The term “digital library” means something different to everyone. The term can used to refer to everything from a website with static pages of miscellaneous links to free resources, or an in-depth, critically reviewed, annotated and subscriber-only set of digital publications in a single discipline, or it can be something much closer to a traditional library: an extended online site with a well-organized set of full-text digital resources, search functions, interactive services, options for personal customization, and ongoing maintenance of the web architecture of the site and its resources. Typically the term is invoked when discussing only one specific resource or service, yet it is used as if it has generic implications for all forms of digital information and for all types of libraries. To understand the challenges of digital libraries, one needs first to define what is meant by the term “library.” The nature of a managed library involves collections selected for a purpose, services to users, and business operations; so a true digital library should encompass all of these dimensions. It needs to be more than one full-test database, or an alphabetical list of journals and links, although many digital libraries start that way. At the same time, it is important to underscore that the provision of some digital materials and services does not mean that the particular library can meet all of its users’ needs through digital information strategies; many libraries are a hybrid of print and digital. This presentation offers an outline of issues and functions to analyze when planning a long-term trajectory for a full scope of digital information services.

 Definitions

 Definitions of the term “digital library” began being articulated and published in the 1990s, as shown in a good overview by Christine Borgmann.¹ These have ranged from the librarianly to the highly technical, and there is still no one way of looking either at this term or at the characteristics of the digital entity. A new book by Karen Calhoun entitled *Exploring Digital Libraries* opens with a chapter focused on this problem of definition, and
how it has changed over time. Calhoun analyzes many technical and service elements and proposes a definition that can reflect current needs and adapt to future trends:

Systems and services, often openly available, that (a) support the advancement of knowledge and culture; (b) contain managed collections of digital content (objects or links to objects, annotations and metadata) intended to serve the needs of defined communities; (c) often use an architecture that first emerged in the computer and information science/library domain and that typically features a repository, mechanisms supporting search and other services, resource identifiers, and user interfaces (human and machine).

(Calhoun 2014, p. 18)

Calhoun goes on to discuss the history, technological frameworks, and the social and scholarly roles of such libraries, and the venturesome opportunities presented by open access and social media.

My own basic definition is that a digital library is a curated interface with clustered, indexed and interoperable full-content resources, and that these resources are directly linked with digital services including delivery, tools for usage, and access to online consultation. But that leaves out so many angles! Complementing Calhoun’s conceptual approach, I will offer some guidelines as to how to think concretely about defining and creating robust digital library services for a particular institution. In this area of discussion, it is not always specified at the outset: is the entire library service (for the institution or project) going to be delivered digitally, or are the digital materials an extension of, or complement to, a set of existing services of more traditional kinds? I refer to the latter as digital collections, and I use the term “traditional libraries” to refer to large collections of physical materials, even though they may be supported through digital services such as acquisitions, cataloging and circulation.

Overview of Planning Elements

This is a high-level outline of the broad categories of planning that might go into designing a comprehensive digital library, something more than just a collection of assorted digital materials. The intent is to lay out the equivalent of a checklist or set of placeholders for discussion, not to provide a detailed implementation plan or operational workflow. The types of library planning elaborated below include collection planning and
selection; acquisitions and resource sharing; academic productivity and research software
tools to accompany digital content; user services and support that can be delivered digitally;
digital systems for library management and operations; cost considerations; and strategies
for phasing a digital library initiative. Throughout, the goal is to develop digital libraries
that are aligned with academic needs; that are holistic in providing resources, user services
and management support; and that are sustainable with regard to costs, infrastructure and
long-term retention of content.

Choosing Digital Collections

A digital library is an assortment of resources that have been explicitly linked,
acquired, licensed or created; it is not random or miscellaneous, it is a curated effort aimed
at meeting the needs of some particular audience. Even before designating the digital
information components for that audience, one must consider what is implied overall by the
subjects those users pursue, and whether that subject scope can be met digitally or might
require a combination of print, digital, media and other formats. Identifying the necessary
materials depends in general on three sets of characteristics about the library’s parent
institution and user population:

1. Subjects covered in the teaching and research at the institution
2. Academic level of the users and of the work being done: introductory,
   comprehensive, or advanced research require different scope and date ranges of
   materials;
3. Resources (budget, staff, technology) available and relationships with partner
   institutions.

Subject is the single most important factor and will determine how many different
bibliographic formats and products must be acquired, how comprehensive can be the digital
coverage, and how much can delivered directly at the local level or must be borrowed or
provided from another university. If the intent is to build a significant percentage of digital
resources to support research in a given field, one needs to consider:

• To what degree are the core journals in that field digitized and over what range of
  years?
• Are the major publishers in the field producing ebooks for current monographs, retrospective works, or both? The scanty availability of ebooks in academic fields is a major drawback to providing digital coverage of many subjects at present.

• Does the subject require the use of film, music, media, interactive modelling or mapping tools, or other formats that exist digitally but that require specialized equipment and expertise?

• Do users need high-quality, peer-reviewed content, or simply quick and informal information? Confusion is widespread among users as to the extent of peer review in electronic journals, whether from commercial publishers or open access sources.

• How much of the course material and texts in the classes at the university are being developed directly by faculty as opposed to being acquired from publishers or external educational sources?

In deciding whether to get print or digital versions of a book or journal, or even both, there are a set of subsequent decisions that depend, again, on funding, space, and consortial relationships. Can a library afford both versions, and even if it can, is it the most efficient use of resources to retain both? This has implications far beyond the present-day users and operations. Considerations include:

• How complete, high quality and reliable is the digital version?

• How many other libraries have print backfiles? Is there an agreement about retaining “last print copies” among these libraries? The development of consortial shared print archives has become a valuable enabler for digital libraries.³

• Is there a digital backup (separate from the current digital files) either via the publisher or, preferably, a third-party service like Portico?

• Is there enough flexibility to use the digital for all academic needs like lending and reserves?

• Is there some unique aspect to the subject – for example, its prominence as a research specialty at the institution – such that the printed versions should be kept regardless of digital availability or user preference?
• Does the library have enough space for older print volumes? This might be in the same library, in an off-site facility, or through consortial facilities – in which case, factors of climate-control, inventory control and delivery also come into play.
• Are there delivery mechanisms to provide rapid digital copies directly to users?

A major challenge at a comprehensive university is how to support a wide range of subjects equitably, everything from medieval history to particle physics. As in the print environment, librarians have to assess the financial and other resources available to them, and develop analytical models to prioritize digital services when the costs and availability may differ greatly depending on the subject.

**Acquiring Digital Content**

For purposes of cost and convenience, a library needs to use multiple concurrent strategies for creating and acquiring digital content. There is an enormous range of business models and producers for digital information, and acquisitions practice has become fairly sophisticated. Most of these transactions involve customizable negotiations and usage rights. Some things might be available at no cost, while others are very expensive; some resources are freely reusable and others have very restrictive copyright permissions. A great variety of items may need to be located including digital texts, images, film and video clips, reference compilations, statistical data, archival records, music, maps, 3-D simulation models and more. In addition to the cost of purchase or licensing, there can be costs and negotiations to clarify and secure the rights to common uses including printing, repository deposit, class reserves, interlibrary lending, simultaneous users and privacy/confidentiality of users. Large libraries and consortia have tried to develop templates and model license language to help other libraries secure these permissions consistently when dealing with publishers.4

To ensure the most flexible coverage of a given field, one should be evaluating the types of resources disseminated via all of the following channels, to the extent these are relevant for a given institution:

• New publications and databases purchased from commercial publishers and vendors;
• New resources distributed at no-cost via the Internet, but that are of high quality and
that merit cataloging and retention (often produced by educational institutions and
government agencies);
• Items and collections digitized from the library’s own holdings, either individually
or as a topical group or concerted project;
• “Mass digitizing” from library collections, through a commercial company or as
part of a consortium (for example, the Google Books Library Project or the Internet
Archive); these projects can be very advantageous in terms of cost, but there may
be little control of content and technical production.
• Unique local content from the entire home institution: faculty lecture notes, course
readings, dissertations, administrative materials, photographs, historical data;
• Collections available through cooperative agreements and regional networks of
other libraries and universities.

Such a diversity of resources risks turning into an online presence that is confusing,
poorly coordinated, and forces the user to learn multiple search protocols and delivery
options. Consolidating and scripting access to digital collections so that they are
discoverable and searchable via a single user interface (not to mention, having a single
interface for managing licenses and fees) is desirable, but can be complex and require a
great deal of local expertise and programming work. Electronic resource management
systems (ERMS) are in the early stages and there is not a standard configuration or
common set of functions in the different systems. For the front end, user-oriented
“discovery platforms” are rapidly becoming the norm for unified searching but they are not
as comprehensive or seamless as their vendors suggest. We don’t even know with any
certainty whether such aggregations help or hinder users, but most libraries try to present
their digital identity through a coherent unique interface.

Digital Tools for Academic Users

The concept of the library is no longer one of a static place where we wait for
people to arrive and they have to use whatever is sitting there, but instead it is a digital and
physical site to foster discovery, interaction and customization. The use of digital
information is so ubiquitous that most people now have sophisticated expectations as to
what they should be able to do with such information and with applications and web sites they encounter. Beyond passive viewing, users want to be able to personalize, re-use and manipulate the information they locate. Digital collections, when purchased from commercial sources, usually include the software features needed to read or interact with the content, for example page-turning, enlarging or reducing images, converting data to charts, or extracting a citation in proper bibliographic format. Additionally, most libraries try to provide a variety of digital software tools for users to manipulate, integrate and reuse digital information independent of its source. There is a reciprocal relationship between what specific digital content is in the library, and which tools may be necessary to work effectively with it; for example, a library with a lot of demographic data sets or maps almost certainly will need geographic information systems (GIS) software embedded for users. There are some essential basic tools often needed to use digital resources even without venturing into any innovative services; each of these may have its own costs and support requirements:

- Citation and bibliographic software
- File management and indexing
- Text and data mining (searching)
- Geographic information systems (GIS) and statistical analysis
- Image zoom and transformation
- Personal digital archiving and data management

Technology compatibility is a challenge when a library provides these tools, in addition to the wide variations in the digital content itself. Hardware, peripheral devices, operating systems and telecommunication networks may all have to be specifically optimized to accommodate the storage space, processing speed and external linkages utilized by the combination of tools and large sets of text and data. Librarians and staff who are assisting academic users need to have knowledge of local platforms, vendor functionality, network protocols and of course the content and scope of the material or data being manipulated. In university libraries and information centers, hybrid job categories are emerging that draw on elements of library science, informatics, web programming, instructional design, and a specific academic domain.
Digital Library Services

Digital content requires a surrounding array of computer-based information management operations to constitute a full digital library. Libraries have been conducting basic services digitally for decades, for example book circulation and searching the catalog. In the era before the Internet, these were run on local systems. Connections to external providers were through direct dial-up or file transfer, and were often asynchronous in terms of transactions and linkages to content. Today, the library website has become the place to offer not only catalogs and full-text resources but customized services and transactions responding to faculty and student needs. Students and faculty expect the web interface and most of the underlying collections to be available twenty-four hours a day, seven days a week, from any location in the world. The web site is the framework defining the “digital library,” however, libraries cannot design their sites expecting a linear progression from the home page. Many users are likely to come to embedded pages for individual journals or databases or delivery options after being routed from Google, Amazon, JSTOR and sites from other universities. Digitally-enabled user services encompass at least the following:

- Search functions for physical and digital books, journals, databases and digital collections, enabled through integrated library system or unified discovery platforms. The immediacy of the “free web” has made users impatient with secondary catalogs and indexes unless every record links to the full text. They expect all connections across layers of content and services to be made seamlessly with one-click, something rarely completely achievable because of the many different sources, licenses, and authorizations involved.

- Class reserve materials: Links to full digital content and to texts, music, and film that may have been custom digitized from local copies. Reserve readings and media are often delivered via a separate digital system and may not be linked to public web sites because of licensing and copyright terms.

- Online transactions for reference assistance, book circulation, delivery of books and digital files to offices, interlibrary loan, instruction, and special acquisitions requests. In some systems this is little more than simple email messaging, in others it may incorporate links to real-time chat, catalogs, management tracking, digital collections, and consortial services.
• Remote access from home, office or offsite locations, via authenticated proxy servers or identity management systems. This is crucial for any library service involving access to commercially-produced books, journals, music and images. A library will likely not be the only campus unit requiring such remote access and thus the negotiation of levels of authentication for different campus systems can delay or complicate the implementation of proxy access to library resources.

• Social media for publicity, news and individual communication; this has proven invaluable for contributing to the user experience and library communication, however, it has also generated a new set of policy and workload issues. Social media representing a formal institutional entity cannot be undertaken casually.

• Institutional repository: Digital archive for deposit, retention and dissemination of faculty publications, student papers, institutional archives and other local documents. The trend now is also to incorporate long-term digital preservation and data management services. Issues of copyright, privacy and security are complex in the repository environment; and while it is often faster to use a third-party hosting service, in that context the above issues are even more acute. It is implicitly assumed that the existence of an institutional or disciplinary repository is the defining characteristic of a digital library, but repositories are not as ubiquitous, as multifunctional or as comprehensive as that suggests. There are policy issues that are more troublesome than the technology when one establishes a repository. Questions to be addressed include whether faculty deposit will be obligatory, which versions can be deposited (pre- or post-publication), how long will documents be retained, will university journals and publishing be an added service on the repository platform, can some documents be restricted, and whether the repository will include university administrative records.

• Consultation and technical assistance for students and faculty who want digitizing and web design for academic projects; in some institutions this is provided by a combination of library staff and the campus computing center. The need to understand the availability and intellectual property restrictions of digital content makes it essential that library expertise be included at some stage.
Digital Infrastructure for Library Operations

A library that is maintaining large physical collections can still have an operational infrastructure that is heavily digital – and the opposite is also true. Maintaining digital resources does not always imply that the business processes of the library are being carried out through software services and online transactions. A comprehensive digitally-supported library operation will have elements such as:

- Workstations, laptops or tablets for all staff including clerical and maintenance staff; and equivalent devices and connectivity outlets for users throughout buildings, linked through wireless networks;
- Budget, personnel, facilities and other administrative transactions conducted through online systems that either connect through the university enterprise systems, or that go directly to external systems and vendors; automated file transfers and feeds to populate library financial and patron databases.
- Integrated library system or library services platform that supports management of digital objects, licenses, vendor interactions, metadata creation, archiving and repository interfaces;
- Technology-enabled spaces for students and faculty including information commons, digital media labs, smart classrooms, video conferencing, visualization walls, and virtual reality “caves.” If located across campus in other buildings, these places should be able to interoperate with digital collections from the library. It might seem contradictory, but rewiring and redesigning older buildings, though costly, is crucial to the advancement of digital libraries in educational settings.
- Software development team: Even if the library purchases software, there is often customization and original programming work for specific applications. The library can collaborate with university technologists, with external vendors and other partners on software development projects, but bibliographic systems and academic content are now so specialized in their formats and protocols that at least some in-house library staff must be dedicated to these functions. Clear definition of mission and goals is needed to keep software development focused on the educational goals and not an elaborate end in itself.
• Systems support: Digital collections support requires on-site staff to address problems with desktop hardware, web sites, mobile applications, enterprise systems, security, vendor links, digital collections functionality, repositories and specialized applications. Staff might be in the library, at the campus data center, or at vendors.

• Long-term preservation of digital information: This is increasingly a concern and something for which it is difficult and expensive to establish the architecture and the local workflows. Digital curation is more than making a back-up copy; it involves format validation, migration and replication, metadata crosswalks, error checking, security, and a long-term hardware maintenance plan. Mechanisms for long-term digital preservation include third-party not-for-profit services like Portico, community-based networks like LOCKSS, and local instantiations of software like Rosetta or DSpace. The Library of Congress’ National Digital Preservation Alliance serves as a coordination point in the U.S. for these discussions.  

• Assessment, analytics and metrics: Software that can be integrated into digital products and websites, tracking system performance and usage patterns; digital or live surveys that solicit user feedback and demographics; programs that can integrate data from multiple internal and external sources and generate reports.

Summary of Digital Library Costs

It’s all too common for librarians to hear comments such as, “Why do we need libraries, everything is free on the Internet!” Even well-informed people within the university don’t always initially perceive the significant expenses incurred in delivering and curating digital academic information. In planning a digital library or even a single digital collection, a “life cycle” approach has to be taken to assess the range of ongoing costs. To consider only the initial acquisition cost is to overlook some significant costs that pertain to providing, using and sustaining the collections. Some costs may be covered by the library, some by other academic units or consortial partners, some by external vendors and publishers, and some by the users. A basic list of costs and issues to be analyzed in each institution includes:

• Acquisitions price and types of channels (vendors, consortia, free sites); one must carefully track which costs are recurring or nonrecurring, whether there are costs to
upgrade or get new versions, or to customize a resource to function with a specific interface or operating system. Add-on membership costs, beyond the specific price of a product, may be needed to participate in consortial buying contracts.

- **Space:** needed for staff and technical operations even if all the users are “remote;” depending on where servers are housed, additional electrical power capacity, climate-control and physical security may be required.
- **Staff salaries and training:** trained staff are needed for online user interaction, acquisitions, contract management, metadata, linking, programming and vendor technical support, and user interface design;
- **Number and types of user devices to be supported:** (this may determine that multiple user interfaces need to be designed, and that additional electrical capacity or server connectivity is needed);
- **User services to be provided:** staff, peripherals or outsourced arrangements to provide digital and physical reproductions, reformatting, transcriptions, images, copyright clearance work, and course integration assistance;
- **Vendor services:** software, upgrades and customization, maintenance contracts, cloud storage, security audits, statistical reporting;
- **Servers, workstations and peripheral hardware:** amount of information to be processed and retained locally or remotely, speed and bandwidth needed for different media, array of external connections to be supported;
- **Preservation:** server storage and costs of archiving software (or fees for an external provider); term of retention of digital files and objects (affects recurring costs); reliability of the source or publisher (affects whether preservation has to be duplicated locally); local archiving versus third-party hosting; intellectual property policies (determined whether additional fees may be needed for archival retention).

On the positive side, there may be opportunities for cost saving or at least cost avoidance. Can building costs be reduced through freeing up shelving space and devoting that to other purposes or campus departments? Can some standing orders for books be converted to an “on demand” patron-driven purchasing model with the result possibly being lower annual costs? Are there discounts for bundled journals or multi-part software purchases? Is there
an opportunity for in-kind cost trading, if the library is a development partner testing new vendor software? Can some low-level physical processing functions be eliminated through using updated digital systems? There is a risk of assuming too quickly that a digital library will save money, however, it is undeniable that some savings are possible if operations are examined in enough detail.

**Strategic and Phased Development**

It is rarely possible to go entirely digital within a short period of time or a single budgetary cycle, but it is possible to develop a multi-year plan that lays out how the digital library will gradually advance in coverage and services from year to year. The plan can demonstrate incremental growth in alignment with such factors as:

- Academic priorities;
- Available institutional budget support;
- Operational readiness and the ability to increase (or reallocate) internal staff, and to expand technological capacity;
- Existing institutional infrastructure and services;
- Special opportunities through partnerships, fundraising and grants.

Almost all academic and professional libraries are digital libraries to some degree, but very few libraries are entirely digital. Those that are given the name, are usually limited to a narrow subject scope and rarely have the full range of functionality that is outlined above. The opportunities of digital information are exciting and vast, and the demand for new academic library services goes well beyond simple provision of digital texts to new services supporting the creation, integration and dissemination of knowledge across many formats. The challenge in the current environment is to define exactly what digital information, services and infrastructure are needed to advance the mission of a given library and of its parent institution, and then to determine how to provide those affordably using the profusion of technologies and resource options available.
3 One venue for tracking these rapidly emerging projects is the Print Archive Network of the Center for Research Libraries, [http://www.crl.edu/archiving-preservation/print-archives/forum](http://www.crl.edu/archiving-preservation/print-archives/forum). [21 September 2014]
5 See for example [http://massdigitization.com/index.html](http://massdigitization.com/index.html) [21 September 2014]