

INTRODUCING OAIS AND DDI INTO AN ON-GOING RESEARCH PROCESS: THE MPC EXPERIENCE

Wendy Thomas
Minnesota Population Center
IASSIST2013
29-31 May 2013

Why are we doing this now?

- ⦿ External demands for documentation and accountability from funding organizations
- ⦿ New priorities around preservation, provenance, process management
- ⦿ Involvement with external discovery systems...each with its own metadata structure
- ⦿ Need to support our researchers with their proposal development and project management

What is the organizational / research structure of the MPC

- ⦿ Research project based
- ⦿ Infrastructure development - on going growth in data products
- ⦿ Secondary data rather than primary collection
- ⦿ Cleaning, integration, harmonization
- ⦿ Focus is on the deliverable product - data access system
- ⦿ Archival responsibility for some data, metadata, and related materials

The MPC as a hybrid institution

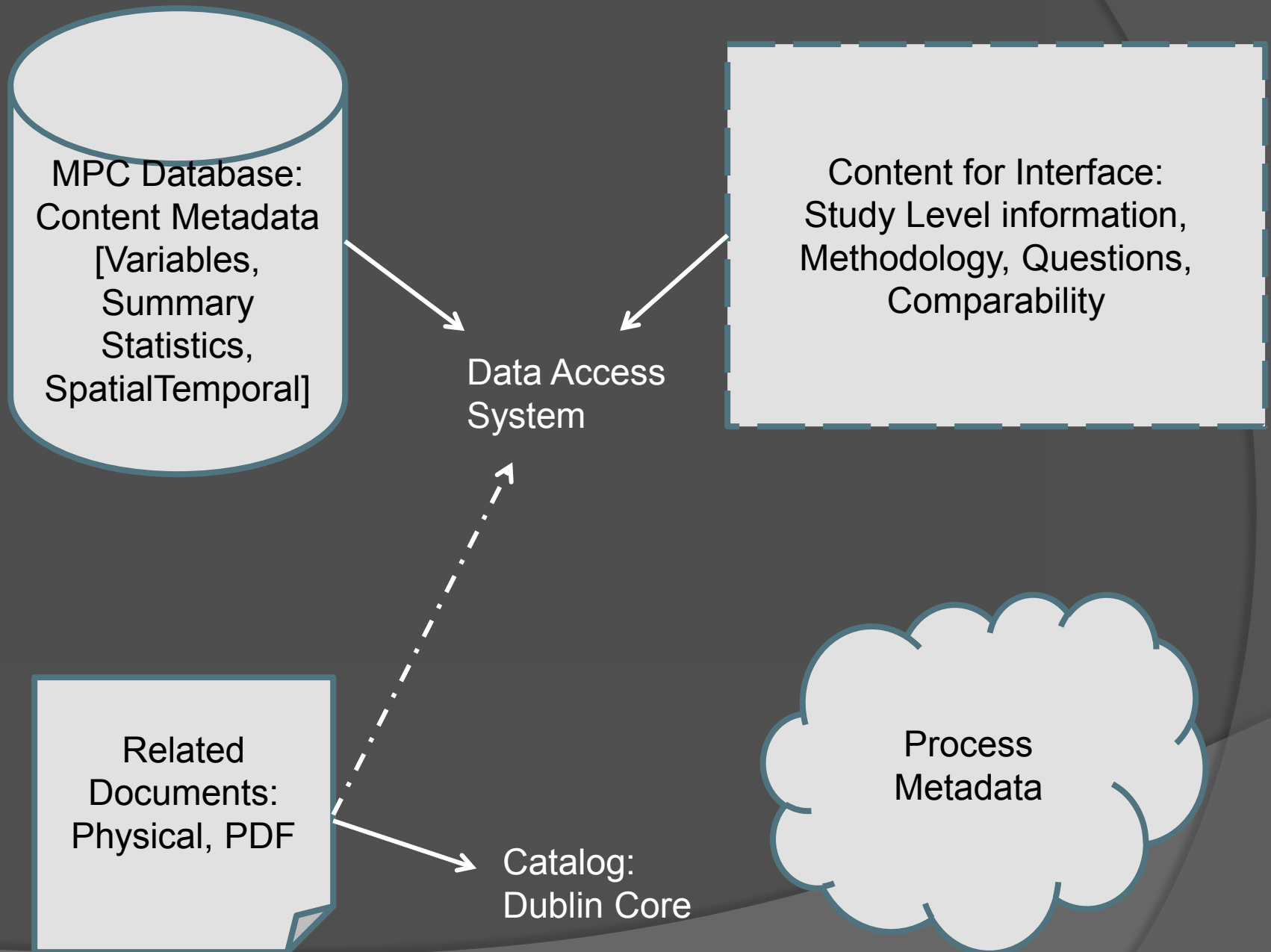
- ⦿ Are we a research center?
 - Modify (integrate and harmonize) rather than collect data
 - Provide the data infrastructure for other people's research
- ⦿ Are we an archive?
 - Archival responsibility for our products
 - Archival responsibility for selected source data
- ⦿ Are we a service center?
 - Provide support for proposal development and implementation
 - Forum for discussion

Management Perspective

- ◉ We need to continue to support our current activities and profile
 - Matching standards to a hybrid institution
 - Assembling an interacting set of standards
 - Need to determine responsibilities on a project-by-project basis
- ◉ Workflows cannot stop – we need to leverage the current processes as much as possible
- ◉ We need to get buy-in by project managers and research staff

Current Data Metadata Structure

- ⦿ Data is held in ASCII fixed format files
- ⦿ Metadata is held in multiple formats
 - Standardized MPC Data Base (microdata and aggregate data)
 - Runs the dissemination and access system
 - Structured text documents
 - Study level information used in user interface
 - Physical and digital images of related materials and original metadata
 - Provenance and Process notes...varied



Current level of standards compliance

◎ Dublin Core

- Use an extended version of Dublin Core Terms to describe related documents and data files

◎ DDI-Codebook

- Original input structure for aggregate data systems
- Output structure for microdata products

(Metadata databases could be mapped to DDI Lifecycle presumably without loss)

The Process

- ⦿ Map out processes and responsibilities
- ⦿ Determine preservation and metadata models to implement and how they interact
- ⦿ Identify current metadata capture points
- ⦿ Provide a uniform structure for metadata capture for each process
- ⦿ Channel into a complete metadata management system
- ⦿ Generate routine metadata output on a regular basis in preservation formats

Understanding our processes and responsibilities

- ◎ Self-assessment and audit plan
 - Required to address TerraPop but it is being conducted with MPC wide activities in mind
 - Based on *Ten Principles for Digital Preservation Repositories* we have recently completed an audit as Step 1 in addressing the 10 principles
- ◎ Provides a basis for developing a work plan to assess metadata related to specific activities, missing metadata, and who to involve in specific parts of the work plan

Headers for a GSBPM-like model

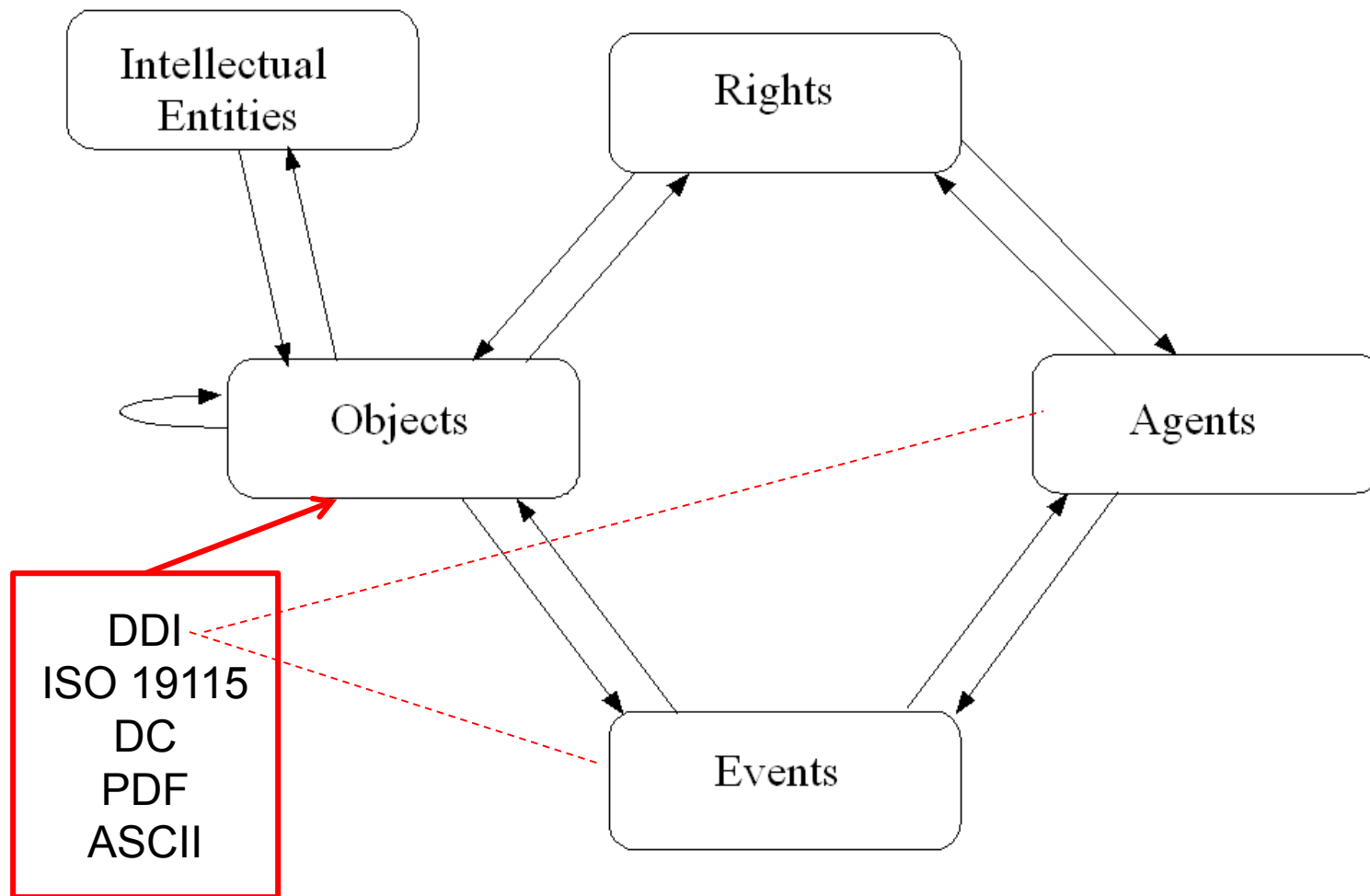
- ⦿ Project Development
- ⦿ Data Acquisition
- ⦿ Clean and Select Data
- ⦿ Content Metadata
- ⦿ Integration / Harmonization
- ⦿ Data Access System
- ⦿ Related Documents
- ⦿ Evaluation

Data Acquisition sub-levels

- ⦿ Identify data set
- ⦿ Develop metadata profile (needs, wants)
- ⦿ Source negotiation
- ⦿ License / MOA
- ⦿ Data acquisition
- ⦿ Metadata acquisition
- ⦿ Related document acquisition
- ⦿ SIP Validation

Model Selection

- Currently going with an integrated model using Premis, DDI, ISO 19115, and Dublin Core
- Working on developing a profile of objects from each that will be supported within the MPC (required/optional) and how they relate to each other
- Determine mapping to external metadata structures we need to interact with



The Issues

- ⦿ Identification of gaps in metadata and determining how to fill them
- ⦿ Involve individuals in resolving metadata capture issues on a process-by-process basis
- ⦿ Minimize time requirements on research staff for analysis activities and process changes
- ⦿ Relaying a sense of the larger picture – why metadata is captured and how it is used – without overwhelming individuals
- ⦿ Develop a means of instituting these practices early in the project proposal stage for future projects

THANK YOU

Wendy Thomas wlt@umn.edu

Questions??