



THE UNIVERSITY OF  
MELBOURNE

Information Services



# Why Digitization Increases the Value of Print Collections

**Baden Hughes**

The University of Melbourne  
[baden.hughes@unimelb.edu.au](mailto:baden.hughes@unimelb.edu.au)

# Overview

- The properties of mass digitization initiatives
- Impact analysis
  - Coverage
  - Integration
  - Local digitization efforts
- Digitization, Libraries and the Long Tail
- Increasing Value of Print Collections

# Background

- Observe the intersection of collection management issues (space, access) with the emergence of a mass digitization agenda driven by Google et al
- Consider the changing nature of the “competition” across multiple dimensions
  - Information discovery mechanism
  - Information delivery platform
  - Information repository

# Mass Digitization – the 30,000 foot view

- What are Google et al doing ?
  - Partnering with libraries and publishers to digitize print materials
    - Google et al bring scaling experience
    - Libraries and publishers bring content experience
  - Blending digitized materials with generic web content
  - Delivering digitized material online in a number of ways largely determined by copyright etc status
    - Full content, limited (a few pages), snippet (search in context), bibliographic reference only
  - Providing links from digitized objects to fulfillment channels such as libraries, book stores
  - Wrapping delivery of digitized materials within the Web 2.0 delivery model, allowing for richer interactions around digitized materials similar to those elsewhere on the web

# Exclusions

- Mass digitization efforts are not homogeneously addressing all print content
  - Materials which are typically reference-oriented rather than “popular”
  - Special collections typically excluded
  - Differentiation around delivery mechanisms based on legal constraints (eg full text vs bibliographic record)
  - High quality but non-archival standard scanning (compromise between quality and scalability)
  - Typically not interested in serial collections
  - Not exhaustive: initial targets are around 12-15% of the estimated 65M titles in print

# Assumptions & Options

- Assuming
  - Libraries will typically not be able to compete with the mass digitization initiatives in terms of resource scaling
  - Libraries are generally interested in digitization as a mechanism to increase information access
- What options do we have ?
  - Join a mass digitization effort
  - Focus local digitization around distinguishing content

# Impact: Coverage

- Which objects in our print collections are going to be digitized under broad coverage initiatives ?
  - Which libraries are currently being engaged by mass digitization efforts ?
  - What public statements have been made about the scope of mass digitization with those libraries ?
  - What are the timeframes for implementation ?
  - How many of the print holdings are likely to fall into which access category once digitized ?
  - How are the print collection holdings of these other libraries similar to, or different from, our local library collections ?

# Outcomes: Coverage

- Different types of libraries are likely to be impacted differently by mass digitization owing to considerably different collections
- Access constraints will act as a high level classification – inclusion of an print item within a mass digitization initiative does not imply the same level of access online as in print
  - Which types of access are being offered for items in your print collection ? Take N items and search for them in a mass digitization repository, and note the type of access available ...
- Understanding the distinguishing aspects of your local collections becomes important, because anything that is approximately unique won't be covered
  - What proportion of your print collection is not available anywhere else, or available in restricted quantities or locations ? WorldCat is a starting point ...

# Impact: Integration

- How can we integrate these remotely held digitized objects into our local catalogues as primary access objects ?
  - Which print items within our collections are available from mass digitization initiatives ?
    - Take N items and search for them in a mass digitization repository, noting presence or absence
  - How can we create robust linkages between our catalogue descriptions and remotely delivered digitized print materials ?
  - What levels of integration support are going to be offered by mass digitization initiatives on along term basis vs being experimental ?

# Outcomes: Integration

- ILMs already support the core functional requirements for integration
  - a separation of description from actual items
  - Multiple instantiations per item, with unique identifiers
  - These are the basic requirements for integrating digitized materials anyway, so major system changes are an unlikely prerequisite
- All digitization projects have a unique identifier per item, but there's no robust public APIs for resolving between identifier schemes (eg ISBNs to system identifiers)
  - Earlier experiences with Google et al show that APIs are typically a second generation feature

# Impact: Local Digitization Efforts

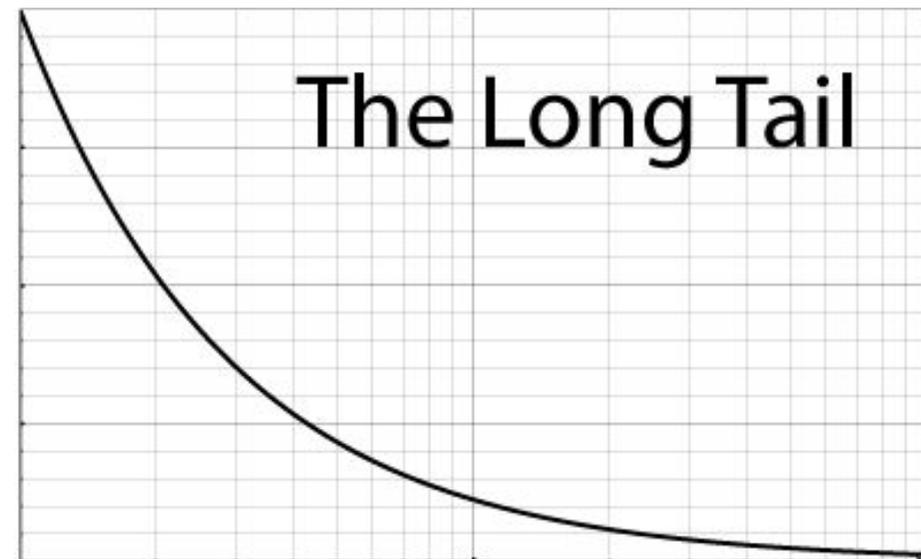
- Is there a place for local digitization efforts ?
  - What are the underlying business drivers for local digitization ?
  - What are the expected outcomes of local digitization efforts ?
  - Which types of resources are targeted and how do these differ from those under mass digitization ?
  - What types of interactions are going to be supported ?
  - What resources are required to be “competitive” ?

# Outcomes: Local Digitization Efforts

- Where is the competitive advantage of local digitization schemes ?
  - Greater focus on locally distinct collections
  - Higher quality digitization practices
  - Leveraging deeper linking options between resources
  - Heterogeneous treatment – customising approach and delivery mechanism on a per item or per collection segment basis
- If you can identify a competitive advantage in any of these areas, then there's probably a positive ROI proposition for local digitization, but only of items which are not covered by more general programs

# The Long Tail Revisited

- “The Long Tail” is the popular name for a commonly recognised feature of statistical distributions (Zipf, power laws, Pareto distributions, general Levy distributions)
- In such distributions, a high frequency population increment is followed by progressively lower frequency population increments
- It is not uncommon for the lower frequency items in total to form the majority of the overall



# Long Tail Economics

- Brynjolfsson, Hu and Smith (2003)
  - observed Long Tail distribution in the sales figures of Amazon.com exploring the relationship between number of sales and sales rankings
  - for Amazon.com, the majority of sales came from obscure items not widely available
- Anderson (2004, 2006)
  - In terms of sales, products in low demand or having low sales volume can effectively make up a market share that rivals or exceeds the relatively few current best selling items assuming a distribution channel is large enough
- Brynjolfsson, Hu and Simester (2006)
  - Impact of information technology in the form of internet search technologies which reduce the cost of item discovery can affect the overall distribution of product sales
- Do these observations hold for print collections in libraries as well ?

# The Long Tail and Collection Management

- From the supply side, the factor which determines a distribution has a long tail is the cost of inventory and storage
  - We already know this is a problem for managing print collections in libraries
- Where inventory and storage costs are insignificant it becomes economically viable to deliver relatively unpopular items
  - Is this a common driver for digitization at the global and local levels ?
- On the demand side, tools such as search engines, recommender software, sampling tools allow customers to locate products outside their local geographic area
  - Consider changing usage patterns of OPACs and convergence of OPAC and other search technologies

# The Effect of Digitization on the Long Tail

- Brynjolfsson, Hu and Simester (2006) further showed that the application of information technology in the form of internet search technologies which reduce the cost of item discovery can affect the overall distribution of product sales, in effect creating an even longer tail compared to traditional product delivery channels
- Digitization makes content discovery and access easier
- Digitization reduces the opportunity cost of every additional item in terms of inventory maintenance and storage

# Digitization and the Long Tail

- Digitization facilitates the rediscovery of long tail conditions in the information need fulfillment space
  - At a certain scaling point, the opportunity cost of storing and delivering an additional electronic item is negligible
    - Many libraries will find that mass digitization approaches offer considerably better economies of scale than they can realise themselves
  - Access is facilitated through multiple services
    - Many libraries will discover a shift to discovery mechanisms provided by third parties over standards-based virtual collections rather than OPACs

# The Long Tail and Libraries

- Libraries are traditionally in the market of information need fulfillment with little consideration for end-user cost recovery
- It has been repeatedly recognised that the economics of library services are affected by long tail distributions
  - Consider: low circulation items and storage impacts
- Digitization extends the long tail paradigm in library contexts in 2 ways
  - reducing the cost of storage and inventory
  - increasing access potential
- Digitization, both local and mass, allow for a new perspective on the economics of information need fulfillment, particularly reducing opportunity costs

# The Increasing Value of Print

- Assuming that mass digitization is successful, then some level of access to a much larger number digitized print resources will be realised
  - Flow on effect from digitization in terms of print access
- The level of access is the important differentiator for retaining and revaluing print collections
  - The general availability of print collections with **unrestricted** access via library collections will continue to offer competitive advantage to libraries despite digitization efforts
    - Analogous to the situation with uncatalogued holdings – “if it can’t be found, then it might as well not exist” becomes “if we can find it but not access it, then is this a tangible improvement”



# Questions, Comments ?

Information Services

