ProQuest: *A Case Study of A Digital Repository*

A look at preservation of Dissertations in the U.S. and Canada

National Science Foundation Grant

NSF awarded CRL a grant to study eight long-lived data repositories.

Case Studies

An in-depth exploration of one particular case (situation or subject) for the purpose of gaining depth of understanding into the issues being investigated.

Case Studies & Repositories

- We don’t know a lot about digital repositories.
- Case studies are not about right and wrong
- We are using TRAC criteria to serve as a framework for a case study.

Our Case Studies Ask

- Key decisions
- Practices developed
- Legal obligations
- Community expectations & needs
- Interoperability/Data integration
- Incentives

Trusted Repository Audit Checklist (TRAC)

TRAC looks at:

1. Repository management
2. Data management
3. Security

Why did we look a ProQuest dissertations?

- A long-lived digital collection
- A unique digital collection
- Primary sources used by CRL members

History of UMI/ProQuest

1938 - Eugene Power starts UMI
1951 - ARL endorsement
1962 - Xerox
1985 - Bell & Howell--ProQuest
2001 - PQE stock sold on NYSE
2005-2007 - Business problems
2006 - Cambridge Information Group

What services does UMI offer?

- A searchable dissertations interface
- Full-text dissertations
- Easy citations
- Many channels for distribution and purchasing of dissertations
- Open Access
- Author copyright services
- A persistent digital archive for dissertations
- Ph.D. Verification

UMI's Designated Community

- Authors
- Universities
- Libraries
- Researchers
- Library of Congress
- National Library of Canada
- Database producers/vendors

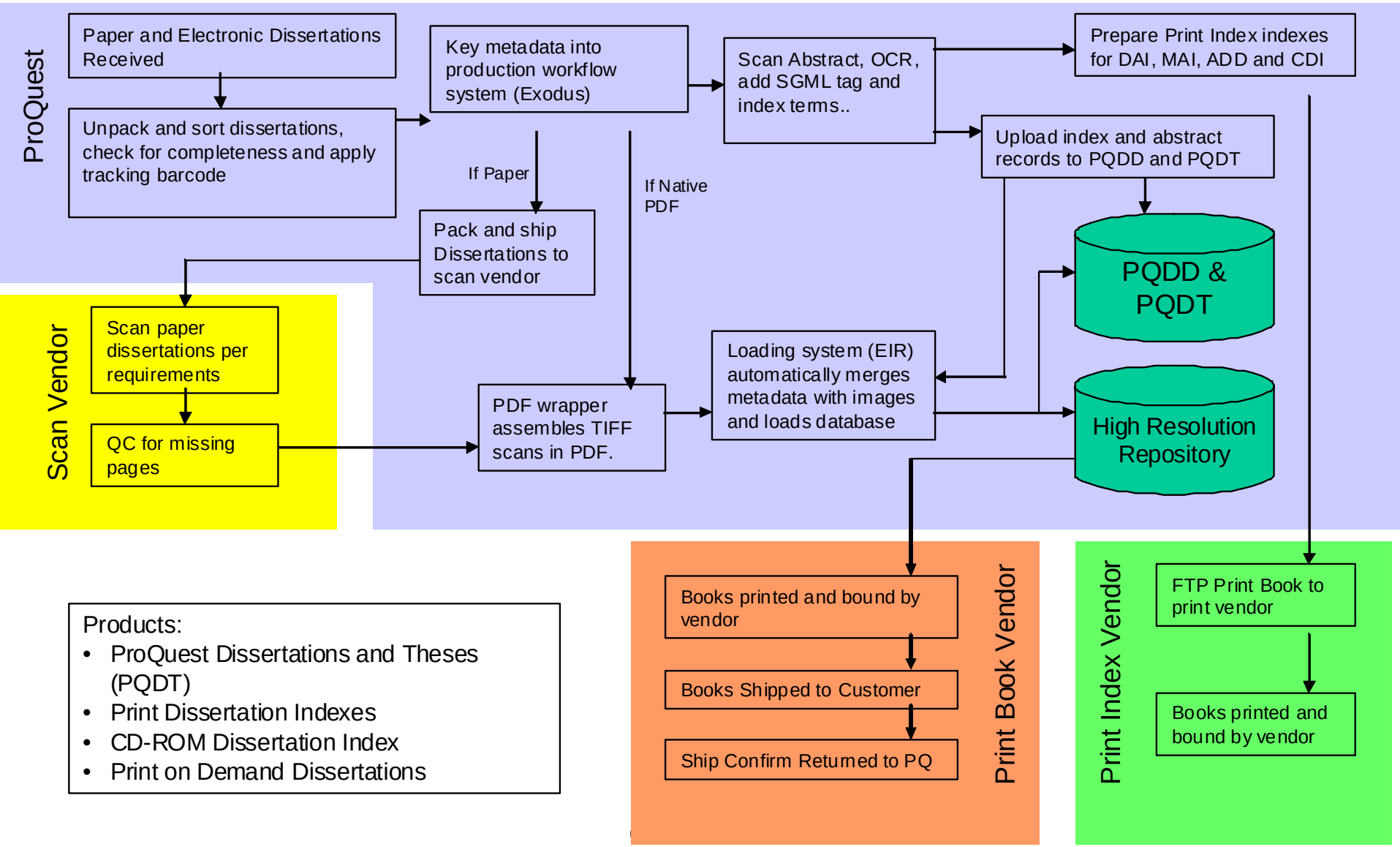
Sources of UMI Income from dissertations

- Subscriptions
- Dissertation Sales
- Authors' Payments

UMI Content

- Uniform content
- Controlled metadata

Dissertation Publishing Process Flow



Products:

- ProQuest Dissertations and Theses (PQDT)
- Print Dissertation Indexes
- CD-ROM Dissertation Index
- Print on Demand Dissertations

UMI's Preservation Planning

- Formed an Advisory Group
- Group recommendations
- Preservation Policy

Preservation Policy

- Two copies of microform
- Stored in climate- and humidity-controlled vaults
- TIFFs and PDFs in the online electronic vaults
- Daily backups
- Offsite storage
- Logs of retrieval failures

Archiving Arrangements

- Microform
- Ann Arbor facility
- Virginia facility
- Iron Mountain

Successful Strategies

- “Edition of one” was a good idea at the time and continues to be.
- The distributed cost model works.
- Collaborating is important: ARL, LC, NLC. etc.
- Uniformity of the product
- Good relations with academic community

Vulnerabilities

- Self-publishing/Open access
- More complexity content (larger size, more formats)
- Dissertations' low market value
- Preservation planning is not transparent

Digital preservation is an ongoing costs for any content provider

Digital Information Lifecycle costs = Acquisition+ Ingest+ Metadata +Storage + Access+ Preservation [1]

[1]Davies, Richard, Life2 Project Manager. "How much does it cost? The Life Project." PowerPoint Presentation for Liber conference. 5th July 2007.

Wrap-Up

No one has all the answers but advocacy is a part we can play

Questions

Center for Research Libraries
GLOBAL RESOURCES NETWORK