



University of Michigan's OAIster Service Provider

Kat Hagedorn

OAIster/Metadata Harvesting Librarian

University of Michigan, DLPS

November 5, 2002



OAIster Overview

- One-year Mellon grant project (one of 7)
- Grants for testing feasibility of using OAI to make metadata accessible to the public
- Digital Library Production Service at UM began work in December 2001
- Publicized as OAIster in February 2002
- Launched as search service in June 2002



Project Highlights

- Any audience
- Any subject matter
- Any format
- Freely accessible
- No dead ends
- One-stop shopping

...retrieving the “hidden web”



Tools We Used

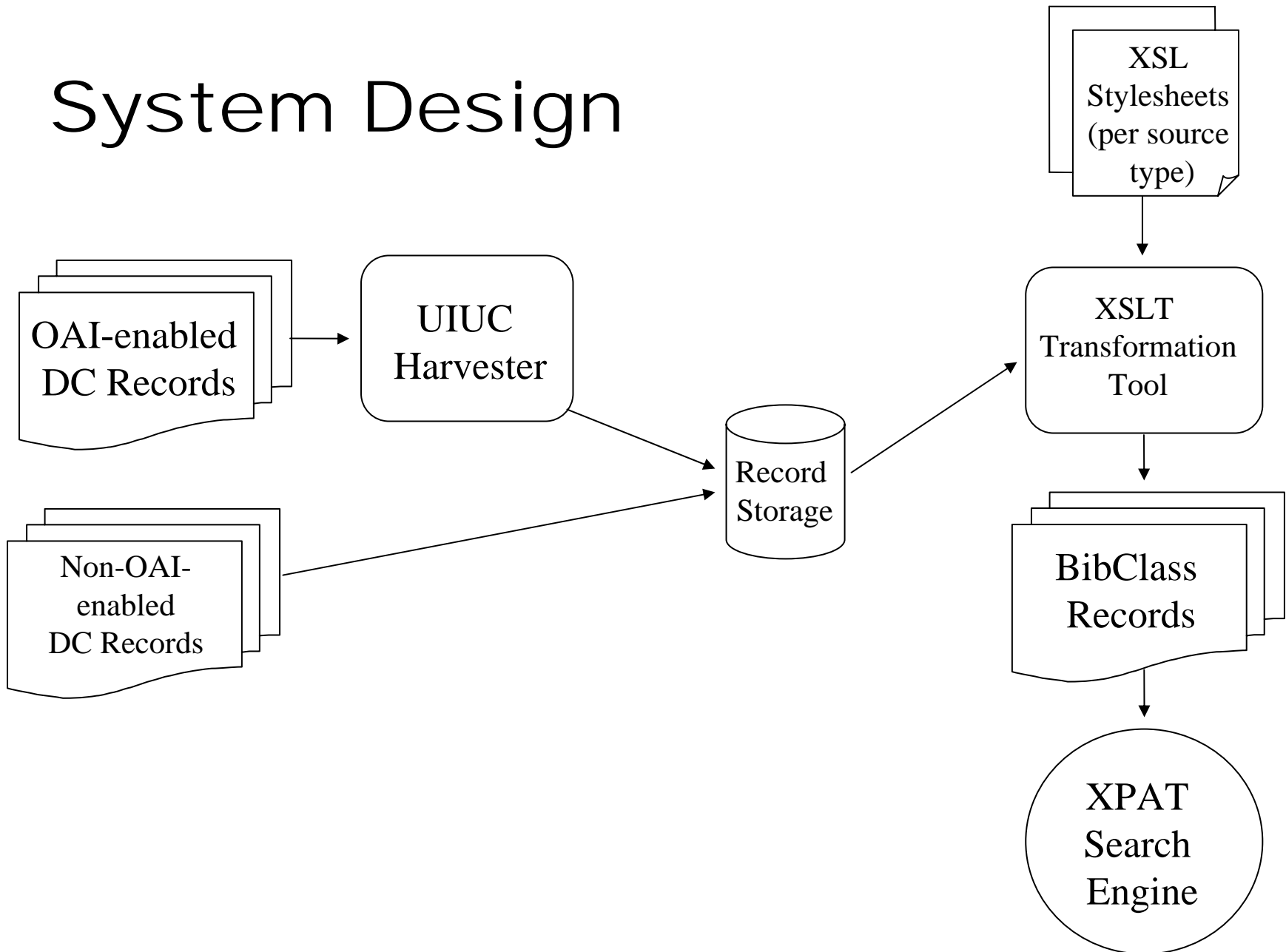
- UIUC Harvester
- Two editions developed; we used Java edition
- Running since March of this year
- Worked collaboratively to iron out kinks



Tools We Developed

- UM DLPS runs DLXS middleware
- Using middleware as a base, developed searchable interface to harvested records
- Also developed a Java-based transformation tool to:
 - Collect harvested records into large files
 - Filter out records that don't have digital objects associated with them
 - Normalize the DC element Resource Type
 - Add institution information
 - Count records and provide quality of data feedback
 - Convert UTF-8 to ISO8859-1
 - Use XSLT to transform DC records into DLXS records

System Design



End Result

- Search service for end-users allowing them to find **944,890 records** from **112 institutions** (as of October 31, 2002)
- Example institutions we harvest from:
 - **Online Archive of California** - manuscripts, photographs, and works of art held in institutions across California
 - **arXiv Eprint Archive** - math and physics pre- and post-prints
 - **Sammelpunkt, Elektronisch Archivierte Theorie** - archive of philosophical publications
 - **British Women Romantic Poets Project** - collection of poems written by British women between 1789 and 1832



Oaister

...find the pearls

[Home](#)

[Search](#)

[View Institutions](#)

[Help](#)

Use this form to find digital resources ([what is a digital resource?](#)) from many institutions ([which institutions are available?](#)).

See [help](#) for search tips.

Search all fields [\(help\)](#)

Use a word or phrase, e.g., diploma*, fancy dress

and ☒ or ☐

Search within particular fields [\(help\)](#)

Use a word or phrase, e.g., diploma*, fancy dress

Title: and

Author/
Creator: and

Subject: and

Resource
Type:

Choose sorting of results [\(help\)](#)

e.g., results in date descending order, from 2002 to 1999

☒ title ☐ author/creator

☐ date descending ☐ date ascending

☐ hit frequency ☐ weighted hit frequency



...find the pearls

[Home](#)

[Search](#)

[View Institutions](#)

[Help](#)

Your search was in **all fields** for **"bungalow"**.

You found **56 records**.

* [Revise your search](#) to retrieve fewer records.

* View your results, starting with **records 1 to 10 of 56**.

Sort by

title



[Next 10 Records](#)

Jump to Records: [1](#) | [11](#) | [21](#) | [31](#) | [41](#) | [51](#)

Record 1 of 56

Title	18 Willingdon Crescent Bungalow : exterior view
Author/Creator	Lutyens, Edwin
Resource Type	Architecture
Resource Format	image/jpeg
Note	This object is held by: University of Minnesota, Dept. of Art History, Visual Resources Center
Note	Keywords: Houses, colonial
URL	http://bigbird.lib.umn.edu/umimage/servlet/DisplayImage/umimages/vrc1009827
Rights	These images may be copyright protected. Please contact the Visual Resources Center for information on use and access restrictions.
Institution	University of Minnesota IMAGES

User Feedback

- 2 surveys, one lengthy and highly publicized before launch, one short and publicized intra-UM after launch
 - Users want electronic journals and online reference materials
 - Users want a comprehensive place to look for online materials
- 2 sets of face-to-face and remote testing
 - Users don't need short and long record formats
 - Users need clearly defined and labeled AND/OR searching options, but found the results clear and easy to understand
 - Users want to sort by title, date, institution, resource format...you name it!
 - Users use OAIster for academic, trustworthy, authentic materials instead of search engines like Google



Statistics

Types of Statistics	Months in 2002			
	June (last 3 days)	July	August	September
Total Number of Accesses	689	8321	3595	2540
Institutions Accessing OAster More Than 20 Times (Excluding UM)	3	12	10	7

Progress and Future Plans

- Improvements to service
 - Better search results access for large results sets
 - Faster sorting, with more relevancy options
- Research questions
 - Relevancy ranking
 - “Best” answers
- Next year
 - Browsing capability
 - Saving/emailing/downloading records
 - More normalizing of data
 - Restricted vs. free access
 - Duplication of records



What Can You Do?

- OAI-enable your data
 - If you're a DLXS customer, this is essentially easy now
 - Make sure your data is UTF-8 / Unicode compliant
 - Provide as much metadata as you can
 - Use standard element tags, e.g., <dc>, not <oai_dc>
 - Develop "sets" for the needs of service providers
- Let us know you're ready to be harvested
- Keep us informed about changes to the harvesting URL, new data available, new contact info
- Collaborate on appropriate harvesting and access to records



Contact Info

- Kat Hagedorn
- UM Digital Library Production Service
- khage@umich.edu
- <http://oaister.umd.umich.edu/>
- For technical info: Mike Burek,
mburek@umich.edu