

Data Curation and Collaboration: Sharing Crystallographic Data through the IDEALS Institutional Repository



Melissa Cragin
Scott Wilson
University of Illinois

Illinois
Digital
Environment for
Access to
Learning and
Scholarship

George L. Clark X-Ray Facility &

3M Materials Laboratory



© 2007, Sarah L. Shreeves
This work is licensed under Creative Commons Attribution-NonCommercial-ShareAlike 2.5 License.
To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/2.5/>



Outline



- Institutional repositories
- Libraries and data
- Data curation education
- Project with the Wilson crystallography lab





Emerging Data Service Models in Research Libraries and Institutional Repositories

- Consultation on data management and curation solutions
- **IR as central repository for faculty data**
- Library Portal – Services for distributed data collections
- Collaboration
 - add services to grant proposals
 - Library develops tools with a research group
- Research Services – provide “end-to-end” services for particular disciplines



Data Challenges for Libraries



- Cleaning up data
- Storage
 - Size and number of data sets
 - Distributed data
- Description of data
 - Can others reuse the data effectively?
- Provenance of data





Data Curation is...

The active and on-going management of data through its lifecycle of interest and usefulness to scholarship, science, and education.

Data curation activities

- enable data discovery and retrieval
- maintain data quality and add value
- provide for re-use over time;

Tasks include

- Appraisal and selection
- Authentication
- Representation
- Archiving
- Preservation



What is IDEALS?



Institutional repository for the scholarship and research in digital form of the faculty, students, and staff of the University of Illinois at Urbana-Champaign.

- Currently in pilot
- Going into production Fall 2007
- A joint initiative between the University Library and CITES with support from the Office of the Provost.

<http://ideals.uiuc.edu/>



What will IDEALS take?

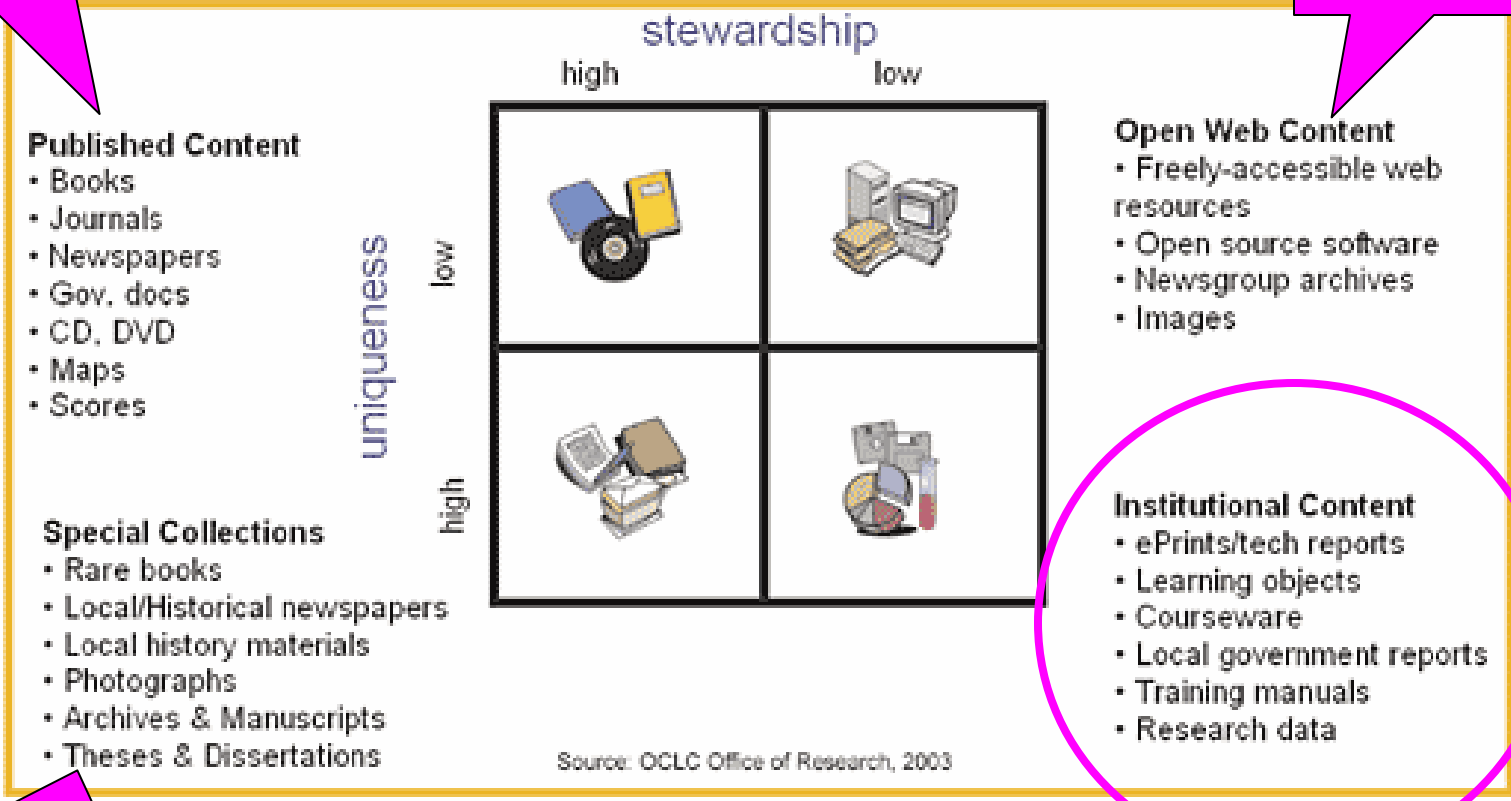


**Journal articles,
books, etc
by faculty**

**Some
scholarly
web sites**

Collections Grid

A framework for representing content



Source: OCLC Office of Research, 2003

**Theses /
Dissertations
Manuscripts**



IDEALS goals....



- Help increase access to published and unpublished research
- Help increase the impact to published and unpublished research
- Provide a persistent, permanent URL for your research
- Preserve research for long term access and use

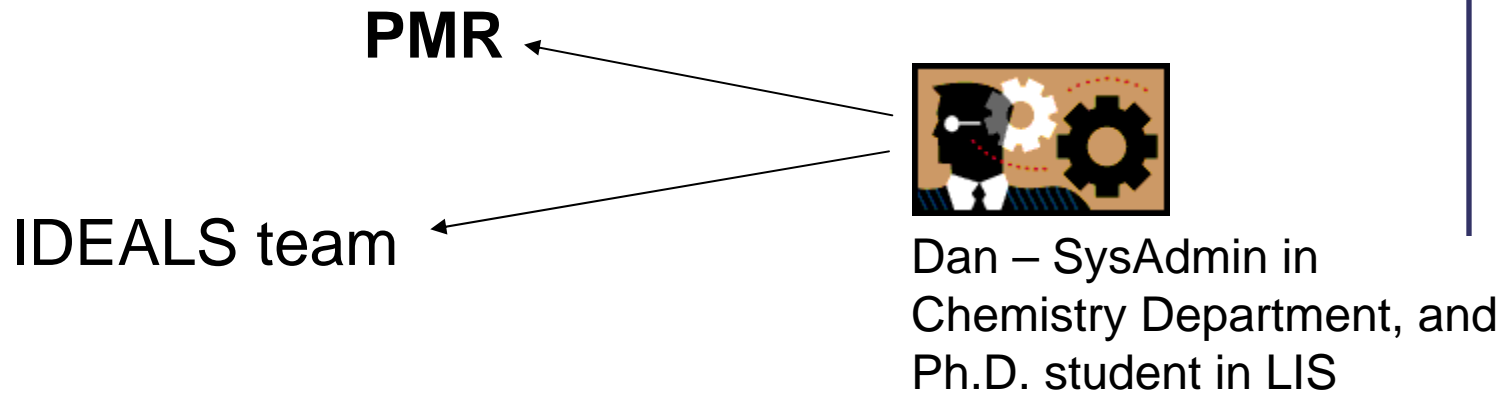


Other services we offer...



- Consultation on copyright issues
- Access restricted items or collections in IDEALS
- Statistics on number of downloads (regular weekly or monthly reports are coming soon...)





Building the UIUC Library – Clark X-Ray Facility Collaboration



The IDEALS team meets with Scott Wilson and Dan at the Lab.

Melissa tags along to learn about data curation needs in crystallography.



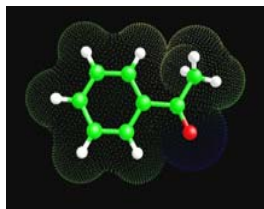
Sharing Crystallography Data at UIUC



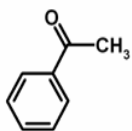
- DSpace for data in open formats
- Open Archives Initiative - Protocol for Metadata Harvesting
- Discovery and Access services
 - persistent, permanent URL
 - Controlled access (embargo)
 - Links to other data resources



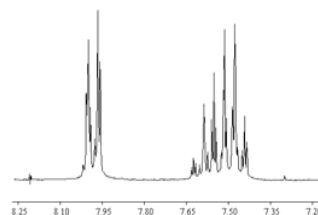
The SPECTRa Model



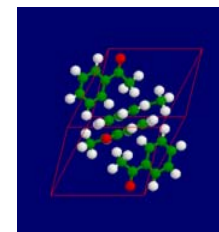
Computational
Chemistry Calculations



2D Chemical
Structures



NMR
Spectra



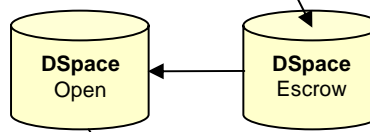
3D X-ray
Structures

SPECTRa Deposit Tools
Create CML, InChI, metadata

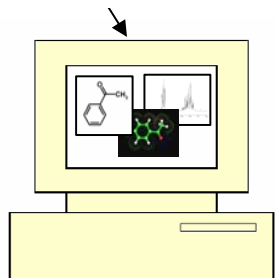
InChI :
InChI=1/C8H8O/c1-7(9)8-5-3-2-4-6-8/h2-6H,1H3

CML :

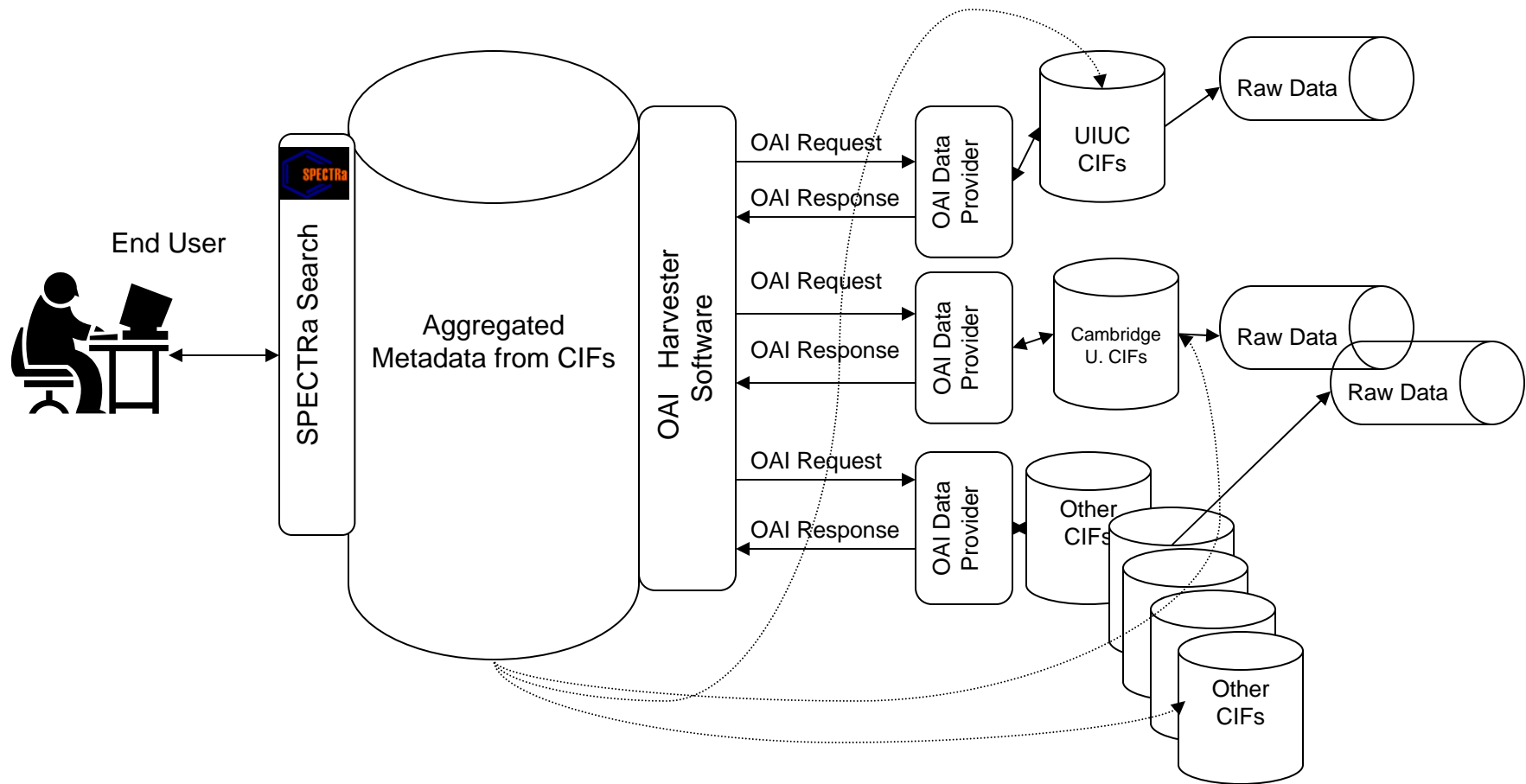
```
<molecule xmlns="http://www.xml.cml.org/schema">
<atomArray>
<atom id="a1" elementType="C" x2="-0.380600" y2="-0.720800"/>
<atom id="a2" elementType="C" x2="-0.381800" y2="-1.548200"/>
<atom id="a3" elementType="C" x2="0.333100" y2="-1.961000"/>
<atom id="a4" elementType="C" x2="1.049500" y2="-1.547700"/>
<atom id="a5" elementType="C" x2="1.046600" y2="-0.717200"/>
<atom id="a6" elementType="C" x2="0.331300" y2="-0.308000"/>
<atom id="a7" elementType="C" x2="1.759600" y2="-0.302000"/>
<atom id="a8" elementType="C" x2="2.475600" y2="-0.711800"/>
<atom id="a9" elementType="O" x2="1.756400" y2="0.523000"/>
</atomArray>
<bondArray>
<bond atomRefs2="a4 a5" order="1"/>
<bond atomRefs2="a2 a3" order="1"/>
<bond atomRefs2="a5 a6" order="2"/>
<bond atomRefs2="a6 a1" order="1"/>
<bond atomRefs2="a1 a2" order="2"/>
<bond atomRefs2="a5 a7" order="1"/>
<bond atomRefs2="a3 a4" order="2"/>
<bond atomRefs2="a7 a8" order="1"/>
<bond atomRefs2="a7 a9" order="2"/>
</bondArray>
</molecule>
```



SPECTRa Search Tools
OAI-PMH Harvesting



Extending Collaborations and Services





DCEP: Components of the new Data Curation Education Program

- Integration of research and practice
- On-line learning option (live, w/ campus visits)
- New curriculum
 - foundations; ontologies; digital preservation
- Practicum and Internship opportunities
- Partnerships with research organizations
- Advisory Committee





Data Curation Education Program

Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign

P. Bryan Heidorn, PI
Melissa Cragin, Project Coordinator

{pheidorn, cragin@uiuc.edu}

Data Curation Educational Program (DCEP)
IMLS – Laura Bush 21st Century Librarian Program
RE-05-06-0036-06



http://www.lis.uiuc.edu/programs/ms/data_curation.html



In Summary - IDEALS



- Library at UIUC has great interest in participating in data curation activities.
- The IDEALS team in particular is actively exploring what roles the Library can have.
- Looking for ways to partner with other libraries that are doing related work



For More Information: IDEALS



Sarah Shreeves
IDEALS Coordinator
sshreeve@uiuc.edu

Tim Donohue
IDEALS Technical Lead
tdonohue@uiuc.edu

<http://ideals.uiuc.edu/>

