

What's the Deal with Data?

Speaker Notes

Presented by Anali Perry and Matthew Harp October 20, 2016
MPLA CALCON 2016 Loveland, CO

Slides 1-2

What's the deal with data?

You've probably heard a lot of "futurists" talk about data, but it's not always clear how data relates to our day to day work in libraries. Why are data important, and what's the big deal? Data are not just spreadsheets and numbers, but come in many different shapes, colors, and flavors! In this presentation, we will give an introduction to data, talk about why it is relevant, and demonstrate how to find and use data in practical situations. We will also provide innovative examples that will inspire you to connect with your colleagues and patrons!

Learning Objectives:

1. Attendees will know what data is
2. Attendees will understand different sources/types of data
3. Attendees will get ideas of how to incorporate data services in their libraries?

Slide 3

What is Data?

Def: facts and statistics collected together for reference or analysis

(placeholder image)

Video, Audio, Spreadsheets, Surveys, Spatial information

Instrument readings, Biological samples... Just about anything

Slide 4

Text and Books

Even books.. Because Text is data, and it can be used for a many things. Text is the raw components the lego bricks that make our our books. Data tells stories

- We have used data without thinking about it, now is more available (net) but more computable for mining and computational analysis and gathered from multiple sources.

- <https://books.google.com/ngrams>
- Everything we use to build knowledge is a form of data
- “Information structured by methodology and organized in ... products that are used as evidence in the research process” - Chuck Humphrey University of Alberta
(<http://preservingresearchdataincanada.net/2013/07/23/data-a-rose-by-any-other-name/>)
- Easier and more available.

Slide 5

Numbers and figures

Just as numbers tell stories. The quantitative stories of our world.

Slide 6

Images

Data can even be images Because they hold so much information time, place, age, size, distance and relationships.

Whether they be from space... or a family album.

[Family album public domain, source: <https://commons.wikimedia.org/wiki/File:Family-album.jpg>]

Slide 7-8

Why does data matter to libraries?

- We help connect people with resources to generate knowledge.
- What we always do - connect people with the resources they need in order to generate knowledge
- Also
 - Privacy
 - Replicability
 - Analysis
 - Visualization

Slide 9

How People use Data?

People may be using in many ways without knowing it.

Slide 10

Data of course are also numbers and figures. A great and timely example of a data scientist using public data to predict an outcome is Nate Silver who used large amounts of free public data to make accurate predictions about who would win the 2012 elections in the United States. Today the same algorithms are in place for the 2016 election that accurately predicted the 2008 and 2012 elections. See <http://projects.fivethirtyeight.com/2016-election-forecast/>

Data sources are from polls which are assessed and rated:
<http://projects.fivethirtyeight.com/pollster-ratings/>

But it's not all serious they also have FiveThirtyEight also looks at sports, science and health, and culture and all of the data and code used to create the results is available on [GitHub](#).

Slide 11-12

Genealogy

Data Visualization - A family tree is a visualization of data

Slide 13

World Bank

Open Data Initiative was launched in 2010 in support of their mission - to end extreme poverty by 2030

All of the data found here can be used free of charge with [minimal restrictions](#). They also provide an Open Data Toolkit; The Open Government Data Toolkit is designed to help governments, Bank staff and users understand the basic precepts of Open Data, then get “up to speed” in planning and implementing an open government data program, while avoiding common pitfalls. <http://opendatatoolkit.worldbank.org/en/>

Also see [Data.gov](#)

Slide 14-15

Application Development

Improved in 2013 in response to an Executive Order which requires that data generated by the government (so various agencies) be made available in open, machine-readable formats, while appropriately safeguarding privacy, confidentiality, and security. Examples that really demonstrated the value of data even before the policy include weather data (which we all love!), and GPS data - which we now use as part of our daily lives!

Also includes [Project Open Data](#), which providing plug-and-play tools and best practices to help agencies improve the management and release of open data, released on GitHub

- <https://www.data.gov/applications>
- <http://data.worldbank.org/products/mobile-apps>

Since the data are open, it's available for anyone to use to develop new tools, often can drill down to local levels

Slides 16-17

Innovative Examples - Citizen Science data creation

- Citizen Science (galaxy zoo, war diary - humanities project where people transcribe not just numbers contributing to research/knowledge)
- Galaxy Zoo started back in July 2007, with a data set made up of a million galaxies imaged by the [Sloan Digital Sky Survey](#). Within 24 hours of launch they received almost 70,000 classifications an hour. In the end, more than 50 million classifications were received by the project during its first year, contributed by more than 150,000 people. The project continues with volunteers classifying galaxies from even more projects: including images from Hubble's CANDELS survey to take ultra-deep images of the universe. Over 85 publications have resulted from Galaxy Zoo projects.
- Shapes, colors & flavors! (foodb.ca)

Slides 18

War Diary

Anali: A big project from the British National Archive - digitizing unit war diaries and using Citizen Historians to beef up the information about the diaries. Volunteers read pages, do

transcriptions, identify types of content and adding tags and other metadata that a computer can't do. Can delve more deeply into getting a detailed picture of what units were going through. Note presentation on Citizen science by Dan Stanton on slide 19

Slide 20

How do you see people using data in your community

Question Audience

Slide 21-22

Where do I find Data?

Looking for actual research data? There are so many places to go listing them is like the truck on slide 21.

- Re3data is an index of over 1,500 research data repositories
- HathiTrust - for Text Mining 3 million public domain books from the HathiTrust Library are currently available for analysis
- Public Library of Science allows you to search Query text content from the seven open-access peer-reviewed journals
- Data.gov
- Databib: <http://databib.org/index.php>
- OAD List of Data Repositories: http://oad.simmons.edu/oadwiki/Data_repositories
- Datacite: <http://www.datacite.org/whycitedata>

Slides 23

Ideas for incorporating data services

Programming can be anywhere from just highlighting open data sources as a resource to you specifically incorporating data into your services and programs. It can range from Children and Teen programs like coding clubs and app development, to programs for students and seniors

- Connecting people with services
- Target specific data needs (i.e. genealogy)
- Citizen Science projects/workshops
- Host Science Fairs

Slide 24

Target specific data needs

We all have different communities - which ones are most appropriate for you?

At ASU, we do a lot of space research. One of our planetary scientists had a Rocks around the World project, where he asked volunteers to mail him rocks from all over to help their group learn more about our own planet, which informs our understanding of other planets!

Slide 25

Science Fairs

Question to audience: Has anyone hosted a science fair?

Data is a key component of the scientific method. Libraries can play a role in teaching good data practices (free available modules are available)

Slide 26

CO Spiders slides 25-26

Spiders are data too!! Denver Museum of Nature and Science - a major regional repository for this taxonomic group. Volunteers collect spiders and send them to the DMNS for identification and storage. They have received over 30,000 specimens. Our home at ASU also has the Hasbrouck insect collection made virtually available or by request in person

Slide 27

What ideas do you have?

THE POINT IS we have resources available to us like never before, free to use, free to analyze, and free to repurpose. Librarians are bridge builders, making libraries a place to cross digital divides, the place where people can not only get access to information but learn about the opportunities to be scientists, innovators and change makers.

What ideas do you have?

- Anything you already working on based on what we've mentioned here?
- Anything got you excited?

- What questions do you still have that may have not been answered?
- Continue the conversation?

Slide 28

Thanks

Not Discussed

- Research Data Management
- Copyright (Data such as text mining is not a form of original expression)
- Using patron data to inform library programming and decisions