NSDL Metadata Repository Quality Analysis

Marcia L Zeng; Gregory M Shreve; Bhagirathi Subrahmanyam
Kent State University NSF DUE-0333572

Metadata Repositories: Loosely-controlled Environments

- Each discrete metadata dataset retains its independent identity
  - — can be used without a union catalog
  - — can exist in two places, may not be identical
- Multiple standards were applied
- Voluntary metadata sharing
  - — no central authority / management
- Records were created by trained and untrained metadata authors

Levels of Quality

- Stand-alone databases/data sets
  - — individual records
- Shared databases / repositories
  - — Individual records
  - — Converted records
  - — Duplication

Measurement Aspects

Completeness
Correctness
— content-info
— format
— input
— browser
— data syntax
— redundancy

Consistency
— data recording
— source links
— ID and identifier
— source description
— search results display
— mapping/integration

Duplication
— intra- & inter-collection

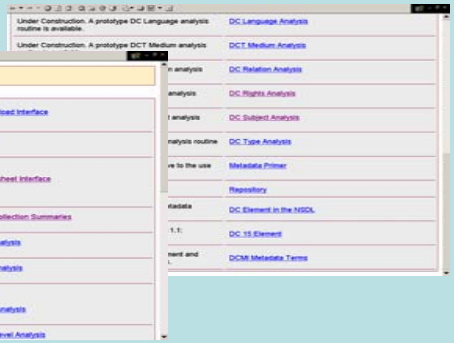
Quality Assurance

Make Policies:
— minimum quality requirements,
— quality measurement instruments,
— quality enforcement policies,
— quality enhancement actions, and
— the training of metadata creators.

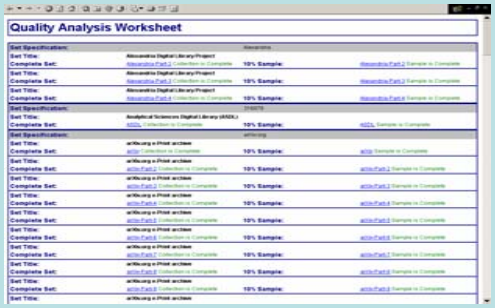
Use Tools:
— Instructions
— Templates for input records, with suggested syntax, vocabularies, and built-in values
— Duplicate-checking algorithm

http://appling.kent.edu/NSDLMetadataQuality/

Project Work Space

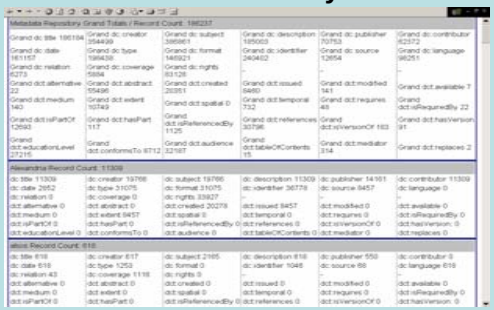


Sampling

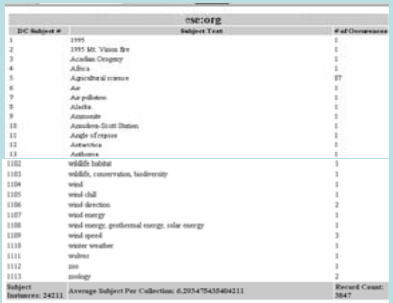
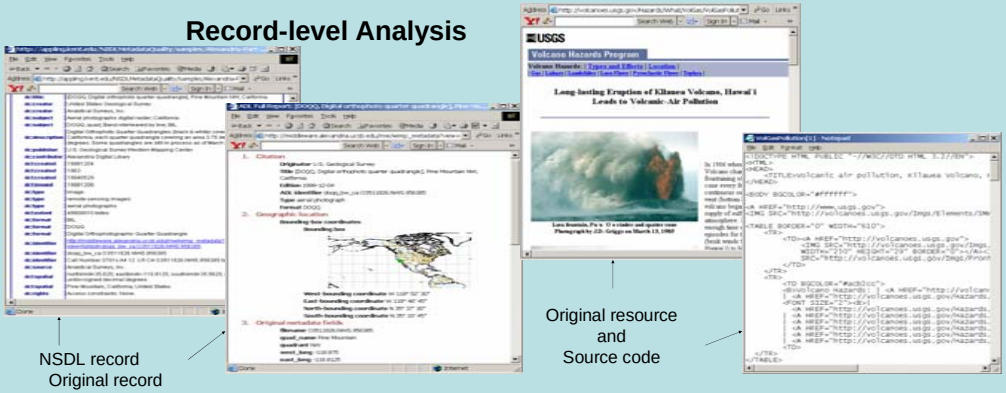


Sample-level Analysis

Collection-level Analysis



Element-level Analysis



Record-level Analysis