



# **Collaborative Metadata Aggregations: CDL's American West Experience + A Whole Lot of Unanswered Questions**

Bill Landis, Metadata Coordinator, CDL

DLF Fall 2005 | Charlottesville, Virginia

Tuesday, 8 November 2005



# American West Project

---

- OAI content harvested from 8 partners (CDL, CDP, Harvard, Indiana, LC, Michigan, Washington, Virginia)
- Topically focused on the American West
- Harvested approx. 300,000 records in 110 OAI sets
- Removed approx. 60,000 records thus far that are out of scope for this aggregation; still more to go



# Goals of Project

---

- Test feasibility of and explore issues surrounding this kind of topically focused OAI harvesting
- Contribute to profession wide OAI metadata best practices
- Assess responses of K-12 teachers (and others?) to faceted, hierarchical browse interface into harvested OAI collection
- Develop tools and services to support harvesting work by UC librarians and others in the context of CDL's Common Framework technology infrastructure



## Best Practices ...

---

- OAI best practices, professional community guidelines for metadata content, etc. - great efforts, but OAI aggregators need to recognize ultimate limits to these approaches
- Need better understandings, keyed to usability and needs assessments, as to why we should bother with OAI aggregations in the first place?
  - Are we able to offer access-system value that Google, for example, can't?
  - Specifically defined audiences make this easier to think about and access



## ... vs. “Get out the shovels”

---

- Importance of specific audiences and access mechanisms to directing aggregator work with metadata remediation and enhancement
- Unavoidable, even in face of the best metadata because we’re frequently aggregating across professional communities of practice (i.e., different data content traditions, standards, and practices)



# American West contexts

---

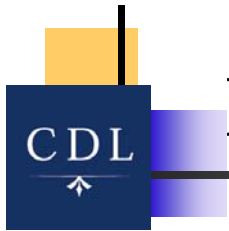
- K-12 (“only want a dozen really good things”)
- Faceted, hierarchical browse inspired by Flamenco project (Prof. Marti Hearst, UCB/SIMS)
- Metadata focus to drive browsing:
  - Dates
  - Topics
  - Geographic locations
  - Genres (hopefully we’ll get there, though this is the least important in the AmWest collection context)



## Long-term OAI Challenges

---

- Assuming that some end users (are we even looking at the right ones?) love the approach we're taking ...
- Assuming we can automate classification on ingest of harvested records, and that the human-intensive process of using clustering to surface highly targeted “bags of words” for a suite of American West topical terms ...



## Long-term OAI Challenges

---

- Does inherent record instability in OAI context bode well or ill for automating the process of refreshing, adding to aggregated collections?
- What are the frameworks in which “micro” (boutique?) OAI aggregations might be more appropriate and successful than our current “macro” approaches (OAIster, CIC portal) or topically “macro” approaches (AmWest)
- Tradeoffs in efforts to remediate/enhance metadata for specific purposes vs. payoffs for academic libraries?
- Who might be doing OAI work? DL staff or line librarians?