“If we built it, will they come?”

-Part Deux!

George Kozak
Digital Library and Information Technology (DLIT)
Cornell University
(gsk5@cornell.edu)
Background - the Initial Report

At the DSpace User Group Meeting held at the Open Repositories Conference 2007 (January 24, 2007), Matthew Connolly of Cornell University did a presentation entitled “If we build it, will they come?” based on a paper that he co-authored with Philip Davis: Institutional Repositories: Evaluating the reasons for non-use of Cornell University’s installation of DSpace (D-Lib Magazine, Volume 13, Issue 3/4, 2007).
The report stated that Cornell’s Institutional Repository was largely under populated and under used by its faculty.

Many of the collections in Cornell’s DSpace implementation were empty or contained very few items.

They postulated that the complex organization of the Communities and Collections discouraged contributions.
Background - the Initial Report

- They interviewed 11 diverse faculty members of the Cornell Faculty and found that there was little knowledge of and little motivation to use DSpace.

- They did a comparison of 7 institutions and found that the scenario described at Cornell University was not unique.

- The outlook for Cornell’s Institutional Repository looked bleak.
Digital Repositories at CUL (a Retrospective)

Development and Adaptation

- CDL (Cornell Digital Library) – RFC#1691 in 1994
  (architecture for storage and retrieval of digital representations for books, journals, etc)

- Early Digital Repositories using Dienst Protocol (Making of America, Historical Math Books)

- **Project Euclid** – DPubS (Digital Publishing System) - open-source software system designed to enable the organization, presentation, and delivery of scholarly journals, monographs, conference proceedings, and other common and evolving means of academic discourse.
Digital Repositories at CUL (a Retrospective)

Development and Adaptation

- Digital Preservation using aDORe – Archiving arXiv, math materials and LSDI (large scale digitization).

- Physics arXiv (Dr. Paul Ginsbaurg, Dr. Simeon Warner, Dr. Thorsten Schwander)

- DigitalCommons@ILR – powered by BePress (the Berkley Electronic Press)

- CUL Media Archive - digital asset management (DAM) interface being built on top of Fedora (to be open source)
In 2002, the Dean of the Faculty at Cornell had a vision:
- “to stimulate a fundamental reshaping and enhancement of the way research universities and their faculties function…” (Cooke, Final Project Report to the Atlantic Philanthropies, 2006, p.2)
- He wanted to create “…an economical vehicle for openly-shared access to formerly inaccessible, but intellectually-rich digital resources…” (ibid.)
Why DSpace was installed

- He envisioned that high quality information could be sustained by a stable economic model using DSpace.

- The Cornell Library was interested in building a knowledge base of print, digital, and other materials using selection criteria that reflected
  - the academic priorities of the University
  - significant research in all areas of study pursued at the University
  - current collection strengths.
Installation (Phase 1)

- Funding for the deployment and maintenance of DSpace at Cornell was initially provided by the Atlantic Foundation.

- Operational responsibility for Cornell’s DSpace implementation was assumed by the Library and funding for basal services was assured through FY 2008.
Installation (Phase 1)

First Community was Internet-First University Press to promote the Dean of the Faculty’s goal to provide open access and to free the faculty from the restrictions of traditional Publishers.

193 communities and 196 collections were created that reflected the structure of the University to provide a framework for Faculty and Staff to deposit items.
Installation (Phase 1)

Outreach

– The Graduate Office was approached to provide a mechanism to allow for online theses and dissertations.

– Graduate students were offered print-on-demand services with Cornell Business Solutions if they used DSpace.

– The Dean of the Faculty began a campaign of convincing faculty and departments to submit content to our DSpace Repository.
Installation (Phase 1)

- We made some code enhancements:
  - Quick Submit Program (one page form with license and verification screens)
  - View counter for items

- We offered to help load materials and provide metadata services

- We offered to provide Digitization services as well
Submit: Describe Your Item

Please fill in the requested information about your submission below. Required fields are marked with an asterisk (*). In most browsers, you can use the tab key to move the cursor to the next input box or button, to save you having to use the mouse each time. (More Help...)

Please enter the name of a file on your local hard drive corresponding to your item. If you click "Browse...", a new window will appear in which you can locate and select the file on your local hard drive. (More Help...)

Netscape users please note: By default, the window brought up by clicking "Browse..." will only display files of type HTML. If the file you are uploading isn't an HTML file, you will need to select the option to display files of other types. Instructions for Netscape users are available.

Please also note that the DSpace system is able to preserve the content of certain types of files better than other types. Information about file types and levels of support for each are available.

Enter the names of the authors of this item below.

* Authors
  e.g. Smith
  e.g. Donald Jr

* Title

Add More Authors
Show Workarea
Installation (Phase 1)

Other selling points to Faculty and Staff:

- guaranteed open access

- Google harvesting to make works more publicly available

- Handles would provide permanent URL’s

- guaranteed storage and web access.
Response to using DSpace from Faculty and Staff was generally poor.

Library suffered some unforeseen staff changes in Library Management that were involved in leading DSpace effort.

Retirement of the Dean of the Faculty left the project without direction.
By January of 2007, when Matthew Connolly and Philip Davis originally wrote their paper, Cornell’s DSpace repository consisted of a mere 2,646 items with 57 empty collections.

To quote Dorothea Salo in *Inkeeper at the Roach Motel*, [http://digital.library.wisc.edu/1793/22088](http://digital.library.wisc.edu/1793/22088), it was obvious that our Institutional Repository had to “adapt or die”
Installation (Phase 2)

- Position of Associate University Librarian for Scholarly Communication and Collections was created and Interim head was chosen.

- AUL set up an Institutional Repository (IR) team of librarians and technical staff founded

- Work was done to create goals and directions
Installation (Phase 2)

Several upgrades were done:
- DSpace 1.2.2 to DSpace 1.3.2 to DSpace 1.4.2
- PostGreSQL 7.3 to 8.3

Renamed Repository from DSpace@Cornell to eCommons@Cornell

Reorganized Collections and Communities and eliminated empty ones.

New User Interface design using CSS and some jsp changes.
Welcome to the Cornell University DSpace Digital Repository!

This repository is open for anyone at Cornell University as a place to capture, store, index, preserve and redistribute Cornell faculty, staff, student or organizational research materials in digital formats.

For information on protecting your rights to your intellectual property Click Here.

If you would like to be included in our repository, please contact us at dspace-admin-lib@cornell.edu

Search

Enter some text in the box below to search DSpace.

Communities in DSpace

Select a community to browse its collections.

Agriculture & Life Sciences
American Indian Program
Archaeology Program
New Design
Installation (Phase 2)

- New name and design generated positive buzz about the collection

- More requests for inputting items materialized

- Different avenues pursued for content
Installation (Phase 2)

- Different sources for content
  - Division of Fluid Dynamics of the American Physical Society videos for arXiv.org
Installation (Phase 2)

- Different sources for content
  - Cornell Cooperative Extension publications
Installation (Phase 2)

Different sources for content

- Technical Reports from NCSTRL Repository and from Operations Research and Information Engineering
Installation (Phase 2)

Different sources for content

– Cornell Alumni Magazine and Cornell Chronicle

historical content
Installation (Phase 2)

- Different sources for content
  - Conference proceedings
Installation (Phase 2)

- Different sources for content
  - Oral Histories
Installation (Phase 2)

Different sources for content
- Web Site Archives
Installation (Phase 2)

- Different sources for content
  - Faculty Works
    - Nimat Hafez Barazangi Scholarly Works
Installation (Phase 2)

- Different sources for content
  - Faculty works
    - Richard Baer on Religion, Education, and the Public Square
Installation (Phase 2)

- Different sources for content
  - Faculty works
    - Billie Jean Isbell Andean Collection
Installation (Phase 2)

- Focus on providing access to materials that are “losing their home”, or previously unavailable on the web.

- Working with Graduate School to mandate electronic submissions of Theses and Dissertations

- Other Avenues such as
  - harvesting our Domain
  - Works seeking publisher
  - Local community works of importance (The Cayuga Lake Watershed Network)
Installation (Phase 2)

(The Cayuga Lake Watershed Network)
Results

- Size of repository tripled in one year from 2,646 items in 01/07 to 8,193 in 01/08.

- As of 4/16/08: 9,436

- Many items added through batch loads (i.e. Tech Reports)
## Results

### Analysis of Items in eCommons@Cornell

<table>
<thead>
<tr>
<th>Year</th>
<th>Theses/Dissertations</th>
<th>Internet-First Press</th>
<th>Library Papers</th>
<th>Research</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>2002-4</td>
<td>72</td>
<td>58</td>
<td>9</td>
<td>35</td>
<td>174</td>
</tr>
<tr>
<td>2005</td>
<td>108</td>
<td>153</td>
<td>82</td>
<td>1345*</td>
<td>1688</td>
</tr>
<tr>
<td>2006</td>
<td>236</td>
<td>111</td>
<td>347</td>
<td>1498**</td>
<td>2192</td>
</tr>
<tr>
<td>2007</td>
<td>283</td>
<td>43</td>
<td>96</td>
<td>3717***</td>
<td>4139</td>
</tr>
<tr>
<td>TOTALS</td>
<td>699</td>
<td>365</td>
<td>534</td>
<td>6595</td>
<td>8193</td>
</tr>
</tbody>
</table>

* 1078 images batch loaded for Cornell Plantations
** 966 NYS Ag Bulletins and 98 Alumni News batch loaded
*** 2932 Engineering Technical Reports, 139 Vet Lab Reports (139) and 145 Cornell Pubs batch loaded
Results

- Number of hits increased by 4 times (recorded over a 6 month period) from 738,624 in 2006 to 3,331,983 in 2007.

- About 70% of hits from Robots

- Hits increased as content increased

- Non-Robot hits were around 130,000 to 150,000, but showed a gradual increase
## Results

### Hits in 2007

<table>
<thead>
<tr>
<th>Month (2007)</th>
<th>Hits (robots included)</th>
<th>Hits (w/o bots)</th>
</tr>
</thead>
<tbody>
<tr>
<td>June</td>
<td>158,143</td>
<td>134,272</td>
</tr>
<tr>
<td>July</td>
<td>271,952</td>
<td>130,736</td>
</tr>
<tr>
<td>August</td>
<td>258,947</td>
<td>146,196</td>
</tr>
<tr>
<td>September</td>
<td>295,181</td>
<td>131,572</td>
</tr>
<tr>
<td>October</td>
<td>382,523</td>
<td>146,869</td>
</tr>
<tr>
<td>November</td>
<td>385,253</td>
<td>134,954</td>
</tr>
<tr>
<td>December</td>
<td>1,579,984</td>
<td>153,727</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>3,331,983</strong></td>
<td><strong>978,326</strong></td>
</tr>
</tbody>
</table>

* In comparison during the same period in **2005: 349,779 (w/bots)**

in **2006: 738,624 (w/bots)**
Downloads increased by 6 times (recorded over a 6 month period) from 277,251 in 2006 to 1,785,477 in 2007.

About 40% of downloads from Robots.
## Results

### Downloads

<table>
<thead>
<tr>
<th>Year</th>
<th>Items</th>
<th>Downloads (w/bots)</th>
<th>Downloads (w/o bots)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2005</strong></td>
<td>1862</td>
<td>303,802</td>
<td>-</td>
</tr>
<tr>
<td><strong>2006</strong></td>
<td>4054</td>
<td>277,251</td>
<td>-</td>
</tr>
<tr>
<td><strong>2007</strong></td>
<td>8193</td>
<td>1,785,477</td>
<td>1,069,912</td>
</tr>
<tr>
<td><strong>2008 (as of 4/15/08)</strong></td>
<td>9436</td>
<td>359,510</td>
<td>215,706</td>
</tr>
</tbody>
</table>
Results

- 65,848 unique IP’s downloaded items from our repository
- Majority of IP’s from outside of our University.
Results

Map of Unique IP’s Accessing eCommons
Results

Map of Unique IP's Accessing eCommons
Results

Greatest Hits through 2007

1. How Do I Do This in ArcGIS/Manifold?: Illustrating Classic GIS Tasks (Lembo, Arthur J. Jr.) [20,342]

2. Nonlinear dynamics and chaos: Lab demonstrations (Strogatz, Steven H.) [20,150]


4. "I Can Do That!" Hans Bethe's First 60 Years at Cornell (Rose Film Inc.) [13,971]

5. Lecture Notes on Fracture Mechanics (Zehnder, Alan) [13,630]

Results

Greatest Hits now

1. Lecture Notes on Fracture Mechanics (Zehnder, Alan) [1,928] *
2. Nonlinear dynamics and chaos: Lab demonstrations (Strogatz, Steven H.) [1,610] *
3. Water Resources Systems Planning and Management: An Introduction to Methods, Models and Applications (Loucks, Daniel P. et al) [1,192]
4. How Do I Do This in ArcGIS/Manifold?: Illustrating Classic GIS Tasks (Lembo, Arthur J. Jr.) [970] *
5. Copyright and Commerce: The DMCA, Trusted Systems, and the Stabilization of Distribution (Gillespie, Tarleton) [775]
6. Basic Fly Tying (Howard, Ronald A. Jr.) [762]
Challenges

Cornell University Library maintains some 10+ digital repositories (DRs) representing a wide range of:
- Architectures
- subject domains
- content types
- Sizes
- Distribution
- Models
- collaboration types
- interoperability standards
- key stakeholders
- goals.

Funding is a chief concern.
Challenges

- Information Policy issues: ownership, confidentiality, privacy, and copyright
- Appraisal and selection of content (content for Preservation?)
- Storage and Network Transmission Costs
- How do we measure the success of our DRs?
What does the future hold?

- Need for a full-time funded staff dedicated to outreach and recruitment for eCommons.

- Mandates such as NIH’s requirement for open access to electronic versions of any peer-reviewed manuscripts arising from NIH funding.

- Harvard proposal to deposit Faculty papers in an open-access repository.
Conclusion

In her article *Innkeeper at the Roach Motel* ([http://digital.library.wisc.edu/1793/22088](http://digital.library.wisc.edu/1793/22088)), Dorothea Salo states that the "’build it and they will come’ proposition has been decisively proven wrong".

Advantages such as handles, storage and access have not attracted faculty participants.
Conclusion

- We are looking at more proactive methods to add to our Repository.

- Faculty and Staff are now seeking us as a place to store their materials.

- Our IR is showing some promising growth as we reposition ourselves with a new look and a new direction.