# "If we built it, will they come?"
## -*Part Deux!*

George Kozak
Digital Library and Information Technology (DLIT)
Cornell University
([gsk5@cornell.edu](mailto: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: <u>Institutional Repositories: Evaluating the reasons for non-use of Cornell University's installation of DSpace</u> (*D-Lib Magazine*, Volume 13, Issue 3/4, 2007).

# Background - the Initial Report

- 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)

# Why DSpace was installed

- 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, <u>Final Project Report to the Atlantic Philanthropies</u>, 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.
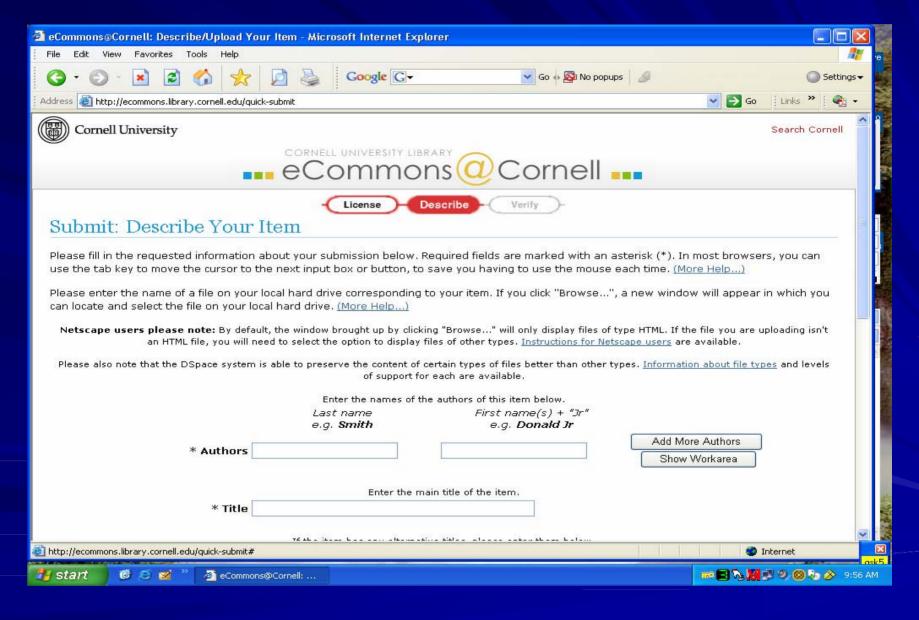
# Installation (Phase 1)

- **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 Respository.
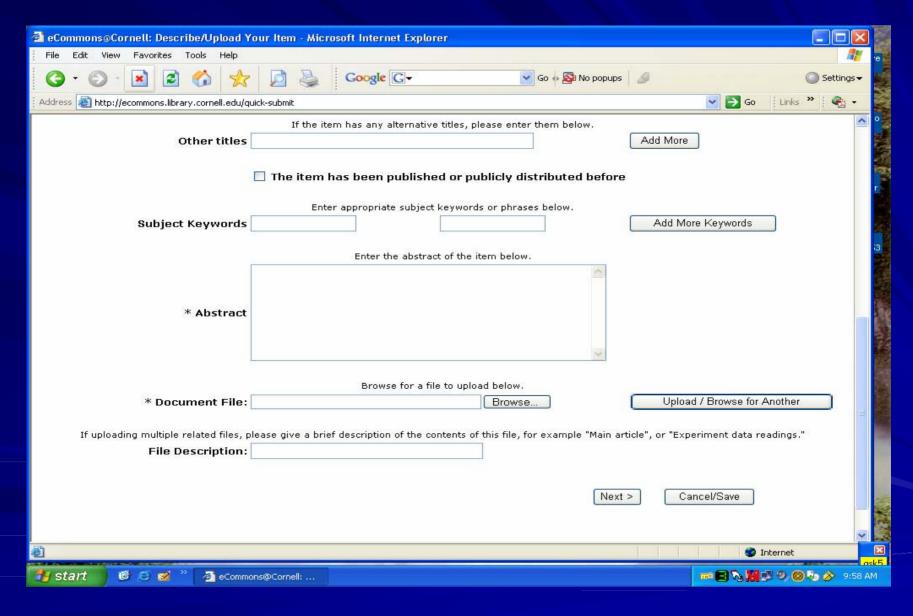
# Installation (Phase 1)

- 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

# Quick Submit

# Quick Submit

# 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.

# Installation (Phase 1)

- 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.

# Installation (Phase 1)

- 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), it was obvious that our Institutional Repository had to "adapt or die"
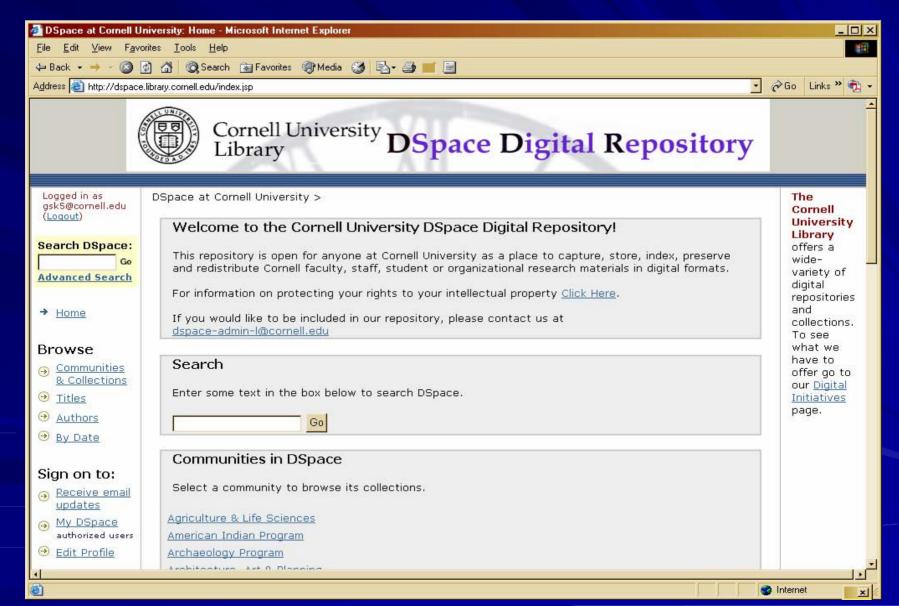
# Installation (Phase 2)

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

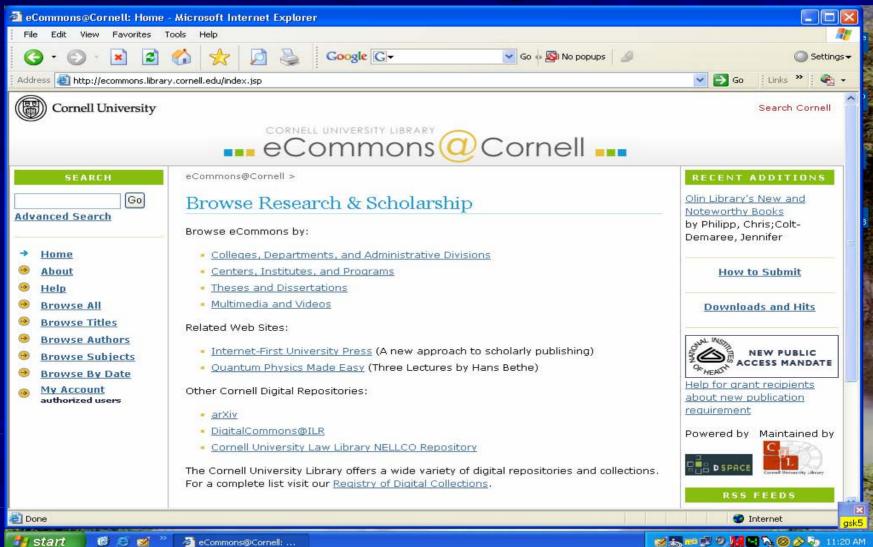- AUL set up an Institutional Repository (IR) team of librarians and technical staff founded

- 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.

# Original Design

# 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)
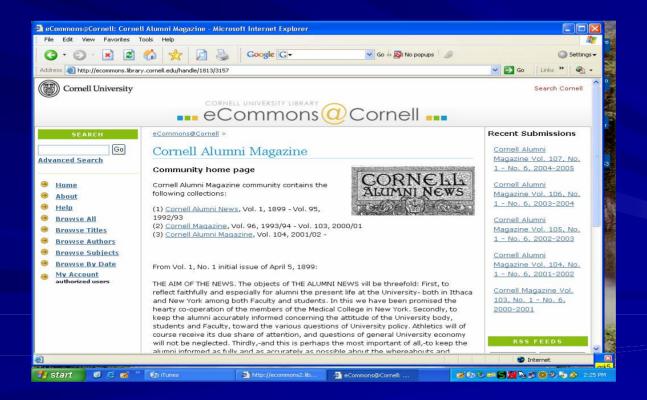
■ **Different sources for content**
 – Cornell Cooperative Extension publications

# Installation (Phase 2)

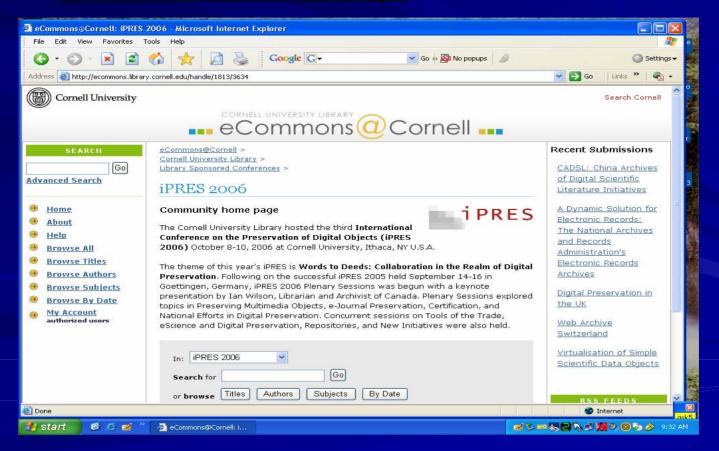■ **Different sources for content**

  – 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 (ie. Tech Reports)

# Results

**Analysis of Items in eCommons@Cornell**

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

* 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 45 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

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

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

in **2006: 738,624  (w/bots)**

# Results

- Downloads increased by 6 times (recorded over a 6 month period) from 277,251 in 2006 to 1,785,477 in 2007.
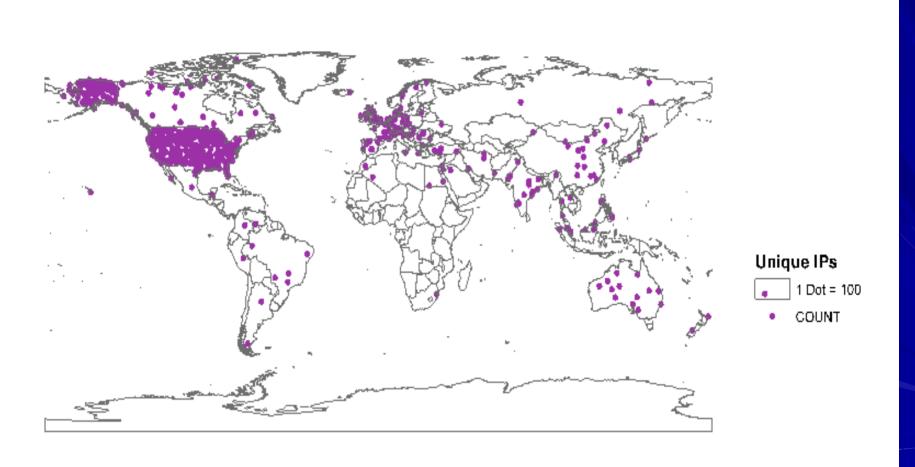
- About 40% of downloads from Robots.

# Results

## Downloads

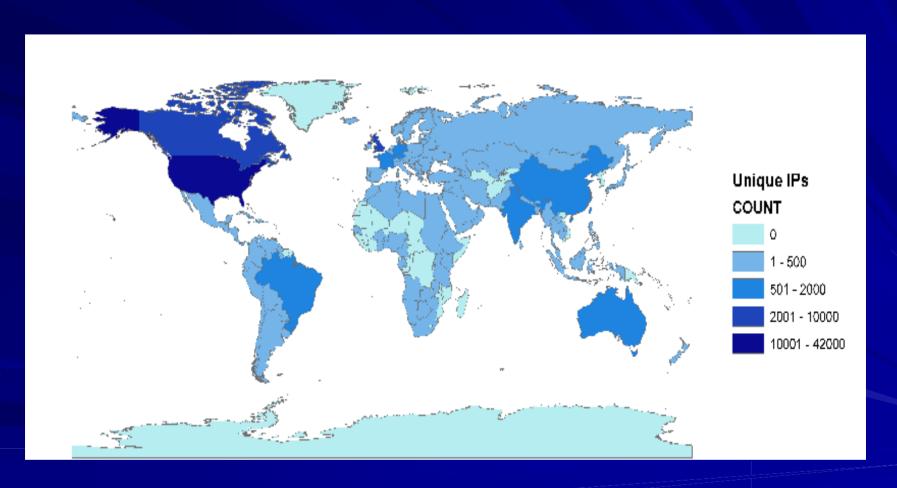| Year | Items | | | Downloads (w/bots) | Downloads (w/o bots) |
|------|-------|---|---|--------------------|----------------------|
| **2005** | 1862 | | | 303,802 | - |
| **2006** | 4054 | | | 277,251 | - |
| **2007** | 8193 | | | 1,785,477 | 1,069,912 |
| **2008** (as of 4/15/08) | 9436 | | | 359,510 | 215,706 |

# 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]

  - 3. Report of the CUL Task Force on Open Access Publishing Presented to the Cornell University Library Management Team August 9, 2004. (Davis, Philip M., et al) [13,988]

  - 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]

  - 6. New Concepts of Cattle Growth (Berg, Roy T.; Butterfield, Rex M.) [11,857]

# Results

- Greatest Hits now

  - 1. <u>Lecture Notes on Fracture Mechanics</u> (Zehnder, Alan) [1,928] *

  - 2. <u>Nonlinear dynamics and chaos: Lab demonstrations</u> (Strogatz, Steven H.) [1,610]  *

  - 3. <u>Water Resources Systems Planning and Management: An Introduction to Methods, Models and Applications</u> (Loucks, Daniel P. et al) [1,192]

  - 4. <u>How Do I Do This in ArcGIS/Manifold?: Illustrating Classic GIS Tasks</u> (Lembo, Arthur J. Jr.) [970] *

  - 5. <u>Copyright and Commerce: The DMCA, Trusted Systems, and the Stabilization of Distribution</u> (Gillespie, Tarleton) [775]

  - 6. <u>Basic Fly Tying</u> (Howard, Ronald A. Jr.) [762]

# Challenges

- Cornell University Library maintains some 10+ digital repositories (DR)

- (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 <u>Innkeeper at the Roach Motel</u> (<u>http://digital.library.wisc.edu/1793/22088</u>), 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.