

# Developing a Digital Library: Scale Requires Partnership

*John McGinty, Marist College Library Director*

## ABSTRACT

---

Digital library technology presents new opportunities for academic libraries to expand access to their collections. A full-scale digital library project involves complex development issues and problems. Collaboration through academic-industry partnerships has emerged as an important strategy for libraries to better leverage resources, minimize risks and increase expertise. The inclusion of faculty and students in the development process further ensures that the system will serve the needs of the academic community locally and remotely. The Marist College multimedia digital library project models the development process and structure of a partnership with a major technology company, information industry vendors and other academic institutions.

---

## Introduction

The focus of this paper will be on collaboration in the development of the large-scale, academic digital library. The principal discussion will concentrate on the idea of academic-industry partnerships as a form of technological interdependence that is essential to ensure that the widest array of information products become available to a campus community with searchability and bibliographic standards of a high order. The relevance of this type of collaboration to libraries was pointed out by Carol Hughes and William Pfannestiel in their ACRL environmental scan of higher education, which emphasized the importance of increasing academic research partnerships with private industry to promote scholarly communication.<sup>1</sup>

The joint development partnership between Marist College in Poughkeepsie, NY and IBM to build a digital library will provide a brief case study to explore some of the most important issues that other colleges and universities must consider in attempting similar projects or collaborative ventures.

The impetus for Marist College to collaborate with an industrial partner comes from its mission to be a technologically advanced comprehensive-level institution in a very demanding market segment: private higher education in the Northeast. It wishes to take advantage of the geographical proximity of a major information technology firm which can provide technical expertise and support for services, software and equipment. IBM stands to gain from Marist by partnering with an institution whose size and organizational structure can facilitate relatively rapid development and testing of a new technology in a college-wide application.

This partnership expanded to include other academic institutions, libraries and vendors to sustain scale-up opportunities in content and database building, access and rights management. Contributions to technological development have come from the Franklin D. Roosevelt Presidential Library, Case Western Reserve University, and the ELIAS Corporation.

## Collaboration

Collaboration can be defined as a formal set of relationships between individuals and organizations in a development process that exists through the cycle of idea formulation, design, product testing and implementation. This form of collaboration relies on a significant input of human energy, resources and commitment over a sustained period of time. The range of possible collaborative ventures encompasses one product/one time partnerships to complex relationships that develop successive products or technologies over a period of years.

The rapid evolution of new information technologies presents fertile opportunities for collaborative efforts in higher education. Collaboration has become an important concept in technological development primarily because of its potential to better leverage resources, including access to critical hardware and software products. Other advantages of collaborative efforts include the minimization of risks, the encouragement of innovative thinking, shared expertise, expanded access by users and improved productivity.<sup>2</sup>

Sectors of the economy under rapid change, such as information technology, require collaborative efforts to develop effective hardware and software solutions because of the significant investment in technical design and problem-solving for a short product life-cycle. An essential advantage in partnering is the ability to cut costs and improve outcomes for each organization by allowing broader (aggregated) resource inputs in the design through implementation stages of a project.

The collaborative model is becoming fairly prevalent in the development of digital library resources. Cornell University has several electronic library projects that have been generated in the last four years. In an issue in *Library Hi Tech* in 1994, the Cornell University Library System highlighted eight digital library projects, all of which are collaborative in nature.<sup>3</sup> Many involve industrial partners including Kodak and Xerox, and publishers such as Elsevier and UMI. Several involve cooperative ventures with other research institutions, foundations and government agencies. It is clear that the scale of large digital projects require testing on a broad scale, a critical mass of documents to digitize, remote access opportunities and financial support and most importantly shared intellectual resources for ideas, development and problem-solving.

Another important digital library project that features extensive collaboration is the JSTOR initiative. This large journal digitization project began in 1993 through the auspices of the Mellon Foundation, involved Princeton University in the idea generation stage, Harvard University provided the hard copy journals for production and the University of Michigan serves as the host computer storage facility for Internet distribution of the digitized articles. Initially, a pilot group of five colleges, including Bryn Mawr, Denison, Haverford, Swarthmore, and Williams tested the use of the electronic resource among faculty and students.<sup>4</sup> Now the benefits of this collaborative effort can be shared more widely as the JSTOR digital journals are being marketed as a digital library component for other libraries.

The Marist digital library project involves three levels of collaboration:

1. Marist faculty, students and staff working together to generate ideas and offer expertise
2. IBM and Marist staff to design, program, test and manage the development and implementation
3. IBM Renaissance Consortium, an inter-institutional partnership which brings together several development projects operating at the second level like Marist, each of a complementary nature. Partners include the Library of Congress, Institute for Scientific Information, Case Western Reserve University, Indiana University, University of Florida System, Virginia Commonwealth University and Virginia Tech

Marist provides the project with management skills, technical support and a broad problem-solving testbed over the lifecycle of the technology. The College generates ideas for applications and design features directly from users and an opportunity to model a holistic solution by involving a cross section of the entire campus in the development and use of the digital library. IBM provides Marist with software, hardware, technical support and management skills. The digital library partners provide additional expertise, problem-solving ability and new functionality to areas of development not possible at Marist.

One of the crucial benefits that has been realized from the collaborative effort of Marist and IBM staff has been that individuals throughout all levels of both organizations have been available to make decisions or generate ideas, not just those individuals responsible for development. This includes faculty members, students, librarians, technical staff and senior administrators at Marist and IBM, all of whom have been responsible for the delivery of specific services or products as part of their job descriptions. This broad inter-organizational collaboration is intended to avoid one of the classic problems of joint development projects: failing to engage the higher levels of a large organization to ensure priorities are met.<sup>5</sup> Economically the product lifecycle costs for development and fixes have been reduced, particularly through the concept of prototyping at the end-user level.

An associated benefit also accrues from the collaborative process: participation in the design enhances the potential for acceptance of the solution and a willingness to implement it. A study in the *Sloan Management Review* on successful information system building in industry suggests that concurrent system development that includes collaboration between engineers, developers and implementors will help insure information system implementability primarily because internally integrated and consistent ideas emerge from the collaborative process that are easy for users to accept because they have been communicated, evaluated, compared and debated in a holistic manner.<sup>6</sup>

Faculty participation in the idea and design phase has been crucial to campus-wide acceptance at Marist, regardless of the barriers of introducing cutting edge technology. Several faculty have put in long hours on teams and in implementing digital course files and developing multimedia content that students would use in their class assignments. Their influence on colleagues and continued input on system re-design have ensured the scale-up of the project meets the teaching and learning needs at Marist.

## **Definition of the Term Digital Library**

A scan of the library literature over the last four years shows there are three terms--digital library, electronic library or virtual library--that are used to describe the building of a large information resource of text, images, numeric data and media to serve the demanding academic user of today. I prefer the term digital library.

The Council on Library Resources provided several definitions of what professionals might consider a digital library to be:

- a collection of materials digitized or encoded for electronic transmission,
- an institution that possesses or an organization that controls such materials,
- an agency that links existing institutions for providing access to electronic information, establishing prices, providing finding aids, and protecting copyright restrictions,
- a consortia of collecting institutions,
- a library that scans, keyboards and encodes all its materials to make the entirety of its holdings electronically accessible from anywhere,
- or simply a library with Internet access and CD-ROM collection.<sup>7</sup>

Michael Hart of Project Gutenberg says an electronic library consists of computer searchable collections which can be transmitted via disks, phone lines, or other media (at a fraction of the cost in money and time as with present day paper media) featuring electronic books that will not have to be reserved and restricted to use by one patron at one time. In other words, all materials will be available to all patrons from all locations at all times.<sup>8</sup>

Jan Olsen of Cornell University says that the basic electronic library provides: resources that are located both locally and remotely, a complex of genres-bibliographic, numeric, text and spatial, a single point of entry to several hundred scholarly resources, with navigational assistance and transparent connections to any resource selected by the user, and high quality user support and instructional services.<sup>9</sup> This is an excellent librarian-centered definition.

At Marist, the concept of the digital library that we have articulated to the campus community is a central repository of large, networked databases containing digitized content in various formats with a single point of entry available anywhere on campus.

## Scale

A digital library by definition will be large--a complex or network of databases. At Marist it is meant to serve the information needs of over 155 full-time faculty, over 3,500 students enrolled in over 1,200 courses per semester.

David Barber in a recent issue of Library Technology Reports wrote about the levels of complexity in building a digital library. He bases the construct on Web-based access, although other browser or interface approaches are equally valid. The levels of complexity are defined as:

---

Level 1: Identification of a collection of digital resources, i.e. links to Web sites

Level 2: Identification of individual resources, e.g. a Web home page created with links

Level 3: Local content management with files, such as a Web page with standalone files

Level 4: Local content management with a program or search engine to access files

Level 5: Local content management with a locally developed program or interface

Level 6: Local content management with a database management system<sup>10</sup>

---

These progressive levels of digital library development establish a clear model for building or scaling functionality. The effort involved in developing more complex functionality presents one aspect of scale, the other aspect would be the amount of content and the number and types of databases that are digitized and organized.

The dynamic of the Marist-IBM partnership presents a good case study for examining the collaborative effort in building a full-scale digital library that functions at a level 6 on the Barber scale.

Marist has taken a systematic approach to the development of the digital library by implementing a four year project in two stages: the first phase modeled the electronic reserve room with a student-centered client interface for easy navigation and a faculty client interface and search engine to structure the digitized content in all media types (a course model) as a foundation technology; the second phase builds a digital library to fully support the curriculum (a research model) to take full advantage of the capacity at the server end.

This plan assumes a scalability by prototyping in incremental steps the building of course content by material types, by subject and by pedagogical approach, i.e. classroom information delivery methods with a staged increase in the number of faculty and students served each semester. This strategic approach has been planned to avoid severe technology backlash, by carefully allocating the available resources, such as staff support and digitization hardware, to meet the expectations of the faculty and students in a realistic manner as new functionality developed and demands for access and performance increased.

After the first two years of the project we have learned that approximately 25 digital items are needed to support a typical course, ranging from a low of 6 items to an aggregate of over 80 items. A digital database to support an entire curriculum will be in the thousands of objects.

We have modeled several of the critical components of the digital library: course reserves, digital journals (that is print journals that are digitized), archives, images, CD-ROM databases (non-index, non-journal) and media files, that is audio and video. The text files are relatively easy to manage and support and scale with predictable results. Images, including photographs, art reproductions, graphs and maps are also manageable. However audio and video files present several significant problems: from standards for access and delivery to storage concerns for scaling to major rights management problems.

The building of a digital library is an event that must be conceived in extended time periods. Most private liberal arts college have been building their library print resources for at least one hundred years. A digital library will take a considerably shorter period of time, since many large databases can be purchased commercially taking advantage of digital economies of scale in distribution, storage and access. However the digitization of original local holdings--that is faculty and student course work and special collections--must be viewed as a re-publishing effort that takes considerable time in planning, selection, scanning, loading , indexing and intellectual property rights management, but will pay dividends for the college curriculum as course material becomes available longitudinally and in more depth.

I contend that a true digital library will have to include locally unique materials produced by faculty and students and library archival materials that no other institution holds. These materials must be made searchable locally and remotely. The library will become publisher of student and faculty scholarly works that have not been previously collected or generally accessible. Thus the original impact of the OCLC Union Catalog on ILL in opening small library collections and relieving the burden on research collections, will be repeated when colleges mount digital collections of unique holdings on the Internet.

The scaling effort to build a digital library rapidly incurs many considerations: number of items to be stored and accessed, the number of databases to be supported, the number of access points to be designed or enabled, the number of users to be supported, the level of functionality to be implemented and the size of the digital items relative to bandwidth capability. Some of these issues are management problems, some are resource questions, and several are technical considerations that must be taken into account. Together they represent a complex problem-solving opportunity that must be carefully planned and executed.

## **Discussion of Project Collaboration**

The digital library development teams formed through inter-institutional consultation at the higher administrative levels to identify individuals who could make significant design and development contributions. Faculty and students came through the impetus of the Academic Vice President, technical staff from Marist were chosen by the Vice President for Information Services, IBM technical staff were identified through an IBM Senior Vice President.

The teams were organized under a specific plan to scale the digital library from an electronic reserve system to a full digital library. The essential functional components were:

1. functional interface for each level of user: student, faculty and system administrator
2. search and retrieval system for accessing stored digital objects
3. capture and digitization routine
4. scalable content database supporting multiple formats
5. network delivery of digital objects

Students, faculty and librarians were heavily involved in designing and developing the client interfaces with the support of programming staff from Marist and IBM. The team of librarians, faculty and technical staff from IBM, Marist and the on-line system vendor have worked extensively to design and implement a system that retrieves multiple digital object types in an integrated method.

Faculty and students joined the development effort to establish feasible teaching and learning objectives that could take advantage of digital technology. They defined critical functionality for a campus-wide system and generated specific design features for a student access interface and a faculty interface for building a digital course file. As the project progressed, faculty developed content and helped test the initial prototype system. Librarians, programmers and networking specialists at Marist and technical staff from IBM joined together to adapt, develop test and implement the basic technology platforms. The librarians played various roles, from assisting with design to establishing critical search and retrieval standards through organizing content and optimizing workflow. Administrators from Marist and IBM resolved resource allocation issues and shared decision-making with the development staff. The team members engaged to develop access to the stored digital objects needed the most demanding skills.

This work continues and is built upon the work of the first team which comprised over 35 individuals. Content creation and digitization took off immediately with librarians and faculty very eager to build a large database of digital items, in diverse formats. The principal problems concerned materials format standards, rights management and network delivery of large objects. The technical support staff have focused most of their efforts on supporting multiple data types and distribution over the campus network. Pure storage and database management issues have been handled well on the project, reflecting campus and IBM expertise.

The strategic decision was made to create a prototype electronic reserve system that could be used during the semester for regular coursework to generate feedback from faculty and students on effectiveness of digital library functionality in support of teaching and learning. The Marist Institutional Research staff developed an extensive assessment program that focused faculty and student experiences into comments and ideas that could be applied to further system improvements.

The controlled prototype phase took place over four semesters and scaled from 8 to 40 courses supporting the different information and digital format types that reflected the breath and depth of the curriculum. The electronic reserve scale-up provided a foundation for building a full digital library.

## **Conclusion**

Librarians seeking to build digital library capability need to consider the scale issues very early in the planning stage.

What size digital library are you planning to create?

What are the campus content access needs?

What are the components that will be in demand?

Who are the most skilled practitioners, users, developers, technical people that can be brought to the project?

What level of functionality can be supported locally or what has to be supplied from vendors?

### **Partners are available to assist in the development.**

Right now many people are excited about the possibilities generated by building digital libraries. Take advantage of the opportunities to work with faculty and students on campus first. They are the source of the best ideas, which will enable you to avoid some of the pitfalls of building librarian-centered systems. Look to collaborate with technical staff on campus next. Buy them dinner and drinks as often as feasible. Seek information industry partnerships. Many vendors are looking for colleges to work with. Be inclusive as much as possible, there are a lot of possibilities out there. Finally the best partners are the academic libraries in your region, get together and start planning how you can work together to reduce costs, build content and increase access.

### **References**

1. Carol Hughes and William Pfannestiel, "Practical visioning for the decade of austerity," *C&RL News* 54: 33 (January 1993).
2. Barbara Gray, *Collaborating: Finding Common Ground for Multiparty Problems* (San Francisco: Jossey-Bass, 1989), p. 21.
3. Jan Olsen, "Introduction," *Library Hi Tech* 12:34 (Issue 47, 1994).
4. Thomas J. DeLoughrey, "Journal Articles Dating Back as Far as a Century Are Being Put on Line," *Chronicle of Higher Education* 43: A30,32 (December 6, 1996)
5. David Reiman and Karen Sendi, "Development Efforts Between High Tech Firms and Academic Libraries: A Case Study of One Library's Experience," *Library Hi Tech* 10: 59 (Issue 39 1992).
6. M. Lynne Markus and Mark Keil, "If We Build It, They Will Come: Designing Information Systems That People Want to Use," *Sloan Management Review* 35:19 (Summer 1994).
7. Mary Agnes Thompson, "Council on Library Resources," *Bowker Annual*, 41st ed., (New Providence, NJ:R.R. Bowker, 1996), p. 293-294.
8. Walt Crawford and Michael Gorman, *Future Libraries: Dreams, Madness & Reality* (Chicago: ALA, 1995), p. 59.
9. Olsen, "Introduction," 36.
10. David Barber, "Building a Digital Library: Concepts and Issues," *Library Technology Reports* 32:615-617 (Sept.-Oct. 1996).