



## Emerging Practices: Integrating Library Content with Educational Technologies

Peter Brantley, Director of Technology, UC California Digital Library

David A. Greenbaum, Director, UCB Interactive University Project

Raymond Yee, Technology Architect, UCB Interactive University Project



# Copyright

---

- Copyright Peter Brantley, David Greenbaum, and Raymond Yee, 2003. This work is the intellectual property of the author. Permission is granted for this material to be shared for non-commercial, educational purposes, provided that this copyright statement appears on the reproduced materials and notice is given that the copying is by permission of the author. To disseminate otherwise or to republish requires written permission from the author.



# Our Talk

---

1. Context and introduction
    - UCB Interactive University: David Greenbaum
    - UC California Digital Library: Peter Brantley
  2. Demonstration of IU Scholar's Box Tool
  3. Architectural and technical overview
    - UC California Digital Library: Peter Brantley
    - UCB Interactive University: Raymond Yee
- A caveat about range of topics and terms
  - Thanks to funders: Mellon, NSF, Hewlett, Department of Commerce



# IU Mission - Who We Are

---

The Interactive University (IU) uses the Internet to open UC Berkeley's knowledge and people to the public, especially K-12 schools.

Our goal is to use technology to democratize the content and community of the campus.

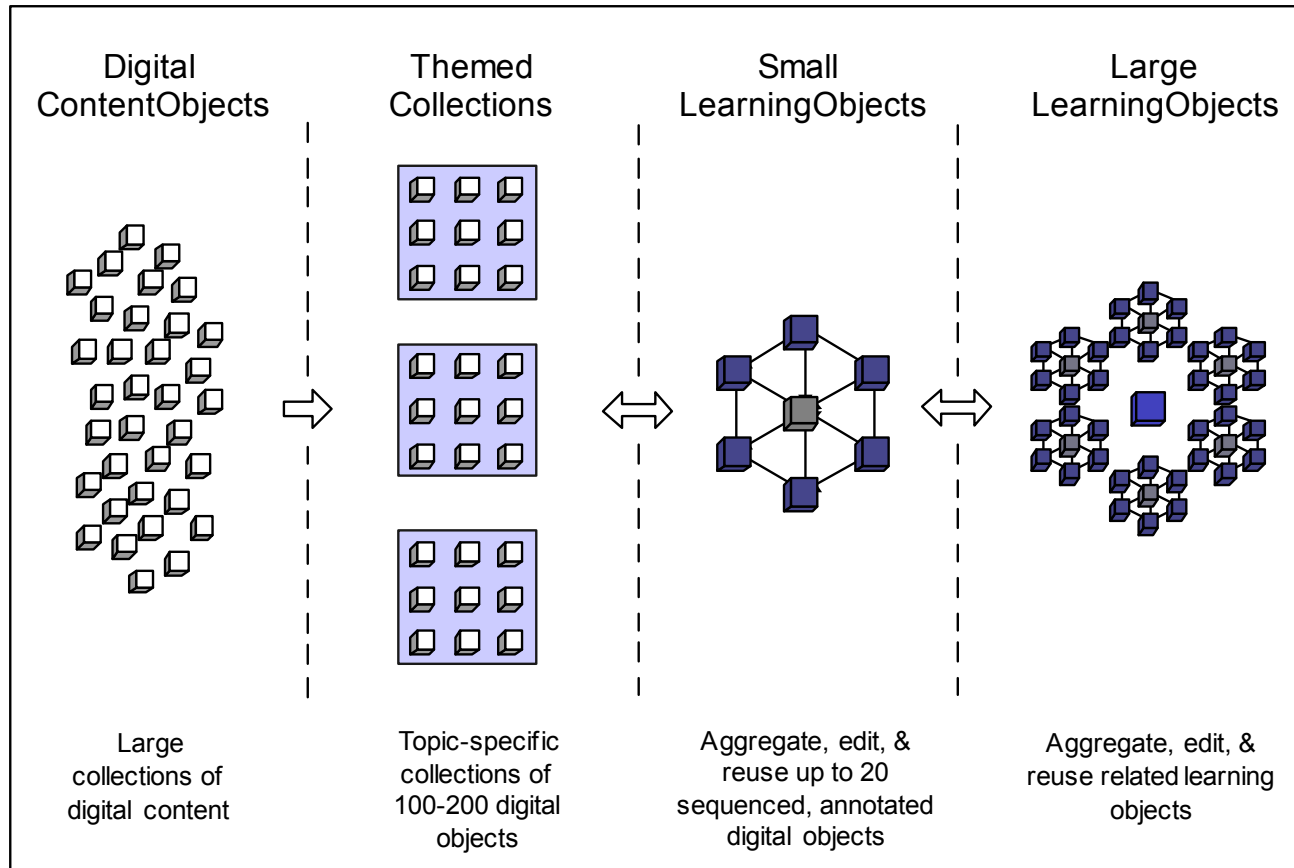
The logo graphic consists of a vertical black line on the left, a horizontal black line below the title, and a cluster of overlapping colored squares (yellow, red, blue) to the left of the text.

# IU Vision

---

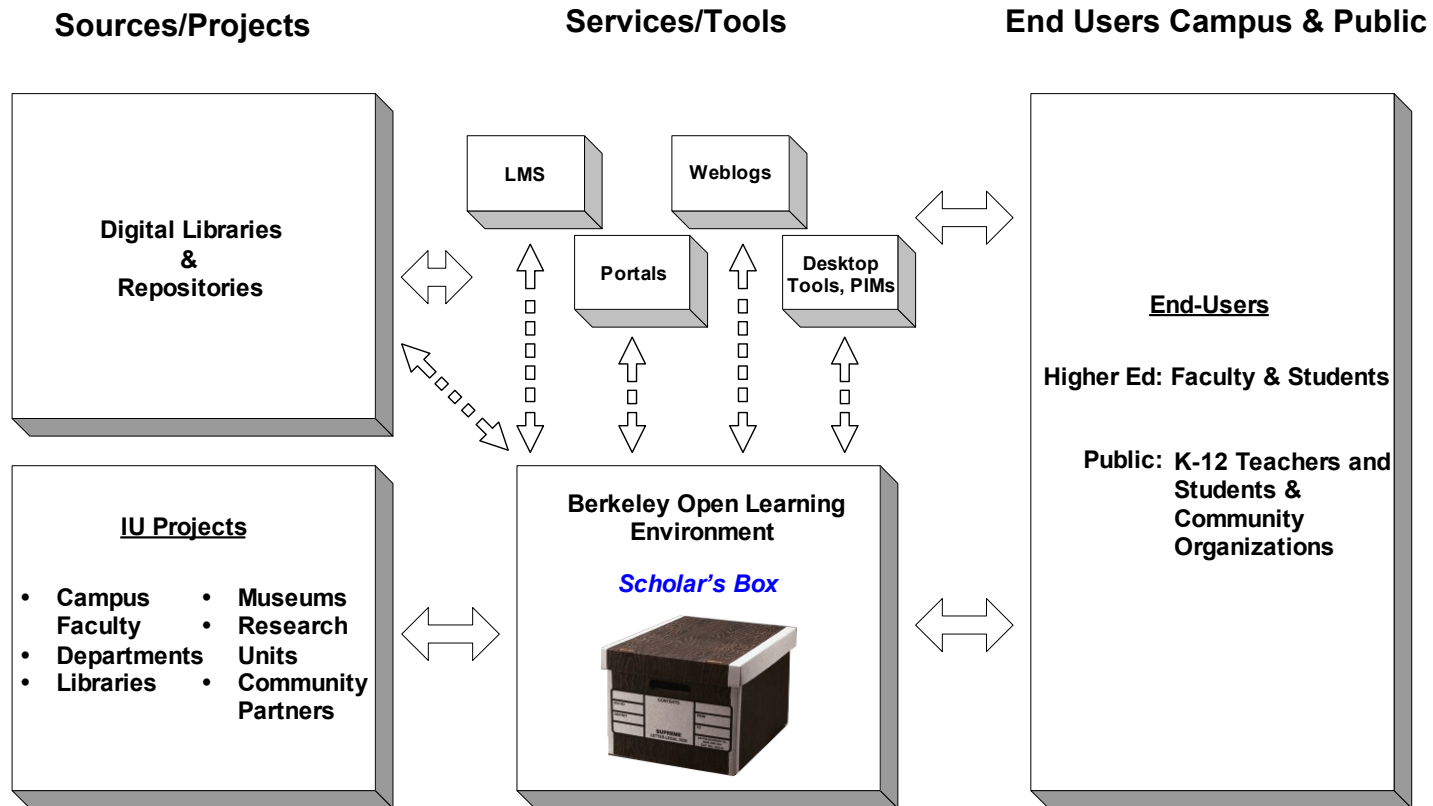
- 250+ Berkeley faculty, departments, museums, libraries, ORUs participating
- IU Projects: collaboratively building learning resources for K-12 and the campus
- Serving thousands of K-12 teachers statewide
- Use Cases (some):
  - Collaboratively building themed collections of teaching and learning resources from university content
  - Building small learning objects from themed collections
  - Using these digital learning materials in instruction
  - Sharing and encouraging reuse

# Content Forms



*Small and structured content and learning objects are beautiful....*

# IU: Opening Content



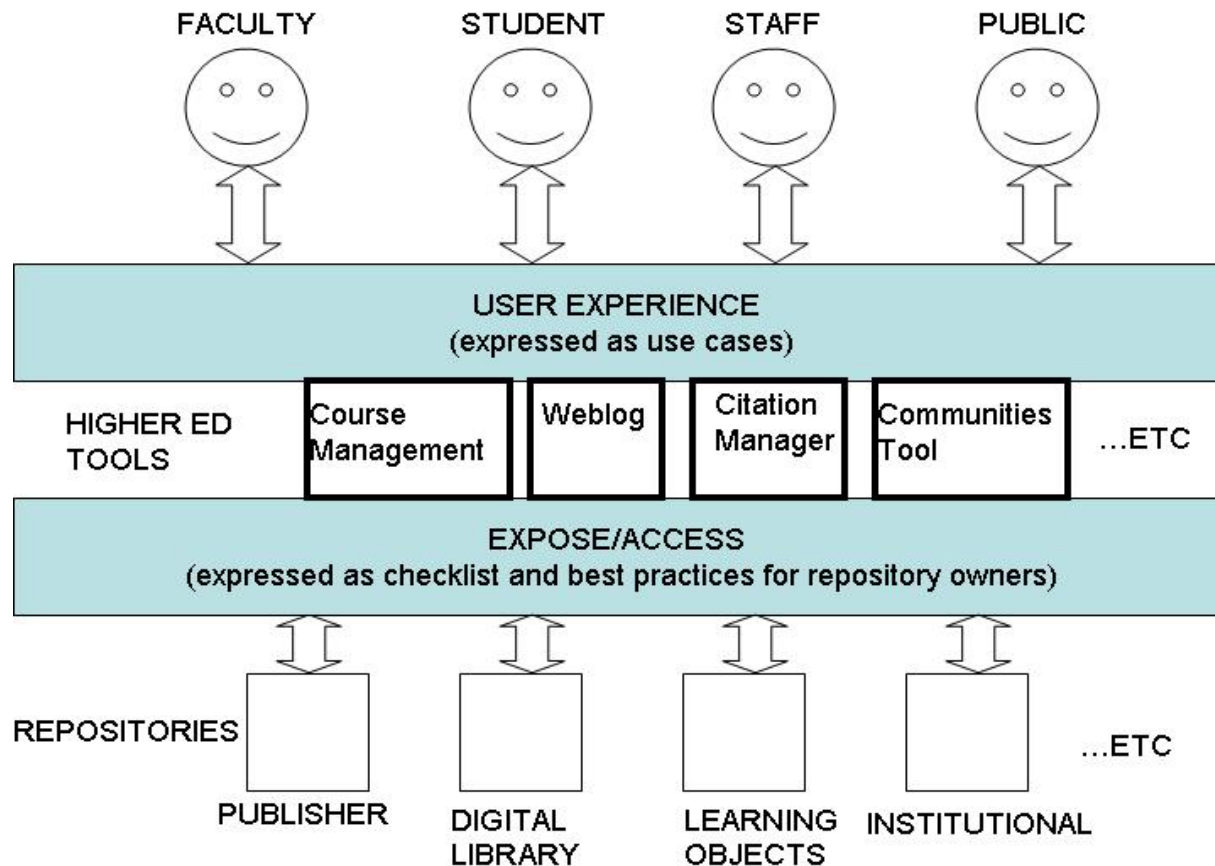


# IU ♥ CDL

---

- Research libraries and museums have extraordinary content, esp. primary source materials
- CDL taking leadership role in:
  - Exploring how to expose digital objects in structured form
  - Working with LMS/CMS and ed. technology
  - Opening materials beyond university
- CDL interested in the broader model of interoperability and partnership with educational technologies ... what is the digital library of the future?

# Framework – A Big Picture





# CDL History

---

- University of California's 11th library
- Founded in 1997 (from DLA)
- CDL users include the UC libraries, UC Press, various museums, and the State of California
- Portal for publicly available content:  
<http://www.californiadigitallibrary.org/>



# CDL Goals

---

- The CDL relies extensively on strategic partnerships and technical innovation. The CDL:
  - **Focuses on digital library collections** by developing, acquiring access to, and persistently managing collections.
  - **Invests in applications** that help campus libraries build meaningful, cost effective, and customized services.
  - **Provides collaborative leadership** in the development of new forms of scholarly communication.
  - **Leverage its investments** to influence the marketplace for scholarly information.



## CDL Produces ...

---

- A major application is Melvyl union catalog
- Has built well-known content portals :
  - eScholarship Editions (*UC Press e-books*)
  - eScholarship Repository (*Bepress*)
  - Counting California
  - Online Archives of California (OAC)
  - SearchLight (*early search interface*)



# Pre-Industrial Age for DLs

---

- Silo-like content portals are “cabinets of curiosities,” heavily utilized, in isolation
- Difficult to integrate with other silos and services
- Do not scale; power law costs for content growth
- MD transforms and content display finely tailored
- Software development becomes baroque, with no cross-app change control or standardization



# Edge of Infrastructure Cliff

---

- DLs can no longer live apart from other systems
- Commune in academic information ecosystem
- “Good Stuff” situated beyond university borders
- Increasing domain/discipline data needs (*i.e.*, the much talked about cyberinfrastructure)
- New development based on distributed discovery, access integration and content re-use/re-purposing



# Scaled Repo Architecture

---

- CDL envisions federated repositories:
  - Individual » (Chandler/Westwood?)
  - Departmental/Disciplinary » (Grid?)
  - College or Campus » (DSpace?)
  - Institutional » (CDL)
- Integrative user tools in development



## CDL ♥ IU

---

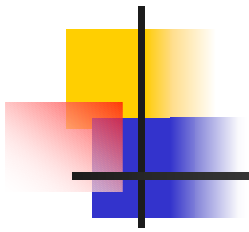
- SB exciting content manipulation tool
- Integration with live, interactive content
- Useful front-end to repositories
- SB can “call” metasearch services
- Disguises some of the messiness
- CDL can assist in content packaging



# IU Scholar's Box

---

1. **Gather** digital content with metadata and context from many sources.
2. **Create** custom collections and learning content: manipulate, annotate, transform.
3. **Share** collections and documents with others. Connect content to other tools.





## Roles for Metadata

---

- Metadata Encoding Transmission Standard (METS) heavily used for DL content
- METS file describes structure of content
- Usually includes other descriptive metadata:  
TEI, MODS, MIX, VRA, EAD, DDI, *etc.*
- Often references display behavior *via* XSLT
- Embracing metadata profiles for content groups



## CDL Profiles

---

- Convenient hamper for supporting admin data
- Might include “Persistence Commitment Policy Classification”
- Hanger for early-stage METS extension rights information
- Could contain guidance for MD transforms for class or collection of objects



# Object Attribute Tangles

---

- Divorcing content from its appearance is an important step in liberating objects for reuse
- We encourage “skinning/slicing” - *i.e.*, local branding by contributors/submitters
- MOAC is really the public CDL web site, but looks different thanks to a style sheet choice



# Increasing Web Services

---

- External content sometimes retrievable *via* SOAP
- Melvyl (Ex-Libris' Aleph) supports XML-based retrieval
- Retrieve item information, online link to content
- Permits SB to integrate with DL objects
- Web services critical element of CDL's new repository architecture for our next-gen DL

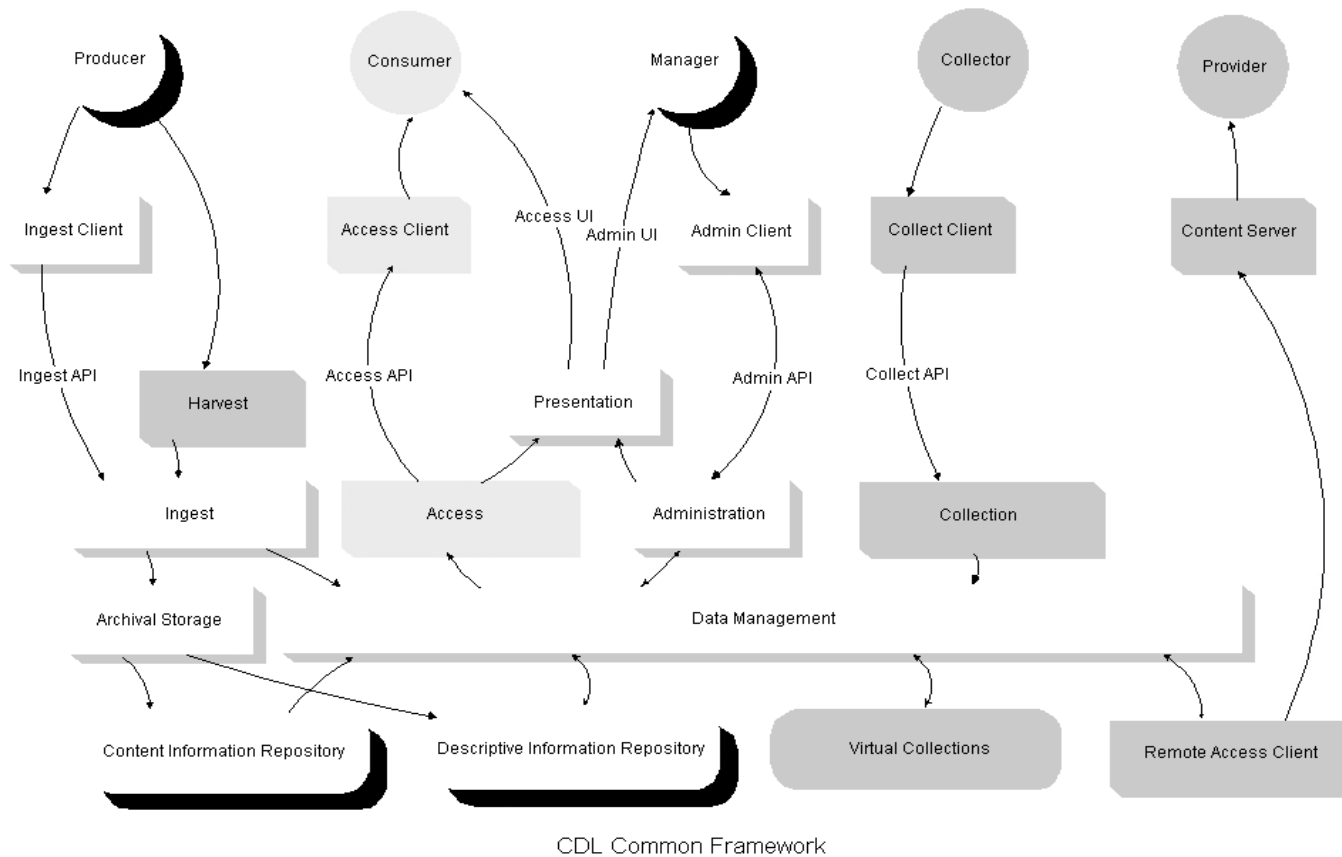


# Foundations of our Future

---

- Creating CDL Common Framework
- Lightweight, XML-based transactions
- Distributed, replicated Grid-based storage
- Repos based on assemblies of modular services for ingest, preservation, discovery, access, *etc.*

# Dig Lib Repository Needs





# Interoperability

---

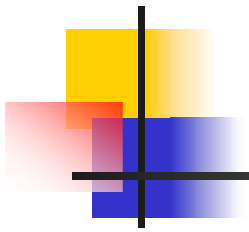
- DL metadata encapsulates but not instantiates actions or triggers (no self-awareness)
- Intelligence for any MD transforms must exist elsewhere in object manipulation services suite
- Repositories need to be able to establish real-time comprehension of interaction capabilities through publication of services



# Intel Repos

---

- Repository systems will gain – if not more intelligence – at least more awareness.
- Capabilities must be obvious and accessible, uniform and well-characterized.
- Our future lies in the services that we provide and the commitments we make, not merely in the content we preserve.





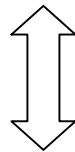
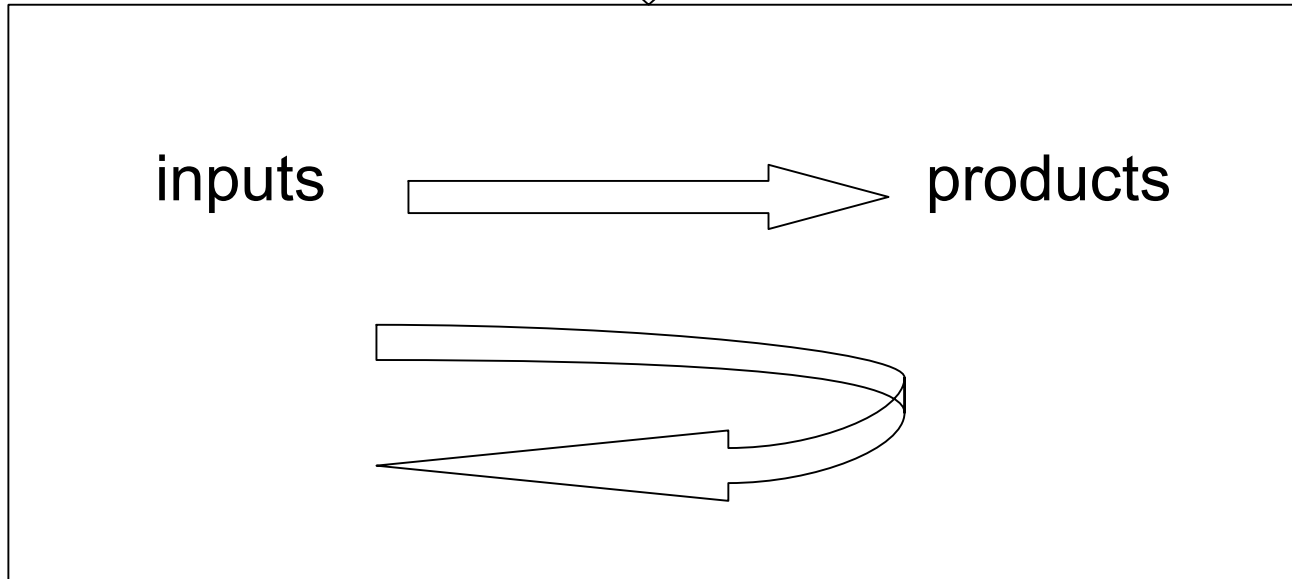
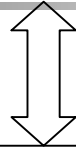
# Functionality of Scholar's Box

---

1. **Gather** digital content with metadata and context from many sources.
2. **Create** custom collections and learning content: manipulate, annotate, transform.
3. **Share** collections and documents with others. Connect content to other tools.

# Technology Architecture

Interfaces (DHTML/web services)



repositories

Data sources



personal documents



# Scholar's Box, fully realized, allows users to

---

- Access any data *source*
- Handle any data *type*
- Incorporate any *service*

The Scholar's Box is aimed to be a hub of interoperability.

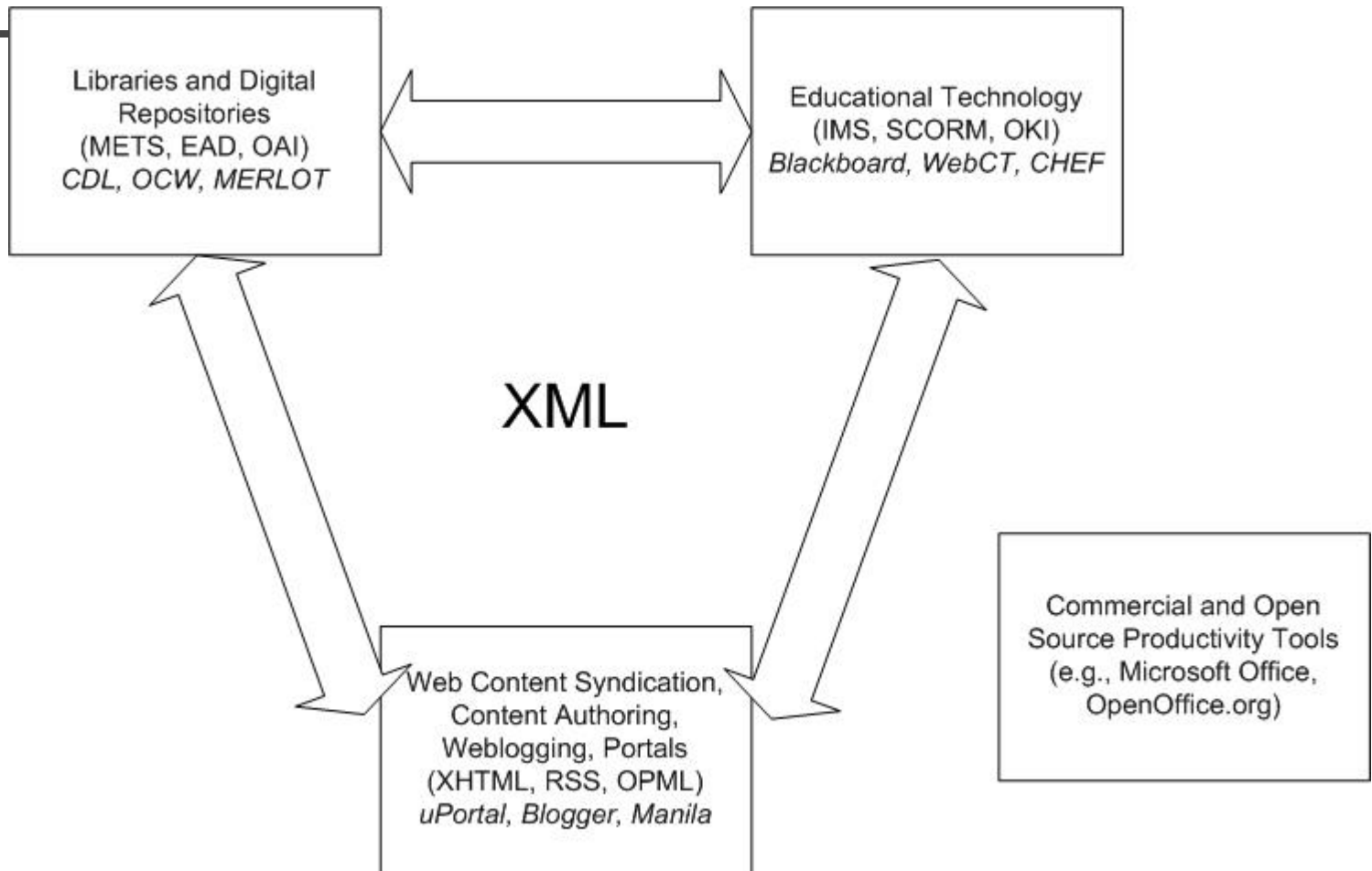


# Our strategy for moving ahead

---

- Focus on specific cases of data inputs and then generalize to others
  - Images
  - References
  - Texts
- Leverage existing infrastructure where possible
- Philosophy of practical interoperability

# Interoperability among domains via XML crosswalks





# Images: Sources

---

- Right now, there are 4 sources (CDL, amazon.com, CalPhotos, local drive)
- What other sources of images to connect?
- How to create a unified model for sources?



# Images from CDL

---

- CDL supports CDL-specific XML search results
- Questions for future implementation:
  - Use full Web services framework?
  - CDL as a metasearch broker?
  - What APIs/formats will CDL want to offer?
  - What would be ideal for the Scholar's Box?
  - Reuse of Scholar's Box components in CDL context?



# Images: User Experience

---

- How to make the flow of transforming images into products usable and natural?
- What products do people really want to make?
- How do people want to handle metadata? attribution?



# Images: Backend issues

---

- Metadata reconciliation/semantic interoperability
- What XML specifications will win out?
- Unified APIs for sources?
- Any workflow engine that intelligently constrains inputs to product transforms?
- Inter-application communication



# Infrastructure: A Wish List

---

1. Unified object models for all data
2. Object dis-aggregator and re-aggregator functionality
3. A “universal canvas” for composing and authoring of multimedia documents
4. Translators among data and metadata schemes (metadata crosswalks)
5. Integrated handlers for many protocols: SOAP, XML-RPC, Z39.50, HTTP, SMTP and inter-application communication
6. Hooks into groupware, P2P and centralized file sharing for sharing of content



# Promising developments

---

- Chandler [[url](#)]
- Microsoft Office 2003 (maybe?)
- OpenOffice.org [[url](#)]
- Multivalent Browser as a universal canvas [[url](#)]
- Fusion of weblogging and RSS and other forms of content syndication



# The Promise

---

*Rip. Mix. Burn.* "Apple, of course, wants to sell computers. Yet its ad touches an ideal that runs very deep in our history. For the technology that they (and of course others) sell could enable this generation to do with our culture what generations have done from the very beginning of human society: to take what is our culture; to "rip" it - meaning to copy it; to "mix" it - meaning to reform it however the user wants; and finally, and most important, "burn" it - to publish it in a way that others can see and hear. Digital technology could enable an extraordinary range of ordinary people to become part of a *creative* process."

– *Lawrence Lessig, The Future of Ideas*



# For more information

---

- IU: [iu.berkeley.edu](http://iu.berkeley.edu)
- CDL: [www.cdlib.org](http://www.cdlib.org)
- Raymond Yee's weblog:  
[iu.berkeley.edu/rdhyee/](http://iu.berkeley.edu/rdhyee/)