NARA Research Prototype Persistent Archives

Reagan Moore, Richard Marciano
San Diego Supercomputer Center
Mark Conrad
National Archives and Records Administration

Data Grid Support for Preservation

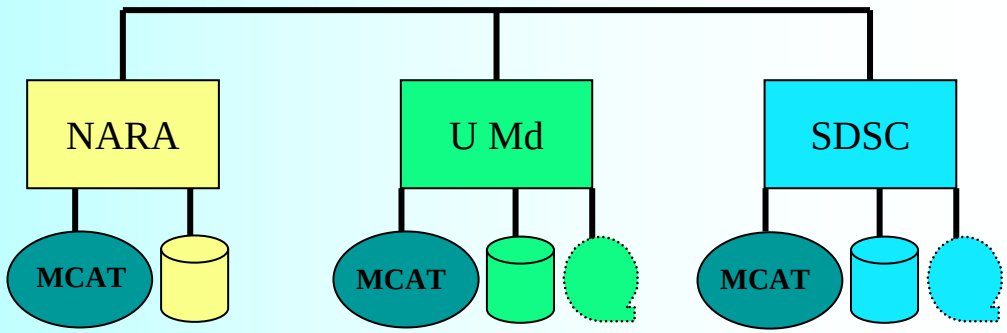
- Authenticity – the assurance that the records are what they purport to be
  - Support for metadata necessary to maintain the provenance of the records
  - Support for maintaining the essential characteristics of the records across transformations
- Integrity - the assurance that the electronic records are not corrupted
  - Support for integrity metadata (audit trails, access controls, checksums, replicas)
  - Support for distributed environments (replication, federation)
- Infrastructure Independence
  - Standard operations across databases
  - Standard operations across storage repositories

National Archives and Records Administration - Research Prototype Persistent Archives

Powerful Platform for Collaborative Research

- Synchronization across zones
- Interoperability across diverse platforms
- Sufficient metadata to ensure complete and authentic records
- Mitigation of risk of data loss
  - Replication of data
  - Federation of catalogs
- Deep archive

Federation of Three Independent Data Grids



A Collaborative Project: Electronic Access Project (EAP)

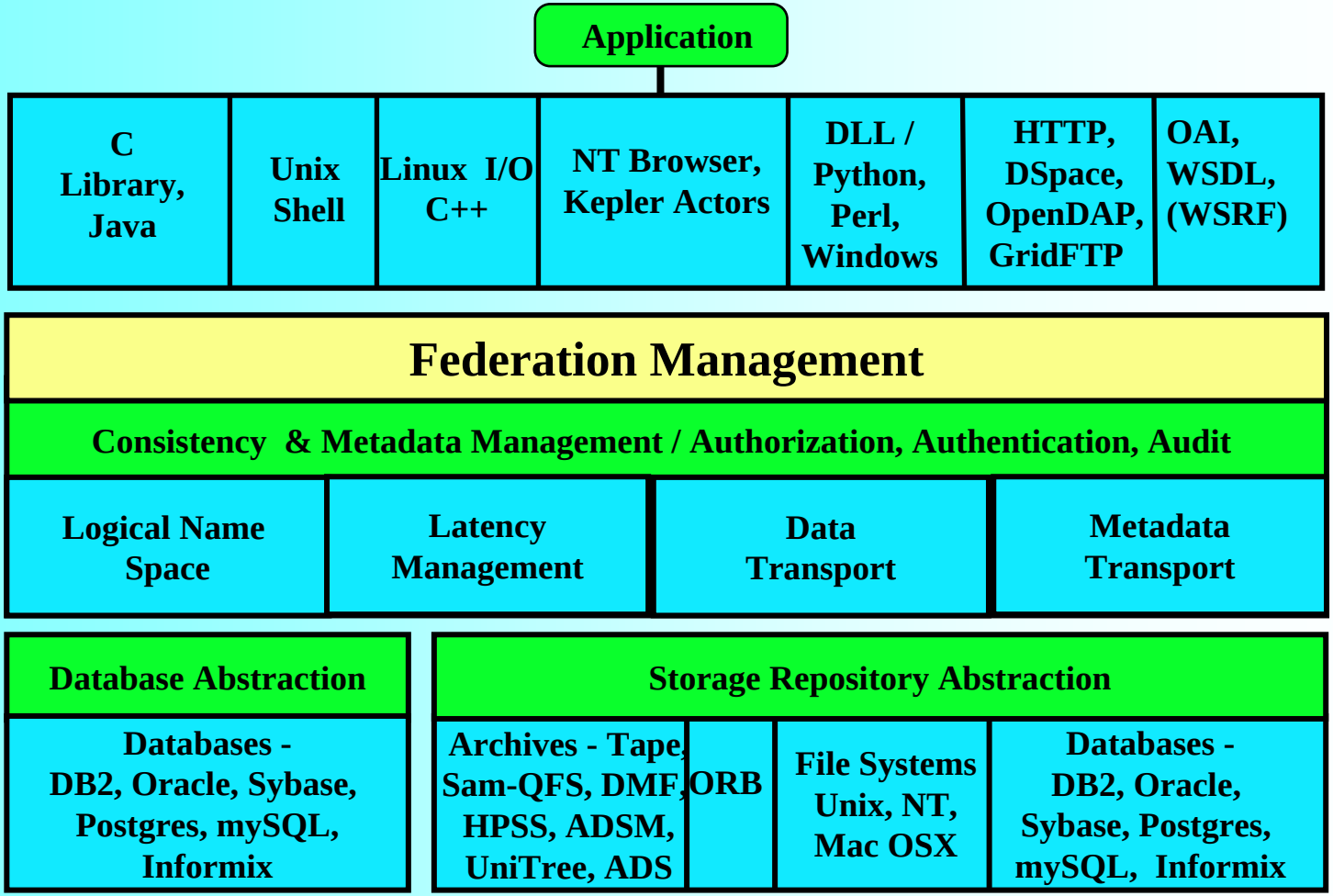
- Electronic Access Project (EAP collection)
  - 123,617 records digitized: 1997-1999
  - Cross-section of NARA holdings across 1056 record series
  - Described in NARA Archival Information Locator (NAIL) database
- Hierarchical Archival Description
  - ISAD(G) / ISAAR(CPF)-like catalog
  - Record Group or Collection / Record Series / File Unit / Item /
  - Separate descriptive metadata for each level of the hierarchy

Challenge

- Reinstantiate the hierarchical EAP collection in the Research Prototype Persistent Archives
  - Validate the metadata
  - Dynamically re-create the metadata hierarchy from the metadata actually present in the NAIL files and build a metadata catalog
  - Link the hierarchical description to the records
  - Provide a way to browse based on the hierarchy
  - Provide access to replicas located in one of the three federated data grids

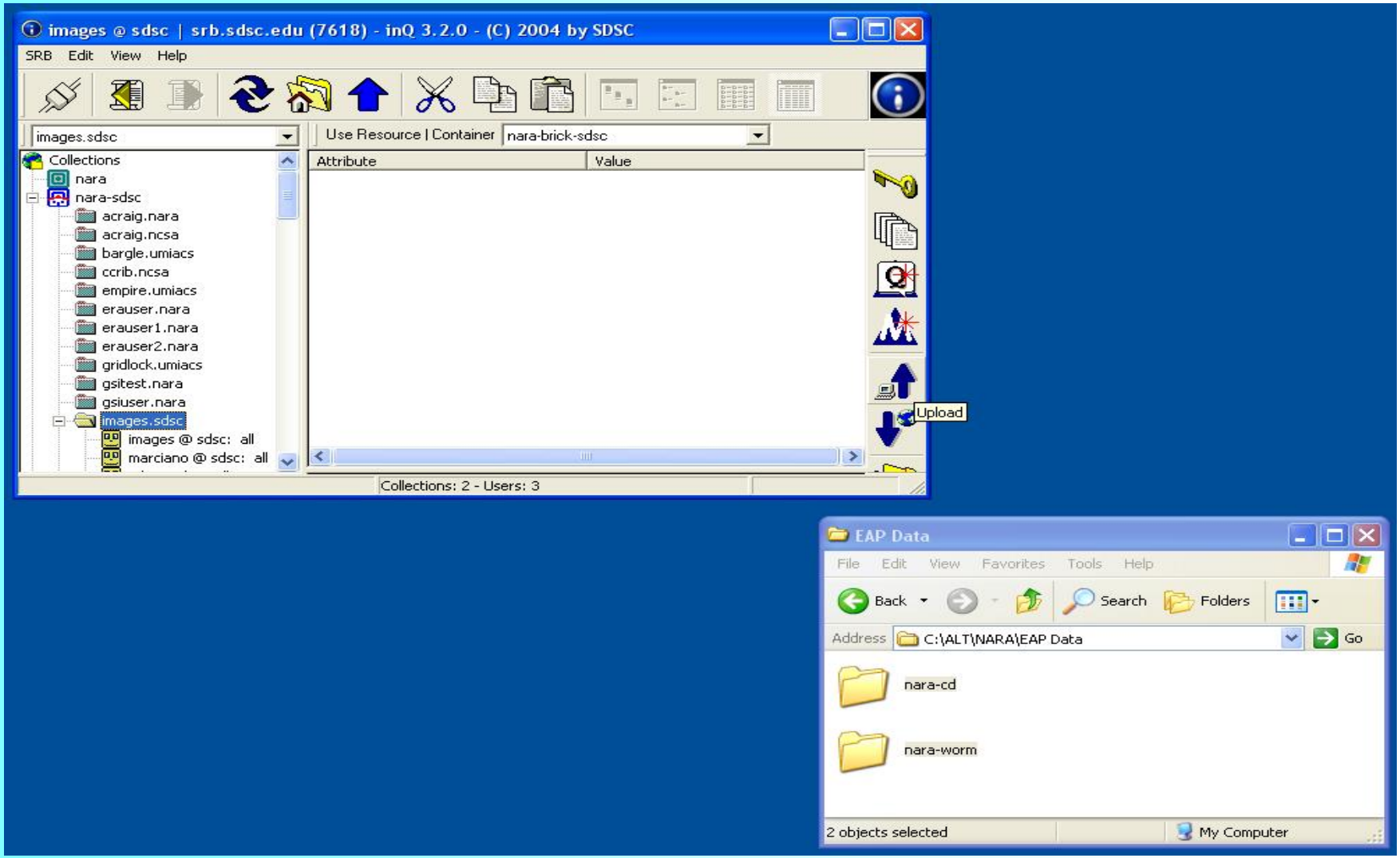
Storage Resource Broker 3.3.1

Preservation Processes (Accession to Access)

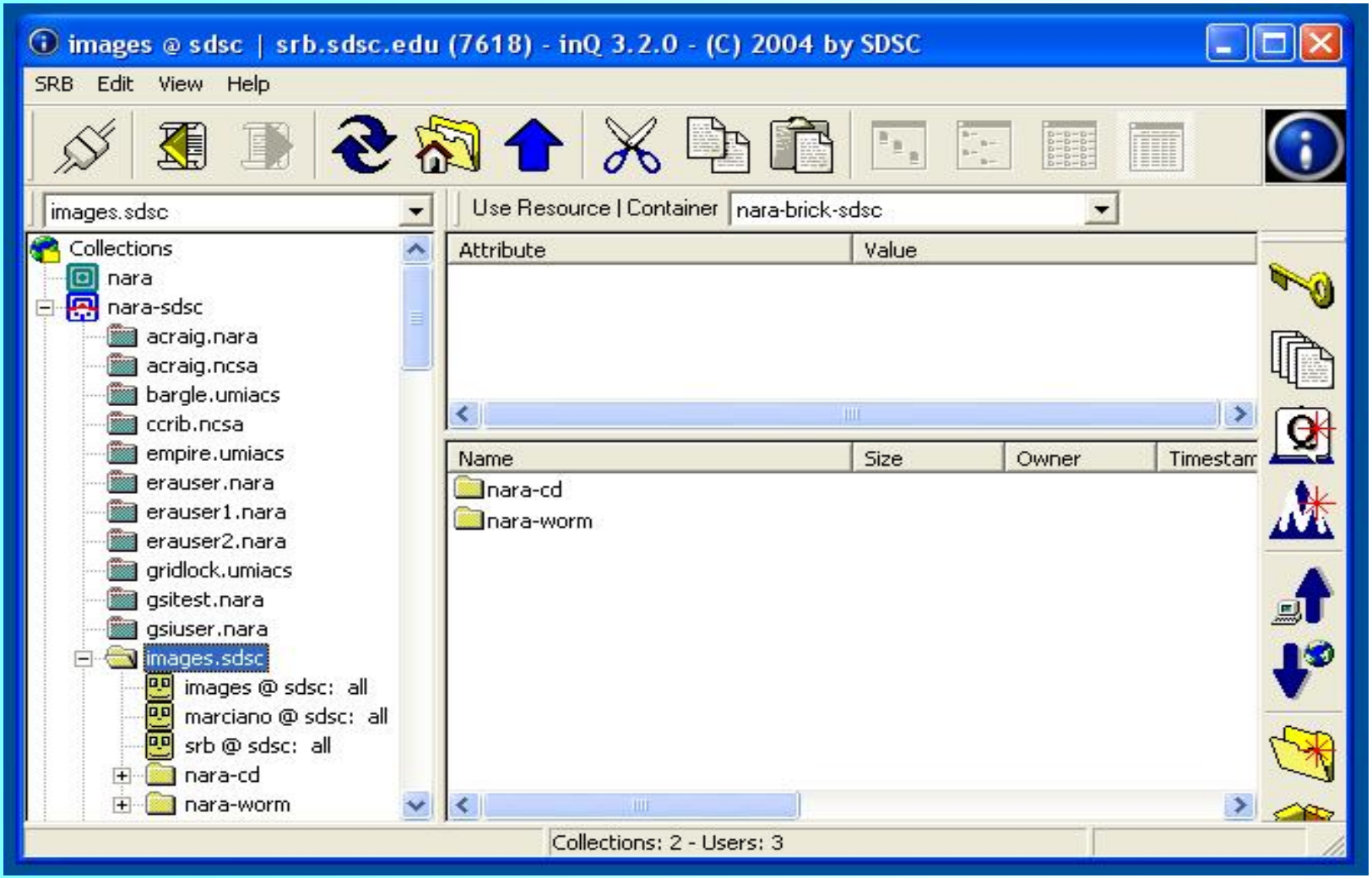


Infrastructure Independence

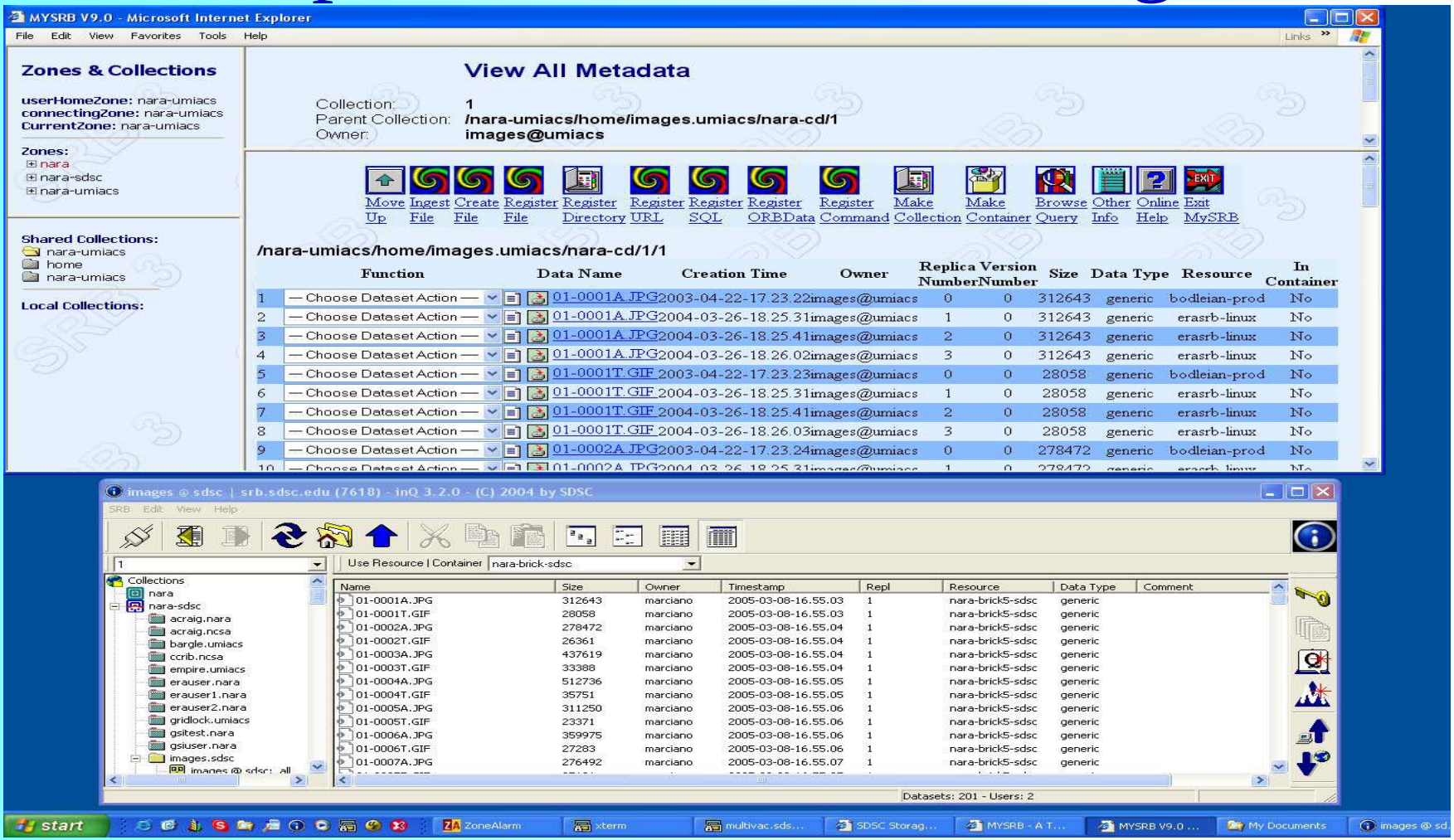
Registration of the data in the grid



Registration of the data in the grid



Federation of data grids
Replication between data grids



Rigorous Validation of Metadata or Records

- Schema-driven validation
- Regular expressions
- Syntactic Validation (well-formed)
- Semantic Validation (valid)