Integrating Digital Libraries and Electronic Publishing in the DART Project

David Millman
Gordon Dahlquist
Brian Hoffman

Columbia University
April 2005

EPIC Background
Electronic Publishing Initiative at Columbia

- 3-way partnership—Columbia Univ. Press, Academic Information Systems, Columbia Libraries
- Publications
  - Columbia International Affairs Online (ciao)
  - Columbia Earthscape
  - Gutenberg-E
- Evolving editorial and technology roles, workflow

DART Background
Digital Anthropology Resources for Teaching

- NSF/JISC funding— "Digital Libraries in the Classroom" program
- Partnership with London School of Economics & Political Science
- Anthropology Departments with Publishing/Educational Technology units
- 2 postdoc Fellows in each Anthropology Dept.—offload teaching load and links to senior faculty in each institution

DART Educational Mission

- To help undergraduate students gain insight into the way in which anthropologists conduct research and draw conclusions

- Improve information literacy of undergraduate anthropology students through use of structured yet unfiltered digital resources

E-Publishing Mission

- To develop a digital library infrastructure that will store digital resources so that they can be used in flexible ways

- To catalogue digital assets embedded within complex learning tools so that they can be used for broader research and/or teaching goals

Case 1: Intro to South Asian Culture

- Online syllabus that links to catalogued digital assets (primary texts, maps, photos, video)

- Teacher builds class assignments around these assets (response to questions, essays on readings, and full research paper)

- Increasing levels of interaction with library materials throughout the semester

Case 2:The Ethnographic Imagination

- The teaching module contains a digitized selection of author's field notes and published book

- Students read both sets of materials and write about the process of transforming the notes into an ethnography

- Increasing understanding of how knowledge is created from data

DART Publishing Environment

- Traditional Roles and Changing Relationships
- Editors/Authors & Publication Process
- Publications & the Library

Digital Teaching Tools and Research Library Resources

- Focus on the relationship between the "closed" world of the classroom and teaching tools, and the "open" world of the library

- Can students explore freely the vast array of research tools available through the Web, while still having an appropriate level of guidance concerning how to select and evaluate the sources that they find?

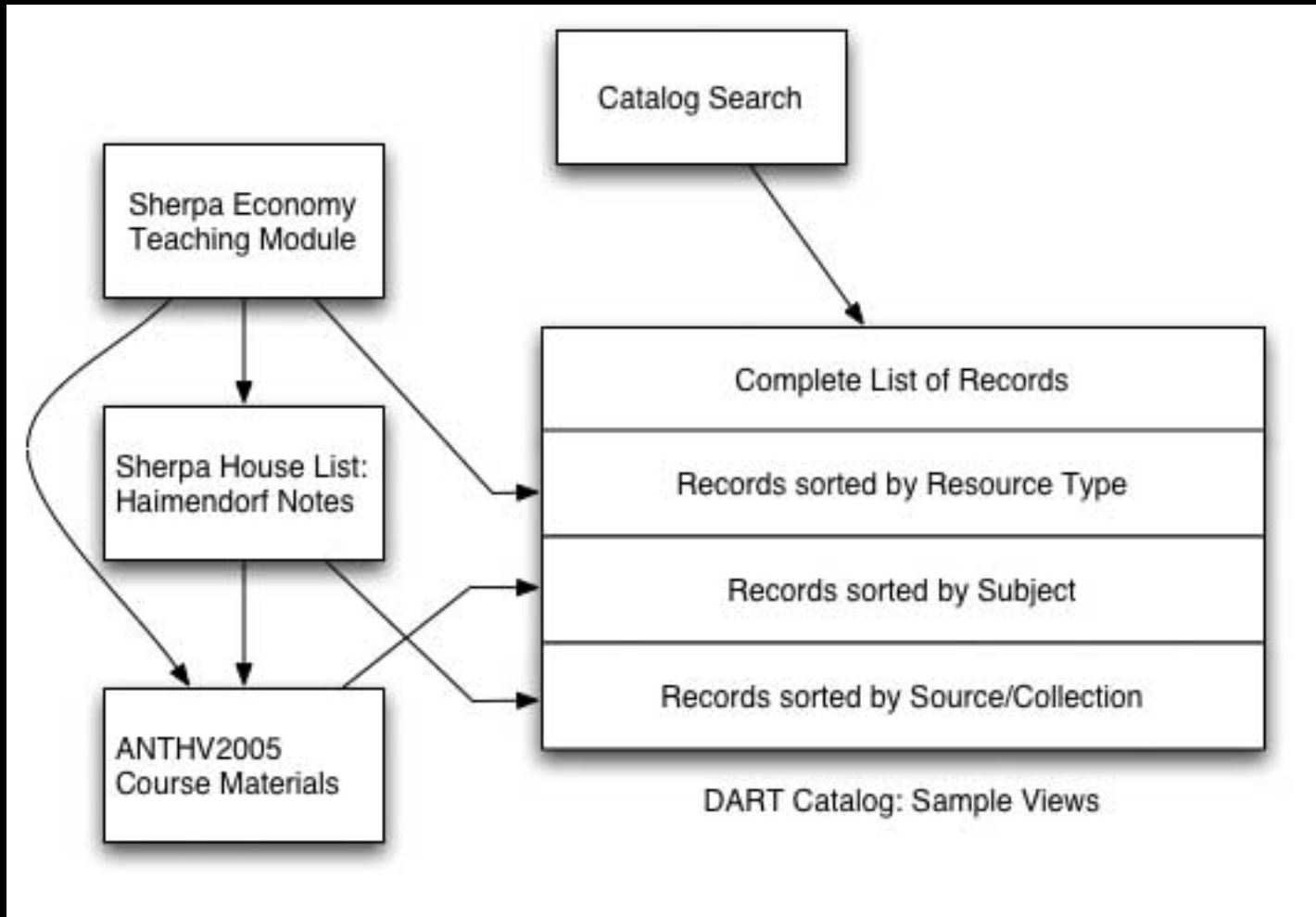# Unlimited Information as Benefit or Obstacle to Learning

- How do we make information meaningful to users with diverse skills and needs?

- Future work will explore how to find the right balance between directed and unfiltered presentation of digital teaching and research materials in electronic publications

Integrating Teaching Tools and Digital Library

Value added from each direction as part of production process

- Non-Hermetic Teaching Tools
- Collection presented within pedagogical context(s)

User Experience

DART Catalog: Sample Views

Technology

- Accommodate different styles for teaching
  - fall '04 (South Asian History & Culture): web browser focus (syllabus navigation)
  - spring '05 (Ethnographic Imagination): digital resource focus (primary source navigation)
  - fall '05 (planning): considering mobile device in DL discovery & retrieval; "Virtual Calcutta" object/software
- Web services import/export
- Access management/Shibboleth
- Metadata: "versions" revisited

Acquisition

Digital South Asia Library
DSAL @ U Chicago

Publishers
& Archives

DART faculty

Cambridge Univ Library
institutional repository

(proposed)
Tibetan-Himalayan DL
thdl @ U of Virginia

mapping

local workflow

OAI   DSpace   Fedora

DART catalog

DART content

Access

OAI

METS

MPEG21/DID

Sakai/OKI

JSR170

IMS/CP

browser
html

library & repository
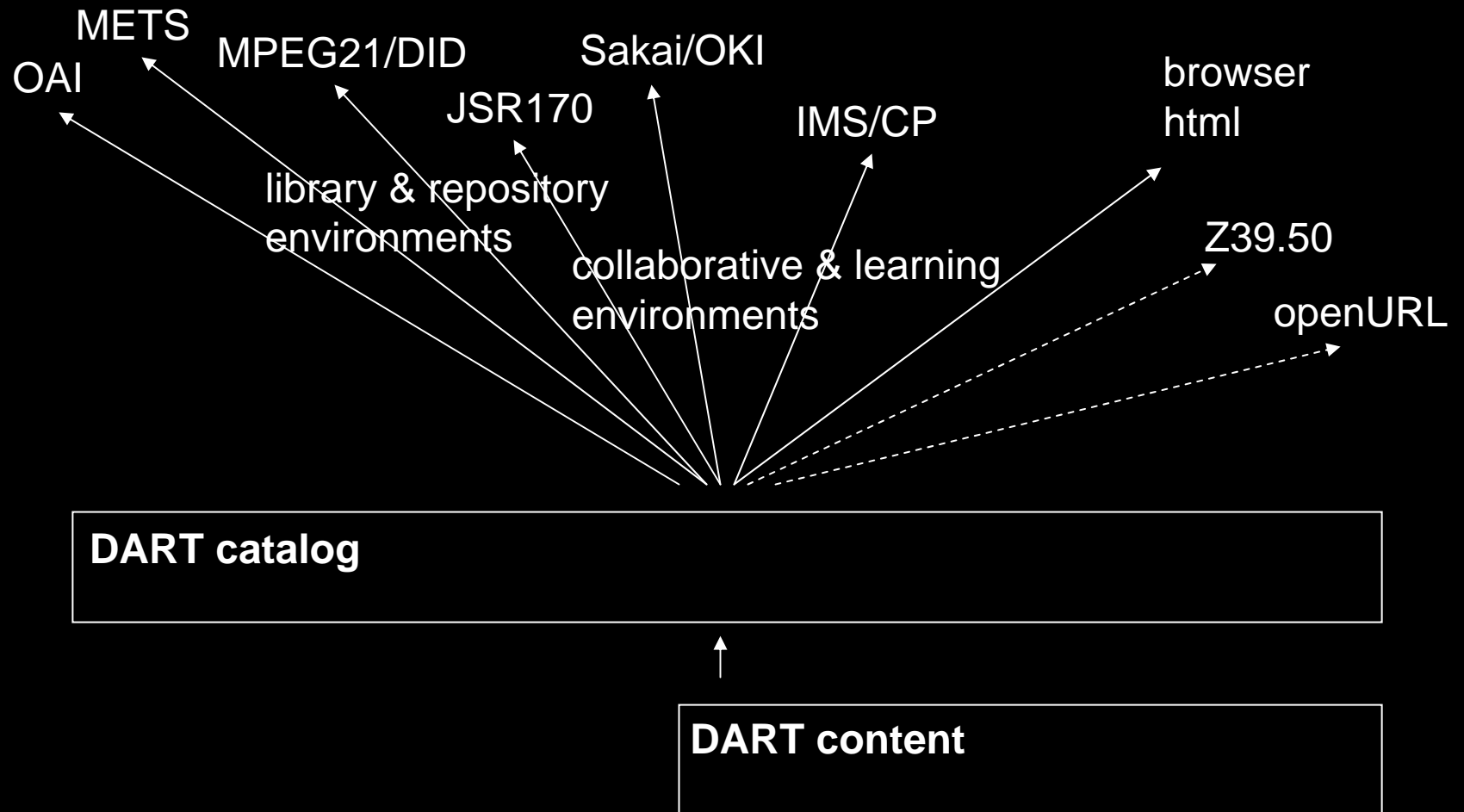environments

collaborative & learning
environments

Z39.50

openURL

DART catalog

DART content

The View from Production

Building DART's e-publishing production cycle

into open archive infrastructure systems

Building Publications

- Structured presentations of digital objects
- Legal presentation of digital objects (rights)
- Presentation through linking or embedding
- One to many relation between locally or remotely stored originals and versions embedded in publications

Examples of Publications

- Slide shows
- Mini-sites for classroom or homework use
- Online syllabi
- Complex page-viewing interfaces (online fieldnotes)
- Interactive games
- Any navigational interface to the digital library (faceted navigation, topic maps, etc.)

Objects within Publications

- Must conform to publication's specifications (e.g., consistent image size)
- Publication-specific metadata (e.g., caption)
- Embedded in a new format (HTML, Flash, Video)
- Objects appearing in a publication called "Assets"

Harvested Assets

- Harvest candidate (metadata) records from open archives and partner institutions
- Identify objects to import: desired assets
- Import bitstreams
- Draft metadata from candidate record (pre-populate fields)
- Edit metadata (catalog from our perspective)

Assets Digitized Locally

- Create digital archival copy (scan, photograph, etc.)
- Original Cataloging
- Store
  - part of preservation strategy

Publication Assembly

- File Modification
  - Crop, detail, resize
  - Reduce, snip, clip, extract
  - Interpret, explain, contextualize

- Presentation Context
  - Associate, locate
  - Incorporate, include, attach
  - Interpret, explain, contextualize

Three Asset Scenarios

Asset 1

- Digitized Map from Digital South Asia Library (http://dsal.chicago.edu)

Asset 1

- Bitstream and metadata copied to DART collection

- Metadata edited by DART editors

- DART bitstream copied and deployed into various publications

- Copies are reduced, cropped, applied with hotspots in photoshop, etc

Asset 2

- Digital video interview with von Furer-Haimendorf (http://www.lib.cam.ac.uk)
- 1.3 hours

Asset 2

- Metadata copied to DART collection
- Metadata edited by DART editors
- Short video clips deployed in various publications
- DART keeps no copy of the original object

Asset 3

- Chapter of *Sherpas Through Their Rituals* by Sherry Ortner

Asset 3

- Bitstream and metadata created by DART
- Re-publication rights secured by DART
- Scanning done by DART
- Archival responsibility assumed by DART

Exposing Items in DART Library to Other Systems

- Complicated relationships between source files and derivations

- Versioning, entropy

- Redundancy and degradation (importing a large file and passing along a small file)

- Even more complicated relationships between source file metadata and derivation file metadata

Expressing Relations Among Versions and Derivations

- DART metadata schema = extension of Dublin Core element set
- **derivedFrom** tag
- Plan to offer OAI harvesters DART schema in addition to OAI_DC
- Now cataloging and tracking derivation information

derivedFrom element

- URI of source file
  - Another DART item
  - An item in an outside system (URI may be download page)
- Date copy was made
- Description of alterations, copy methods, purpose, etc.
- Analogous to OAI provenance tag
  - OAI provenance : metadata :: derivedFrom : bitstreams

OAI provenace

- Describes metadata provenance
- Assumes fixed object, mobile metadata
- 0 provenance tags for a copy made for the purpose of alteration and incorporation
- Problem of metadata
  - Source metadata used to "seed" derivation metadata
  - Can't record this kind of provenance through OAI provenance

Exposure of Others' Metadata

```
<!—Record 2: a record harvested from Chicago, representing an object in the -->
<!--DSAL library, as EXPOSED by DART-->
<record>
    <header>
        <identifier>oai:lib.uchicago.edu:ta013</identifier>
        <datestamp>2004-10-08T18:50:13Z</datestamp>
        <setSpec>dsal</setSpec>
        <setSpec>dsal:hensley</setSpec>
    </header>
    <metadata>
        <oai_dc:dc>
                <identifier>http://pi.lib.uchicago.edu/1001/org/dsal/ima...</identifier>
                <title>Gate into Taj grounds</title>
                ...
        </oai_dc:dc>
    </metadata>
    <about>
        <oai_dc:dc>
            <dc:publisher>The University of Chicago Library</dc:publisher>
            <dc:rights>No rights to the use of these...</dc:rights>
        </oai_dc:dc>
        <provenance>
            <originDescription harvestDate="2004-10-08T14:10:02Z" altered="false">
                <baseURL>http://dsal.uchicago.edu/</baseURL>
                <identifier>oai:lib.uchicago.edu:ta013</identifier>
                <datestamp>2004-10-01</datestamp>
                <metadataNamespace> OAI... </metadataNamespace>
            </originDescription>
        </provenance>
    </about>
</record>
```

Exposure of DART's Metadata

```xml
<!--Record 3b, metadataPrefix = dart_xdc -->
<!--A record representing an object in the DART digital library that is a derivation of
  the object represented in Record 2, exposed with DART metadata (an extension of dublin
  core that includes work-derivation information-->
<record>
    <header>
        <identifier>oai:dart.columbia.edu:dart0023</identifier>
... </header>
    <metadata>
        <dart_xdc xmlns:dart_xdc=...>
            <identifier>https://dart.columbia.edu/main/DART-0023.html</identifier>
            <title>Photograph of Gate Into Taj Grounds</title>
            ...
            <derivedFrom>
                <description>This image was resized to 700 by 800 pixels,
                  and cropped around a sketch at the corner of a notebook...</description>
                <sourceObject>
                    <identifier>http://pi.lib.uchicago.edu/1001/
                       org/dsal/images/hensley/ta013</identifier>
                    <datestamp>2004-10-07T06:05:04Z</datestamp>
                </sourceObject>
            </derivedFrom>
        </dart_xdc>
    </metadata>
</record>
```

Open Publications?

- Potential for Publication-based harvesting
- "Dissolve" a publication into a set of de-contextualized digital objects
- Many points of alignment between publication and archival processes
- Publications can supply as well as re-purpose archived material

dart.columbia.edu