Special Collections in Our Digital World

The Future Is Now

Jackie Dooley

Program Officer

OCLC Research

LIBER workshop: Heritage Collections in the Digital Future

Barcelona, 28 June 2011

Overview

1. OCLC Research
   - Who are we?

2. OCLC Research survey results (2009)
   - Digitized special collections
   - Born-digital special collections

3. Digitization
   - Our published work to date

4. Born-digital special collections
   - Our latest project

OCLC Research: *Who we are*

- ~45 FTE with offices in Ohio, California and the UK
- Sponsored by OCLC and a partnership of research libraries around the world that share:
  - A strong motivation to effect *system-wide change*
  - A commitment to *collaboration* as a means of achieving collective gains
  - A desire to *engage internationally*
  - Senior management ready to provide *leadership* within the transnational research library community
  - Deep and rich *collections* and a mandate to make them accessible
  - The *capacity* and the will to contribute

1. OCLC Research: Current portfolios

- Mobilizing unique materials
- Research information management
- Metadata support and management
- Infrastructure and standards support
- System-wide organization
- User behavior studies and synthesis

- … & the OCLC Research Library Partnership

2. OCLC Research survey results (2009)

- Digitized special collections

- Born-digital special collections

Taking Our Pulse: The OCLC Research Survey of Special Collections and Archives

# Key Findings

- The top three "most challenging issues"
    - Space
    - Born-digital
    - Digitization

- Tough economy renders "business as usual" impossible; 75% of library budgets diminished

- User demand for digitized collections is insatiable

- Management of born-digital materials is in its infancy

Top education and training needs

1. Born-digital materials: 83%

1. Information technology: 65%

2. Intellectual property: 56%

3. Cataloging and metadata: 51%

Special collections involvement in digitization projects

Large-scale digitization

Pie chart:
- Projects completed, 38.0%
- Intended in future, 36.2%
- No plans, 17.8%
- Not sure, 8.0%
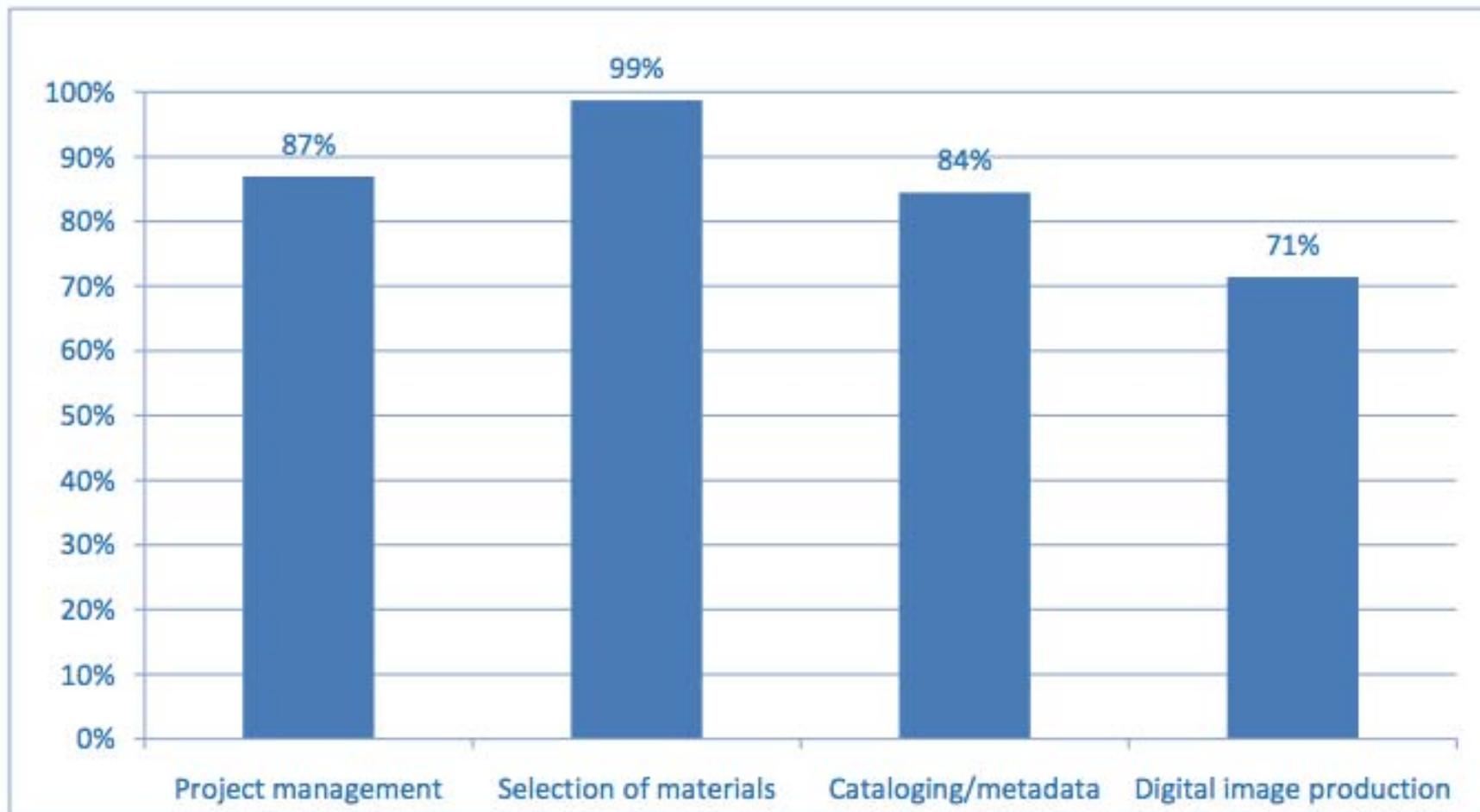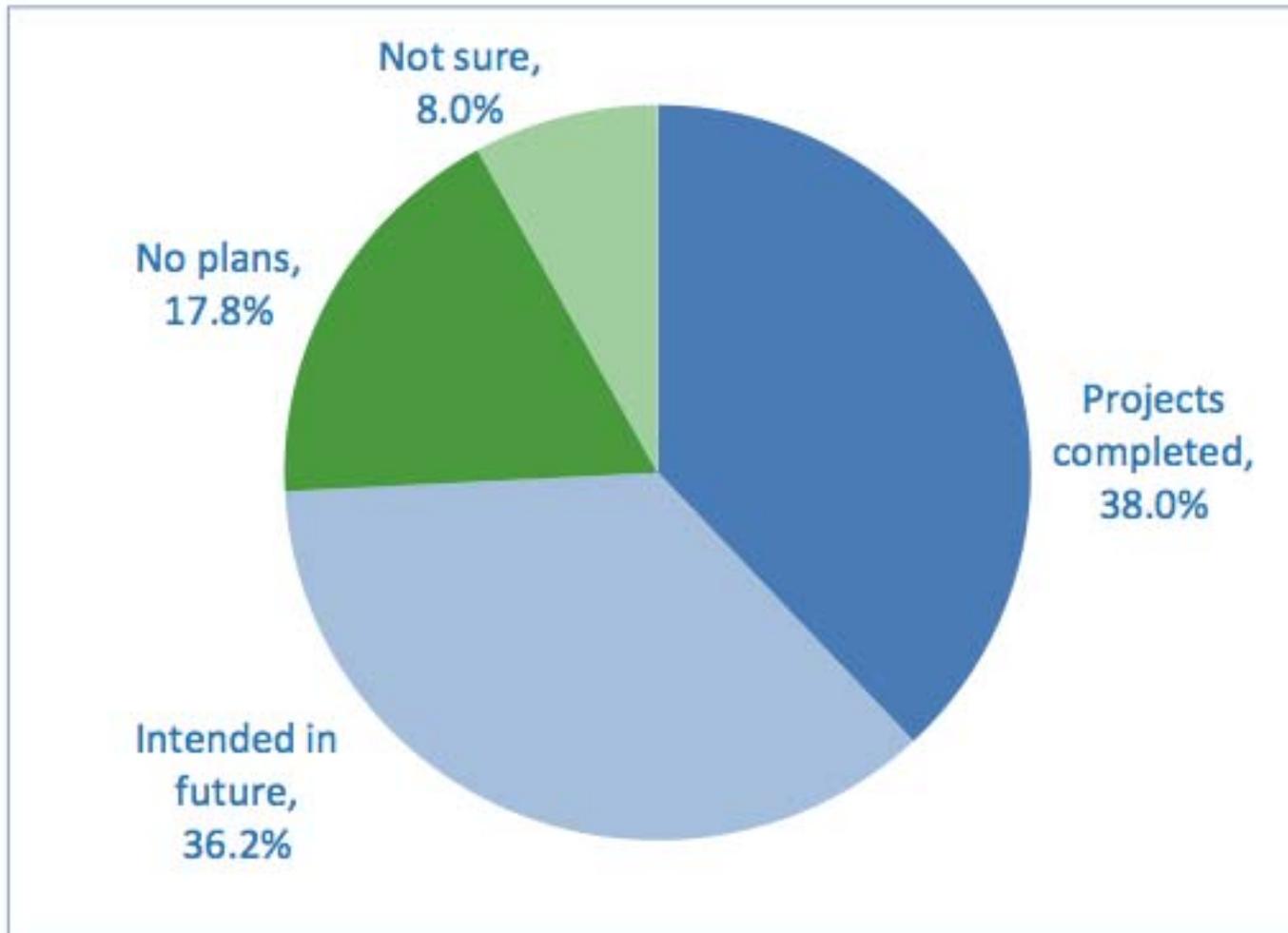
Definition: Systematic reproduction of entire collections using streamlined production methods that account for special handling.

Born-digital archival materials

In a nutshell ...

- Undercollected
- Undercounted
- Undermanaged
- Unpreserved
- Inaccessible

Born-digital archival materials

- Holdings reported by: 35%
  - Percent held by top two libraries: 51%
  - Percent held by top 13 libraries: 93%

- Digital materials currently held by: 79%

- Assignment of responsibility for born-digital management made by: 55%

- Education/training needed by: 83%

- We surmise that collecting is generally reactive, sporadic, limited.

Born-digital: Assignment of responsibility
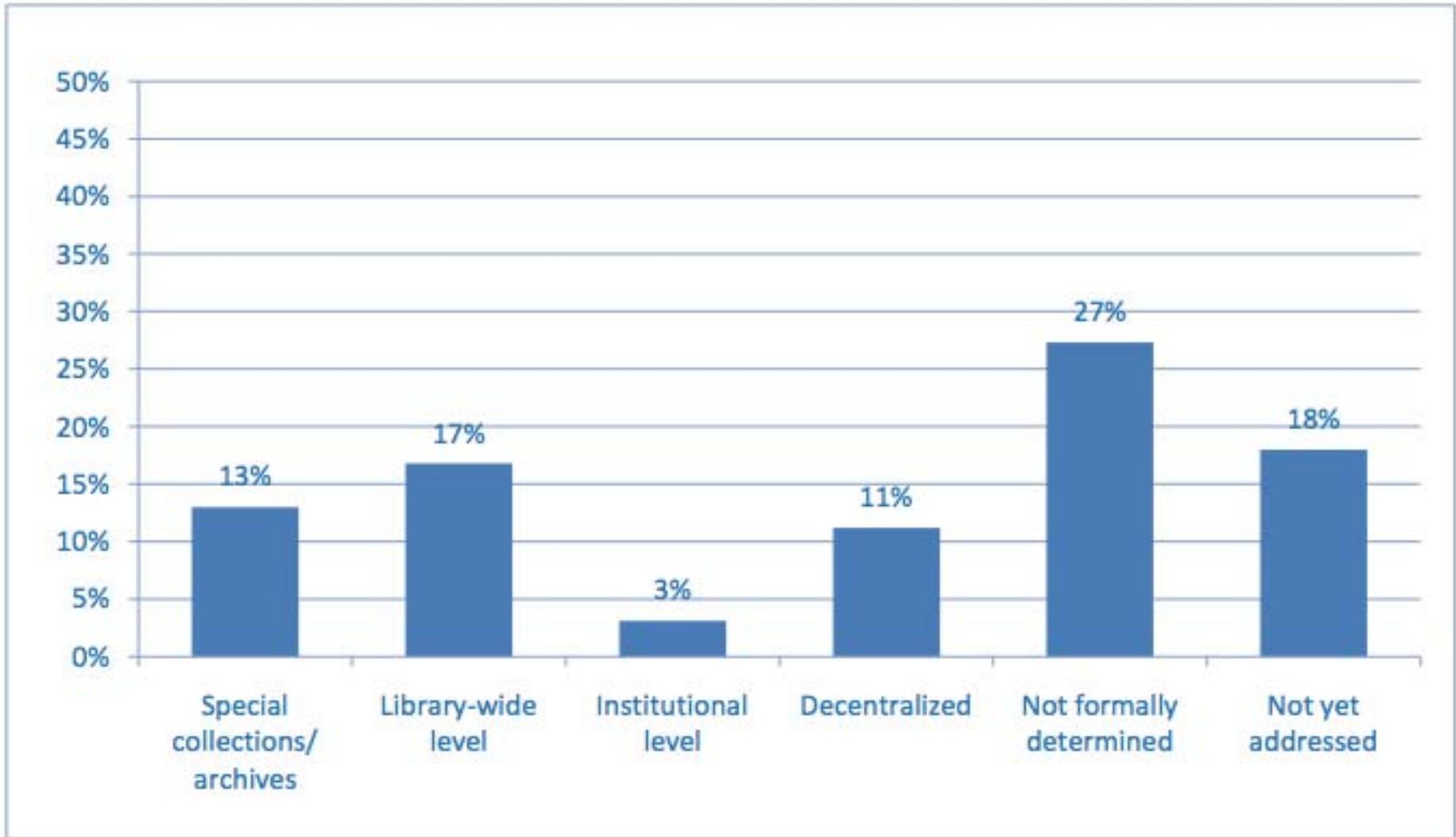
Born-digital: Materials already held
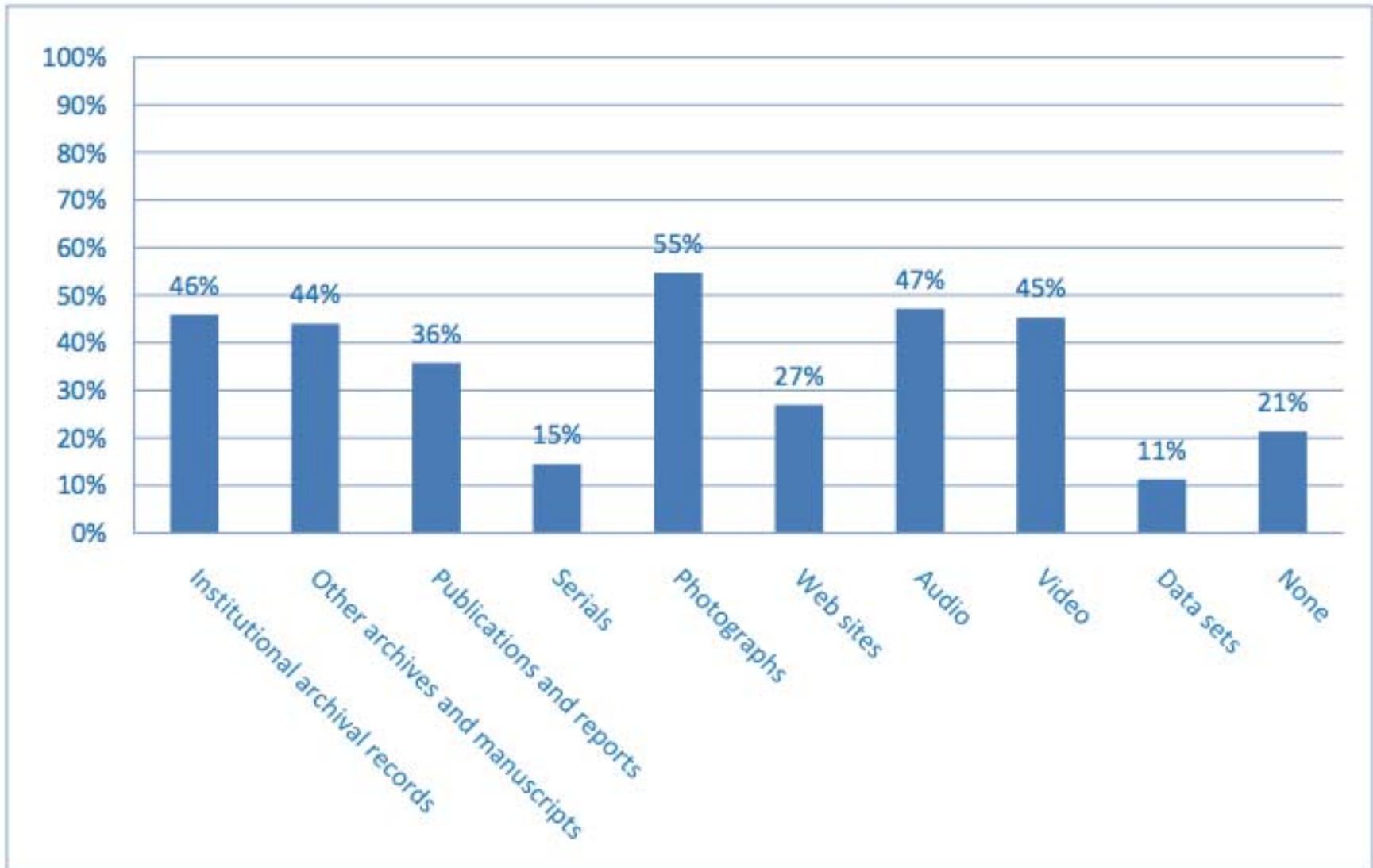
Born-digital: Impediments
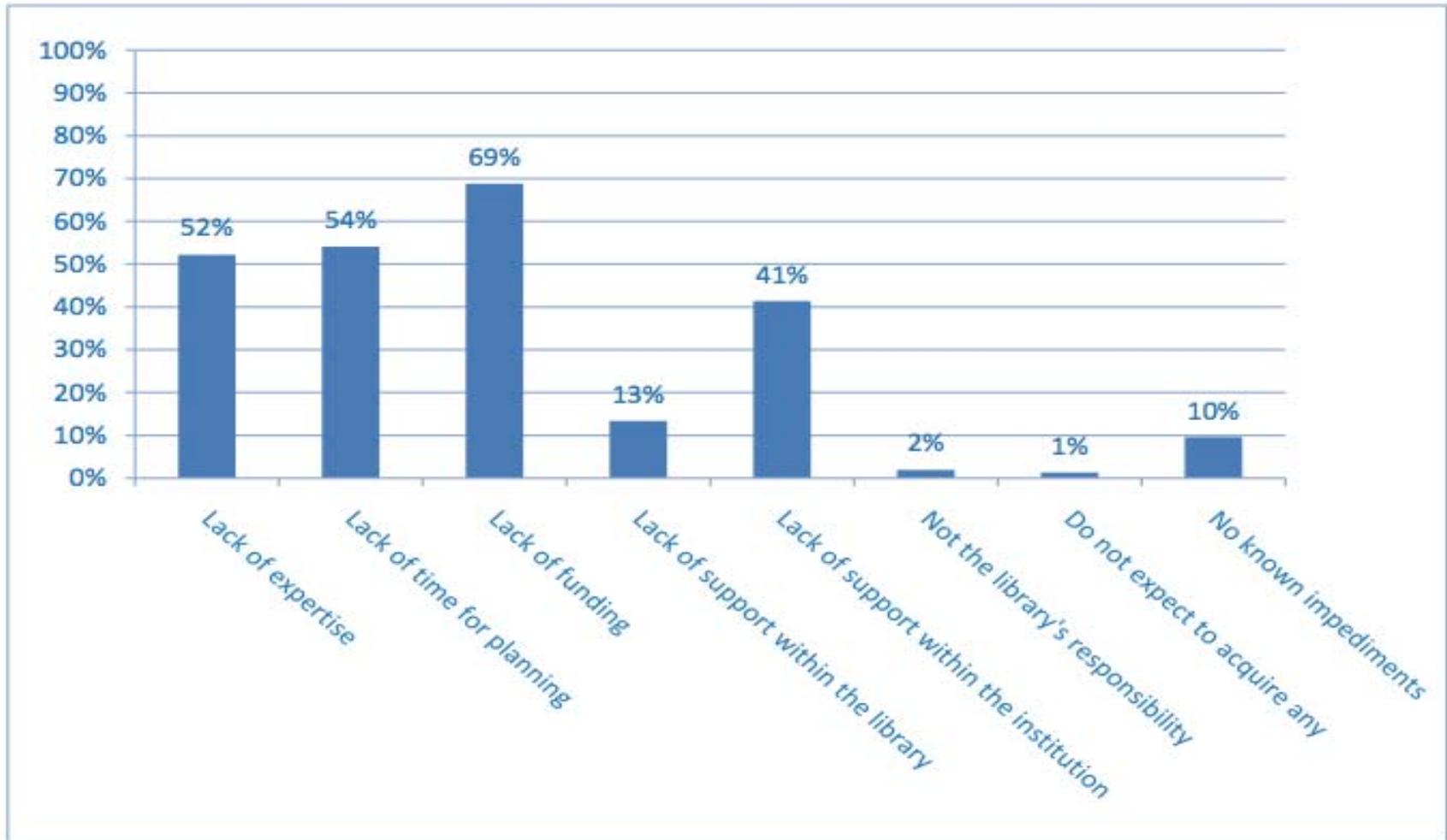
Note: Respondents could check all that apply.

Recommended actions: Digital special collections

- **Digitization**
  - Develop models for large-scale digitization
  - Scope the existing corpus of digitized rare books to aid identification of new projects

- **Born-Digital Archival Materials**
  - Identify relationship between "special collections" and "born-digital"
  - Outline "baby steps" approach to creating a program for management
  - Develop use cases and cost models for selection, management, preservation

3. Digitization: Our Work to date

- Digitization is more for **access** than for preservation
- Balancing **risk analysis** and access, with regard to copyright and privacy
- Guidance for **commercial** partnerships
- **Scaling up** capture of archival formats
- Building the digital library with researcher-initiated digitized content

Shifting Gears: Digitization for Access

Shifting Gears:
Gearing Up to Get Into the Flow

Ricky Erway and Jennifer Schaffner

Program officers
OCLC Programs and Research

A publication of OCLC Programs and Research

OCLC
The world's libraries.
Connected.

Shifting Gears: Recommendations

This project catalyzed libraries to switch from highly selective "boutique" digitization projects to thinking about ways to scale up.

- Access vs. preservation
- Quantity vs. quality
- Selection has already been done
- Do it ONCE (then iterate)
- Programs, not projects
- Archival description practices as model
- Discovery happens elsewhere
- Partnerships and funding

"Good Terms": Improving Commercial-Noncommercial Partnerships for Mass Digitization

ARTICLES

D-Lib Magazine
November/December 2007

Volume 13 Number 11/12

ISSN 1082-9873

Good Terms - Improving Commercial-Noncommercial Partnerships for Mass Digitization

A Report Prepared by Intelligent Television
for RLG Programs, OCLC Programs and Research

Peter B. Kaufman and Jeff Ubois
Intelligent Television

Executive Summary

In 2007, OCLC Programs and Research engaged Intelligent Television to study the partnership agreements between cultural institutions and for-profit companies for the mass digitization of books and other media. This report presents the findings of that study.

Libraries have been digitizing portions of their collections for more than twenty years, but recent opportunities to work with private partners, such as Google, Microsoft, and others, on mass digitization has opened up possibilities that were unimaginable just a few years ago. Private funding, commercially developed technology, and market-oriented sensibilities together may generate larger aggregations of digitized books far sooner than the library community had dreamed possible. There are many efforts underway to assess various aspects of these partnerships; this paper focuses on the terms in mass digitization agreements that affect research-community-centered outcomes.

The libraries and other cultural institutions that private companies first approached saw significant potential in these overtures; they were diligent in seeing that near-term local needs were met. Only when it became clear that a significant number of these partnerships were underway did the library community as a whole begin to think about the overall impact of these business relationships on the future of scholarship.

When we fantasize about that future, we imagine a single way to search all digitized books, journals, and other media; a combined index of all the full texts that will enable research that is otherwise

Detail of illustration by Dan Beard in Cosmopolitan Magazine, December 16, 1893. Public domain

Good Terms: Recommendations

Negotiate for:

- Limited confidentiality
- More complete deliverables
- More open access allowing the development of new applications
- Less restricted distribution
- Responsible treatment of usage data
- Limited duration and survivability

Well-intentioned practice for putting digitized collections of unpublished materials online

United States Air Force. "DAYTON, Ohio (10/27/2009) -- Members of Boy Scout Troop…at the National Museum of the U.S. Air Force" Public domain

Well-intentioned practice: Balancing rights & access

This project identified strategies for analyzing and developing acceptable risk behaviors and recommending rights clearance and digitization practices for libraries and archives.

- Select collections wisely
- Use archival approaches to make decisions
- Provide take-down policy statements and disclaimers to users of online collections
- Prospectively, work with donors

Rapid capture: Faster throughput in digitization of special collections

Rapid Capture:

Faster Throughput in Digitization of Special Collections

Ricky Erway

Senior Program Officer
OCLC Research

OCLC

A publication of OCLC Research

Rapid capture

This project investigated digital capture of special formats *at scale*. Case studies describe various methodologies, including in-house capture, outsourcing solutions, hardware (scanners, cameras, materials handling tools, and lighting equipment), staffing, and throughout rates.

- Audio
- Correspondence
- Medieval manuscripts
- Microfilm
- Mixed archival materials
- Photographs
- Scrapbooks and albums
- University records

Managing User-Initiated Digitization in Special Collections and Archives
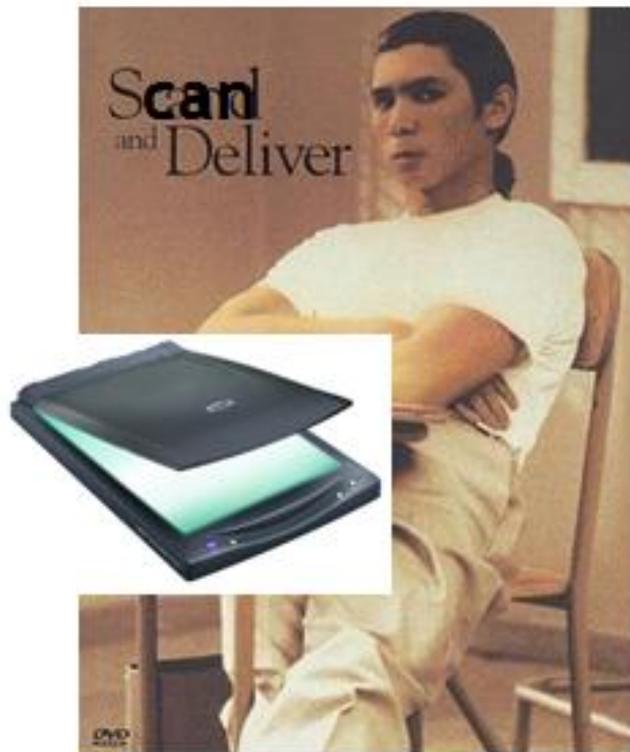
Scan and Deliver:

Managing User-initiated Digitization in Special Collections and Archives

Jennifer Schaffner
OCLC Research

Francine Snyder
Solomon R. Guggenheim Museum

Shannon Supple
University of California at Berkeley

OCLC

Managing user-initiated digitization

Sample workflows for creating digital versions of original materials in response to users' requests, ranging from a simple method that requires minimal staff resources to an elaborated workflow that facilitates addition of the files to the institution's digital collections.

Simplify each stage in the process

- Review
- Decide
- Scan
- Deliver

Born-digital special collections: Our latest project

- Types of material

- Skills and expertise

- Initial steps

Born-Digital Materials are …

- *Digital art*
- *Digital documents*
- *Digital manuscripts*
- *Digital media publications*
- *Digital photographs*

- *Dynamic data*
- *Electronic records*
- *Static data sets*
- *Websites*
- *… and more*
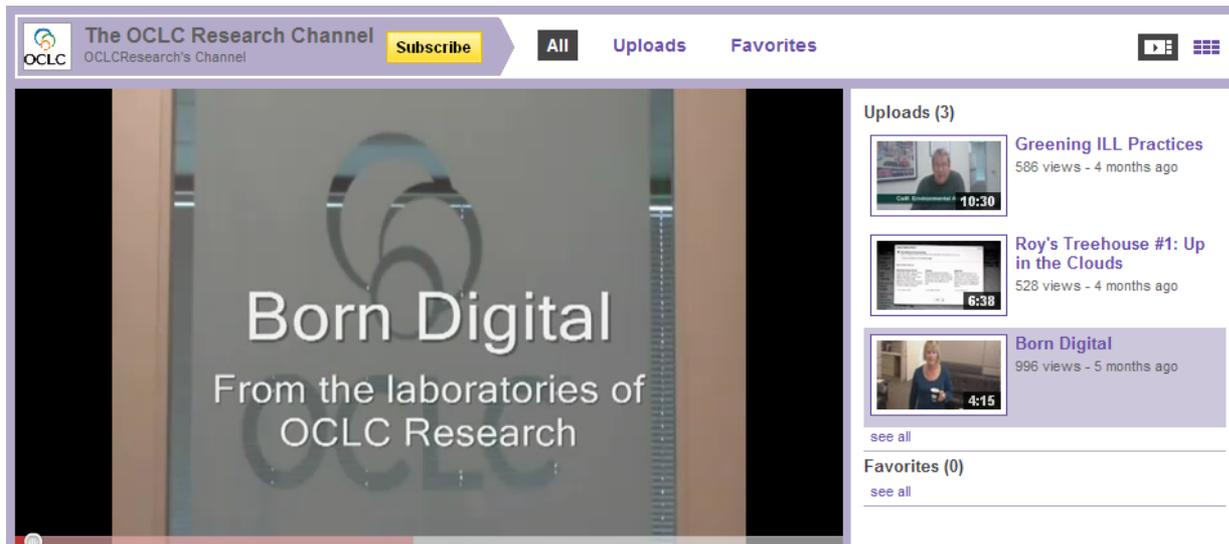
Skills and expertise

This new project focuses on effective management of born-digital materials as they intersect with special collections and archives practices in research libraries. We will:

- Explore the array of **skills and expertise** held by special collections librarians and archivists that are crucial to effective management of many types of born-digital materials.

- Describe how those areas of expertise do or don't pertain to various types of born-digital materials.

Initial steps

A third element of the new project will outline some of the most **basic initial steps** that an archives can take to begin managing born-digital materials.

Planning

- Connect with key stakeholders
- Consider key elements of donor negotiations

Managing

- Inventory what you have
- Transfer from physical media to secure storage
- Learn "do no harm" file management steps

Thank you

Jackie Dooley

dooleyj@oclc.org