Appendix F

Site visit questionnaire and responses

Questionnaire form used for site visits

Site Visits:

Brandeis University

Harvard University Physics Library

Northeast Massachusetts Regional Library Systems, Danvers, MA

Tufts University

University of Rochester

Questionnaire for repository site visits, Sabbatical research Spring 2005

Location:
Contact name, email, and phone:
Date(s) visited:

1. Please describe your rationale for pursuing an institutional repository project. Please include your vision of what your IR will be, how you define an IR and what you expect to get out of it. Please describe any evaluation of IR software you have conducted and conclusions reached.

2. What constituencies of the university were involved in the project? Which one(s) was the driving force behind the project? Has this proven to be a good choice? Why or why not? Were there constituencies who, in hind sight, ought to have been involved in the project that weren't there? Possibilities might include institutes or museums.

3. Please describe the resources (equipment, staffing, and expertise) that were needed to implement the repository. Are there additional resources that have been added to the initial resources? Please describe in detail. For example, where are the staffing resources coming from, what is their expertise level, has additional training been needed, what partnerships have been developed, is the workload centralized or distributed, who fills the role of content providers and managers?

4. What major issues needed to be addressed prior to or in conjunction with the startup of the repository? Some examples might include populating the repository with content, copyright, intellectual property rights, integration with integrated library system, course management system, harvesting by Google.

5. Have you developed plans to ensure the long-term sustainability and growth of the repository? Please address finances, maintenance, and staffing issues for the original project as well as current and future growth. Are you willing to share these plans? Are there aspects that have not yet been addressed in your plans?

6. If you were to do this over again, knowing what you know now, how would you change the conception, implementation, and ongoing operation of the repository? Are there resources that could have been made available that could have made the project go more smoothly?

7. Please describe your ongoing development path and any additional resources it will take to get you there. What are some of the main issues and trends to be on the lookout for in the next 3-5 years? What do you think your IR will look like in 5 years?

8. Other comments

Details of site visit

Hardware:

Operating System:
Software:

Technical expertise needed:

Common features of a repository generally include the following elements

- content in digital form in a wide variety of types (text, audio, video, images, data sets)

- community focus, which determines what is included in the repository. The community members are the authors and copyright owners of the deposited content.

- institutional support, requiring collaboration across an organization, including long-term financial support to ensure that the content is preserved and maintained

- durable, permanent content.

- access to content by a broad audience, a community-shared alternative to local storage of content, fosters serendipitous discovery across disciplines.

Does your repository have all these features? Which ones have proven to be the most critical for success? Does the repository have additional features to add to this list?

Core functions which are essential for a repository

- material submission, some way for the author to deposit material, provide for editing to assure quality of content, conversion to archival format such as PDF

- metadata application, such as author, title and descriptive information and administrative data such as date and time of submission

- access control, or digital rights management, to provide for controlled access to the repository content. Even if the entire community has access to the content, there needs to be a way to restrict the ability to add, delete, edit, and approve content.

- discovery support, usually a search engine that supports browsing and full-text searching of the content

- distribution and dissemination of content to enable display and download capabilities

- preservation, some mechanism(s) for the content to be preserved and retrievable over time, including a persistent documentation identification system.

Are there additional core functions that need to be added to this list?

Questionnaire for repository site visits, Sabbatical research Spring 2005

Location: Brandeis University
Contact name, email, and phone: Susan C. Pyzynski, pyzynski@brandeis.edu, 781-736-4697
Date(s) visited: phone interview on June 24, 2005

1. Please describe your rationale for pursuing an institutional repository project. Please include your vision of what your IR will be, how you define an IR and what you expect to get out of it. Please describe any evaluation of IR software you have conducted and conclusions reached.

   Brandeis had 2 projects of images: lithographs and a digital slide library (ongoing). They are doing one class per semester for faculty use and wanted something that would support more than just images. They wanted something that would work with other Ex Libris products so they chose DigiTool, which they are in the process of upgrading this summer. DigiTool is searchable by MetaLib patron directory services so allows for single sign on. They were the first DigiTool site to be harvested by OAIster and Google – an IMLS grant.

2. What constituencies of the university were involved in the project? Which one(s) was the driving force behind the project? Has this proven to be a good choice? Why or why not? Were there constituencies who, in hind sight, ought to have been involved in the project that weren't there? Possibilities might include institutes or museums.

   Library initiated. Worked with the Anthropology Dept.

3. Please describe the resources (equipment, staffing, and expertise) that were needed to implement the repository. Are there additional resources that have been added to the initial resources? Please describe in detail. For example, where are the staffing resources coming from, what is their expertise level, has additional training been needed, what partnerships have been developed, is the workload centralized or distributed, who fills the role of content providers and managers?

   They are using Dublin Core metadata.
   With DigiTool, there is a web client so can access from anywhere.
   In Version 3, staff use a GUI client to connect to database.
   They use the ALEPH authority files.

4. What major issues needed to be addressed prior to or in conjunction with the startup of the repository? Some examples might include populating the repository with content, copyright, intellectual property rights, integration with integrated library system, course management system, harvesting by Google.

5. Have you developed plans to ensure the long-term sustainability and growth of the repository? Please address finances, maintenance, and staffing issues for the

original project as well as current and future growth. Are you willing to share these plans? Are there aspects that have not yet been addressed in your plans?

6. If you were to do this over again, knowing what you know now, how would you change the conception, implementation, and ongoing operation of the repository? Are there resources that could have been made available that could have made the project go more smoothly?

   Yes, they would make the same choice again. Previously, wouldn't have said so.
   They don't have the infrastructure needed to go with open source.
   DigiTool is standards-based so they could pull out into open source if they wanted to do that at a later time.

7. Please describe your ongoing development path and any additional resources it will take to get you there. What are some of the main issues and trends to be on the lookout for in the next 3-5 years? What do you think your IR will look like in 5 years?

8. Other comments

   Html pages were still a problem when migrated to ALEPH 16.
   University-wide application
   Time – less time than traditional
   Expertise – low fewer people involved, not dealing with acquisition metadata. Slide librarian does own cataloging. Graduate students do regular work. Need understanding of technology for digital collections.
   Need to develop policies for collections
   There's reluctance among staff to do more, change to digital. There is a cross campus digital assets committee.

Details of site visit

Hardware:
Operating System:
Software: DigiTool v3.0
Technical expertise needed:

Common features of a repository generally include the following elements

- content in digital form in a wide variety of types (text, audio, video, images, data sets)
- community focus, which determines what is included in the repository. The community members are the authors and copyright owners of the deposited content.
- institutional support, requiring collaboration across an organization, including long-term financial support to ensure that the content is preserved and maintained
- durable, permanent content.
- access to content by a broad audience, a community-shared alternative to local storage of content, fosters serendipitous discovery across disciplines.

Does your repository have all these features? Which ones have proven to be the most critical for success? Does the repository have additional features to add to this list?

Core functions which are essential for a repository

- material submission, some way for the author to deposit material, provide for editing to assure quality of content, conversion to archival format such as PDF
- metadata application, such as author, title and descriptive information and administrative data such as date and time of submission
- access control, or digital rights management, to provide for controlled access to the repository content. Even if the entire community has access to the content, there needs to be a way to restrict the ability to add, delete, edit, and approve content.
- discovery support, usually a search engine that supports browsing and full-text searching of the content
- distribution and dissemination of content to enable display and download capabilities
- preservation, some mechanism(s) for the content to be preserved and retrievable over time, including a persistent documentation identification system.

Are there additional core functions that need to be added to this list?

Questionnaire for repository site visits, Sabbatical research Spring 2005

Location: **Harvard University** Physics Research Library using DSpace.
Contact name, email, and phone: Michael Leach, leach@physics.harvard.edu, 1-617-495-9038

Date(s) visited: June 21, 2005

1. Please describe your rationale for pursuing an institutional repository project. Please include your vision of what your IR will be, how you define an IR and what you expect to get out of it. Please describe any evaluation of IR software you have conducted and conclusions reached.

At Harvard, the Science Libraries are within the departments so they get questions from the faculty. 18 months ago, the science librarians got together to decide which software, spent 2-3 months reviewing, meeting, and discussing. Not many options back then. They decided on DSpace. He recommends looking at the PALS report.

Commercial solutions offered interesting services. However they decided they couldn't afford them.

Important elements of an IR: community structure for departments, research groups, etc; OAI protocol for harvesting; extended Dublin Core (can't do everything within DSpace); core set of objects to start with (research articles, theses, journals, datasets, video, learning objects)

Their intent is to provide a service that supports the majority.

Dublin Core is useless with datasets (1-2 terabytes of data just in the dataset. Working on new initiative with MIT to address this need, a module-like system.

2 What constituencies of the university were involved in the project? Which one(s) was the driving force behind the project? Has this proven to be a good choice? Why or why not? Were there constituencies who, in hind sight, ought to have been involved in the project that weren't there? Possibilities might include institutes or museums.

Science Libraries and faculty. Five libraries are participating in the pilot (done by this fall): Physics, Organismic, Molecular, Cell Biologies, Astrophysics, Geosciences.

There are two research groups in each department participating. They are gathering sets of information, creating community spaces.

Now discussing publicity, linking from their pages to the IR.

When asked a followup question about dealing with faculty who say they already have the functionality of an IR covered, ask them the following:

Is it searchable, through Google and Google Scholar? Are their materials harvested by them?
What happens if / when you retire, move to another institution, etc? If they go to another site, they could have two IRs seamlessly connected.

Describe to faculty the higher visibility that results from participating in the IR. Stress the impact and

citation factors.

3. Please describe the resources (equipment, staffing, and expertise) that were needed to implement the repository. Are there additional resources that have been added to the initial resources? Please describe in detail. For example, where are the staffing resources coming from, what is their expertise level, has additional training been needed, what partnerships have been developed, is the workload centralized or distributed, who fills the role of content providers and managers?

The Science Libraries are tapped into strong IT departments within each department, not the library IT department. Michael has some IT skills, especially UNIX.

The libraries expect jobs to change in the next 3-4 years and will re-allocate staffing as this happens.

Currently they have ½ fte from various libraries working on the IR. Most are from Technical Services, one from public services.

They are analyzing workflow, benchmarks, timing processes etc with the expectation to streamline workflows.

4 What major issues needed to be addressed prior to or in conjunction with the startup of the repository? Some examples might include populating the repository with content, copyright, intellectual property rights, integration with integrated library system, course management system, harvesting by Google.

One problem they encountered: Harvard's General Counsel office wouldn't allow a click through license. They still need to have faculty, etc sign a piece of paper stating that they have the rights to the material they are depositing. This includes rights to images, graphs, etc

Content is another issue: Project ROMEO – green and gold publishers. Can take these articles and put up directly into the IR. No need to get permissions.

Learning Objects are a different topic.

Address need for a logo for marketing purposes – a complex issue. Needed to deal with Publications Office. Branding

5 Have you developed plans to ensure the long-term sustainability and growth of the repository? Please address finances, maintenance, and staffing issues for the original project as well as current and future growth. Are you willing to share these plans? Are there aspects that have not yet been addressed in your plans?

They have a Steering Committee, which is comprised of the 5 science library directors, the e resources librarian (licensing expertise), the social sciences data set director of their IT unit, humanities / social sciences CJK librarian. SC is developing best practices, conducting training sessions such as on Dublin Core metadata.

The IR is a library-driven model. The IR will be sustainable within "our realm". The libraries are providing the metadata creation. Don't expect the faculty to do this work.

Libraries work with their own authorities.

Semantic web capability

Now looking at workflows and efficiencies. Refocusing their work to the new framework. Lessen some things, start new things.

They will need to broaden out to the entire campus.

6. If you were to do this over again, knowing what you know now, how would you change the conception, implementation, and ongoing operation of the repository? Are there resources that could have been made available that could have made the project go more smoothly?

Have a longer test period of the technology itself, such as Apache, Tomcat, etc. Tomcat is key and is the most complicated part.

Problem: Wish they'd known more about the handle system within DSpace. The servers are behind firewalls. Handle.com does not work well with firewalls.

Upgraded to the new beta version of the software. Earlier versions had allowed through various codes of html. No longer available with new version but MIT will fix this.

They couldn't figure out how to remove an object from the repository, eventually did. Now this procedure is a best practice.

They are adding to the documentation – needs lots more work.

7. Please describe your ongoing development path and any additional resources it will take to get you there. What are some of the main issues and trends to be on the lookout for in the next 3-5 years? What do you think your IR will look like in 5 years?

They predict that journals will be completely electronic within 3 years. Those staff will be retrained from binding, claiming, etc and relocate to support the IR.

He expects that within 4-5 years, they will need to relocate the repository to the Libraries IT Department.

LOCKSS model
Virtual journals
Other collections
Niche communities over multiple IRs

8. Other comments

Michael has developed a canned presentation that other librarians use.

They are now building relationships around the campus:
I Commons group which provides course support
Provost's Office
Give presentations to IT and the Librarian Community

They anticipate multiple IRs at Harvard. Need to determine how to link together
Libraries are the best equipped to do this work – Librarians are curators of all sorts of information.

Details of site visit

Hardware:

Operating System:
Software:

Technical expertise needed:

Common features of a repository generally include the following elements

- content in digital form in a wide variety of types (text, audio, video, images, data sets)

- community focus, which determines what is included in the repository. The community members are the authors and copyright owners of the deposited content.

- institutional support, requiring collaboration across an organization, including long-term financial support to ensure that the content is preserved and maintained

- durable, permanent content.

- access to content by a broad audience, a community-shared alternative to local storage of content, fosters serendipitous discovery across disciplines.

Does your repository have all these features? Which ones have proven to be the most critical for success? Does the repository have additional features to add to this list?

Core functions which are essential for a repository

- material submission, some way for the author to deposit material, provide for editing to assure quality of content, conversion to archival format such as PDF

- metadata application, such as author, title and descriptive information and administrative data such as date and time of submission

- access control, or digital rights management, to provide for controlled access to the repository content. Even if the entire community has access to the content, there needs to be a way to restrict the ability to add, delete, edit, and approve content.

- discovery support, usually a search engine that supports browsing and full-text searching of the content

- distribution and dissemination of content to enable display and download capabilities

- preservation, some mechanism(s) for the content to be preserved and retrievable over time, including a persistent documentation identification system.

Are there additional core functions that need to be added to this list?

Questionnaire for repository site visits

Location: Danvers, MA **Northeast Massachusetts Regional Library System**, as an example of consortial CONTENTdm site,

Northeast Massachusetts Digital Library

Contact name, email, and phone:

Greg Pronevitz, Regional Administrator, greg@nmrls.org, 1-888-326-7772 x15

Scott Kehoe, Consultant/Trainer, scott@nmrls.org, x1-888-326-7772 x16

Date(s) visited: June 21, 2005

1. Please describe your rationale for pursuing an institutional repository project. Please include your vision of what your IR will be, how you define an IR and what you expect to get out of it. Please describe any evaluation of IR software you have conducted and conclusions reached.

Seven years ago, historical societies, special libraries, academic libraries – vision to get together to do things. NMRLS provided the stimulus and the funding (through Massachusetts Board of Library Commissioners). Bought large 4 mb scanner. Greg (administrator and coordinator), Scott (technical support and trainer), and Susan (talented web designer) developed a scanning lab for the region, then started a 3 year grant with the Essex National Heritage Area pilot project, ECHOES (Essex County Heritage Online Exhibits) using maritime, industrial and other categories of materials. Scanning lab at NMRLS –lightly used, possibly due to large time commitment to learn equipment and software.

Weren't a lot of choices of software available when they started looking 3 years ago.

A primary consideration was that CONTENTdm offered a hosted version. They don't have a lot of bandwidth to serve up images, also don't have a large server environment to host such a service. CONTENTdm had scaled pricing based on your content.

Cost and hosted version were the major considerations.

DiMeMa was the third party that developed the software and they'll be taking back the hosting of the sites. Greg's impression is that this initiative has not been as successful as OCLC had hoped. OCLC is giving back to DiMeMa. Lots of upgrades. Nice improvements to the software.
These projects are very time consuming and demanding. Takes a talented person, a trained tech person, art student – design with database relationship expertise.
This project is not the main focus of NMRLS.
They received an LSTA grant – got a vendor to do scanning and metadata creation, Backstage Library Works scanning services. It's been great. However since Backstage has been hired, they've become very popular so work timeline has slipped.

Important work – coordinating with vendors, with libraries.

The original plan was to install client software at some of the larger sites. However, it's not practical for most libraries. One library may do this soon but time is a large issue. They gave one of their CONTENTdm licenses to Backstage so they could input content directly. With CONTENTdm, you have a template. Can change input form names. Dublin Core metadata.

Established a small committee from the membership (NOBLE rep Ron Gagnon, NELINET rep Amy Benson) to develop standards for the NMDL site. Public relations campaign to all Essex Co. Done programs on scanning and preservation. Developed contacts outside the standard membership such as town officials. When working with Peabody museums, discovered they use Past Perfect, an all-in-one museum software package. Includes records management features, forms for digitizing permissions, connects images with metadata, functionality for donor lists, etc. Has a web design built into it. Endicott College uses it. Has a Dublin Core to MARC output feature, image output. Peabody Institute has used this. They have heard that once a collection has a large size, it becomes unwieldy to use.

As a model for features, etc, it is excellent to look at and talk with vendors such as OCLC, ExLibris, etc
Currently, their version of CONTENTdm doesn't allow zoom in/out.
Ledger pages – lot of playing with contrast, size, trying to make large enough to be readable but not too large for download. Didn't OCR so need adequate metadata.
Pixel size, file size important considerations for public use.
For those who want to be hosted somewhere, the file size, etc is a consideration
Images – full size would be 1100 pixels on long side, 270 kb maximum file size
Screens normally 1024 x 624 – keep this in mind

Having a good vendor helps a lot. They know questions to ask, what works, what doesn't, etc
Vendor sent examples of photos before and after – here's scan before, here's scan after sharpening, contrast, etc. Preservation issues. However, want materials to look good, be usable
They have the original digital masters, scan in TIFF, not online. The working files are jpeg for sharpening, etc.

Backups of images: CD is what people want. Some are going to DVD but no archival format with DVDs. Currently there are too many types.

Demo of NMDL:

- town images of Lawrence and Gloucester. Used Simmons Grad students for this work. Posted the jobs on the Simmons job bulletin board. Catch is time and money
- ECHOES
  Lawrence – buttons, plates, images
  Thumbnail and large images – feature of CONTENTdm
  Diary – took digital photo – jpeg, not OCR'd so not accessible. Backstage will OCR. For printed page - $.99 per page. For handwritten page - $8.99 per page
  Topics: Maritime, Industrial, Early Settlement, etc. Have titles and thumbnails on page. Or could just display titles of images.

Can have side by side views
Every word is keyword searchable
Can create own favorites, links, etc

Metadata (only took information from the back of pictures)

Title

Subject
Type of material
Copyright holder

Backstage scanned in front and back of photo and had to hand key names of people. Backstage knew the technical aspects of the software so made it much easier to get the work done.

Ipswich Public Library: Antiquarian Papers

Used Paint Shop Pro, or could use Photoshop, and ran in batches

Need to consider: page size for printing / downloadability, readability

Banded each image with ownership

CONTENTdm also allows watermarking

2. What constituencies of the university were involved in the project? Which one(s) was the driving force behind the project? Has this proven to be a good choice? Why or why not? Were there constituencies who, in hind sight, ought to have been involved in the project that weren't there? Possibilities might include institutes or museums.

The CONTENTdm choice was an excellent one. Interest from the membership was very strong. The website gets lots of hits. However it needs to be part of something bigger. Amy Benson from NELINET wrote a technical white paper for Boston Public Library. They recommend using Amy as a consultant when starting up a digital library.

Driving force behind project: NMRLS membership and NMRLS staff. Membership was eager to jump on the project and support it.

Collections need a lot more added value. Excellent example: project main street. Narrative history built in with images. Has more of an exhibits feel. NMRLS has lots of images but not the context. Need partners to develop this, possibly Tsongus Center.

3. Please describe the resources (equipment, staffing, and expertise) that were needed to implement the repository. Are there additional resources that have been added to the initial resources? Please describe in detail. For example, where are the staffing resources coming from, what is their expertise level, has additional training been needed, what partnerships have been developed, is the workload centralized or distributed, who fills the role of content providers and managers?

Staffing: NMRLS staff for scanning, metadata, and web design phases. Using the MassCat model for policies, library staff responsibilities, etc.

Need a full time person to manage the digital library. Needs experience and aptitude to coordinate and make sure work gets done. Share wealth and share knowledge. Track vendors. Know lingo, be able to scan. Partners have to be committed. Need to have funding to get started. Use standards!

4. What major issues needed to be addressed prior to or in conjunction with the startup of the repository? Some examples might include populating the repository with content, copyright, intellectual property rights, integration with integrated library system, course management system, harvesting by Google.

OAI protocol crucial. CONTENTdm is harvested by Google, OAIster. Need to be found, be searchable.

Stepping away from doing scanning, have members do it themselves
Don't take responsibility for the masters
TIFF, backed up to gold CDs
Don't make claim to be the repository of digital masters

Digital preservation – large issue. It's money, it's time.
If don't take care of these things, data disappears

Other large issue is copyright.
Members sign a permission form. They sign off that they have the rights to sign this form, right to the material. If there's a challenge, they will take down the image. If there are requests to use an image, they get forwarded to the owner (owning institution).

Who receives these requests? Where do they get permissions?

Keep a history of the project! Where the forms are, etc. Keep records!

5. Have you developed plans to ensure the long-term sustainability and growth of the repository? Please address finances, maintenance, and staffing issues for the original project as well as current and future growth. Are you willing to share these plans? Are there aspects that have not yet been addressed in your plans?

Vision is to join statewide group with support by membership. Statewide portal for digital images using OAI and will include other projects in the state. Want to be part of bigger repository. Membership is on board with this vision. Boston Public Library is interested in managing this project. Greg is currently serving on a committee that is writing bylaws, developing standards, etc.
Statewide planning grant, will count on distributed expertise. Find a way that it's beneficial for groups such as CWMARS, NOBLE, WMRLS
Hosts metadata in MARC
Decide what output of search will be. Make those mandatory metadata.
Don't look down nose at MARC – can have better metadata due to subject heading
Next step: another small grant from ENHS. Digitization of genealogical materials. This time asked members to contribute 10% to scanning costs. $4000 per year for 4000 images, 1 gig space.
Need to have sustainable funding, revenue generators – sell images online. Currently don't do this.
Maine Memory Network is doing this. Society of American Archivists.

Started up with a $5000 grant with matching funds, $4200 per year in years 2-4 over what have in labor costs (already in NMRLS budget). $27,500 2 year LSTA (almost a dollar per image to scan, metadata, and putting online – inexpensive company.
Outsourcing is the way to go with a given project. Dedication for a given period of time is really important!

6. If you were to do this over again, knowing what you know now, how would you change the conception, implementation, and ongoing operation of the repository? Are there resources that could have been made available that could have made the project go more smoothly?

Put cost sharing up front.
Get true buy-in from the beginning by getting members to pay 10% of scanning and metadata cost
Annual maintenance fee – needed to add this to the budget for all members to share
To make it a success, libraries need to pay every year. Start up costs could be free but tell them "here's the ongoing cost that we need to share" Get libraries to participate in metadata creation. Compare to sharing

cataloging costs and active participation in cataloging process.

7. Please describe your ongoing development path and any additional resources it will take to get you there. What are some of the main issues and trends to be on the lookout for in the next 3-5 years? What do you think your IR will look like in 5 years?

See above

8. Other comments

Exhibits, more coherent information
Trying in school curriculum. Bring in small towns' information and figure out how to input into statewide site. Getting state support and commitment from Boston Public Library, MBLC, sponsorships and foundations
Sustainable model is tricky. Don't know what the cost will be in 5 years.

Questionnaire for repository site visits

Location: Tufts University
Contact name: Eliot Wilczek

Anne Sauer
Interim Director, Digital Collections and Archives and Interim University Archivist
Voice: 617-627-2696
Fax: 617-627-4650
anne.sauer@tufts.edu

Eliot Wilczek
University Records Manager
Voice: 617-627-4588
Fax: 617-627-4650
eliot.wilczek@tufts.edu

Robert Chavez
Technology Manager
Voice: 617-627-4347
Fax: 617-627-4650
robert.chavez@tufts.edu

Date(s) visited: June 03, 2005

1. Please describe your rationale for pursuing an institutional repository project. Please include your vision of what your IR will be, how you define an IR and what you expect to get out of it. Please describe any evaluation of IR software you have conducted and conclusions reached.

Had a core of objects, photos, and Tufts archival objects. Brought in major vendors. CONTENTdm, IBM and various vendors (2001)
Tech support and infrastructure to support – need to articulate to IT folks.
Fedora – need to be disciplined – articulate patience. Got up and running in a year with lots of variety of content

2. What constituencies of the university were involved in the project? Which one(s) was the driving force behind the project? Has this proven to be a good choice? Why or why not? Were there constituencies who, in hind sight, ought to have been involved in the project that weren't there? Possibilities might include institutes or museums.

DCA was the driver with lots of interest from AT (Academic Technology) Learned how to articulate value, etc of IR from AT. AT out there working with faculty, making sure folks get on board. DCA and AT saw the future need as opposed to faculty administrators. DCA and AT provide vision. Is good choice. Archivists think of management of and preservation. AT attuned to faculty needs. Should have brought in rest of library.

3. Please describe the resources (equipment, staffing, and expertise) that were needed to implement the repository. Are there additional resources that have been added to the initial resources? Please describe in detail. For example, where are the staffing resources coming from, what is their expertise level, has additional training been needed, what partnerships have been developed, is the workload centralized or distributed, who fills the role of content providers and managers?

Fedora requires you to build what you need. It has a core architecture with modules, such as OAI search. You can customize to your own needs – plug in architecture. Tufts uses Sun with Solaris 9. Fedora runs on any platform with PCs or Macs. Digital Collections and Archives (http://dca.tufts.edu)

has been the library unit that's been responsible for the IR. They partnered with Academic Technology, the developers for academic side of Tufts are there, to build the repository and digital library, it took 2 ½ fte; 5 people contributed time. IR was organized by DCA. Expertise needed: java programmer, XML and knowledge of how to build web services. Systems administration: library and AT coordinated, now have worked out an agreement with University IT.
Large hardware commitment. There are 3 parts to Tufts Computer Services. Academic Technology (AT), Systems hardware (USG), and OP (Systems management). DCA's relationship with AT is informal. However they developed service level agreements with USG (systems hardware) and OP (systems management)
Development of repository system and digital system that runs on top of it is in AT; content creation and management with DCA. AT is developing applications for faculty.
Questions / issues. May be different needs
What goes in; how is it managed, preserved.
Levels of services:
- Archival
- Warehouse
- Preservation and storage

Policy work:
Started with developing infrastructure. Content from 2 groups was first. Some content is archival. Perseus Project and Art History Project.

Levels of preservation support:
Archival: context, content
At the other end: bit stream preservation not functionality or context

Active use repository
Faculty publications and art history – people working on these. Research, class work, comments by other faculty.
Requires transactional management.
Cyclical review, may not stay in repository – appraisal work.

Issue: data authentication – need for this. Data sets and researchers.

See the Open Archives site for ingest guide.

4. What major issues needed to be addressed prior to or in conjunction with the startup of the repository? Some examples might include populating the repository with content, copyright, intellectual property rights, integration with integrated library system, course management system, harvesting by Google.

Content migration issues. 1 ½ years for regularization of content and bringing up to current standards.

5. Have you developed plans to ensure the long-term sustainability and growth of the repository? Please address finances, maintenance, and staffing issues for the original project as well as current and future growth. Are you willing to share these plans? Are there aspects that have not yet been addressed in your plans?

Not yet

6. If you were to do this over again, knowing what you know now, how would you change the conception, implementation, and ongoing operation of the repository? Are there resources that could have been made available that could have made the project go more smoothly?

More tech people
Digital resources archivist, archivist for reference and collections
Bottlenecks of content preparation and then getting to ingest phase
Fedora architecture, tech support for department of 5. NHPRC project person. DCA usually has 1-2

grant people on staff. Students and interns – have 10-12 over summer who do text markup.

7. Please describe your ongoing development path and any additional resources it will take to get you there. What are some of the main issues and trends to be on the lookout for in the next 3-5 years? What do you think your IR will look like in 5 years?

Getting more content pipes
Establishing policies
Work on architecture
NHPRC grant
Accessioning
Access
Active workspace as opposed to just repository

8. Other comments

Digital Library (Staples) centered around some sort of subject collection.
Tufts – generic delivery and access system for objects in the repository.
IR – holdings of digital output of University. Output that meets some sort of criteria.

Don't underestimate your tech needs.
Learn how librarians and archivists communicate with IT types
New England Archivists meeting session
Learn each other's vocabulary
Steven hardard?
Have someone on staff who can speak both languages
Have people know each other in advance

Policies:
- collecting
- Donating
- Collection policy on faculty papers

NHPRC Electronic Records Project – Tufts and Yale – developing an ingest guide
Synthesize electronic records preservation
Research with digital library repository research
Test potential of Fedora to serve as the architecture for an electronic records preservation system.

Formats
- Text - TEI with xml
- Images – TIFF with JPEG thumbnails
  - PDF
- Map – type things
- Working on audio with oral history project

Ingest process is command line
- Work is done by one of developers or Rob

** need to develop web interface for ingest
- VTLS and Rutgers (digital highway project)

Fedora users group conference – web site
Njdigitalhighway.org
Ingest / workflow tool

Details of site visit

Hardware:
Operating System:

Software:
Technical expertise needed:

Common features of a repository generally include the following elements

- content in digital form in a wide variety of types (text, audio, video, images, data sets)
- community focus, which determines what is included in the repository. The community members are the authors and copyright owners of the deposited content.
- institutional support, requiring collaboration across an organization, including long-term financial support to ensure that the content is preserved and maintained
- durable, permanent content.
- access to content by a broad audience, a community-shared alternative to local storage of content, fosters serendipitous discovery across disciplines.

Does your repository have all these features? Which ones have proven to be the most critical for success? Does the repository have additional features to add to this list?

Core functions which are essential for a repository

- material submission, some way for the author to deposit material, provide for editing to assure quality of content, conversion to archival format such as PDF
- metadata application, such as author, title and descriptive information and administrative data such as date and time of submission
- access control, or digital rights management, to provide for controlled access to the repository content. Even if the entire community has access to the content, there needs to be a way to restrict the ability to add, delete, edit, and approve content.
- discovery support, usually a search engine that supports browsing and full-text searching of the content
- distribution and dissemination of content to enable display and download capabilities
- preservation, some mechanism(s) for the content to be preserved and retrievable over time, including a persistent documentation identification system.

Are there additional core functions that need to be added to this list?

1. basic search keyword
   exact phrase search is an option
   type of search: all, full text, metadata – all part of basic search
   sort by: creator, collection date

2. Advanced search and / or search functionality
   Exact phrase
   Metadata (3 boxes to choose to fill in): all, creator, collection, date, description, organizations, people, places, topics, title, type
   Sort by: - same as above

3. Collection search
   Keyword or exact phrase

Ingest / workflow system – Rutgers, Tufts / Yale
University records, preservation, ingest guide
Indiana – moving images and sound files
Northwestern – digital collections want to keep separate

Encyclopedia of Chicago
Imaging system that connects these disparate together
RDF metadata text, images, maps
To connect all together city directories

VTLS Australia – Red Hat business model
Polished ingest tool – open source but sell support
Arrow project – Australian consortium arrow.edu.au/
**Contact Susan Pyzynski at Brandeis using DigiTool

Delivery applications are "disposable"
Content sitting separately in repository
Interface – Generalized discovery tool, fairly generic representation for everything in repository, for general public display
Can change application layer without changing content.
Collections management / service decision
AT could be the group to develop web service to display this to specific groups
Access restrictions still an issue

Fedora 2.1 will have access XACML
Oasis site role issue permissions
ID management system
http://fedora.info

Questionnaire for repository site visits

Location: University of Rochester
Contact name, email, and phone: Susan Gibbons,
Date(s) visited:

1. Please describe your rationale for pursuing an institutional repository project. Please include your vision of what your IR will be, how you define an IR and what you expect to get out of it. Please describe any evaluation of IR software you have conducted and conclusions reached.

Charles Phelp met Steve Harnett at a conference in the mid 90s. He'd always been interested in the economics of publishing, faculty e-archive; e theses and dissertations. He dropped this on the library. Susan was the Digital Initiatives Librarian. Grant with e books. In 2000 ETD conference.
Research in 2001. Started with e prints version 1; part time unix person. Faculty with lots of digital materials / research. Students with e theses – what would thesis look like. Bepress was just on market – publishing tool. $80K said thanks but no thanks. Endeavor with Encompass; Ex Libris with DigiTool; DSpace contact. Going for Mellon grant. Tech support and training.
Personal donation from Provost $25K server, memory, etc. Created permanent programmer position. Have a unix programmer part time. Bought development server. Content group – functions for pages. Usability group (librarians). Design group – section 508, graphic designers / programmers.
DSpace public interface flunked usability. Meanwhile quietly adding content. IT completely independent.
**"Work practice study" – go into work environment and observe for a long time – Nancy Foster. Then make a technology that works for them, led to IMLS grant to study faculty work, what they need and want. Multi-disciplinary team.
Authoring tool to share work as do research. "Finished work" didn't resonate with them. Faculty are peacocks – all about me. Faculty – no time. Resent anything that takes their time. Self-archiving not a priority for them.
Researcher pages!! – point of pride with number of download counts.
Changes to code – now on branch. Not yet part of core code. Focus now on content. Each discipline has own sets of tools. Streaming audio and video – getting started with Eastman School of Music.
Relatively successful at getting content. Went to dept and institute content – low hanging fruit. Shadow open market committee reports
Need to get library staff familiar with what DSpace is and what it can do.
ETD – no policy yet. Student deposited thesis and got a contact cold from publisher. CRL News – Sept 05
Interviewing own colleagues.
**Next step – how do students do their work. Last semester followed 4 students. Next step going to a dorm. Transcripts, facilities (where do they study) Reference interview thing. Web products. What can we build.

Vision – will be 3 voices based on who we're talking to use one of these insititutions – will be showcase of research; faculty – this is my stuff I want to share; library – collection development. Conscious recognition that going beyond what's available outside to include what's inside the institution. (eprints at Cal Tech)

2. What constituencies of the university were involved in the project? Which one(s) was the driving force behind the project? Has this proven to be a good choice? Why or why not? Were there constituencies who, in hind sight, ought to have been involved in the project that weren't there? Possibilities might include institutes or museums.

Provost was the driving force and the library. Would love to have IT involved. If had a better relationship. Too slow in including other campuses – they didn't really understand at the time what an IR was. They've very different needs (audio, images, at Eastman) Conflicts in how spend time. Want faculty ownership of IR. For museums, might create another instance, would have to think about it more.

3. Please describe the resources (equipment, staffing, and expertise) that were needed to implement the repository. Are there additional resources that have been added to the initial resources? Please describe in detail. For example, where are the staffing resources coming from, what is their expertise level, has additional training been needed, what partnerships have been developed, is the workload centralized or distributed, who fills the role of content providers and managers?

Full time programmer; 50% Susan as DI librarian at start. As created something broadened involvement for policies, usability testing; assigned 1/3 time of a bibliographer to fill DSpace – entire collection development routine. Recognize self-archiving isn't going to happen any time soon. As start new collection, consult with them. Key terms from faculty, add generic subject terms. Create abstract. DSpace student who does scanning, deposits, etc. Susan checks their work. ** Click through license for depositing material, covers copyright. Faculty retain copyright. They give non-exclusive right to distribute, to migrate to future. Expertise on what's appropriate content covers from subject librarians and from faculty. Historic perspective of documents – share this with faculty.
Partnerships with other 2 libraries been difficult. Rochester is de-centralized. Workload distributed.

4. What major issues needed to be addressed prior to or in conjunction with the startup of the repository? Some examples might include populating the repository with content, copyright, intellectual property rights, integration with integrated library system, course management system, harvesting by Google.

Whose content goes in? Students or just faculty. Outsiders? Conference on campus. If sponsorship of faculty members, student papers if supported by faculty such as theses.

Question about e portfolios, especially for faculty tenure process. Faculty are the judges of quality of student work.
Access control – didn't want IR to be dumping ground of faculty. Needs to be viewable by others
What format you can deposit whatever you want, we support open source. Faculty hold own copyright.
No integration with ILS – someday want to do federated searching.
CMS – no. ITS owns that and there's no cooperation.
Harvesting by Google – yes. Faculty love that.
Download code on each item – modified code to get true count of downloads not harvesters.

5. Have you developed plans to ensure the long-term sustainability and growth of the repository? Please address finances, maintenance, and staffing issues for the original project as well as current and future growth. Are you willing to share these plans? Are there aspects that have not yet been addressed in your plans?

Growth – yes and have SAN. Preservation piece is what don't know. In same boat with MIT, Harvard, etc.
Finances – part of regular infrastructure and staffing.
Cost of getting up and running is easy – customizing is where the costs are.
Programmer and graphic designer.
How much work from librarians – collection development, catalogers.
At least one person equivalent divided up in the library staff. Student hours to do grunt work – scanning, easy deposits, etc. Collection on department web site to IR; format conversion from word to pdf, etc
Working on:
Faculty nervous about violating copyright. Sketchy understanding of open access. Pro-actively talking with faculty abou this. If they're published …
Green literature – very complicated.
**Missing piece - necessary. Hard to do. Expertise needed!
**Librarian has to be someone very savvy with licenses. One area where faculty defer to librarians.
ROMEO, SHERPA; then contact publisher / read license. Then go to Science Direct, Web of Science, ISI, etc to find out where our faculty are publishing.

6. If you were to do this over again, knowing what you know now, how would you change the conception, implementation, and ongoing operation of the repository? Are there resources that could have been made available that could have made the project go more smoothly?

Strategy for hiring programmer. Working in open source environment very different fro regular. Person needs to be really good at perl – let's see your portfolio, talk to employers, clients. Need engaging, community-focused person
**Now would use results of anthropology study. Concept of repository is right where to go.

UR Research.

7. Please describe your ongoing development path and any additional resources it will take to get you there. What are some of the main issues and trends to be on the lookout for in the next 3-5 years? What do you think your IR will look like in 5 years?

Hoping to get code into core code to get what faculty want into code such as loading more than one file at a time. Will be more portfolio focused. Will be building more tools that will integrate with their workflow. As can save to hard drive, want to be able to map to IR.

8. Other comments

Details of site visit

Hardware:
Operating System:
Software:
Technical expertise needed:

Common features of a repository generally include the following elements

- content in digital form in a wide variety of types (text, audio, video, images, data sets)
- community focus, which determines what is included in the repository. The community members are the authors and copyright owners of the deposited content.
- institutional support, requiring collaboration across an organization, including long-term financial support to ensure that the content is preserved and maintained
- durable, permanent content.
- access to content by a broad audience, a community-shared alternative to local storage of content, fosters serendipitous discovery across disciplines.

Does your repository have all these features? Which ones have proven to be the most critical for success? Does the repository have additional features to add to this list?

Core functions which are essential for a repository

- material submission, some way for the author to deposit material, provide for editing to assure quality of content, conversion to archival format such as PDF
- metadata application, such as author, title and descriptive information and administrative data such as date and time of submission
- access control, or digital rights management, to provide for controlled access to the repository content. Even if the entire community has access to the content, there needs to be a way to restrict the ability to add, delete, edit, and approve content.
- discovery support, usually a search engine that supports browsing and full-text searching of the content
- distribution and dissemination of content to enable display and download capabilities
- preservation, some mechanism(s) for the content to be preserved and retrievable over time, including a persistent documentation identification system.

Are there additional core functions that need to be added to this list?