

Digitization in libraries: balancing tradeoffs

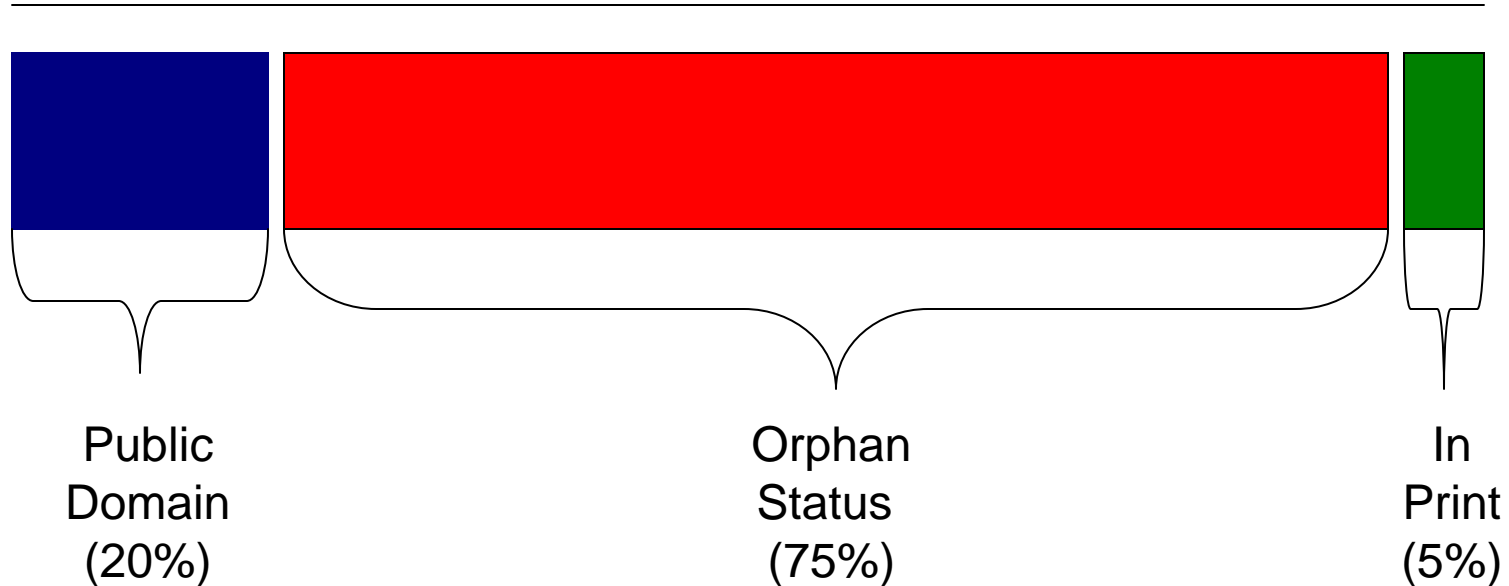
Perry Willett
University of Michigan
pwillett@umich.edu

The Context

- Mass digitization: Google, OCA, Million Books Project
- Hundreds of thousands of books online already, millions in the next few years.
- Will soon become an expectation in research, even if only to be able to search within restricted item

The orphan work problem

32 million unique titles in all US libraries



Source:

Tim O'Reilly <http://www.lib.umich.edu/mdp/symposium/presentations/oreilly.pdf>, slide 25

Set goals and priorities

What do you want to accomplish with your digital projects?

- Preservation
- Access
- Fund raising
- Grants
- Advertising important collections
- Online exhibits
- Facilitate research
- Scholarly communication
- Print-on-demand
- Collection priorities

Audiences

- Researchers
- College Students
- K-12 Education
- General Public (genealogists)

Rights/Access

- First questions to ask:
 - Are these materials in the public domain?
 - If not, who owns the rights?
 - Can you get permission to digitize?
 - Can you make it freely available?
 - Is it worth it to you if you have to restrict access?
- This issue is central to every digitization project
- Get advice/approval from campus general counsel's office

Rights on the digital object

- In the US, copyright is based upon original contributions
- In the US, there is no copyright on photographic facsimiles of works in the public domain
- What restrictions will you put on copying the digital objects?
 - Downsampling, watermarking, page-by-page access

Funding

- Some base funding is required
- Collection funds are generally the most fungible
- Grant funds require lots of matching costs, may shape/warp priorities, require long-term commitment once grant expires, contain start-up/shut-down costs

Technical considerations I

- What formats do you want to work with?
Moving from easy to hard:
 - Non-hierarchical descriptive metadata
 - Still images
 - Text
 - Audio
 - Video

Technical considerations II

- Text as special case:
 - Page images w/OCR?
 - Plain text?
 - Encoded text w/page images?
 - Encoded text w/o page images?

Technical considerations III

- Technical training needed
- Someone on your staff will need to understand the process, the standards, the technical specifications
- Get to know the vendors and what they can provide: good source of information on new developments, formats, standards

Digitization decisions

- Inhouse / Outsource?
- Costs in maintaining equipment, software
- Need for local expertise:
 - To understand specifications
 - For materials that can't be sent out
- What restrictions on outsourcing do you have?
 - Geographic? Size of collection? Handling?

Standards

- So many to choose from
- Places to go for advice:
 - Digital Library Federation <http://www.diglib.org>
 - Northeast Document Conservation Center (School for Scanning)
 - Provincial or federal granting agencies
- Develop and document local best practices: file naming, identifiers, directory structures, repository, access

Metadata formats

- Metadata—descriptive, administrative, technical
 - An expensive part of the process
 - Should be a factor in setting priorities
- Again, a multitude of acronymic choices: MARC, MODS, MIX, DC
- Uses, reuses, audiences

Tradeoffs

- Accuracy/Volume
- Access/Preservation
- Decide on the goals: what is most important to you?
- Decisions will impact cost, workflow, staffing, speed

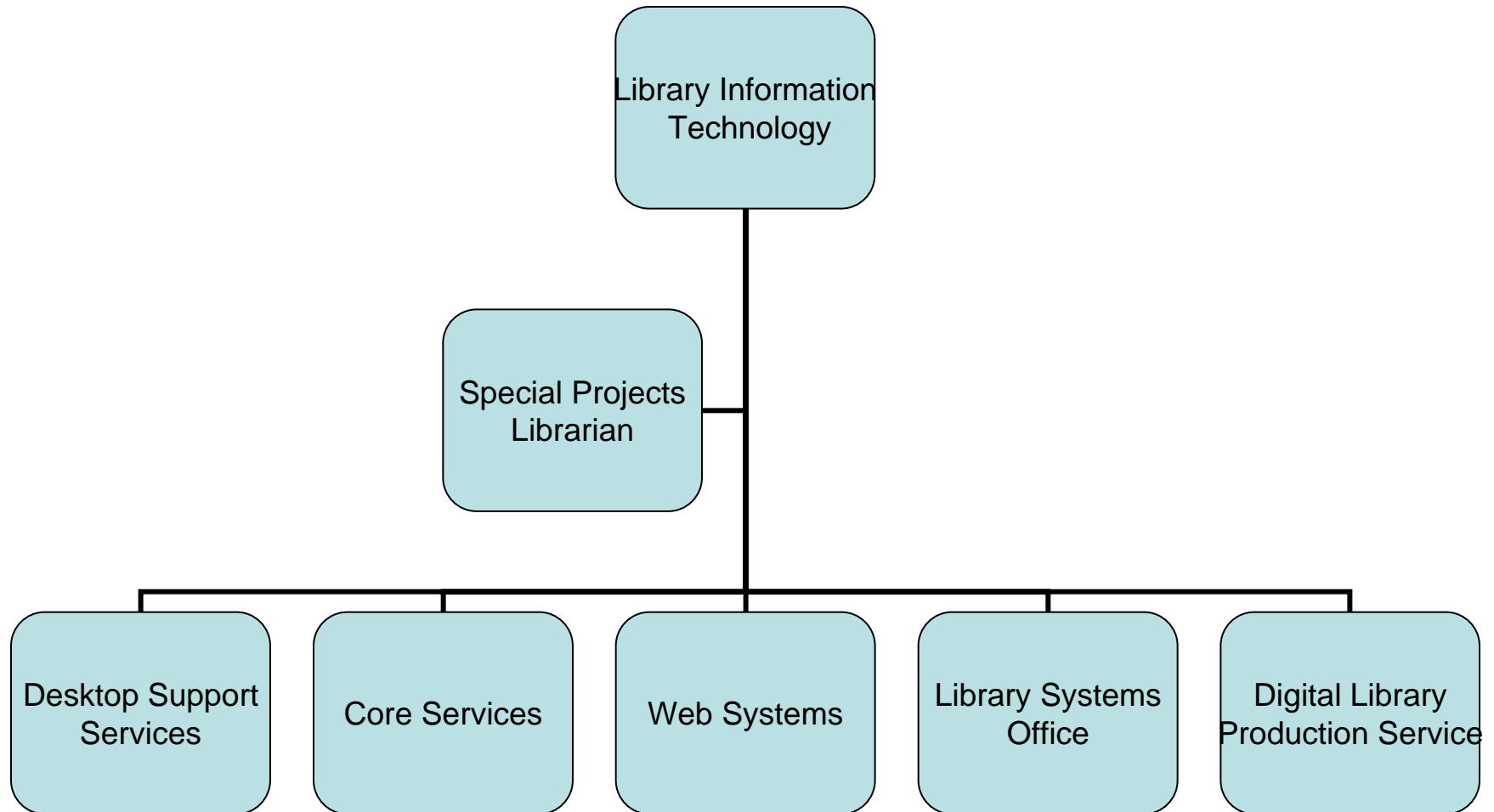
Online access

- What functionality is needed?
- What kind of collections/formats?
- Who are your primary audiences?
- What resources/staffing are available?
- How much customization do you require?
- Hardware? System admin? Maintenance?
- What are the choices?
 - Off-the-shelf: Endeavor Encompass, Exlibris DigiTool, ContentDM
 - Home grown
 - Open source tools: Greenstone, DLXS (sort of)
 - Hosted elsewhere?

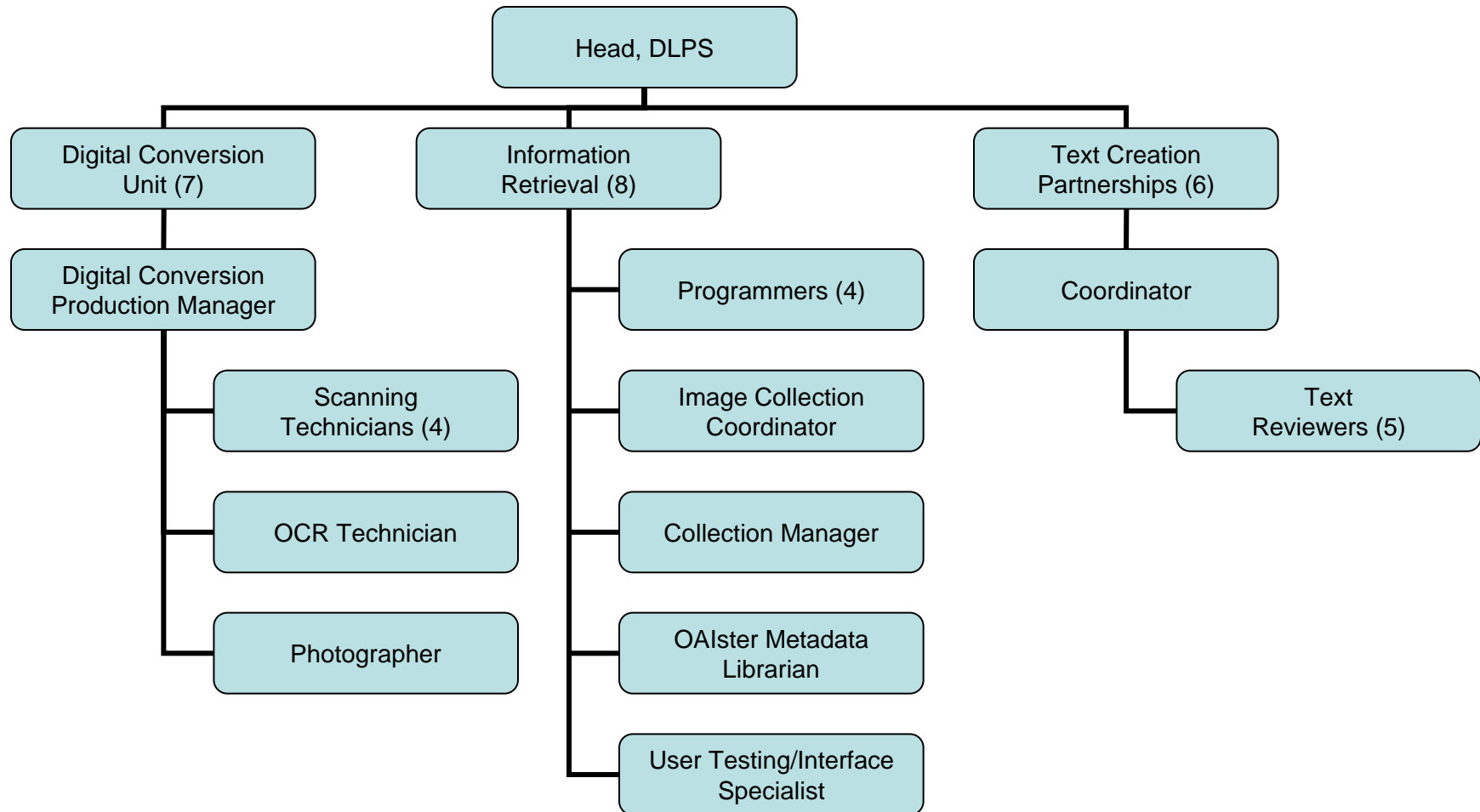
Long term issues

- System maintenance
- Growing collection/ “Finished”
- Error correction
- Reference assistance
- Staffing, funding, infrastructure, commitment
- Need to plan for the long term
- Learn from mistakes

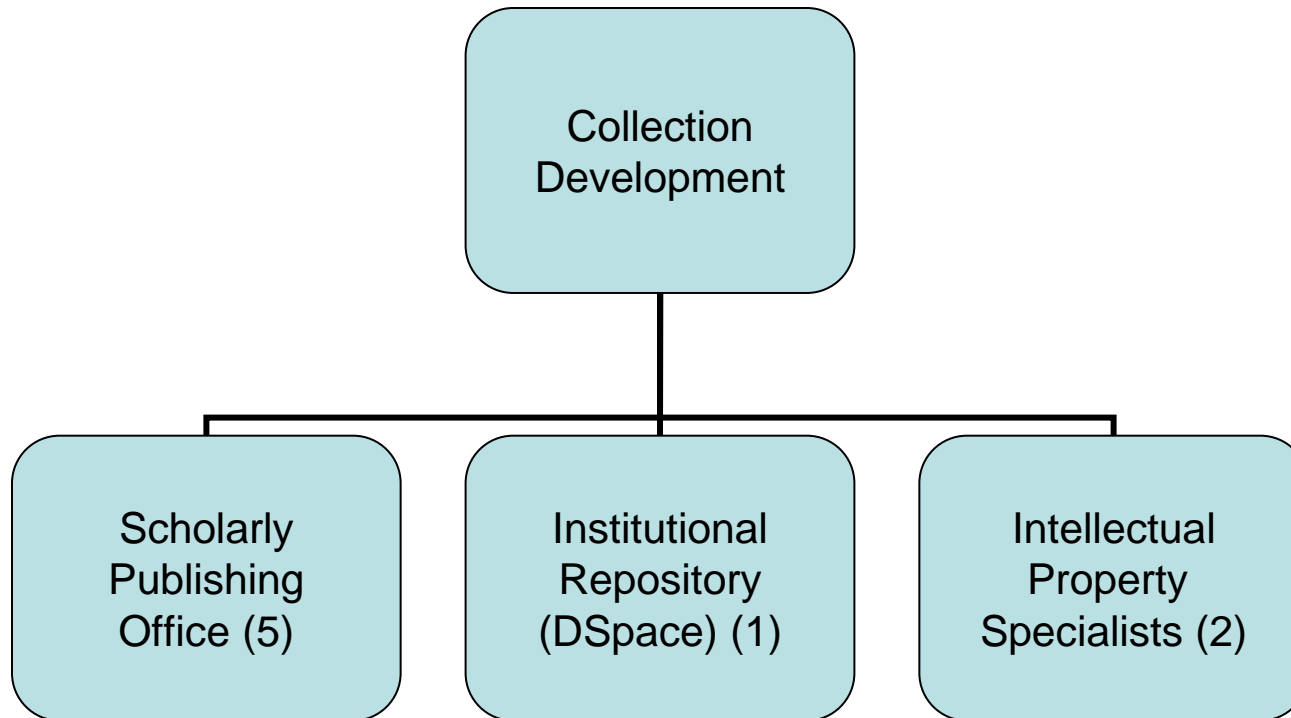
UM Organization



DLPS Organization



Related digitization staff



Others: Metadata Librarian, Catalogers in Technical Services

Google Project

- Cataloging
- Preservation
- Circulation
- Shelving
- Selectors
- Public Services

Strategies

- Collaboration
- Consortia
- Share expertise, collections, costs