Planning and Implementing a Digital Library Project

Kristine Brancolini, Director
Jenn Riley, Metadata Librarian
Indiana University
Digital Library Program

www.dlib.indiana.edu
Workshop Outline

- Introductions/Review Handouts
- Background on IU Digital Library Program
- Statewide digital library planning: Indiana Digital Library Summit
- Library Services and Technology Act (LSTA) funding for digitization: New standards and guidelines
- Define/Plan/Write Project Application
- Do the project
- Evaluate the project
- Workshop evaluation
Definition of Digital Library

A digital library is a networked collection of digital objects – text, still images, moving images, sound, data – with arrangement, search features, and metadata that allow for discovery and presentation, supporting research and teaching, and with attention paid to architecture, persistence, longevity, and digital preservation.
Definition of Digital Library

A digital library is a networked collection of digital objects – text, still images, moving images, sound, data – with arrangement, search features, and metadata that allow for discovery and presentation, supporting research and teaching, and with attention paid to architecture, persistence, longevity, and digital preservation.
What is a Digital Library Project?

- Digital conversion for the purpose of online access to an information resource
- Creation of metadata for future digital conversion
- Tools for creating online resources

It is not only creating a web site. It is not only scanning to place photographs or other images on your web site. The goal to create and sustain “good” digital collections.
Building “Good digital collections”*

- Interoperable – with the important goal of cross-collection searching
- Persistent – reliably accessible
- Re-usable – repositories of digital objects that can be used for multiple purposes

A collection can be defined as a selected and organized set of digital materials (objects) along with the metadata that describes them and at least one interface that gives access to them.
Achieving “Good Collections”

- Curated, cohesive group of materials
- Adherence to standards and best practices
- We must do things the way that others do them
- We must use commons systems or create metadata in ways that allow us to share

The Internet means that we are all in it together – large and small. Digital libraries are by nature collaborative and national/international.
Indiana Digital Library Summit

- Convened in December 2003 by Barbara Maxwell, Indiana State Librarian
- Goal: “To bring together representatives from all types of Indiana libraries, as well as archives, historical societies, and allied cultural organizations. The purpose was to begin discussing how those groups can work together on the common issue of digitizing unique Indiana resources.”

http://www.statelib.lib.in.us/www/isl/diglibin/
Mission Statement

The Indiana Digital Library, a collaboration of Indiana libraries, museums, archives, and related cultural organizations, enables access to Indiana's unique cultural and historical heritage through a variety of digital formats and free distribution over the Internet. The IDL seeks to support and enhance education and scholarship for all Hoosiers for increased understanding of Indiana's past and its role in preparing for Indiana's future.

-- Approved 5/19/2004
Indiana Digital Library Developments

- Indiana State Library is going to license ContentDM for all libraries in the state.
- Indiana University plans to submit a separate grant to ISL to create a portal of Indiana-related content, using OAI metadata harvesting.
- Indiana University and Indiana State Library submitted an IMLS grant to study statewide digital library financing and sustainability.
Principles

- Adherence to national standards and recommended best practices
- Leverage existing digital library expertise in the state
- Recognize that most smaller institutions cannot build their own digital library infrastructure; partners could offer hosting
- Explore regional digitization centers and centralized hosting of digital content
New LSTA Requirements (2005)

- Local, unique content with statewide interest
- Relates to existing digital content
- Follows standards and guidelines established by Indiana Digital Library working groups
- Has P-16 educational component or potential for development
- Collaborative projects given preference

Our goal is to give you the tools to meet these requirements and recommendations; refer to the “Overview and Guidelines” document for specific information.
Most important point --

Make sure that there is a good match between the components of the project plan you developed and the LSTA categories of funding, priorities, and evaluation criteria.

You may have an excellent idea for a digital project, but be sure that it fits the requirements for LSTA funding.
LSTA Grant Summary 2005

- 24 grants were submitted – 12 were funded; list on web site: http://www.statelib.lib.in.us/www/isl/diglibin/LSTA_digital_grantFY05.pdf
- Most were from academic or public libraries
- Total requests: More than $500,000
- Amount available: $250,000
- For the first time, proposals were reviewed by an experienced committee
High Priority Subject Matter (2006)[1/2]

- Theme: “Personal narrative, life in Indiana”
  - Does it identify Hoosier traits?
  - Does it inform our understanding of Indiana history?
- New in 2006
- Content areas targeted in 2005 are still a priority
High Priority Subject Matter (2006) [2/2]

- Famous Hoosiers
- Underground Railroad
- Native Americans of Indiana
- Transportation -- regardless of time period, including canals, railroads, National Road
- Indiana at War
  - Civil War -- all aspects
  - Indiana in World War I
  - Indiana in World War II
- Communication -- regardless of time period, including early newspapers
- Architecture -- regardless of time period
Changes in 2006

- Partnership was always given priority, but in 2006 it will be required if you are not experienced.
- Website includes examples of proposals
  - Vigo County for experienced libraries
  - Jasper County Public Library for inexperienced libraries
- Priority given to libraries that contribute matching funds
- The project plan must address sustainability; how will you support this resource once the grant funding ends?
- You must list URLs for completed projects or those of your partner,
- Grant limit is $40,000
Define the Project

- **Why** you are undertaking the project?
- **For whom** you are undertaking the project?
- **What** you want the project to achieve?
- **How** you will achieve it?
- **When** you will achieve it?
- **How** will you measure success?
Plan the Project

- Plan activities that will help you develop and implement your project's solution
- Develop an evaluation approach to determine whether your project's goals are being met
- Create a schedule based on the activities and the evaluation approach
- Identify and assign appropriate personnel and material resources
- Develop a budget for your project
- Decide on appropriate sources of funding and in-kind contributions to close any resource "gaps"
Planning Activities

- Selection of content
- Select appropriate standards and guidelines
- Write the proposal
- Timeline
- Budget preparation
  - Categories of expenses
  - Calculating costs
- Sustainability planning
Selection Basics

- Purpose
- Copyright/Intellectual Property
- Audience
- LSTA guidelines from the Indiana State Library
- Selection Priorities
- Other considerations for any project
Purpose/Need

- You must be able to explain succinctly what you want to do and why; define need from users’ perspective.
  - Why is it desirable to digitize these materials?
    - Preservation
    - Access
    - Both
  - What will users be able to do with these materials that they cannot do now?
- What is the significance of the materials you want to digitize? Why these and not others?
Copyright

- What is the copyright status of this material?
  - Public domain
  - Pre-1923
  - Not renewed
- Unpublished material; author deceased <1932
- Controlled by your library/archive
  - Transferred by deed of gift
  - Copyrighted by your institution
- Permission secured
- Situation unknown or murky; not necessarily bad
Audience

- Who currently uses these materials?
- Who is the primary intended audience?
- Is there a secondary intended audience?
- In general, do you anticipate that this resource will be used by a large number of people or a very specialized or small number of users?

Indiana Digital Library is interested in educational and cultural use; genealogy projects should have an audience that is broader than genealogy researchers.
Important Points/ LSTA Priorities

- All projects require a time commitment from permanent staff – even if you have external funding.
- Selecting one project means that others must be rejected.
- Digital projects require a long-term commitment to sustain the resource.
- Local, unique content with potential statewide interest.
- Develop some tentative ideas then see if you can partner with another library; this will strengthen the proposal for LSTA funding.
Selection Priorities

- Significant collection or materials
  - Current audience or potential audience
  - Popular collections
  - Can partner to create the collection
- Meets the Indiana Digital Library collection development guidelines
- Item-level information or metadata available
- Complements previous projects – creation of a critical mass
Writing the Proposal

- Start with the selection criteria for the grant program, in this case LSTA Digitization Grants from the Indiana State Library
- Write about the significance of the project, then move to the technical plan
- Be prepared to change your project once you begin the planning process
- Assemble resources that can help you with your project planning and proposal writing

Indiana State Library Website:
http://www.statelib.lib.in.us/www/isl/diglibin/
General Proposal Outline

- Description of the Project
- Goals and Objectives
- Plan of Work
  - Intellectual access
  - Digital conversion
  - Network delivery and access
  - Preservation and maintenance of files
- Evaluation
- Impact on the community
- Primary staff
- Timeline
- Budget and budget narrative (explain elements)
LSTA Application 2006 [1/3]

- Part 1. Project Summary – brief; must fit into space on cover sheet
- Part 2. May attach extra sheets.
  - Need – not YOUR need but your USERS’ need for the project.
  - Activities – Be very specific! This was one of the biggest problems last year; too general about what the applicant would actually do with the money. More about this section later…
LSTA Application 2006 [2/3]

● Evaluation – Two aspects
  ● What will be the output from your project?
  ● What will be the impact of your project?
● Awareness
  ● How will you advertise your project?
● Timeframe
  ● List each activity and how long you estimate it will take to complete; June 30, 2006-July 30, 2007
LSTA Application 2006 [3/3]

- Sustainability of the Online Resource
  - When the grant ends, how will you keep the digital collection up and running? Storage and delivery issues.

- Budget
  - Provide a narrative explaining each category of funding requested
  - Include a statement of matching funds; what is your institution providing?

- Previous Digital Projects
  - URLs for your project(s) and/or your partner’s
Purpose of the Proposal

- Communicate clearly regarding what you want to do and why it is important
- Persuade the selection committee that you have the qualifications and experience to complete the project as described
- Request adequate funding, with a budget that reflects the work you must do to complete the project; make sure that the budget matches the project plan – not too high or low
Project Timeline

- Month-by-month activity outline; it will change once the project begins, but it’s a starting point.
- Leave time at the beginning of the project to organize.
- If you need help estimating the time it will take to do the digitization, contact other libraries with experience or consult sample proposals.
- Be sure to build in quality control.
- Everything takes longer than you think it will.
- Hiring people takes time.
Sustainability of the Online Resource

- Primarily refers to digital content management system; if you ask for a ContentDM license in the grant, how will you pay for it when the grant ends?
- If you ask for a server, how will you replace it?
- If you don’t have anyone handling technical support on staff, who will handle upgrades and maintenance?
- If you have these resources, describe them.
- The biggest point is that you have thought have these issues. If you don’t have a good plan, reconsider applying this year.
Budget Elements

- Equipment – be specific; justify the choice of scanner, for example
- Software
- Supplies – don’t nickel and dime
- Necessary training – don’t pad this
- Personnel – not full-time appointed staff; can fill in time for part-time staff
- Services
  - Metadata creation
  - Digitization
  - Lesson plans or other services
  - Your partner could be one of your service providers!
Benefits to Partners

• For a large and/or experienced partner
  • Access to complementary content
  • Opportunity to share what we have learned
  • Opportunity to learn from others

• For a small and/or inexperienced partner
  • Access to complementary content
  • Opportunity to share what we have learned
  • Opportunity to learn from others
Healthy Partnerships

- Be sure to develop a clear understanding of what each partner will contribute; doesn’t mean you digitize each other’s content. Takes many forms.
- Contributions do not need to be equal
- The proposal may be your only written agreement or you may need a separate memorandum of agreement
- Be flexible and realize that the project may have to change to meet both partners’ goals
- A vendor is not a partner; don’t confuse the two
Project Implementation

- Digitization
- Dealing with text
- Metadata
Digitization in context

- Can be one of the easier parts of digital projects but still requires careful planning.
- You don’t want to have to re-do digitization later – do it right the first time!
- If it’s done poorly your whole project will suffer.
Materials for Digitization

- Still Images
  - Printed text
  - Handwritten text
  - Photographic materials
  - Newspapers
  - Maps
  - Sheet music
  - Etc.!
- Audio
- Video
Capture once, use many

- Master file
  - Captures all “important” information
  - For long-term storage and later use
  - Little or no processing done
  - Sometimes called an “archival” file
  - Still images almost always use uncompressed TIFF
  - Captured from earliest generation practical

- Derivatives for specific uses
  - Web viewing
  - Printing
  - Publication
Quick technical primer

- Resolution
- Bit depth
- Color representation
- Reflectivity and polarity
Resolution (1)

- Often referred to as “dpi” or “ppi”
- RATIO of number of pixels captured per inch of original photo size
  - 8x10 print scanned at 300ppi = 2400 x 3000 pixels
  - 35mm slide (24x36mm!) scanned at 300ppi ≈ 212 x 318 pixels
Resolution (2)

- Higher is not always better
- Scan at highest resolution necessary to achieve your stated purpose, no higher

chart from Cornell’s online digital imaging tutorial: <http://www.library.cornell.edu/preservation/tutorial/conversion/conversion-03.html>
Resolution (3)

- Common choices
  - Photographic materials: 4000 pixels on longest side
  - Text: 600 ppi
- Read carefully best practices documents listed on your handout before deciding what resolution you will scan at
Bit depth

- Refers to how many colors are represented in an image
- Common choices
  - 1-bit (bitonal), for printed text
  - 8-bit, for b/w photographs, handwritten text
  - 24-bit, for color materials
Color representation

- **RGB**
  - Scanners generally have sensors for Red, Green, and Blue
  - Each of these “channels” is stored separately in the digital file
  - 8 bits for each of 3 channels = 24 bit color
- **CMYK** (Cyan, Magenta, Yellow and Black) is used for high-end “pre-press” printing purposes
<table>
<thead>
<tr>
<th>Reflectivity and polarity</th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reflective</td>
<td>Paper</td>
<td>Negative film</td>
</tr>
<tr>
<td></td>
<td>Photographic prints</td>
<td></td>
</tr>
<tr>
<td>Transmissive</td>
<td>Slide film</td>
<td>Negative film</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Use the right equipment (1)

- There is no single piece of digital imaging equipment that can be used for all originals!
- Your goal is to produce high-quality, re-usable images. A $1000 flatbed scanner cannot do this for many types of originals.
Use the right equipment (2)

- Loose-leaf paper, 8.5” x 11” or smaller
- Paper sheets up to 11” x 17”
- Larger paper sheets
- Bound materials
- Brittle materials
- Negatives & slides
A good imaging workstation

- PC: $750
  - 2.8GHz processor
  - 512MB RAM
  - 80GB hard drive
  - DVD+/-RW drive
- 17” or larger CRT monitor, good quality: $750
- Flatbed scanner for medium- to large-format photographs & negatives: $1000
- Scanning and image management software: $500
Produce locally or outsource?

- Local production
  - With your own equipment
- With an experienced partner institution
- Outsource
  - To reputable vendor for a fee
- In any case, YOU manage the process

Factors to consider
- Nature of materials
- Number of items
- Condition of items
- Value
- Homogeneity of materials
- Funding sources (if you are doing the project over time, with multiple funding sources, may not be able to outsource)
Quality control

- Essential part of every digitization project
- Objective criteria
  - Can be automated
  - Can check all items
- Subjective criteria
  - Require human checks
  - Must sample
Dealing with text

- Scanned page of text is just a picture; it’s not searchable
- Methods of converting to searchable text
  - Optical Character Recognition (OCR)
  - Double-keying
- Text encoding (e.g., in TEI) is yet another level of processing
Metadata

- What is metadata?
- Types of metadata
- What you need metadata for
- Descriptive metadata
- How do I pick a metadata format?
- Crosswalks
- Good practices for metadata
- Metadata creation
What is metadata?

- “Data about data”
- “Structured information about an information resource of any media type or format.”

Types of metadata

- Descriptive metadata
- Administrative metadata
  - Technical metadata
  - Preservation metadata
  - Rights metadata
- Structural metadata
- Cataloging… AND MUCH MORE!
What you need metadata for
Descriptive metadata

- Purpose
  - Description
  - Discovery
- Some common general schemas
  - MARC
  - Metadata Object Description Schema (MODS)
  - Dublin Core
- MANY domain-specific schemas
How do I pick a metadata format?

- Genre of materials being described
- Format of materials being described
- Nature of holding institution
- Robustness needed for the given materials and users
- What others in the community are doing
- Dublin Core can be a good choice, but consider all options
- More information on [handout]
Crosswalks

- For transforming between metadata formats
- Mapping from more robust format to less robust format effective; mapping from simpler format to more robust format less so
- Good practice to create and store most robust metadata format possible, then create other views for specific needs
- Often need to tweak a generic crosswalk for a specific implementation
Good practices for metadata

- Use library cataloging principles whenever possible
- Enter one value per field; repeat fields when necessary
- Clearly describing original vs. digitized item
- Clear relationships between records
- Plan for interoperability
Metadata creation

- Spreadsheets
- Databases
- XML
- Library catalogs
- Digital library content management systems
  - ContentDM
  - Greenstone
Storage

- Storage of master files
  - Local server hard disks (with backups)
    - May not have enough space
  - Optical media (CD-R, DVD-R)
    - Longevity questions
    - Need good file naming scheme and a way of keeping track of which files are on which discs
  - Digital preservation repository service
    - e.g. OCLC Digital Archive
    - Very new option, not much experience yet
  - Have partner institution store for you

- Storage of derivative files
  - Typically stored in whatever system is used for delivery
Access

- LSTA projects must produce content freely accessible to the public via the WWW
- Web design
  - Adhere to standards
  - User input
  - Look for models
Two approaches to delivery

- Basic web pages
  - Pros: Easy to do
  - Cons: Difficult to maintain, no searching, no dynamic browsing

- Digital library software
  - Pros: Powerful functionality for searching, browsing, and managing content
  - Cons: Can require high level of technical skill, can be expensive
Digital library delivery software options

- Off the shelf software
  - CONTENTdm
  - Integrated Library System (ILS) vendors, e.g. ExLibris Digitool

- Open source software
  - Greenstone

- Build your own
  - Web application on top of relational database such as FileMaker or Microsoft SQL Server

- Use a digital collection hosting service
  - e.g. ILS vendors, OCLC-hosted CONTENTdm

- Partner with another institution
Some characteristics of a good delivery system

- Easy to use and accessible interface
  - For patrons and catalogers/administrators
- Customizable user interface
  - Can integrate into your web site
- Persistent URLs for individual items
  - So that users can link to them
- Can easily import/export objects
- Can share information about collections with others
  - Support for Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)
Evaluation

- In grant proposal
  - Outline evaluation plan
  - Cover how project will impact end-users
- At end of project
  - How well you met project objectives
  - How well your online resource meets the needs of users
Evaluation Activities

- Review of the project's goals
- Data collected from the project's evaluation approach
- Interpretation of the data
- "Lessons learned"
- Recommendations for improving future projects
Digital States

- A discussion list for participants in statewide digital library planning.
  http://lists.mdch.org/bin/listinfo/digistates
- Web site with links to resources and information about statewide digital library efforts.
- Links to state web sites:
  http://www.mtsu.edu/~kmiddlet/stateportals.html
- Digital Best Practices (University of Washington):
  http://digitalwa.statelib.wa.gov/newsite/best.htm
More information

- Kris Brancolini: brancoli@indiana.edu
- Jenn Riley: jenlrile@indiana.edu
- http://www.dlib.indiana.edu
- http://www.statelib.lib.in.us/www/isl/diglibin/
Advanced Digital Library Project Planning

Kristine Brancolini, Director
Jenn Riley, Metadata Librarian
Indiana University
Digital Library Program
www.dlib.indiana.edu
Outline

- Review of major points regarding the Indiana Digital Library
- LSTA guidelines and changes from 2005
- Moving from novice to expert
- Sources of grants/Advice
- Small group discussion of project ideas/draft proposals
- Wrap up – How could we help one another?
Mission Statement
-- Indiana Digital Library

The Indiana Digital Library, a collaboration of Indiana libraries, museums, archives, and related cultural organizations, enables access to Indiana's unique cultural and historical heritage through a variety of digital formats and free distribution over the Internet. The IDL seeks to support and enhance education and scholarship for all Hoosiers for increased understanding of Indiana's past and its role in preparing for Indiana's future.

-- Approved 5/19/2004
Principles

- Adherence to national standards and recommended best practices
- Leverage existing digital library expertise in the state
- Recognize that most smaller institutions cannot build their own digital library infrastructure; partners could offer hosting
- Explore regional digitization centers and centralized hosting of digital content
New LSTA Requirements (2005)

- Local, unique content with statewide interest
- Relates to existing digital content
- Follows standards and guidelines established by Indiana Digital Library working groups
- Has P-16 educational component or potential for development
- Collaborative projects given preference
Most important point --

Make sure that there is a good match between the components of the project plan you developed and the LSTA categories of funding, priorities, and evaluation criteria.

You may have an excellent idea for a digital project, but be sure that it fits the requirements for LSTA funding.
LSTA Grant Summary 2005

- 24 grants were submitted – 12 were funded; list on web site: http://www.statelib.lib.in.us/www/isl/diglibin/LSTA_digital_grantFY05.pdf
- Most were from academic or public libraries
- Total requests: More than $500,000
- Amount available: $250,000
- For the first time, proposals were reviewed by an experienced committee
High Priority Subject Matter (2006) [1/2]

- Theme: “Personal narrative, life in Indiana”
  - Does it identify Hoosier traits?
  - Does it inform our understanding of Indiana history?
- New in 2006
- Content areas targeted in 2005 are still a priority
High Priority Subject Matter (2006) [2/2]

- Famous Hoosiers
- Underground Railroad
- Native Americans of Indiana
- Transportation -- regardless of time period, including canals, railroads, National Road
- Indiana at War
  - Civil War -- all aspects
  - Indiana in World War I
  - Indiana in World War II
- Communication -- regardless of time period, including early newspapers
- Architecture -- regardless of time period
Changes in 2006

- Partnership was always given priority, but in 2006 it will be required if you are not experienced.
- Website includes examples of proposals
  - Vigo County for experienced libraries
  - Jasper County Public Library for inexperienced libraries
- Priority given to libraries that contribute matching funds
- The project plan must address sustainability; how will you support this resource once the grant funding ends?
- You must list URLs for completed projects or those of your partner,
- Grant limit is $40,000
Indiana Digital Library Developments

- Indiana State Library is going to license ContentDM for all libraries in the state.
- Indiana University plans to submit a separate grant to ISL to create a portal of Indiana-related content, using OAI metadata harvesting.
- Indiana University and Indiana State Library submitted an IMLS grant to study statewide digital library financing and sustainability.
Moving from Novice to Expert

[1/3]

- Learn from past projects
  - Keep detailed records about what worked and what didn’t
  - Track your time on task, so you know how long each activity took – not always a perfect match with the next project, but a good start
  - Keep careful financial records; budgeting is one of the hardest aspects of project planning
- Seek advanced training if necessary
Moving from Novice to Expert [2/3]

- Create strategic partnerships
  - Choose partners who will help you increase your knowledge
  - Find partners who are more expert in some area
- Share your experiences in publications and presentations – this will push you to learn more
- Seek challenging projects; once you “master” some aspect, try something new
Moving from Novice to Expert [3/3]

- Outsource routine work; once it is no longer challenging, try to find someone else to do it.
- Consider developing a project with national appeal and going for larger grants or a different kind of grant – but be realistic.
- Build upon past projects – digitize more of a larger collection or a complementary collection
Major grantors to IU Libraries and Cultural Heritage Institutions

• Institute of Museum and Library Services
• National Science Foundation
• National Endowment for the Humanities
• U.S. Department of Education
• National Archives and Records Administration
• Andrew W. Mellon Foundation
More Information About Grants

- Digital Library Brown Bag presentation from January 25, 2006, “Grant Opportunities and Grant Writing for Digital Library Projects. Slide available online:

  www.dlib.indiana.edu/workshops/bbspring2006.htm