Five Colleges
Shared Digital Collections - Phase Two

Final Report
December 8, 2010

R2 Consulting LLC
www.r2consulting.org

Ruth Fischer
Rick Lugg
Elna Saxton

© 2010 R2 Consulting LLC
Contents

I. Introduction and Observations .................................................................................................. 3

II. The 5C Commons and Un-Commons .................................................................................. 9
    To the extent possible, expand the Commons
to include commercial images and streaming A-V ............................................................. 9
    Promote and embrace the work of the SCIPP/SMC ........................................................... 10
    Establish a Digital Preservation Task Force ........................................................................ 10
    Prioritize a single shared Discovery Layer ....................................................................... 11
    Develop a shared vocabulary related to locally digitized content ........................................ 11
    Adopt shared policies, standards, and best practices related to locally digitized content .......................................................... 11
    Develop a shared platform/interface for user assigned (social) tags ........................................ 14
    Consider deeper analysis of the current circumstance .......................................................... 15

III. Collaboration Opportunities ............................................................................................... 16
    Establish the Commons first ............................................................................................... 16
    Define “readiness” uniformly across the Five Colleges ......................................................... 16
    Establish a shared digitization services for large-scale conversion work ............................. 17
    Consider a Five Colleges “digital bunker” ............................................................................. 17
    Pursue additional targeted collaborations ......................................................................... 18
    Adopt shared systems and tools as opportunities present themselves ............................... 18

IV. Closing .................................................................................................................................... 19

Appendix A: Sharing Digital Media for Educational Use Across the Five Colleges:
Intellectual Policy Considerations ............................................................................................... 20

Appendix B: Inventory of Collections, Projects, Initiatives ..................................................... 23

Appendix C: Inventory of Services and Tools ............................................................................. 37

Appendix D: Inventory of People/Expertise ............................................................................. 41

Appendix E: Glossary of Acronyms .......................................................................................... 44
I. Introduction and Observations

In November and December 2010, R2 Consulting LLC (R2) spent 15 days considering locally-produced digital collections across the Five Colleges (5C) consortium. This was the second phase of a two-part project. Phase One, completed immediately beforehand, considered licensed and purchased electronic resources. In the initial thinking of the Five Colleges Librarians Council (FCLC) and R2, Phase Two was posited to be the smaller project. As we began to identify people to be interviewed and local digital projects to be considered, however, it quickly became clear that the scale of activity dwarfed our expectations. Because there are so many and such varied initiatives, it became apparent that the analysis would have to be much reduced in scope. Upon consultation with FCLC, this was agreed.

As in the first phase, R2 reviewed extensive documentation provided in advance, conducted seven days of onsite interviews, drafted a report, and returned to Amherst to present recommendations. In some respects, our work benefited from Phase One, although the cast of characters and range of activities were markedly different. Recommendations in the first phase revolved around the concept of the Five Colleges Commons, with emphasis on a shared discovery layer and shared infrastructure for management of commercial electronic resources. These ideas were in part intended to increase the number of staff hours available to support locally-produced digital content—the Un-Commons. A shared discovery layer will enable users to search across both Commons and Un-Commons content in a single context. These suppositions, along with more deliberately shared vocabulary and more standardized practices, also inform our thinking on Phase Two.

Our modified objectives, then, were these:

- To assess the level and variety of demand for local digitization, description, discovery, preservation, and related processes. This information-gathering was focused on library-driven initiatives, with campus-wide activities captured only to the extent the library is involved in or aware of them.

- To understand the distribution of expertise, tools, systems, standards, and practices across 5C.

- To identify possible areas for collaboration, while recognizing that locally-produced content is an area where individual libraries seek to distinguish themselves.

- To solicit opinions from those directly involved in these activities regarding collaborative opportunities, gaps in existing programs and services, and scalability.

- To identify current and potential barriers to collaboration

Our dominant impression, after more than 40 hours of interviews and meetings, supplemented by at least one ream of non-digital documentation: the generic term “digital initiative” covers too wide a range of activities to be useful. Much more specificity is needed to understand this diverse domain, because to a large degree all collections, classroom tools, and scholarly output are now generated in or converted to digital form. As a result, “digital initiatives” can include projects in a wide range of topics and formats:

- Archives (institutional history, yearbooks, course catalogs, student newspapers, photographs, film)
• Special Collections (manuscripts, photographs, multi-media, finding aids, rare books)

• Image repositories (to support classroom work)

• Audio and video streaming (to support classroom work)

• Institutional repositories (student and faculty work in text, photos, audio, video, web, including honors papers, Div III and other capstone projects)

• Academic departments (lecture capture, Moodle/Sakai content, training videos, course syllabi, even a digital rendering of a historical clothing collection for a Theater department)

• Research data (20-year longitudinal study, videos, interview transcripts, surveys, email archives, data and the code that runs the data)

• Museum collections (not part of this project, but mentioned by several people)

Some of this variation is charted in Appendix B. Many more initiatives, unknown to the library, are likely underway in academic departments. In short, everyone seems to be digitizing, describing, and attempting to present something. Many items require conversion to digital format, but many others are born digital. Each has its own set of challenges. And while digitization itself is a relatively easy process, associated tasks such as creation of descriptive and administrative metadata, development of interfaces, and web-based presentation are much more demanding and time-consuming.

Even with library-driven initiatives, the boundaries with other campus entities are often not clear. Management of digital assets on behalf of a college or university is a complex and evolving challenge, and there are many stakeholders and competing priorities. Much remains to be learned. Some obvious areas of overlap exist within individual institutions, between the Library and:

• Campus IT, whose operations are also grappling with data storage, archiving, and conversion of content.

• Academic departments, where students and faculty are creating, obtaining, and using digital content in classroom instruction, and submission of completed work.

• Educational technology, where Moodle and other courseware intersect with student portfolios, library content, etc.

• Records Management, which sometimes overlaps with Archives content and responsibility.

• Professional information management, which may have no organizational home, and which may encompass faculty and administrative email archives, as well as archives of blogs, wikis, and individual web sites.

These are just examples, of course. But they serve to highlight how much work remains to be done, and the complexity of assigning roles within the institution. For the library, the fuzziness of these borders throws some essential questions into relief. What is or should be the role of the library in digital asset management? This must be considered not only in terms of content, but also in terms of competency.
What distinctive expertise can the library bring to its own campus on these issues? How does or could library expertise with metadata standards, file naming conventions, and data preservation intersect with domain expertise that may be elsewhere? These questions are especially challenging in the realm of research data, but are also relevant in the areas outlined above. Good strategic thinking with campus partners will be critical here, with numerous iterations, before clarity around roles can be achieved. But the libraries clearly have much to bring to the table.

In the immediate, there is some excellent work underway in all five libraries. To cite a single example, the UMass project called Credo, under development for more than a year, provides a beautifully crafted interface to several repositories of unique content. Most ambitious of these is DuBois Online, which will include 200,000 high-resolution images, 100,000 rich MODS/XML metadata records, and 5,000 TEI transcriptions. The technical infrastructure, based on Fedora and MODS, is highly scalable, and will enable UMass to extend this sort of treatment to many other collections.

Across the Five Colleges, we were introduced to all manner of fascinating content, including a Virginia Woolf display at Smith, which combined an annotated copy of the American version of *To the Lighthouse* and some of her letters with the Web site related to a conference of Woolf scholars at Smith in 2003—highlighting the potential for interaction between textual and digital. At Hampshire, we became aware of films, performances, and papers produced by students in fulfillment of their so-called “Div III” requirements, representing both unique content and a unique conversion and rights management challenge.

In all 5C libraries, the staff engaged in digital initiatives consists of motivated and talented people who are conscious of contributing to an important transition in librarianship. It was energizing and inspiring to spend time with them. There is clear vision and ambition in Amherst’s plans for its extensive Special Collections and Archives. Mount Holyoke has taken a leadership role on the “born digital” institutional information issue. Overall, a sense of humor and humility pervades the proceedings, with every library except UMass laying claim to being “the furthest behind the curve.”

Despite these protestations, there has been significant adaptation to and prioritization of digital initiatives by individual libraries:

- UMass formed a Digital Strategies Group to coordinate activities, provide an overview, and to establish deliberate direction for its several Working Groups.

- Amherst has just created a new Digital Programs Department, in part to enable greater library participation in campus-wide planning and activities, but also to increase visibility of Special Collections and Archives content.

- Mount Holyoke recently created a Digital Assets and Preservation Services (DAPS) group, charged with overseeing its DSpace and CONTENTdm instances, as well as digitization and metadata production.

- Smith has for several years had a Digital Services Team (DST) that oversees and advises on content, metadata, and discovery issues.

- At Hampshire, although there is no formal structure in place, participation in digital initiatives is broad-based, drawing from the Web team, Technical Services, Reference, Advanced Media, and Archives.
These responses highlight each library’s recognition of the increasing importance of digital activities, the need to obtain expertise and to scale up operations to meet the demand. By creating a discrete organizational identity for digital initiatives, management and integration with other library operations become much more possible. These are signs of healthy adaptation.

There have also been a number of successful Five Colleges-wide collaborations to date, with more in progress. These have generated a good deal of practical experience, and can be seen as building blocks for additional collaboration:

- The Intellectual Property Policies for Shared Media Committee has focused its attention on standards, policies, and best practices to enable sharing while retaining local autonomy. Their work strikes us as so useful that we’ve included their “briefing document” as Appendix A to this report, to be certain everyone is aware of it.

- The two-day DEDCC-sponsored workshop on Digital Preservation that took place in June 2010 was repeatedly cited as an example of the education and training that are possible collaboratively.

- The DEDCC’s recent Digital Preservation Task Force proposal shows vision and initiative, and has identified an enormous problem that exists in all 5C libraries, and for which no library yet has an adequate solution.

- The Five Colleges Archives Digital Access Project (FCADAP), originally built in 1996, enabled online access to the consortium’s combined archival collections in women’s history. An October 2010 appeal to Mellon seeks to update and expand this work.

- Five Colleges Archives and Manuscripts online finding aids project resulted in shared EAD templates, and a Five Colleges Archives MARC Cataloging guide—both of which are still widely used. The Archivists also agree on best practices for creating Dublin Core metadata.

- There has been some collaboration on shared content (i.e., beyond the Commons content identified in Phase One), especially with image databases such as Archivision, Scholars Resource, Antequam, and ARTstor (including some experimentation with its Shared Shelf capability).

- Generally, Five Colleges enjoys a growing pool of expertise. The small cadre of digital specialist is growing steadily, and there appears to be a genuine willingness to share knowledge. This will help with cohesion going forward.

Despite these successes and new efforts, however, there remain many unsolved problems. This is unsurprising, given the fundamental nature of this transition, and the scale and variety of activity. In fact, some of these problems may be amenable to collaboration. A high-level list of impending challenges might include:

- There are many more digital projects on campuses than the libraries are aware of. There is a need for comprehensive information gathering, both campus-wide and Five Colleges-wide. Some of this is already underway, e.g., with a campus survey planned at Smith.
• Commitment to existing processes, systems, and infrastructure: Each library has already invested significantly in its own digital asset management infrastructure. There are at least ten separate systems and versions of systems dispersed across the Valley: Fedora, CONTENTdm, BePress Digital Commons, several versions of DSpace and Luna Insight, DigiTool/Rosetta, etc. Each of these has its own process and metadata requirements. Little of the work done on one campus is directly applicable to another. At present, there are more disincentives than incentives for collaboration on hosting and describing content.

• Among other factors, the differences among systems and infrastructure reflect differing philosophies on some key topics:
  - Open source vs. commercial software solutions - some libraries have resources and skills to handle open source, others must rely more on “turnkey” solutions
  - Build vs. buy – a variant of the above
  - Hosted vs local installation
  - System imposed metadata standards – MODS vs Dublin Core (DC)

• Collaboration may extend toward campus initiatives rather than toward other libraries. Because on-campus partners face problems similar to the library’s, and because activities overlap, there may in some cases be more compelling reasons to align collaboratively with those activities than to seek collaboration with other libraries. While this need not be an either/or proposition, it does potentially limit library-to-library cooperation.

• Collaboration may need to extend beyond the Five Colleges horizon to be effective. While our focus has been on 5C collaboration, there are certainly initiatives in which participation in a broader regional or national effort may make more sense. Digital preservation may be an example, although that remains to be determined.

• Born-digital institutional publications (such as course catalogs at some institutions) pose an entirely different challenge than print publications that have been converted. While Mount Holyoke has begun to lead in this area, there is at present not widespread awareness of the problem. An effective program here will require additional resources to build and maintain.

• Technological and organizational “readiness”: This was one of the most interesting discussions of the week onsite. In some respects, existing departments, working groups, and committees focused on digital initiatives serve almost as research and development operations. With some exceptions (like DuBois Online), they have not achieved much scale. More importantly, they are largely segregated from traditional library functions such as cataloging and reference. The degree of readiness will depend to some extent on how fully the full staff can embrace the primacy of digital content, and how successful library management is in retraining and redirecting staff.

• Related to the last point, integration of traditional library operations with digital initiatives will be essential. Assuming that shared management of the Commons material has the desired effect of reducing workloads, there will be staff hours freed for redeployment to digital initiatives. This may require adaptations, such as shifting MARC catalogers to Dublin Core, MODS, VRA Core, and other metadata schemas. It may also require increased specialization
within reference and public services functions, as individuals are designated to master more specialized databases and interfaces.

- The prevalence of grant-based and project-based funding complicates staffing and sustainability. This, too, is related to organizational and technical readiness. Because it is relatively difficult to establish new permanent positions in some libraries, long-term temporary funding has been used. This has enabled new skills and more hours to be brought into the operation. But given the strategic priority assigned to unique content, these functions and skills are critical to the organization’s future. These positions need to be made sustainable, and traditional permanent positions must be converted as opportunities arise.

- Finally, competition may be something of a factor. This may not occur so much in regard to content: unique is unique. But it may surface in the way digital initiatives are managed, and also in the perception that individual library efforts can actually be hampered by necessity to collaborate and communicate.

In the end, the most promising opportunities for collaboration present themselves in areas where solutions are not yet in place, and where investment will not first have to be undone in order to work together. Most of these are already under discussion or were suggested more than once by interviewees. These will be described more fully in the next two sections of the report.

As we hope is clear in the preceding pages, R2 was very much impressed with the skill and energy that have already been engaged in these areas. Owing to the patience and generosity of our interviewees, we learned a great deal in a relatively short time. We appreciate the many meetings, phone calls, and emails endured on our behalf. As always, librarians and library staff are a pleasure to work with—informed, forthcoming, and at times opinionated. Thanks to all for participating so fully.
II. The 5C Commons and Un-Commons

The 5C Commons, as described in the Phase One Report, assumes the establishment of a shared collection of commercial e-resources. A critical corpus of the most heavily used e-resources would be accessible to all members of the 5C community, through an identical discovery experience. For this core, management tools and tasks would be shared, and institutional branding would be replaced by the 5C moniker. The potential benefits to this approach are predicted to be significant, both in terms of a more uniform user experience, and in terms of more efficient technical/back-room management.

The hope is that by reducing staff investment in the management of heavily-used commercial resources, more staff time will be available for local and unique ones. Materials held by a single institution, then, become the 5C Un-Commons; more specialized content for fewer, more advanced users. R2 assumes that most collection and resource specific tasks in the Un-Commons will remain fully within the domain of individual libraries. These are the assets that effectively differentiate each institution and for which they may at times compete.

Nevertheless, many of the same discovery principles should be applied to the Commons and Un-Commons alike. R2 believes there are important ways in which the 5C Libraries can and should collaborate with regard to the management of locally digitized content. Specific design features of online exhibits and myriad descriptive and interface decisions will distinguish digital collections in exciting and useful ways, but the benefits of a uniform discovery experience persist, especially when so much of the content is intended to be open to all. In order to achieve a consistent and intuitive user experience, it will be important for the 5CL to adopt at least some of the same terminology, policies, and standards, and to begin to consider local Un-Commons initiatives and activity within the context of the consortium. In short, though content and infrastructure may differ, discovery and key practices should be similar. Because the Commons and Un-Commons are on a continuum, and because many users will operate on both ends, recommendations in this section of the report mirror those offered in Phase One.

To the extent possible, expand the Commons to include commercial images and streaming A-V

It is now clear that the overall scope and design of the R2 Project was imperfect. From the start, there was a lack of clarity about how to incorporate workflows related to the acquisition and management of commercial images, and streaming A-V. They did not fit especially well in either phase of the project. In the end, those issues did not get the attention they deserve. More is warranted. Nevertheless, we can offer a few obvious observations and recommendations:

As most readers know, Haivision’s Furnace is installed at Amherst. In addition to meeting local needs for course-related streaming video, Amherst extends this service to Mount Holyoke faculty. Other 5C libraries assume it to be cost prohibitive, which may need to be confirmed.

Four of the five libraries subscribe independently to the Naxos Music Library, which includes more than 85,000 tracks of recorded classical, jazz, world, folk, and Chinese music. This is a standard commercial e-resource that should probably be part of the Commons. Likewise, ARTstor’s commercial images, Scholars Resource, and Archivision’s non-profit architectural images (hosted on ARTstor), should be part of the Commons, and managed centrally.

Three of the five libraries subscribe to AP Photo Images – including Hampshire, Smith, and Mount Holyoke. This resource is one for which multiple subscriptions may be appropriate. Commercial images do not conform to article delivery models, and extended access to the UMass patron base might be too expensive.
If we understand correctly, Antequam (local scholar images) is a 5C resource which is managed in two different ways. UMass and Smith store the Antquam images in LUNA and the 5C hosts them in ARTstor. A consistent 5C approach would likely benefit users.

Most locally captured and digitized images should be managed according to the principles appropriate to other unique content. There does appear to be some redundancy within the Valley with regard to copy images (pictures taken of pictures in printed library resources), and departmental/faculty slide collections. Again, more investigation is needed to identify opportunities for increased efficiency and improved service, but many people see opportunities here for collaboration.

Promote and embrace the work of the 5CIPPSMC
R2 had the opportunity to meet with the 5C Intellectual Property Policies for Shared Media Committee, and was immensely impressed with the approach being taken. We were disappointed, however, to learn that the work of this committee had to date remained unfamiliar to many. We have therefore included their July 2010 Briefing Document as Appendix A to help increase awareness. In early November, the group’s “Guiding Principles” were in draft mode, but may now be available for review.

The goal of this committee is to draft a set of 5C guidelines for sharing digital educational collections across the Valley. This work is envisioned to establish an intellectual property umbrella, under which resources can be made available without the overhead of establishing collection and/or format specific usage rights. An inventory of rights types has been developed to include purchased, licensed, copyright exception asserted, copyright owned/granted, public domain, and unknown; with images, audio and video examples for each. R2 believes this work to be invaluable, insofar as it eliminates legal confusion and emboldens the inclination to share. This collaborative policy development is precisely what’s needed; comprehensive analysis, consultation with legal professionals, clear and applicable standards, adopted by all. Bravo!

Establish a Digital Preservation Task Force
In Phase One, R2 offered a largely unfavorable review of the 5CL Committee structure, suggesting that the investment of staff time and has not yielded sufficient returns. Too many committees have overlapping agendas and very few decisions are made in unison. Without an overarching vision or 5C mandate, a culture of non-conformance prevails. With regard to the management of commercial e-resources, each library is operating entirely independently and R2 went so far as to recommend a complete overhaul of the committee structure, once a new strategy for collaboration has been established.

However, the situation regarding locally digitized collections and digital preservation is considerably less developed. None of the libraries has a fully realized preservation program; no one has adequate digital storage for the long term. Rather than each library striking out on its own to solve these problems, now is the time to identify state-of-the-art, cost effective solutions that can be implemented by all. In February 2010, DEDCC proposed a Digital Preservation Task Force and in August 2010, drafted a charge for the proposed group. R2 applauds the size limit (7 members), the time frame (action plan within a year), and the practical approach described. If, in addition, the expectation is for full 5C compliance with eventual guidelines and suggestions, we recommend that the FCL accept this DEDCC proposal and move forward expeditiously.
Prioritize a single shared Discovery Layer

In the Phase One Report, R2 recommended adoption of a single 5C discovery layer, as a way to solve the problems of the ALEPH Union View. If the 5C decides that a Commons should be built, we believe that a shared discovery layer is vital to the unification of the user experience and will shape many underlying, collaborative workloads. A priority for the Commons, a single discovery layer will offer at least as much benefit to the Un-Commons. In general, unique and locally digitized collections are enormously “silied”, meaning that users must know quite a bit about where to search for some of the most valuable resources. OPAC and web searches retrieve special collections resources inconsistently at best. 5CL must implement a primary search tool that retrieves relevant content regardless of format; published commercially or locally.

This, then, may be the single most important R2 recommendation. It has huge positive implications for the Commons and the Un-Commons; it will leverage a lot of cataloging and metadata creation that has already occurred; and it will inspire activity throughout the valley – to reveal heretofore “hidden” resources. Once the tool exists, we believe that staff and librarians will learn to take advantage of it, in infinitely creative ways. All of this, of course, presumes the availability of harvestable metadata, which is discussed below.

Develop a shared vocabulary related to locally digitized content

As with commercial electronic resources, there is a fair bit of variation among the 5C libraries with regard to local terminology. The adoption of shared policies and guidelines will, of course, facilitate coherence around shared language, but at present, even the most basic terms are not used uniformly. For example, Institutional Repository or IR, typically connotes a collection of local (institution specific) scholarship, but as far as we can tell, this definition is not fully shared within the 5CL. We have no doubt that other key terms like digital library, subject portal, online exhibition, etc. are used inconsistently as well.

Adopt shared policies, standards, and best practices related to locally digitized content

The 5C Intellectual Property Guidelines mentioned above should be followed with others. Consistency of practice will increase the uniformity of the 5C user experience, and joint adoption of shared policies and standards will eliminate the need for all five libraries to invest in the research and development process, which can be extensive. After all, standards are not intended to be institution specific.

Some work in this area has already occurred, at both the 5C level and within individual libraries. For example, in March 2007 the 5C Archives Committee published Guidelines for Dublin Core Metadata Best Practices and cited the North Carolina ECHO template and the Collaborative Digitization Program. Once such guidelines/policies are adopted, some attention should be paid to their implementation.

Electronic records management

When one library undertakes the work of establishing best practices and/or standards, it would make sense for others in the Valley to monitor and support the effort and to avoid “reinvention of the same wheel.” One immediate example is with regard to the grant-funded project currently underway at Mount Holyoke. MHC Archives and Special Collections recently received an NHPRC grant to develop “sustainable workflows and processes to accommodate institutional records that are born-digital”. With this grant, the library will hire a staff person to establish four basic workflows, to accommodate electronic records ingest, processing, storage, and
description. Since none of the 5C libraries has an especially good solution for electronic records management, the potential benefit of this work could be leveraged throughout the valley. Better consortial awareness is really all that’s necessary.

In addition, there are a couple of obvious areas in which the development and adoption of shared standards will benefit the 5CL and its users: metadata and preservation.

**Metadata**
The key here is to ensure that 5CL metadata is OAI-PMH compliant (Open Archives Initiative Protocol for Metadata Harvesting); available for retrieval via standard discovery tools. Metadata must be interoperable. That is, storable and retrievable in many systems without losing integrity.

The UMass Metadata Working Group has been working since June 2009 to “provide direction for and coordination of metadata creation and management with the goal of interoperability.” They began with a survey designed to identify and evaluate metadata practices within the UMass Libraries. This brief description of best practices is taken verbatim from their summer 2010 report:

> Best practices for digital collections metadata have emerged over the last decade, and it is imperative that they become a part of any metadata creation. Managers of any digital collection should know its audience, as this will decide to some extent suitable levels of description. Structural and content standards (including controlled vocabularies or thesauri) appropriate to the scope of the collection should be used. It is also important to select syntax standards and technologies (e.g. digital asset management systems, servers, etc.) that allow for the most technical interoperability. Adherence to these standards will enable easier migration and repurposing of data. Lastly, all decisions made related to metadata should be documented.

Beyond the general guidelines mentioned above, shareable metadata must abide by the following tenets:

- **Records should have an appropriate amount of content and levels of granularity necessary for indexing and display**, e.g. the record should indicate the vocabularies used and it should be clear whether a record is providing collection or item level description.

- **Practice should be consistent between records in a collection**. This makes it possible to normalize values for large groups of records when needed. One should be able to predict what fields will be in a record, what vocabularies are used for a field, field punctuation, and granularity of indexing.

- **There should be coherence within a record**. It should be easy to ascertain what the record is describing, with values in appropriate fields, only one value in a field, and type of resource should be clear.

- **Records should make sense outside of their native environments**. Data that applies to all objects in a collection should be in the record (including the
Their report includes careful descriptions of the (UMass) libraries’ digital initiatives along with the strengths and weaknesses of collection specific metadata practices.

The group’s metadata guidelines are exceptionally well developed, and incorporate considerable detail including “metadata elements by requiredess”, and “metadata elements a-z”. Here again, R2 is highly impressed with what’s been accomplished here and very correctly, there is nothing “UMass” about the guidelines. R2 recommends, therefore, that the work of this group be reviewed, and to the extent possible, re-used at the 5C level. The real value of this UMass contribution should not be underestimated, and could/should be balanced by other libraries’ contributions of different ilk.

**Persistent identifiers**

During the course of our discovery process, we were told that Robin Wendler (Metadata Analyst from Harvard) had at some point suggested that the FCL start collaborating on a single system of persistent identifiers for digital content. It is a key component of the infrastructure needed for digital preservation, and not something that has been tackled in a consistent way. Individual systems often assign control numbers, but it is unclear what will happen if/when the libraries migrate to new systems. R2 believes this idea is worthy of further consideration.

**Preservation**

In April 2009, UMass also established a Digital Creation and Preservation Working Group. The intent is to:

- inventory current practices
- establish guidelines for creation of digital content in various formats
- develop a policy framework for digital preservation
- make recommendations on strategic planning for digital preservation
- evaluate and develop strategies for long-term access to digital objects, for example migration, refreshing, etc.

This strikes us as an appropriate charge for the 5C Digital Preservation Task Force, and is in fact, very similar to that proposed by the DEDCC. Here again, it may be possible to leverage the work going on in individual libraries to benefit the consortium, especially if a strategic decision is made to invest jointly in some of the underlying infrastructure.

Best practices in digital preservation are, of course, still evolving, so it will be critical for the 5CL to stay abreast of national level initiatives including but not limited to:

- DLF Aquifer: [http://quod.lib.umich.edu/a/aquifer/](http://quod.lib.umich.edu/a/aquifer/)
- CLOCKSS: [http://www.clockss.org.clockss/Home](http://www.clockss.org.clockss/Home)
- TRAC Certification: [http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf](http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf)
• American History Online: [http://www.americanhistoryonline.org/](http://www.americanhistoryonline.org/)
• Hathi Trust: [http://www.hathitrust.org/home](http://www.hathitrust.org/home)

Just this week, we learned about a new project undertaken by Ithaka S&R that will be of interest over the coming months. Ithaka’s “Sustainability of Digital Resources” research program has published several reports which can be found online: [http://www.ithaka.org/ithaka-s-r/research/ithaka-case-studies-in-sustainability](http://www.ithaka.org/ithaka-s-r/research/ithaka-case-studies-in-sustainability).

Their new project will be focused on ways that funding agencies and private foundations influence the sustainability of the digital resource projects they fund. “Drawing on in-depth interviews with program officers and other stakeholders in the UK, US, Canada, and Western Europe, this research seeks to share best practices in planning for the long-term impact of digital resources.” These findings will be available in spring 2011.

**Develop a shared platform/interface for user assigned (social) tags**

The benefits of user assigned tags are especially apparent in relation to locally created content. Wikipedia says:

* A tag is a relevant keyword or term associated with or assigned to a piece of information (a picture, a geographic map, a blog entry, a video clip etc.), thus describing the item and enabling keyword-based classification and search of information.

* Tags are usually chosen informally and personally by item author/creator or by its consumer/viewers/community.

* Tags are typically used for resources such as computer files, web pages, digital images, and internet bookmarks.

R2 suggests that 5CL may increase the findability and use of unique content by introducing social tagging functionality sooner rather than later. It will be especially helpful to enlist the assistance of domain experts who can add tags even after the record is “live.” If a 5C discovery layer is implemented, and shared metadata policies have been embraced, it may be appropriate to introduce a single platform for controlled social tagging.

A few related articles of interest:

• Tag Clouds and Old Maps, Rainer Simon et al: [http://stko.psu.edu/lstd2010/paper%201.pdf](http://stko.psu.edu/lstd2010/paper%201.pdf)

• Tagging Tagging: Analyzing User Keywords in Scientific and Bibliography Management Systems, Markus Heckner et al: [journals.tdl.org/jodi/article/download/246/208](journals.tdl.org/jodi/article/download/246/208)


• Cultural Difference in Image Tagging, Wei Dong and Wei-Tat Fu: [http://appliedcogsci.vp.uiuc.edu/admin/upload/1271908810Image_tagging_CHI_100110.pdf](http://appliedcogsci.vp.uiuc.edu/admin/upload/1271908810Image_tagging_CHI_100110.pdf)
Consider deeper analysis of the current circumstance
During the course of the project, R2 struggled to assimilate the information gathered about the digital initiatives underway within the 5C Libraries, and saw just the “tip of the iceberg” beyond the libraries. It may be appropriate for 5C, Inc. to commission a comprehensive survey of digitization activities across all five campuses. Once the data is gathered and analyzed, it may be possible to determine the appropriate role for each library or for the 5C Libraries collectively.

In the meantime, it may be useful to develop a series of inventories that compile information about 5C Library activities and expertise related to digitization, digital collections, and preservation of digital content. Symmetrical, comparable information might make it possible for the 5CL to identify patterns of activity, redundancies, service gaps, and opportunities to cooperate.

To demonstrate this idea, R2 has created three inventories; each only partially populated:

- Collections, Projects, Initiatives – see Appendix B
- Services and Tools – see Appendix C
- People/Expertise - Appendix D
III. Collaboration Opportunities

As suggested earlier, collaboration across the Un-Commons may offer fewer opportunities than the Commons. In the previous section, we suggested extending the Commons approach to some additional streams of content. Here we outline a few more ideas for collaboration, most of which are already under consideration, at least in some form, by the Five Colleges. Because there has been so much investment in individual library solutions, and because so much infrastructure is already in place, these recommendations tilt toward areas where problems are as yet not fully addressed, or where collaboration will cause minimal disruption to existing operations. Though they are relatively few in number and in some cases modest in scope, there is significant progress possible if they are adopted.

Establish the Commons first

In order to free staff capacity to address some of the recommendations below, it will be necessary to implement at least some of the recommendations suggested in Phase One. For example, if MARC records for e-journals are eliminated, some cataloging hours would become available to produce Dublin Core, MODS, or VRA Core records. Those hours might also be used for MARC cataloging of finding aids or creating metadata crosswalks. Once a single consortial instance of SFX is adopted, additional highly skilled hours can be made available in support of locally-produced digital content. In short, the rationale behind many recommendations to build the Commons is to create additional capacity for the Un-Commons.

Obviously, establishment of the Commons will take significant time and effort, which will mean that these hours will not be available for a year or more. This may limit progress on Un-Commons ideas temporarily, but will also have the advantage of buying time for training and organizational readiness.

Define “readiness” uniformly across the Five Colleges

As suggested in the introduction, digital initiatives at present are very much in the hands of specialists. To make this work scalable and sustainable, it must ultimately be integrated much more fully into each library organization, and into the consortium as a whole. Orientation and support need to evolve from project-based, grant-funded basis to ongoing operations. In other words, it needs to become a mainstream activity, part of the workday of most library employees.

This will require a comprehensive program of education, training, and retraining for technical skills, new metadata standards, and perhaps even productivity expectations. It will take time for everyone to understand and accept that unique content has become the library’s priority. People will need to be shown how they can contribute, and how important the technical services “engine” is to scaling up these initiatives. As several interviewees noted, digitization is easy, but several projects are “choked up” on metadata production. There is need for EAD encoding, high-volume production of Dublin Core, VRA, and MODS records and doubtless others will emerge.

As individual libraries wrestle with this transition, a shared definition of “readiness” is needed. This might take the form of a list of competencies, supplemented by an estimate of the anticipated volume of work required in each competency. It also might take the form of a required training curriculum, in which individuals are trained and certified in specific tasks. The important point is to develop a shared approach, so that consistent performance can be delivered throughout the Valley.
Establish a shared digitization services for large-scale conversion work

“We are all scanning stuff”, as someone said. Outside of UMass, scanning operations are relatively small-scale and project-based. Scanning itself requires minimal training and scales easily. There are ambitious plans in some locations to digitize very large bodies of material from archives or special collections. Although UMass has a relatively large-scale operation, it is based on grant funding and serves only those projects. In short, there appears to be significant potential demand for digitization, but existing local solutions are inadequate.

It is important to distinguish between large-scale programmatic digitization and smaller-scale ad hoc digitization that is done in response to patron requests. Most of the patron-driven work needs to remain supported locally (the exceptions being items that are scanned from the Bunker). But we suggest, based on numerous comments heard during our visit, that Five Colleges establish a shared digitization center for large-scale projects. This center would support not only digitization of text, but reformatting of audio-video content as well. It would be built as a high-volume, production-oriented operation, supporting digitization and reformatting projects from all five libraries (and potentially beyond).

Such a facility might be located at the Bunker, if proximity to physical collections seems useful. It might also be combined with a consortium-wide conservation and book repair operation. Wherever it is ultimately located, materials for digitization would need to be trucked to and from that location. Adequate safeguards would need to be in place for archival and special collections materials, but the general opinion was that this would be possible. It is also conceivable that such an operation could be outsourced to a commercial entity such as Crowley if that proved more attractive. The bottom line is that investment in and use of this expensive infrastructure could be shared among all five libraries, and conceivably with other libraries or with campus partners.

Consider a Five Colleges “digital bunker”

This idea surfaced more often than any other. Preservation and storage of digital objects is a complex area, with many unknowns. Every library believes that some material is currently at risk, despite some participation in community-wide solutions such as LOCKSS and CLOCKSS. Since this is a complicated and widely-shared problem, and because individual solutions are currently inadequate and likely very expensive, this is fertile ground for collaboration. It is also an area where other campus entities, such as academic computing, educational technology, bursar, registrar and others are facing the same problem.

Any shared digital storage and preservation solution would have to accommodate many formats, support data authentication, and be certified as a trusted repository. The library community has a number of such standards, including TRAC (Trusted Repositories Audit and Certification), and the OAIS (Open Archival Information Systems) Reference model. To the degree that this is a libraries-only enterprise, those will likely be adequate. However, if institutional records containing personal or financial information are part of the mix, a whole new group of standards will also be necessary.

In some respects, it is quite interesting to think of this as a libraries-only initiative. Not only would it solve a problem for Five Colleges libraries, but the service might be extended to other libraries—much as use of the print journals archive in the bunker already is.
Pursue additional targeted collaborations

While the current technical and organizational infrastructure(s) may not lend themselves to collaboration on many fronts, the Five Colleges Libraries have an excellent track record on more focused collaborations. The Five Colleges Archives Digital Access Program (FCADAP), dating back to 1996, consolidates access to consortium-wide archival content in a single interface. While its subject scope is limited, it has realized the potential of working together in a coordinated manner. The current proposal to Mellon to update and expand FCADAP is another step in that direction. Similarly, the Five Colleges Archives Committee created and agreed to a template-driven approach to EAD finding aids that has improved consistency across the collections.

At present, the Intellectual Property Policies for Shared Media committee is laying the groundwork for a tightly-focused successful—and adaptable—collaboration in its area of responsibility. The work just beginning at Mount Holyoke on born-digital institutional publications may be another example. It seems likely that many more such small-scale, targeted collaborative efforts are going on regularly, and that this approach should be encouraged. It is possible to make a great deal of progress this way, without necessarily resorting to large-scale strategic initiatives. In the same vein, there may be some 2C, 3C and 4C collaborations—e.g., around support for DSpace—that might make sense.

Another potential area for collaboration is with regard to data management requirements imposed by grant funded research. For example, those researchers applying for NSF (National Science Foundation) funding will soon need to describe the way in which they will expose the data they collect. These requirements are part of a broader move by federal funding agencies to emphasize the importance of community access to data. The Libraries could step in here, to facilitate conformance in a unified way.

Likewise, there is an open access mandate for NIH funded research. This is the language of the law, taken from the SPARC website ([http://www.arl.org/sparc/publications/an-open-access-mandate-fo.shtml](http://www.arl.org/sparc/publications/an-open-access-mandate-fo.shtml)):

> The Director of the National Institutes of Health shall require that all investigators funded by the NIH submit or have submitted for them to the National Library of Medicine’s PubMed Central an electronic version of their final, peer-reviewed manuscripts upon acceptance for publication to be made publicly available no later than 12 months after the official date of publication: Provided, That the NIH shall implement the public access policy in a manner consistent with copyright law.

Here again the 5CL could jointly facilitate conformance on behalf of researchers in the Valley.

Adopt shared systems and tools as opportunities present themselves

It is eminently clear from the charts that accompany this report that there is little common ground at present regarding the systems that support institutional repositories and special collections. It’s entirely possible that local needs for local content will mean that this situation persists indefinitely. Nonetheless, any time a migration or change is contemplated, it would serve the consortium well if the library making the change considered the possible benefits of a more fully shared infrastructure.
IV. Closing
We’ll say again, recommendations from outsiders can have enormous value but they have limitations. This report is based on only a few days’ immersion in 5CL’s processes, systems, and culture. We are certain that we have mistaken some of what we heard and saw, and that our recommendations will need scrutiny by those of you closest to the situation.

As always, we are hopeful that our ideas will inspire local conversations that would not otherwise occur. During the course of this project, R2 has had the privilege to seek an overview, to paint a big picture – something that has not often been attempted within the 5CL. We appreciate having had the opportunity and are immensely respectful of the work that has occurred and that continues to occur in the Valley. Thank you.