

Wrangling the Megalith:

Mapping the Data Ecosystem of the
Harvard Library

Mark E. Shelton, MLS, Ed.D.
Assessment Librarian
Harvard Library

Southeastern Library Assessment Conference
November 16, 2015

- Relationships
- Time Travel
- Defining the word “The”
- The Human Brain
- The libraries at Harvard University

It's Complicated

- 73 Libraries employing 931 staff across 12 schools, Radcliffe, and the Harvard Library
- Each one offers different services, open different hours, and supports different citation tools
- Over 20 million volumes, 14.5 million titles, 400 million manuscripts, 2 million photographs, 180,000 serials, over 175 TB of digital files
- And we spend over \$120 million dollars

If you do not believe
me

- Student Tuition Dollars pay for us
- If it is complicated to us, how complicated is it to those who use the library?
- Others benchmark against us
- Do we really know what we are getting for our time, staff, money, effort?
- We claim we are out to achieve certain things
- Our worth is more than just vanity numbers

So why do we care?

- We have to figure out who we really are
-figure out where we are
-link where we are to where we want to go
-decide our story is more than just one-dimensional
-decide our story has bigger implications
 - our students, our faculty's, our own success
-map our own data, link it together, and link it to what we claim is our values, motivations, purpose

So what do we do?

- Data comes in many forms
- It comes from many sources
- It can be looked at in many different ways
- It is more than just the standard stuff
- We have to think beyond the silos
- The data is linked, we have to follow the path

Examining the data ecosystem

- If we are going to pull it all together....
- We have to think about those linkages....
- Where do you see them??
- Mapping the data is a TAXONOMY problem!!

A mapping model

- Model is the result of...
 - Examining all the sources of data
 - Attending meetings where assessment is a topic
 - Talking to the people gathering their own data
 - Reporting ARL Statistics, IPEDS, ACRL
 - Gathering data from ITHAKA, MISO, COFHE, internal surveys, etc.
 - Responding to lots and lots and lots of questions

How did I get here

Collection

Object

Library

Loot (aka - Finances)

Activity

People

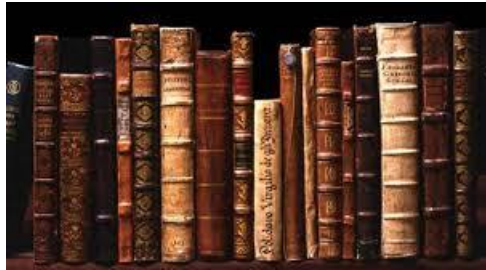
Systems

Events

COLLAPSE

- Archives and Special Collections portal lists 59 different digital collections
- Many of the physical libraries house a variety of different collections
- We have collections within collections!!!
- Some collections represent content located at multiple locations, physical and digital
- Harvard has 1,364 defined circulation collections

Collection



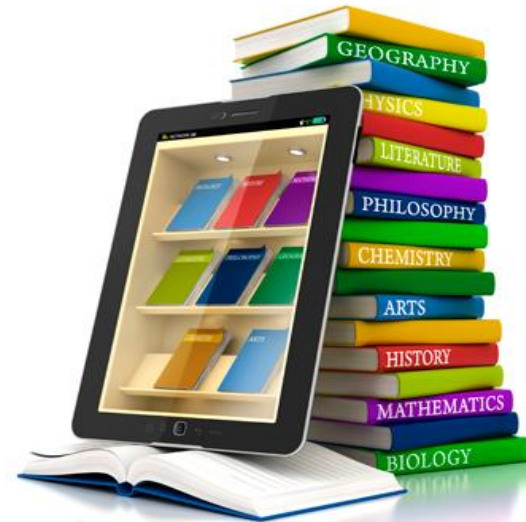
Object

Physical places -

associated with schools, rooms, carrels

113 circulation locations

Digital places -

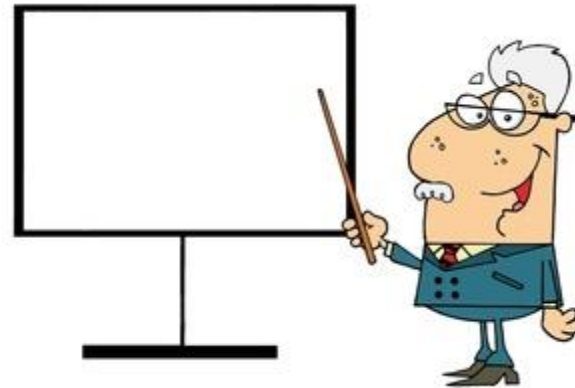


Library

- Budget
 - Projected
 - Real
 - Strategic Alignment
- Revenue
 - Sources
- Expenditures
 - To whom
 - How much
 - What for



Loot (Finances)



Harvard has 78 different borrower types, and 53 different borrower statuses

Person

- Internal Systems
 - ALEPH/HOLLIS, DASH, DRS, PDS, OASIS, VIA, SFX, VERDE
- External Systems
 - All that digital content (ebooks, journals, databases, etc.)
 - Citation tools
 - Collaboration tools (ILL, Borrow Direct, etc.)
 - Content acquisition (EBSCOHost, Gobi, etc.)

Systems

- Reference Transaction
- Circulation Transaction
- Turnstile
- Re-shelf
- Article Download
- Instruction Session
- Web view
- Twitter post
- Book Purchase



Event

- COLLAPSE represents the top level
- Each one of the categories has to be broken down further.
- Consider a turnstile

PID	TURNSTILE	DATETIME	DATE	YEAR	MONT H	WEEK OF YEAR	DAY OF WEEK	HOUR	LIBRARY	Count
219412	Wdnr Mass Turnstile (mag)	Sep 1, 2015 1:33:17 AM	Sep 1, 2015	2,015	9	36	2	1	Widener	1

- Or MARC record for an object

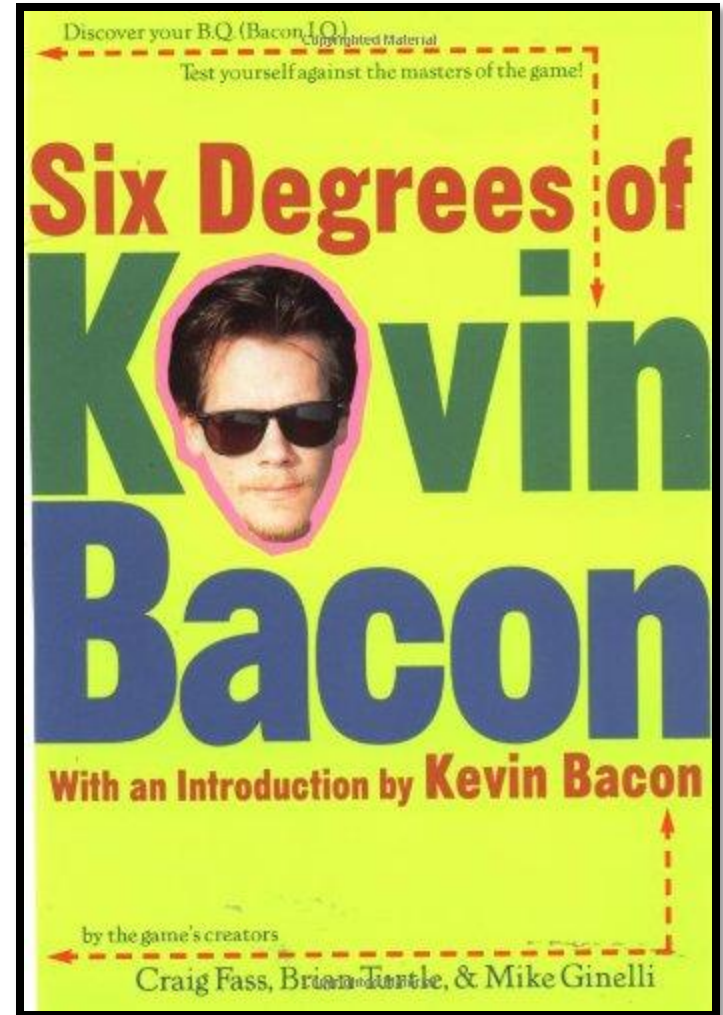
The next level

- The source of the data
- The format of the data
- Method of extracting the data
- What fields are the linkage points?

Next Level cont.

- From the model, we should be able to fully link up all of our data (or at least a vast majority of it)
- There is six degrees (or more) of separation between points.
- This allows us to explore the data from a variety of angles

Kevin Bacon

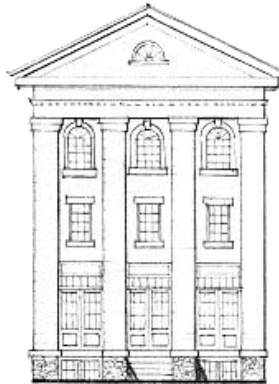
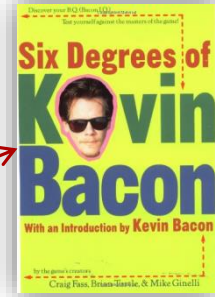




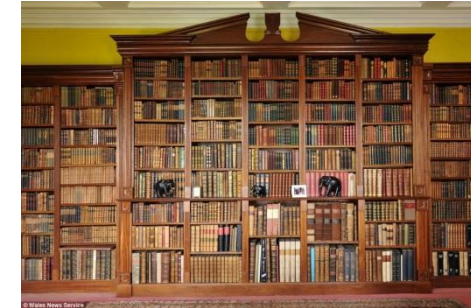
Person

Events

Object



Library



Collection



Loot

Activity



A Simple Example

- How many students from XY school use books purchased with endowment AB?
- Will posting an article about the BIG BOOK collection increase the number of circulations to grad students or increase the number of re-shelving?

And we can answer
questions like...

Harvard Library has . . . **FIVE** Strategic Objectives and
TEN Strategic Priorities

*Current University Capital Campaign has
SEVEN Aspirations*

Each school has its own mission, vision, and strategic priorities

We can link ourselves
to ...

- Decide on a tool that will allow us to interact with and visualize the data
- Determine where the data will be stored
- Ingest and link the data according to the model
- Build the dashboard - one public, one private - based on the model
- Do the work to get other (secure/private/restricted) data like student data connected to our data

Where are we and where do we go from here?

- Can you see this model being applied to your institution?
- What are your areas that may be much less complicated than ours.
- What are your examples that show the linkages?

Your Institution



Questions
