

**Geospatial Multistate
Archive and Preservation Partnership
Metadata Comparison**

Cindy Clark, Utah's Automated Geographic Reference Center

Glen McAninch, Kentucky Dept. for Libraries and Archives

Steve Morris, North Carolina State University, Library Archive

For Preservation and Discovery, Data must have Metadata that tells

- ▶ Who: creator of data,
- ▶ What: title and description of data,
- ▶ Where: geographical extent of data,
- ▶ Why: reason the data was created,
- ▶ When: when the data was created,
- ▶ How: how the data was created.
- ▶ Definition of data attributes.

Overview of Two Geospatial Metadata Standards

- ▶ Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata
 - ▶ Version one (1994) mandated for use by federal agencies in 1995
 - ▶ Descriptive metadata, plus some administrative and technical
 - ▶ Extensive use at state level, spotty use at local level
 - ▶ Problem: content standard without an encoding specification
 - ▶ FGDC profiles: ESRI, NBII, Remote Sensing, etc.
- ▶ ISO Standards
 - ▶ ISO 19115: Geospatial Information – Metadata (2003)
 - ▶ ISO 19139: Geospatial Information – Metadata – XML (2007)
 - ▶ North American Profile of ISO to replace FGDC CGDSM
 - ▶ Not yet widely implemented in the U.S.

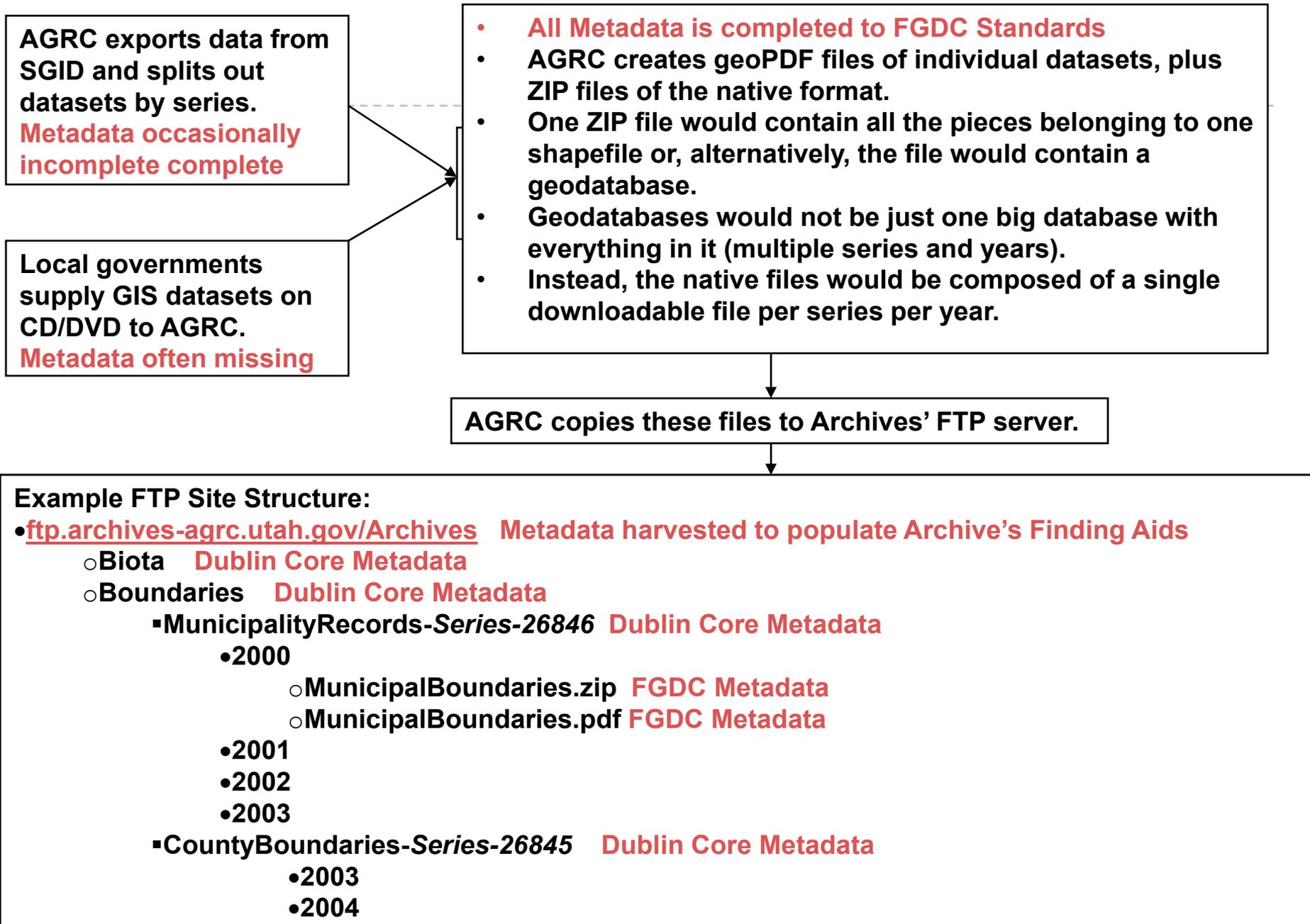
Overview of Dublin Core Archival Metadata Standards

- ▶ The Dublin Core Metadata Initiative (DCMI) is an open organization engaged in the development of interoperable online metadata standards.
- ▶ DC provides a simple and standardized set of rules for describing electronic resources online in ways that make them easier to find.
- ▶ Created to be a broad metadata capture tool.
- ▶ METS was looked at, but deemed to costly and time consuming to implement.

North Carolina GDAP Metadata Experience

- ▶ Existing geospatial metadata often needs:
 - ▶ Remediation – to fix errors or omissions
 - ▶ Normalization – to adhere to a standard structure
 - ▶ Synchronization – so that the data at hand matches the metadata
- ▶ If no metadata then:
 - ▶ Can build minimal metadata using templates and auto-extraction
 - ▶ Lose key information such as data quality, lineage, data dictionaries
- ▶ Automating metadata for repository ingest
 - ▶ Raster data is easy – large sets of consistently structured files
 - ▶ Vector data is hard – each dataset is a different story
- ▶ Lots of desired additional administrative and technical metadata elements not accommodated by FGDC CGDSM

Draft of Utah's GIS to Archives Data Flow



Kentucky Metadata Workflow into DSpace and iRODS Environment

DSpace

Database with Dublin Core Descriptive and Administrative Metadata

Metadata & content entered by agencies using template and modified by Archivist

Single item & batch ingest into DSpace by Archivist

iRODS

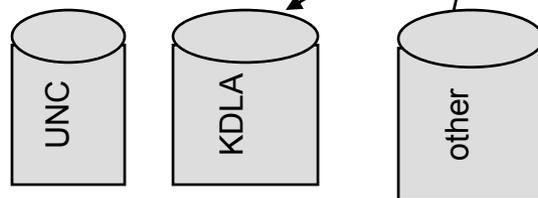
Database with Administrative & Preservation Metadata

Content Files

Batch metadata extraction using iRODS rules

Preservation metadata from iRODS rules

Distributed Storage Layer



Kentucky Metadata Questions

- ▶ Can we depend on useful data being imbedded in files by creators?
- ▶ What is the value of metadata automatically created by software and extracted into system metadata fields by tools such as iRODS?
- ▶ What standards can be used for administrative metadata created by systems (Dublin Core, METS or none)?
- ▶ How can we structure metadata for various formats and types of records (geospatial, e-mail, web records etc.) when integrated in a common archival system

Discussion/Questions