

---

# Capturing research outputs at the University of Cambridge: experiences with DSpace

*Peter Morgan*  
*Project Director, DSpace@Cambridge*  
*&*  
*Medical Librarian*  
*Cambridge University Library*



# Outline

---

- the Cambridge institutional context
- DSpace@Cambridge project
- SPECTRa project
- points to ponder



---

# The Cambridge institutional context



# University of Cambridge

---

- collegiate university, c.1209-
- 23 Faculties in 6 Schools
- >17,800 students (11,700 u/g, 6,100 p/g)
- >5,000 academic and academic-related staff
- 81 Nobel Prize laureates since 1904
- decentralised, democratic governance structure
- separate library service and computing service
- tripartite library system (>100 libraries)
  - University Library (legal deposit library)
  - faculty & departmental libraries
  - college libraries



# Cambridge University Library

---



© Cambridge University Library



# The Library's policy context

---

- digital preservation research and practice
  - CEDARS and CAMiLEON projects
- scholarly communication
  - SPARC Europe
  - advocacy for self-archiving of research papers
- library collections (born-digital, digitized)
- local archiving
  - e-theses
  - University Archives
- visibility and credibility among scientists
- legal deposit of UK digital publications



---

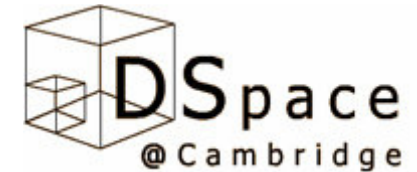
# DSpace@Cambridge



# DSpace@Cambridge

## - project outline

---



- January 2003 to August 2006
- funded by Cambridge-MIT Institute
- collaboration between Cambridge (Library + Computing Service) and MIT Libraries
- project vision and goals
  - identify and respond to user requirements
  - establish institutional repository
  - develop DSpace software
  - support adoption and use of DSpace in the UK
  - create a sustainable business plan





# Potential material

---

- scholarly papers (advocacy campaign)
- library collections (born-digital or digitized)
- learning materials (interactive, multimedia)
- research materials (texts, images, video, etc.)
- e-theses
- datasets
- administrative records
- e-journals & e-books
- websites



# Internal market survey

- questionnaire (Web and paper) in 2004
- >1,800 academic research staff surveyed
- 250 (13%) replied
- 48% definitely/probably would use DSpace
- 17% definitely/probably would not use DSpace
- respondents asked to rank 11 suggested benefits...



# Benefits of using institutional repository?

---

- respondents rated each of 11 benefits on a scale of 1-5
- the three top-ranked benefits were:
  1. 48%: "DSpace will provide long-term storage of research material"
  2. 34%: "DSpace will allow me to restrict access to items that I wish to archive..."
  3. 32%: "DSpace will make it easy for other people to search for and locate my work"



# Assumptions

---

- submitters provide their own Dublin Core metadata
- the University Library reserves right to validate, correct, or reject metadata
- copyright remains with the owner
- the repository is licensed to make copies for dissemination and preservation
- "open access" is the default status, but restricted /closed collections may be agreed



# End-of-project metrics (1)

- 17 DSpace communities
- 35 collections containing:
  - peer-reviewed literature
  - reports
  - e-theses
  - images
  - digital video
  - data
- some closed access collections
  - commercially valuable material (TIFF files)
  - restrictions on Intellectual Property Rights
  - work in progress



# End-of-project metrics (2)

- total number of items: 179,691
- total number of registered users: 190
- 27,000 visits per month
- greater visibility of linked research materials: Social Anthropology websites now ranked 5<sup>th</sup> in "Webstats4U" website list of the top 1,000 UK Higher Education sites



# Main project results

- business case approved by University
  - 5-year funding programme for DSpace@Cambridge *service*
  - 4 full-time posts (2 Library, 2 Computing Service)
- DSpace@Cambridge an integral part of the University's official information strategy
- significant input to DSpace Open Source code base
- Cambridge developer on DSpace Committer group
- hosted first European DSpace conference in 2005
- decision to focus next on work that embeds repository use in research practice, leading to...
- further research projects
  - SPECTRa (deposit tools for chemistry research data)
  - other partnerships with JISC-funded projects
  - ?SPECTRa-T (text- and data-mining from chemistry theses)



---

# SPECTRa

(Submission, Preservation, & Exposure of  
Chemistry Teaching and Research data)





# SPECTRa



- 18-month project partnership between  
Cambridge University Library (lead site)  
Cambridge University Chemistry Dept  
*and*  
Imperial College London - Chemistry Dept  
Imperial College London - Library
- in collaboration with ebank-UK
- funded by JISC (Joint Information Systems Committee) Digital Repositories Programme



# The problem



- machine-understandable data is needed for:
  - eScience
  - Semantic Web
  - re-analysis through informatics
- Open Data is not the same as Open Access
  - OA licences often don't address reuse and redistribution of data
  - publication destroys information
  - Open Data can co-exist with non-OA publications



# The problem (continued)



- publication of data requires extra work
- lack of infrastructure for handling Open Data
- many publishers are unconvinced about / antagonistic towards OA
- therefore many researchers are also unconvinced about / antagonistic towards OA
- lack of exemplars to illustrate benefits
- lack of practical tools to overcome obstacles



# SPECTRa's tasks



- survey of researchers' requirements in crystallography, computational chemistry, and synthetic chemistry
- development of customised Open Source tools as part of researchers' workflow to enable deposit of, and access to, Open Data using DSpace institutional repositories



# Data embargo



- chemists often need to hide research data from competitive research groups or for commercial reasons
- publication of chemical structures must be embargoed until the chemist
  - publishes work involving those structures, or
  - moves on to a different line of research
- researchers are unlikely to deposit their data without adequate recognition of concerns
- need for an escrow process to manage release of Open Data



# The SPECTRa workflow

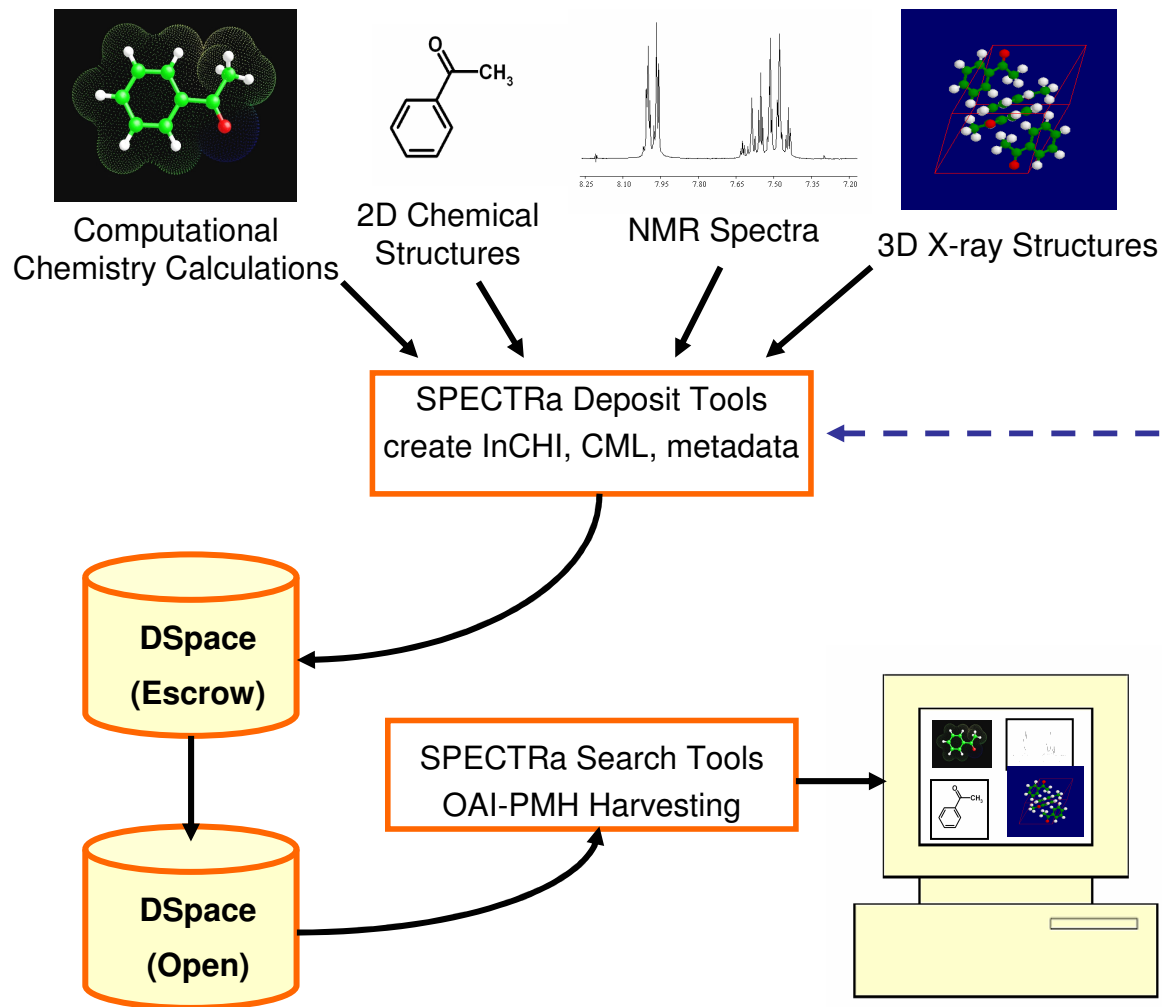


- capture selected data from chemistry workflows in open format (JCAMP, MOL, CIF)
- add context-specific metadata
- add persistent identifiers (InCHI)
- deposit in closed "escrow" repository
- manage the release of data into open Institutional Repository by agreement
- data can then be searched and harvested



# SPECTRa

(Submission, Preservation & Exposure of Chemistry Teaching & Research data)



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>
```



---

# Points to ponder





# Points to ponder...

---

- find out what your researchers really need
- every research community is different
- build solutions to the problems they identify
- be prepared to look beyond peer-reviewed papers
- consider both open- and closed-access collections
- consider federated repository structures
- demonstrate the added value a repository can provide in supporting their research

but

- what works for Cambridge may not work elsewhere...



---

# Thank you for listening

Peter Morgan  
pbm2@cam.ac.uk

DSpace@Cambridge: [www.dspace.cam.ac.uk](http://www.dspace.cam.ac.uk)

SPECTRa: [www.lib.cam.ac.uk/spectra/](http://www.lib.cam.ac.uk/spectra/)

