FGDC, Meet the DDI

Adding Geospatial Metadata to a Numeric Data Catalog

Julie Linden
Yale University
Outline

• StatCat: numeric data catalog
• Why add geospatial metadata?
• FGDC/DDI crosswalk
• Implementation
StatCat

- ssrs.yale.edu/statcat/
- Partnership between Social Science Library and Social Science Research Services
- Records for Yale’s Social Science Data Archive, CDs, subscription databases, Internet sources
- Subset of DDI elements in relational database
Why add geospatial metadata

- Use of GIS increasing at Yale
- Current searching for geospatial data is inadequate
- GIS and data searches closely tied
- Existing GIS metadata could be imported
- FGDC standard for geospatial metadata
Interface goals

- Could restrict search to numeric, geospatial, or both types of data
- Eventually can link directly from StatCat record to datasets on server
Data Documentation Initiative

- Document description
- Study description
- Data files description
- Variable description
- Other study-related materials
Geospatial metadata standard

- Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata: http://www.fgdc.gov/metadata/csdgms/
- ESRI Profile: http://www.esri.com/metadata/esriprof80.html
FGDC CSGDM

- Metadata
- Identification Information
- Data Quality Information
- Spatial Data Organization Information
- Spatial Reference Information
- Entity and Attribute Information
- Distribution Information
- Metadata Reference Information
- Citation Information
- Time Period Information
- Contact Information
FGDC, meet the DDI

- Compared FGDC/ESRI elements to our DDI subset to see what obviously mapped
- How would FGDC structure map to StatCat database structure?
- FGDC doesn’t treat “files” the same way as the DDI or StatCat
- “Entity and attribute” information didn’t map to “variables.”
Brief detour into XML...

• XML is good at:
  – Describing hierarchical data
  – Exchanging data between independent sites in a highly structured manner

• Since both standards are in XML, why not present XML records rather than to select elements from each for inclusion in common database?
...and back to original plan

• Decided to go with database
  – Amount of time it would take to start from scratch would greatly delay inclusion of geospatial metadata
  – FGDC and DDI elements did not have to correspond exactly; just needed a minimum set of elements to facilitate data discovery and evaluation.
Database modifications

• Included some FGDC elements that had no DDI equivalent:
  – bounding coordinates
  – “type” – to distinguish (and thus search on) geospatial and numeric data
• Added “Data Sources” and “Record Information” tables
• New database design
Near-term next steps

- Updating database tables
- Writing parser to import records
- Redesign search interface and results screens
Long-term next steps

• Link study description to live data sets, including documentation and software setups.
• Interoperability among catalogs
• Spatial queries
• Variables table
• Geodatabases