

Digitization of Libraries

Digital Revolution



What Do These Changes Mean...

For libraries?

For librarians?

For library patrons?

For communities?

MYTH: Digitization Will Kill Libraries



Who needs a library full of books if I can just download a digital copy to my Kindle?

- Instant access, without traveling to library
- Full-text search

Yet, eBooks sales are now in decline and paper book sales are on the rise.

And (transition to next slide...) libraries are full of

TRUTH: Modern Libraries Adapt to Serve Community Needs



How have libraries adapted?

Have they been successful? Why and why not?

Digitization of Libraries

- > Digitization of Information Retrieval and Knowledge Organization
- > Digitization of Content

Timeline: Information Retrieval and Cataloging

1960s: Computers arrive to save us all!

- Information retrieval with complex and expensive Boolean queries
- MACHine-Readable Cataloging (MARC) first developed

1980s: Online Public Access Catalog (OPAC) on LANs

2000s: Rise of the Internet

- Web-based catalogs, natural language search, and UI affordances
- Crowdsourcing metadata

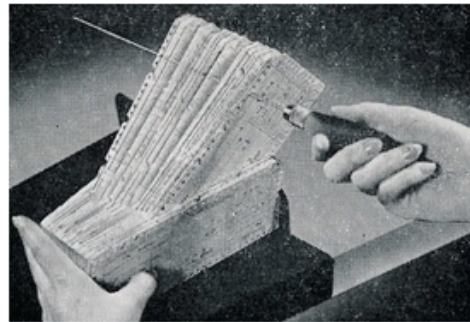
Information Retrieval

Computers and databases support faster, more comprehensive search.

- > But you still need to know how the system works.

- > And queries can be expensive.

Librarians have a very hands-on role in this process.



Information School
UNIVERSITY of WASHINGTON



Highly Structured Boolean Queries

```
DTYPE("CLINICAL TRIAL*" OR  
"CONTROLLED CLINICAL TRIAL" OR  
"MULTICENTER STUDY" OR  
"RANDOMIZED CONTROLLED TRIAL" OR  
"EQUIVALENCE TRIAL") OR  
MESH.EXACT.EXPLODE("CLINICAL  
TRIALS AS TOPIC")
```



[MEDLINE Search Fields](#)

The librarian really needs to know how the system works and how the corpus is structured, classified, and labeled.

MARc Machine Readable Cataloging (MARC)

Digital format for describing items in library catalogs

Developed by Henriette Avram in the 1960s

Became international standard in 1971

MARC 21 and MARCXML still used today



MARC Record										
Fixed Fields -- Record type: BKS										
Type	a	Elvl	i	Srce		Audn		Ctrl		Lang
Blvl	m	Form	r	Conf	0	Biog		MRec		Ctry
		Cont		GPub		LitF	1	Indx	0	
Desc	a	ills	#	Fest	0	DtSt	s	Date1	2012	Date2
LDR		00895cam a22002891a 4500								
001		223								
003		CONS								
005		20140402175013.0								
008		120125s2012	nyu#	r		000	1	eng		



Current view: Complex storage format and outdated technology – lots of decisions seem to be around saving disk space. But MARC records contain lots of detailed and granular data and the system is in place.

Joe's story: He had a library patron ask him for all of the Japanese newspapers in the Carleton library collection. Joe knew the MARC code for Japan in the country field and the MARC code for newspaper in the Type field. Put those two together and you have all of the Japanese newspapers.

Note that this system is for librarians, not patrons.

Henriette Avram: American computer scientist.

Online Public Access Catalog (OPAC)

OPAC systems first introduced in 1970s

Dynix software introduced in 1983

> Dominated libraries for almost 20 years

Self service computers in libraries

Library patrons are empowered to search by themselves



Web-Based Library Catalogs

- > Available from home and at library
- > More user-friendly search interfaces
- > Natural language search
- > Similar to Google, which is also competition



Natural Language Search



Natural Language Search

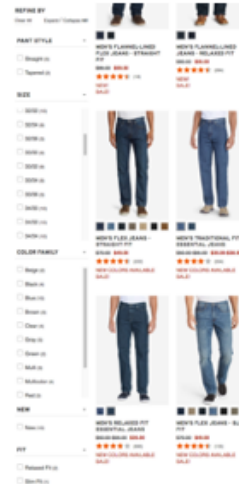
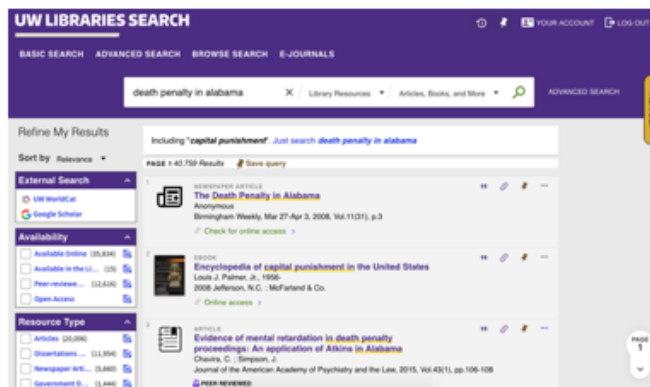
“death penalty in Alabama”

- > [Seattle Public Library](#)
- > [Google](#)
- > [UW Library](#)



Search has topic, but not author or title

User Interface Affordances



Crowdsourcing Metadata

Folksonomies and Social Tagging

del.icio.us

LibraryThing
What's on your bookshelf?

Information School
UNIVERSITY of WASHINGTON

The Seattle Public Library

COMMUNITY ACTIVITY

Comment (28) Quotes (2) Summary (1) More

★★★★☆ Oct 23, 2018 **hacker**
This book uses a series of case studies to explore the practice of incarceration in the United States. It is written by a lawyer, Bryan Stevenson, who works with Death Row prisoners in Alabama. Reading this book was an eye-opening experience that left me in an almost constant state of outrage from cover to cover. I would have even questioned much of it if it weren't for the well-documented notes in the back. I have since gone on to watch several talks and speeches he has made about his work and I have found him to be a smart, gentle, compassionate man who has really interesting things to say about justice and mercy in society. The book is not overly technical or filled with legal jargon. It is very readable and packed full of information.
O'Jah 91 17

★★★★★ Aug 23, 2018 **Waterford**
Great book. This is a must read for learning more about the injustices of the legal system, Bryan Stevenson is a truly incredible person.
O'Jah 91 17

★★★★★ Jun 10, 2018 **joachimrow**
Stevenson should run for public office. He is someone who really supports the underrepresented, poor people in our country. The prison system, esp. the private prison systems in the southern US, systematically apprehend and prosecute people to the fullest of the law often disregarding evidence of innocence. Learning the devastation of juvenile and adult lives brought tears to my eyes. Stevenson has committed his life to bringing justice to those falsely accused or given sentences in

LISTED [+ Add](#)

STAFF LISTS THAT INCLUDE THIS TITLE

- Death Penalty No More
- Race & Social Justice Nonfiction and Biography for Teens Grades 6-12
- More West Region Staff Picks
- If you like "Evicted"
- Chelsea's staff picks

[Show more](#)

LISTS THAT INCLUDE THIS TITLE

- Books by Recommendation
- Murphy Wilmet
- Prison Industrial Complex
- Bellevue Friends of the Library Book Discussion Titles for 2018-2019
- A Peek into Their Lives

[Show more](#)

TAGS [+ Add](#)

narrative nonfiction (2) non-fiction (2)

alabama (1) best of 2014 (1)

black lives matter (1)

[Show more](#)

SIMILAR TITLES [+ Add](#)

- The New Jim Crow

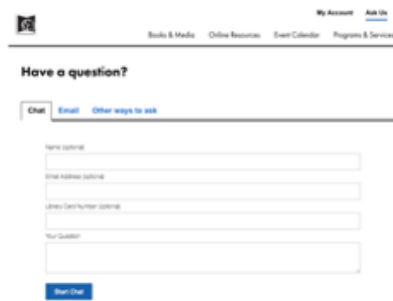
W

This is what the DeZelar-Tiedman reading was about.

People (no experts or librarians) tag content and others can use those tags to find information. It is like crowdsourced cataloging.

- No controlled vocabulary.
- No hierarchy.
- Reflects natural language
- Consensus typically emerges for "aboutness" of resources

Digital Communication with Patrons



My Account Ask Us

Books & Media Online Resources Event Calendar Programs & Services

Have a question?

Chat Email Other ways to ask

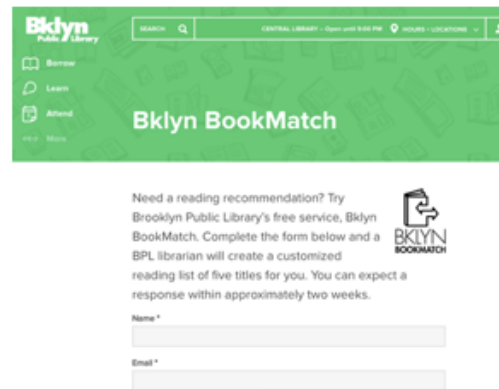
Name (optional)

Email Address (optional)

Library Card Number (optional)

Your Question

Start Chat



Bklyn Public Library


SEARCH

CENTRAL LIBRARY - Open until 9:00 PM HOURS LOCATIONS

Browse Learn Attend More

Bklyn BookMatch

Need a reading recommendation? Try Brooklyn Public Library's free service, Bklyn BookMatch. Complete the form below and a BPL librarian will create a customized reading list of five titles for you. You can expect a response within approximately two weeks.



Name *

Email *

<https://www.bklynlibrary.org/bookmatch>

Questions

Information retrieval has moved from experts using complex systems to more user-friendly search.

> Is that a good thing?

Talk about the librarians evolving role in helping people find information.

- How to formulate a search
- Knowledge about key resources and journals
- Google-style searches are better for finding a result or two, not for exhaustive research
- Joe's story about MARC records and Japanese Newspapers

Questions

Information retrieval has moved from experts using complex systems to more user-friendly search.

> Is that a good thing?

> Should cataloging move in the same direction?

Make sure that the cataloging discussion covers Hjørland and the question of whether you need to catalog at all if you have full-text search.

Timeline: Digital Content

1930: The “Readie”

1967: Hypertext Editing System

1971: Project Gutenberg digitizes first eBook

1993: Launch of Mosaic, the first browser

1995: JSTOR digital library for academic journals

2000s: Rise of the Internet

2004: Launch of Google Books

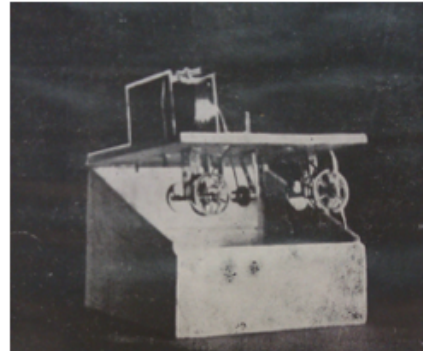
2007: Launch of Amazon Kindle



Transition: But digitization isn't only about information retrieval and cataloging. The content itself is increasingly digital. What does that mean?

The “Readie”

- > Proposed in 1930 by Bob Brown after watching his first “talkie”
- > *“To continue reading at today’s speed, I must have a ... simple reading machine which I can carry or move around, attach to any old electric light plug and read hundred-thousand-word novels in 10 minutes if I want to, and I want to.”*
- > The Readie would “allow readers to adjust the type size and avoid paper cuts.”



Talkie = movie with sound

Hypertext

- > One of the most popular formats for digital content because of links.
- > 1967: Ted Nelson & Andries van Dam develop Hypertext Editing System.
- > 1989: Tim Berners-Lee prototypes hypertext system call WorldWideWeb.
- > HTML becomes the *lingua franca* of the Internet.



Mention Borges short story and Vannevar Bush's memex

Note that HTML/Hypertext is a way to deliver content , it is also key for information retrieval for two reasons – 1) moving from one resources to the next via links and 2) Machines using links to understand how documents are related.

Photo from 1969 “shows HES on an IBM 2250 Mod 4 display station, including lightpen and programmed function keyboard, channel coupled to Brown's [IBM 360](#) mainframe.”

The First eBook

- > In 1971 Project Gutenberg digitized the U.S. Declaration of Independence
- > Used mainframe computer on ARPANET
- > Not just an image, but digital text.
- > Project Gutenberg is still going, with 57,000 items – mostly public domain books.



Talkie = movie with sound

The First Internet Browser

- > 1993: Mosaic Browser launches
- > Evolved into Netscape Navigator
- > Free to download and (relatively) use to use.
- > Internet content is available to the masses*
- > Browsers continue to evolve: Internet Explorer, Firefox, Chrome



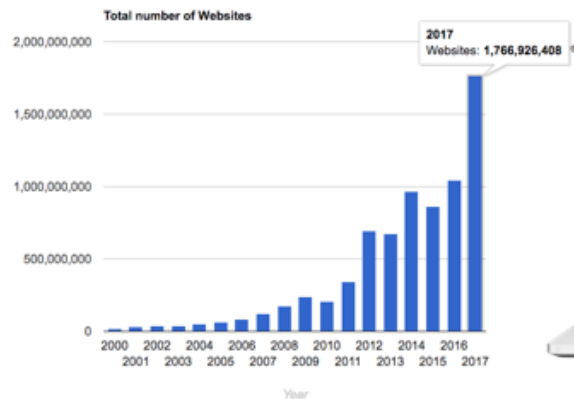
Academic Journals Online

- > 1995: JSTOR launches
- > Late 1990s: Academic journals move their content online.
- > Subscription model: libraries spend a **lot** of money for access to journals.
- > Open Access



I couldn't find a definitive answer about which academic journal was online first, but JSTOR made a bunch of academic content available online early in the game.

Online Content Explodes



There's a ridiculous amount of content online.

- Some is unique content that isn't in library collections.
- Some is digital versions of traditional library content, like journals and periodicals.
- Some is freely available, some requires payment – often libraries pay these subscription fees on behalf of patrons.

So, with all of this content online, much of it free, why do we need library content?

Google Books

- > Started in 2004, partnering with university and public libraries
- > Scanned over 25 million books
- > Used OCR (Optical Character Recognition) to support full-text search
- > Widely criticized for copyright issues
- > Also criticized for scanning and metadata errors
- > Despite winning lawsuits, no longer very active



Copyright: Went outside of public domain

eBooks

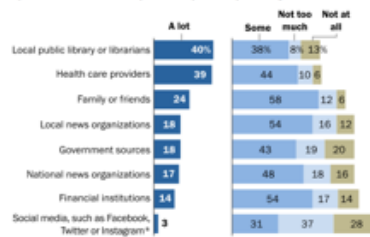
- > Amazon Kindle launched in Nov 2007
- > eBook sales rose steadily over the next decade
- > eText format supports full-text search and it formats like a book
- > Many libraries now offer eBooks as part of their collections



Libraries, Technology, and Trust

Americans are not fully trusting of information from key sources. Libraries and health providers top the list

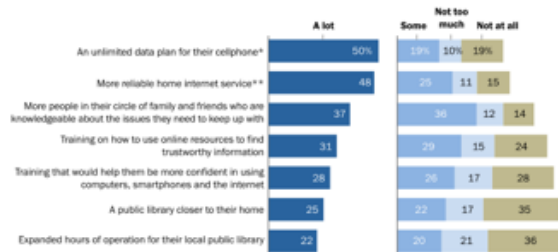
% of U.S. adults who trust information from the following sources ...



*Among social media users
Source: Survey conducted Sept. 29-Nov. 6, 2016.
"How People Approach Facts and Information"
PEW RESEARCH CENTER

Mobile and home internet connectivity top people's wish list for aids in getting information to make decisions

% of U.S. adults who say the following things would help them in making decisions ...



*Based on cellphone users.
**Based on those who use the internet at home.
Source: Survey conducted Sept. 29-Nov. 6, 2016.
"How People Approach Facts and Information"
PEW RESEARCH CENTER

Questions

> How have libraries adapted as more content has moved online and into eBooks?

Cost of internet connection and computer
Paid access to journals
Guidance about what's good
Traditional library content that isn't online

Another question: How do you look for internet-native content AND traditional library content at the same time?

Questions

- > How have libraries adapted as more content has moved online and into eBooks?
- > Have libraries succeeded?

Cost of internet connection and computer
Paid access to journals
Guidance about what's good
Traditional library content that isn't online

Another question: How do you look for internet-native content AND traditional library content at the same time?