Worlds Collide

A Repository Based on Technical and Archival Collaboration

Erin O'Meara, Electronic Records Archivist
Gregory Jansen, Software Developer
University of North Carolina at Chapel Hill

Open Repositories 2010, Madrid

Carolina Digital Repository (CDR)

University-wide initiative to serve as the digital archive for members of the University of North Carolina at Chapel Hill community

Strategic goals:
- Ensure UNC digital scholarly output is safe and accessible into the future.
- Provide a digital preservation platform for born-digital special collections and university records.

Context and Challenges

Repository committee met for two years before project team was formed in 2008
- There were no specific software requirements.
- TRAC checklist and draft policies were our primary guide

Lessons learned from other IRs:
- no "roach motel"
- focus on services needed by faculty and curators
- use collaboration between library administrators and faculty to position the repository within the institution
- use collaboration between curatorial staff and technologists to make it work

Collaboration is not always pretty...

- Who's the boss?
- Large, traditional institutions do not foster collaboration
- Staff creating repositories must take more initiative

Our Approach: Speaking the same language

Programmers, librarians and archivists worked together to implement standards:

- OAIS was our common language.
- Defined METS profile for workflow and submission.
- Defined PREMIS record content and storage model.
- Defined MODS for description.

...standards act as glue within the larger structure comprised of workflows and processes

Our Approach, cont'd: Prototypes

Object models tested against diverse pilot collections:

- digital special collections
- electronic theses and dissertations
- academic conference materials
- journals

But..

"It is not a question of how well each process works, the question is how well they work together." - Lloyd Dobens

Many iterations later...

example object model.dia

PREMIS persistence model.dia

premis object.dia

IR modelling.dia

repository domain model.dia

work indexing data model.dia
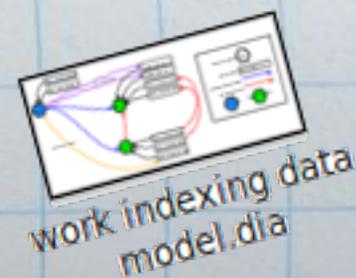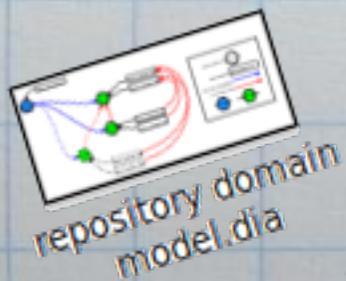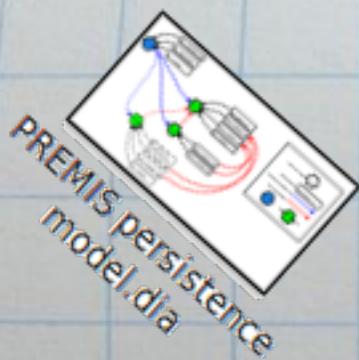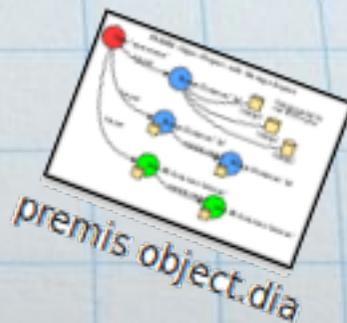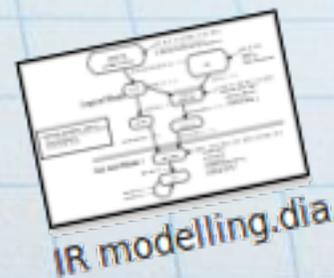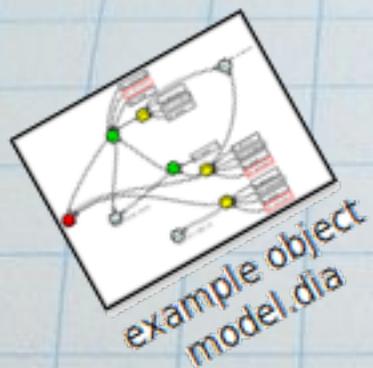
System Architecture
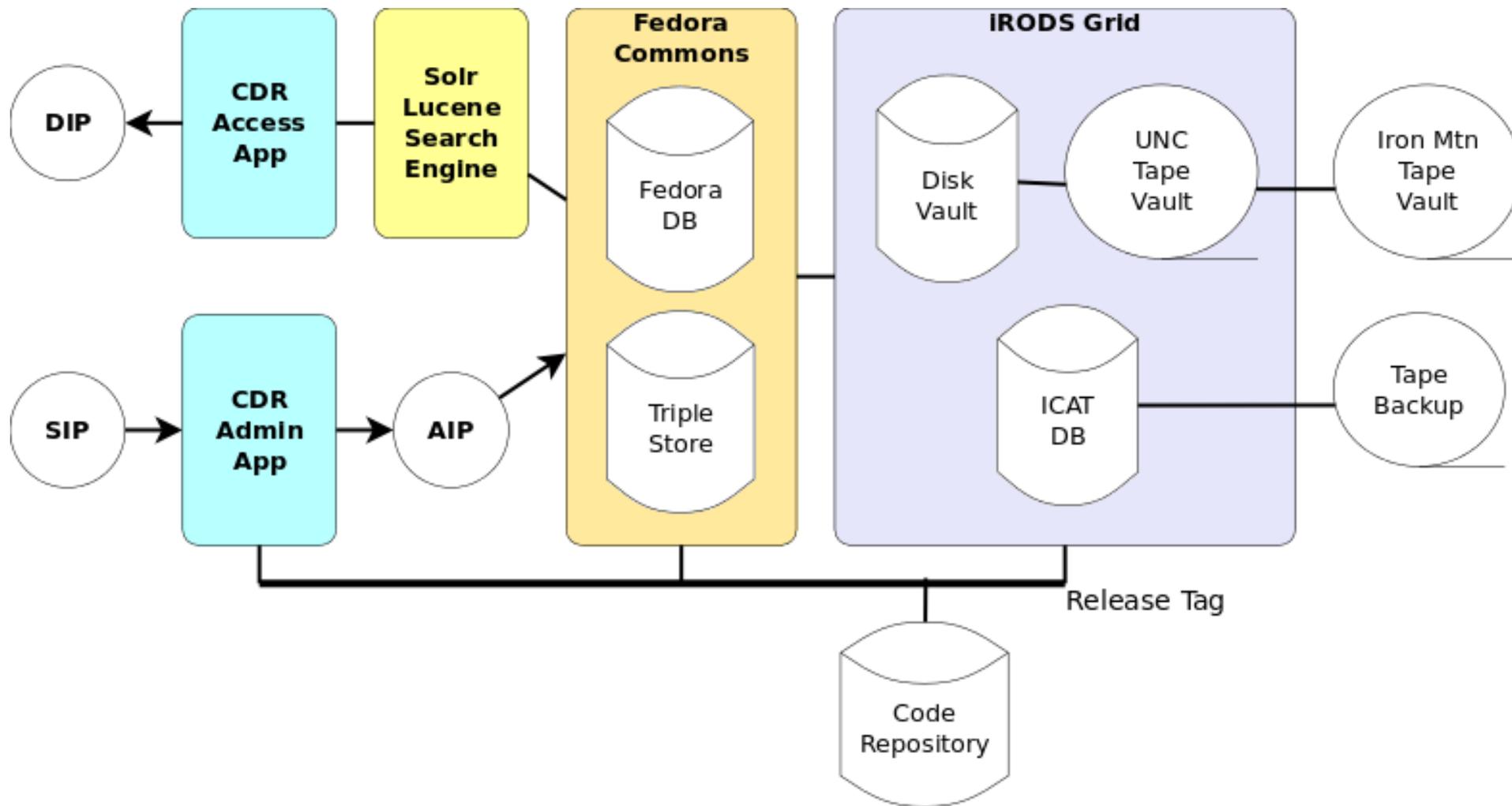
Mix-In Object Models

- Folder hierarchy
- Preservation
  - PREMIS events
  - PREMIS reports
- Content or data shape
 - Disk Archive
  - File
  - Image
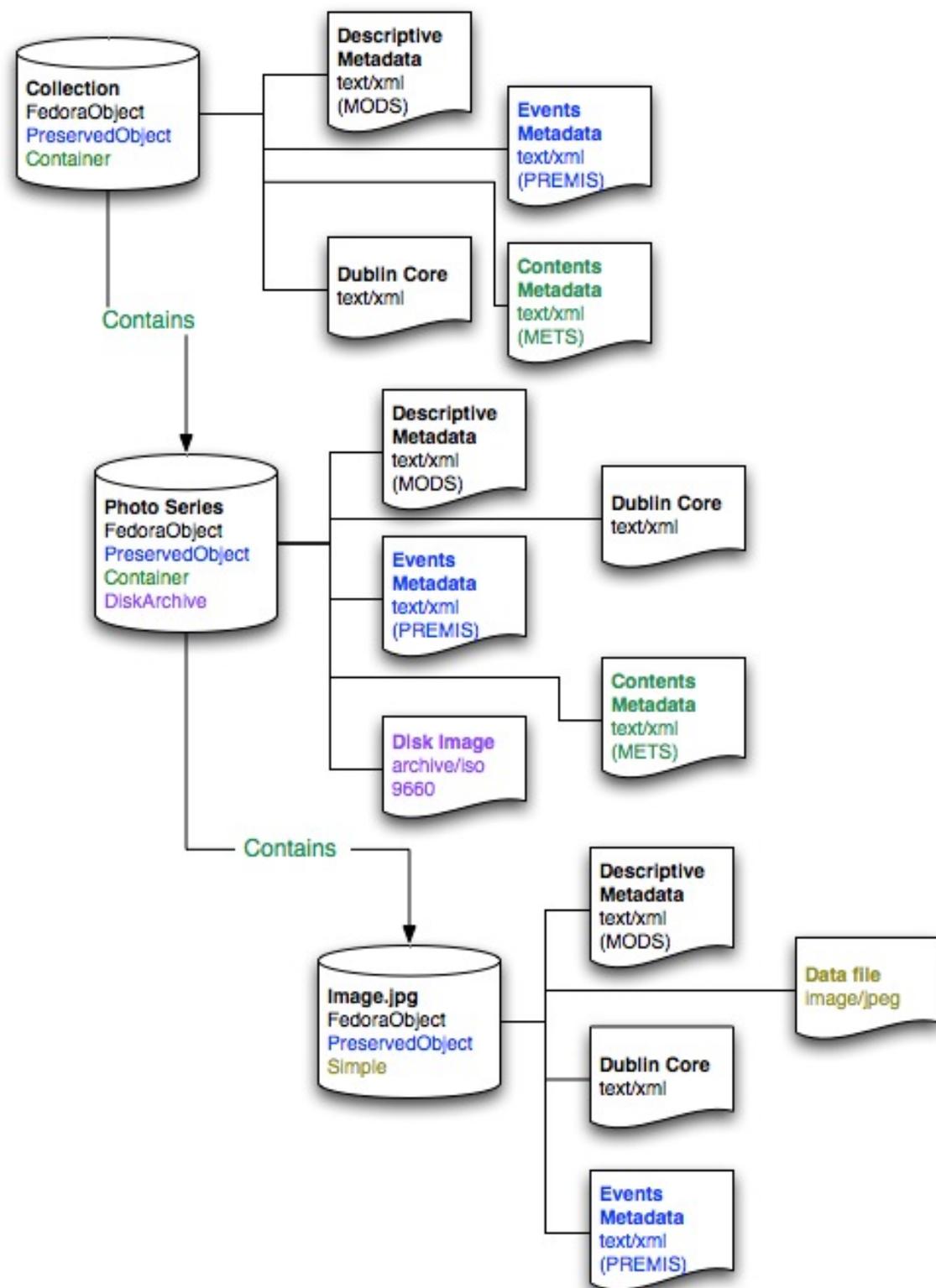
Tough Decisions:

Hierarchy is represented as a single model and relationship.

Metadata may differentiate Containers, such as *folders, sub-folders and series.*

PREMIS reports created dynamically from the current state of an object.
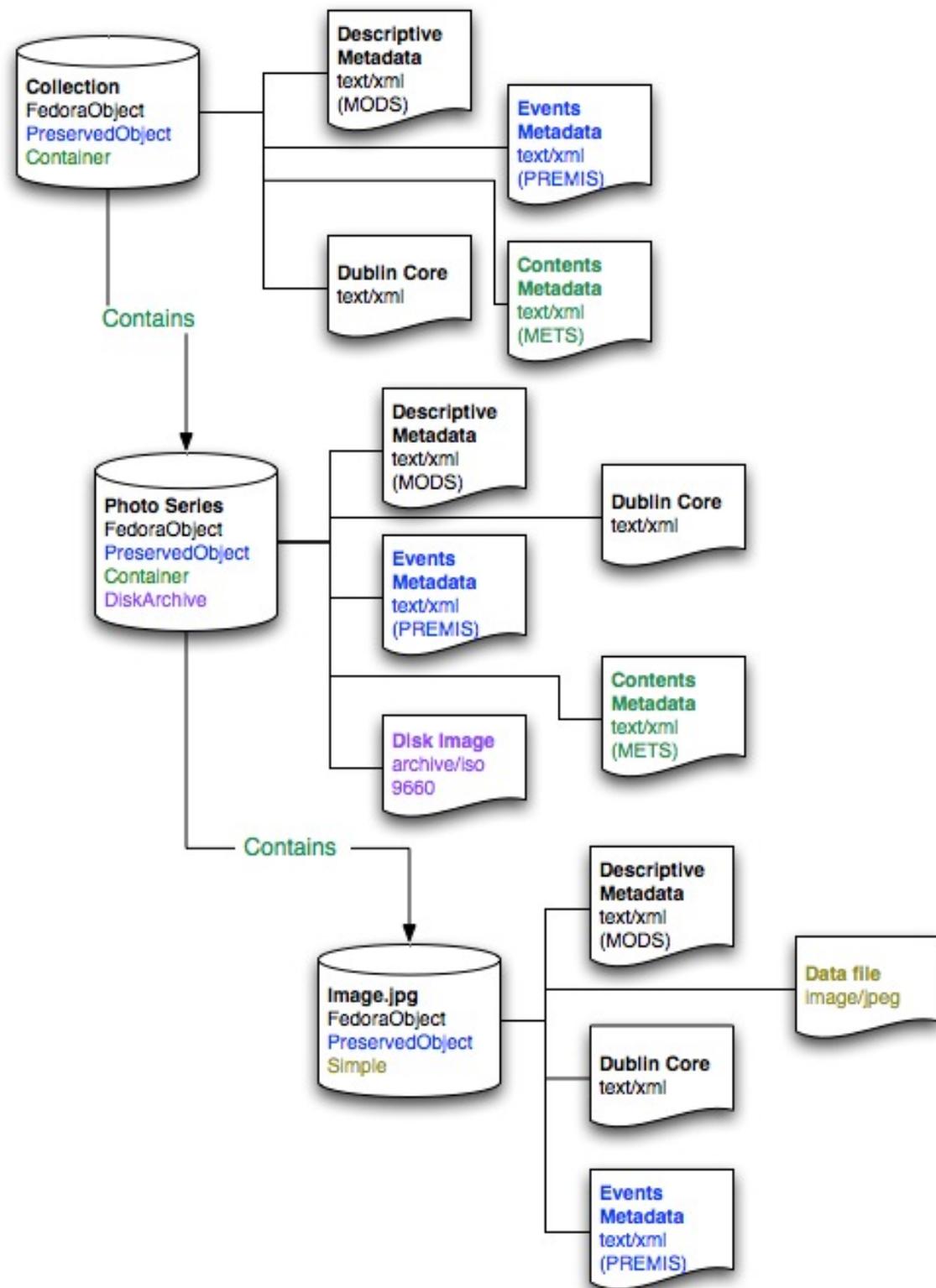
Outcomes

Mutually understood implementations of MODS, METS and PREMIS.

Archival principles incorporated into software:
- Archival bond: Maintaining inextricable links between related objects in a collection.
  - migration durable xrefs via UUID identifiers
- Ensuring provenance of materials
  - recording pre-ingest and repository-mediated events
- Hierarchical description of archival materials
- Appraisal activities as a non-linear function
  - user-centric workbench, over step-by-step workflow, where curator can appraise, arrange and describe

Conclusion

In April, the CDR had it's soft launch http://cdr.lib.unc.edu

Success was possible through:
- Strong working relationships between technical and curatorial domains and

- A mutual desire to solve issues of digital preservation

"Coming together is a beginning. Keeping together is progress. Working together is success."
- Henry Ford

References

Pearce-Moses, Richard and Susan E. Davis. Proceedings of New Skills for a Digital Era Colloquium. Washington, DC. 2006. Available at, http://www.archivists.org/publications/proceedings/NewSkillsForADigitalEra.pdf

Salo, Dorothea. "Innkeeper at the Roach Motel." Library Trends 57:2 (Fall 2008).