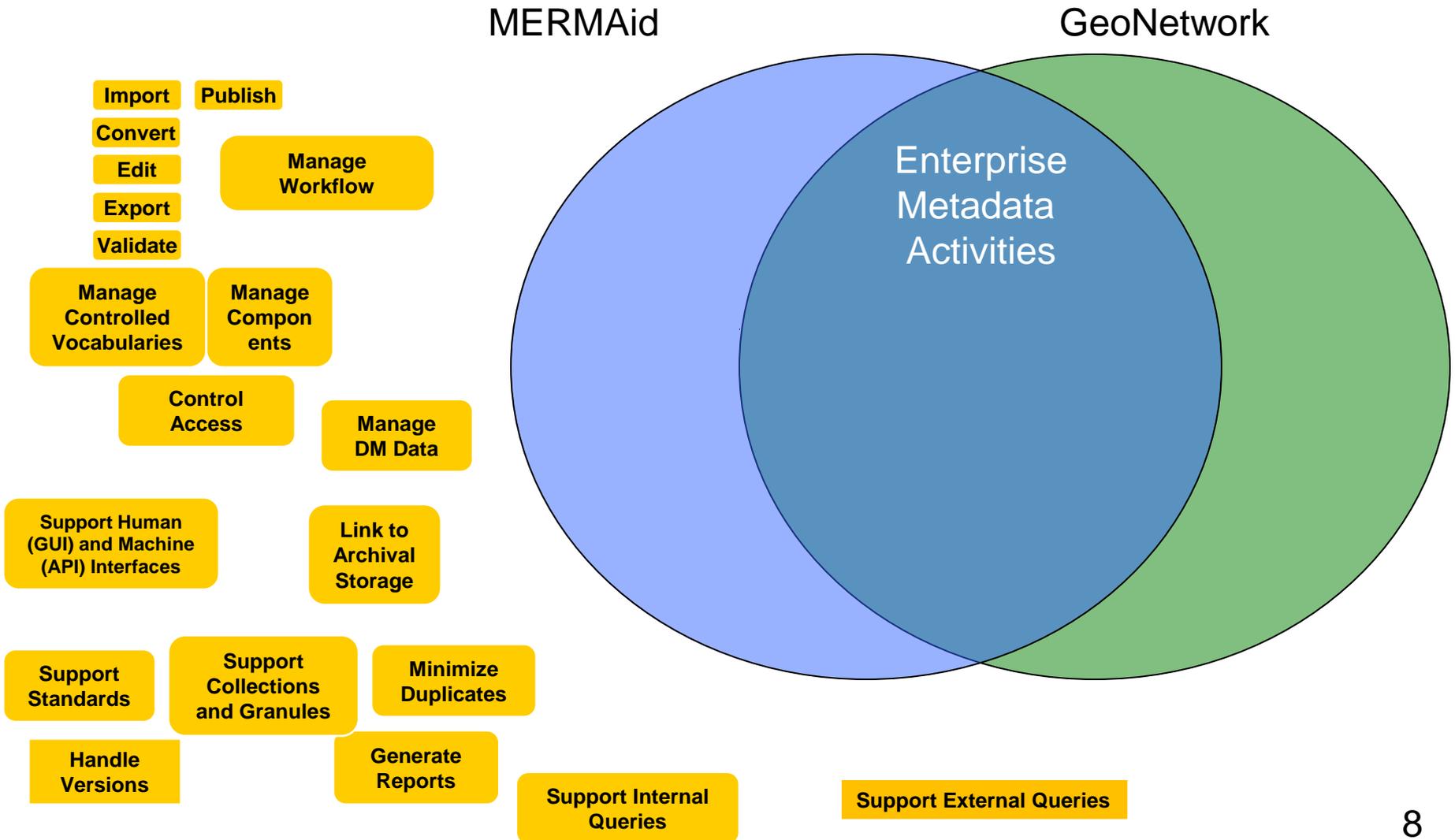
Support Human (GUI) and  
Machine (API) Interfaces

Support  
Standards

Support Collections  
and Granules

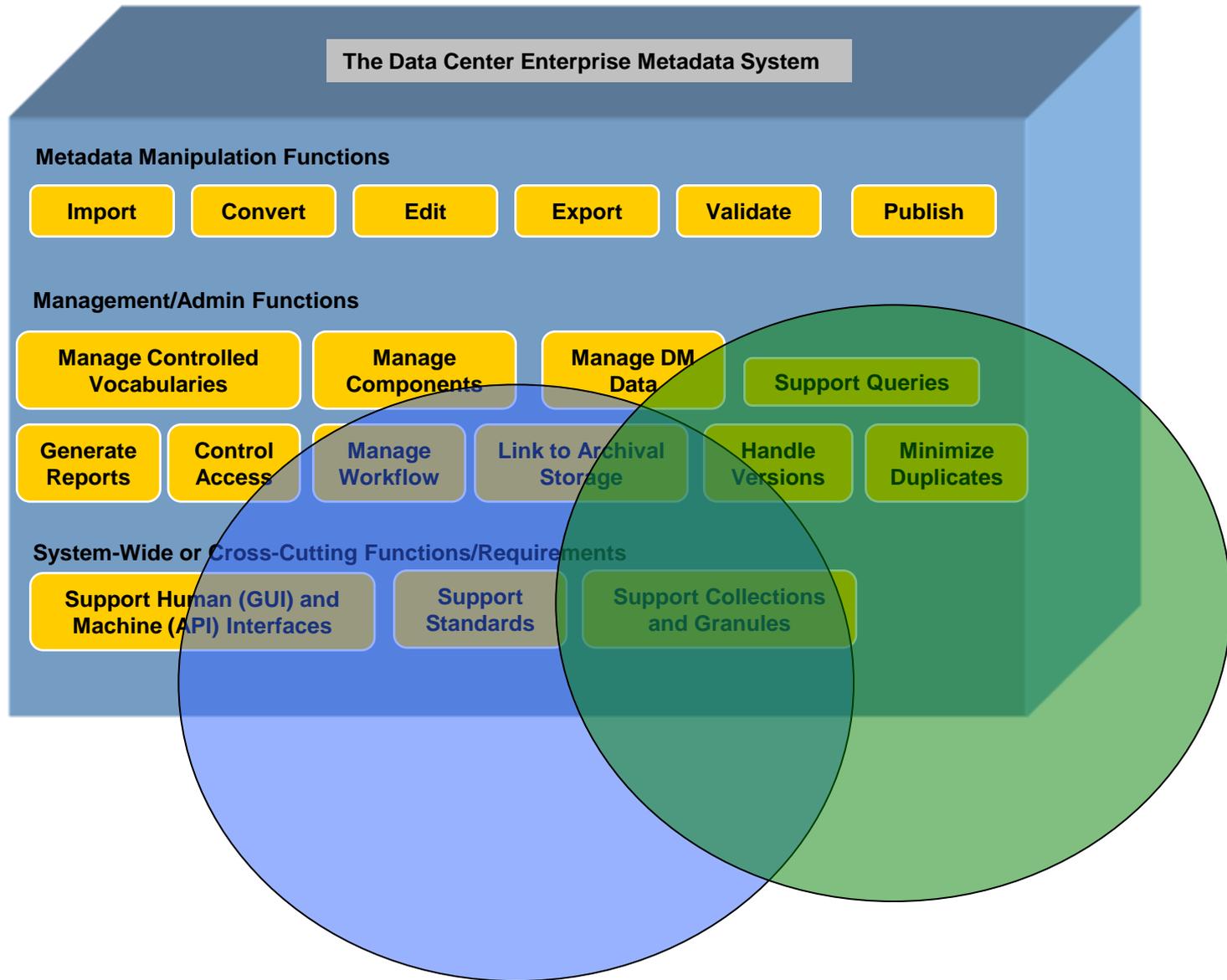


# The Metadata Enterprise



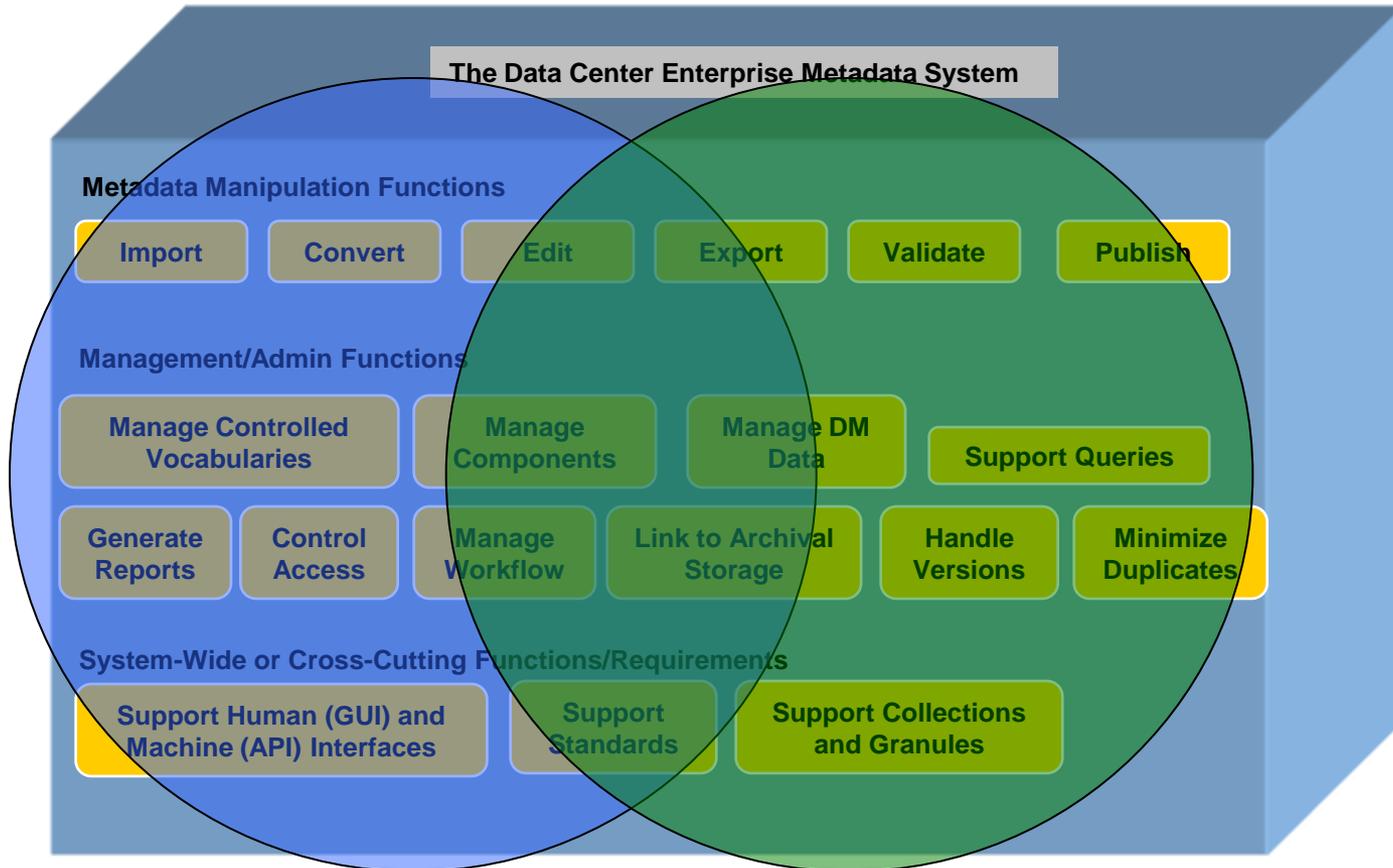


# How to get from this state...





# ... to this state?





# Requirements: Import

Import

- Description: The system must have the ability to import external metadata information into the enterprise system
- Requirements:
  - Import multiple standards/profiles
  - Support multiple ingest formats (Text, HTML, XML)
  - Support Batch Import
  - Support multiple languages/character sets



# Requirements: Convert

Convert

- Description: The system must be capable of converting metadata between different representations
- Requirements:
  - Convert between standards and profiles
  - Convert between formats (text, XML, etc.)
  - Convert in as “lossless” a way as possible



# Requirements: Export

Export

- Description: The system must be capable of exporting metadata to various representations
- Requirements:
  - Exporting must support standards
    - ISO 19139 XML
    - FGDC
    - MARC XML
  - Exporting must support preconfigured views
    - XML (non-19139)
    - Text
    - FAQ
    - HTML
  - Exporting must support User-defined custom output
    - style sheets, selectable fields, etc



# Requirements: Validate

Validate

- Description: The system must be capable of validating metadata information
- Requirements:
  - Must validate against specific standards/profiles/user needs
  - Spell-checking must be possible
  - Must validate against vocabularies
  - Must support Batch Validate
  - Validate Incomplete Record, and those without components
  - Easy to understand validation error messages
  - Validation must be integrated into system (real-time, inline)
  - Fuzzy and custom validation
  - Must support validation of heavily nested record sets (to user defined levels)
  - Must validate URLs, positions (points and bounding boxes), domains, ranges, dates



# Requirements: Publish

Publish

- Description: The system must be capable of publishing metadata information to external entities
- Requirements:
  - Must support Z39.50 / ISO 23950
  - Must be able to publish to Web Accessible Folder
  - Must support Open Archive Initiative Metadata Service
  - Must support ArcIMS Metadata Service
  - Must be easy to add publishing destinations
  - Must have ability to select publish destination on a per data set basis
  - Must be able to publish in multiple formats (including custom and new)
  - Must be able to un-publish/request retraction from external catalog (if external system can do it)
  - Must be able to automatically re-publish based on change detection (could be complicated by components)
  - Support DOIs (Durable, or Digital Object Identifiers) for publishing datasets



# Requirements: Edit

Edit

- Description: The system must provide interfaces for manual creation and editing of metadata information
- Requirements:
  - Supports Reuse / Access Values in Other Records
  - Edits In Multiple Languages
  - Saves Incomplete Record
  - Auto-populates some Metadata Elements
  - Multiple authors can edit
  - Supports Web-based user interface including WebDAV clients
  - Supports a range of users (beginner to expert)
  - Has ability to Batch edit/create
  - Includes both human and machine interfaces



# Requirements: Manage Controlled Vocabularies

## Manage Vocabularies

- Description: The system must be able to manage controlled vocabularies
- Requirements:
  - Must have ability to use controlled vocabulary lists
  - Supports thesauri
  - Vocabularies are integrated across the system (e.g., drop down lists)
  - Supports controlled vocabulary terms as complex objects with identifiers



# Requirements: Manage Components

## Manage Components

- Description: The system must be able to manage (CRUD) and link components
- Requirements:
  - Maintains unique identifiers
  - Resolve complete records from components
  - Search components – return composites
  - Validate components
  - Support REST interface to components
  - Lots of administrative details, orphans/dups



# Requirements: Support Queries

Support  
Queries

- Description: The system must be capable of responding to search queries
- Requirements:
  - Supports both internal and external searches (clearinghouse function)
  - Must support Catalog Service for the Web (CSW)
  - Must support SRU (Search-Retrieval via URL)
  - Enables free-text searches
  - Enables searching by fields (spatial/temporal searches)
  - Supports logical searches (AND/OR, etc)
  - Supports both human and machine interfaces (RSS, CSW, ...)
  - Includes admin “who did what when” info
  - Provides Google-style “did you mean?” searches precise/fuzzy (could be expensive)
  - Leverages ontologies
  - Support result ranking...
  - Save searches
  - RSS Feeds



# Requirements: Manage DM Data

Manage  
DM Data

- Description: The system must be able to manage OAIS “Data Management Data” (OAIS RM page 1-9)
- Requirements:
  - Must support ability to track where records are published
  - Every component within the catalog must have a unique identifier
  - System **must** handle preservation process history data, system stats, etc.



# Requirements: Generate Reports

Generate  
Reports

- Description: The system must be able to generate reports
- Requirements:
  - Both routine, regularly-scheduled and custom ad-hoc reports must be possible
  - Supports reports on descriptive metadata, administrative reports (DM Data) and summaries
  - Supports client- and server-side configurable outputs



# Requirements: Control Access

Control  
Access

- Description: The system must be capable of controlling access to the system and the metadata it contains
- Requirements:
  - Supports users external to NOAA
  - Automated password reset/recovery
  - Provides variable levels of access, from viewing only up to editing for records and components
  - Could integrate with NOAA “single sign on”
  - Filtered views on publication



# Requirements: Manage Workflow

Manage  
Workflow

- Description: The system must be aware of and able to manage metadata workflows
- Requirements:
  - Supports customized workflows
  - Support archive approval process
  - Includes validation steps and gates
  - Linkable to NOAA Records Appraisal Process
  - Linked to `gov.noaa.nodc:publish/unpublish` of approved accessions/datasets



# Requirements: Link to Archival Storage

[Link to Archival Storage](#)

- Description: The system must be capable of linking metadata records to AIPs in Archival Storage
- Requirements:
  - Maintains unique identifiers
  - Supports many-to-many relationships
  - Works in a bi-directional fashion
  - Maintains authoritative relationship between metadata and filesystem



# Requirements: Handle Versions

Handle  
Versions

- Description: The system must be capable of handling multiple versions of metadata information
- Requirements:
  - Earlier versions must be viewable and time-stamped
  - Accounting records maintained for every version
  - Linked to import (“Accession Compare”, replace, update, compare, merge)
  - Recorded on publish
  - Impact on components?



# Requirements: Minimize Duplicates

Minimize Duplicates

- Description: The system must be capable of detecting and reducing the presence of duplicate and near-duplicate information
- Requirements:
  - Must be capable of detecting duplicates across all components (e.g., within vocabularies, across metadata records, etc.)
  - Can communicate with external systems
  - Compares constituents of components AND unique identifiers
  - Launches into merge/compare
  - May require human input / validation



# Requirements: Support Human (GUI) and Machine (API) Interfaces

Support Human (GUI) and  
Machine (API) Interfaces

- Description: The system must consist of both human interfaces and machine-to-machine interfaces
- Requirements:
  - Must span the system
  - Supports event-based actions
  - Supports RESTful interfaces
  - System functions can be treated as externally available services
  - System can use external services



# Requirements: Support Standards

Support Standards

- Description: The ability across the enterprise to support various standards and profiles really well
- Requirements:
  - ISO 19115 Geospatial Metadata including NAP and other profiles
  - ISO 19110: Feature Catalog
  - ISO 19115 Part 2: Extensions for imagery and gridded data
  - ISO 19119 Geographic Information – Services
  - Extensions for Remote Sensing (RSE)
  - FGDC CSDGM (FGDC-STD-001-1998)
  - Biological Data Profile of the CSDGM
  - Metadata Profile for Shoreline Data of the CSDGM
  - Ecological Metadata Language (EML)
  - GCMD DIF
  - MARC XML
  - Dublin Core
  - ...



# Requirements: Support Collections and Granules

Support Collections  
and Granules

- Description: Ability of the system to link collections and granules, including products and their constituent originals
- Requirements:
  - Maintains relationships between components
  - Link to a Rich Inventory system
  - Maintain links between products and archival objects