Harnessing the power of data

JARED HOWLAND, BRIGHAM YOUNG UNIVERSITY
REBECCA L. LUBAS, CLAREMONT COLLEGES LIBRARY
Jared Howland
Collection Development Coordinator, Brigham Young University

Rebecca L. Lubas
Associate Dean, Claremont Colleges Library
Rebecca L. Lubas
Associate Dean, Claremont Colleges Library
The Power of the Collective: Shared Collections

REBECCA L. LUBAS, CLAREMONT COLLEGES LIBRARY
The Claremont Colleges Library

- A unique consortium: seven colleges, one library!
- The “Goldilocks” Library: large enough to have resources for innovation; small enough to be nimble
- Vit@lization of collections and spaces
- The Active Print Program – print is still core
CLAREMONT COLLEGES LIBRARY • VITAL
A Shared Strategy

• One facet of the Active Print Program
• Commitment to the long term scholarly record
• Commitment to partnerships with libraries of all types
• Access to increased collections for our users
One Collection, Two Shared Print Programs

- Analysis, Uniqueness, and Overlap
- SCELC
  - First cohort of 14 libraries
  - Holdings in network, in California, in North America
  - 15 year commitment
- HathiTrust
  - First cohort of 50 libraries
  - 25 year commitment
Impact

• Collection decisions
• Master Plan decisions
• Resource sharing
Next Steps

• Local awareness
• Data-driven
  – Resource Sharing policies
  – Asset management policies
• Second cohort
• The collective collection grows!
Jared Howland
Collection Development Coordinator, Brigham Young University
Small Data Management

Jared Howland
Collection Development Coordinator
Brigham Young University
* Not the military archive, but it's a pretty cool picture, right?

https://cdn.thinglink.me/api/image/759538569067888642
T. Claudii A. Fulvi T. Ivli L. Fulvii L. Nonii Q. Lolli L. Fulvii
M. Manndri L. Vsti U. Urbani D. P. Vici
V. Victoris F. Est A. Anti.
BIG DATA\(^1\)

[D]ata sets that are so large or complex that traditional data processing application software is inadequate to deal with them

SMALL DATA\(^2\)

[D]ata that is small enough for human comprehension

\(^1\)https://en.wikipedia.org/wiki/Big_data
\(^2\)https://en.wikipedia.org/wiki/Small_data
Data Problems

- Capturing
- Storing
- Analyzing
- Searching
- Sharing
- Visualizing
- Querying
- Normalizing
- Updating
- Securing
- Privacy
- Transferring
- Storing

Diagram with keywords related to data problems.
Data Problems

- Capturing
- Storing
- Transfering
- Updating
- Securing
- Privacy
- Normalizing
- Querying
- Sharing
- Visualizing
- Analyzing
- Storing
- Searching
**DATA MANAGEMENT**

<table>
<thead>
<tr>
<th>Data governance</th>
<th>Reference &amp; master data management</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data architecture, analysis, &amp; design</td>
<td>Data warehousing &amp; business intelligence management</td>
</tr>
<tr>
<td>Database management</td>
<td>Document, record, &amp; content management</td>
</tr>
<tr>
<td>Data security management</td>
<td>Metadata management</td>
</tr>
<tr>
<td>Data quality management</td>
<td>Contact data management</td>
</tr>
</tbody>
</table>

**MASTER DATA MANAGEMENT (MDM)**

[C]omprises the processes, governance, policies, standards, and tools that consistently define and manage the critical data of an organization to provide a single point of reference

2. [https://en.wikipedia.org/wiki/Master_data_management](https://en.wikipedia.org/wiki/Master_data_management)
<table>
<thead>
<tr>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Free</td>
<td>1. No data set search tools</td>
</tr>
<tr>
<td><a href="https://education.github.com">https://education.github.com</a></td>
<td></td>
</tr>
<tr>
<td>2. Desktop tools</td>
<td>2. More training required for new users</td>
</tr>
<tr>
<td>3. Version control</td>
<td></td>
</tr>
<tr>
<td>4. Easy to update data sets</td>
<td>3. Difficult to organize</td>
</tr>
</tbody>
</table>

GitHub
https://github.com
<table>
<thead>
<tr>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Open source</td>
<td>1. Need technical expertise to set up or</td>
</tr>
<tr>
<td></td>
<td>pay someone to host</td>
</tr>
<tr>
<td>2. Searchable data sets</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2. No version control</td>
</tr>
<tr>
<td>3. Rudimentary visualizations</td>
<td></td>
</tr>
<tr>
<td>4. Generated and customizable</td>
<td></td>
</tr>
<tr>
<td>metadata per data set and/or</td>
<td></td>
</tr>
<tr>
<td>per file</td>
<td></td>
</tr>
</tbody>
</table>

https://ckan.org
Collection development data management at BYU

For the purposes of managing our collection development data Brigham Young University, we use three branches for our data that we treat independently and will never merge. We have a central repository for all of our data with three branches. The first is named raw and contains all of our data in its raw state before we make any changes to it. Our second branch is named normalized and includes all of the changes we make to normalize and clean our data to get it ready for reliable analysis. Finally, our last branch is named projects and is used to store the projects that use our various data sets.

Projects often require using multiple data sets and our projects branch includes a folder for each project and, subsequently, each project folder includes the transformed data sets in individual folders.

All folders within all branches of our repository include metadata about the data sets to properly document both the
Collection development data management at BYU

For the purposes of managing our collection development data Brigham Young University, we use three branches for our data that we treat independently and will never merge. We have a central repository for all of our data with three branches. The first is named raw and contains all of our data in its raw state before we make any changes to it. Our second branch is named normalized and includes all of the changes we make to normalize and clean our data to get it ready for reliable analysis. Finally, our last branch is named projects and is used to store the projects that use our various data sets. Projects often require using multiple data sets and our projects branch includes a folder for each project and, subsequently, each project folder includes the transformed data sets in individual folders.

All folders within all branches of our repository include metadata about the data sets to properly document both the
Collection development data management at BYU

For the purposes of managing our collection development data Brigham Young University, we use three branches for our data that we treat independently and will never merge. We have a central repository for all of our data with three branches. The first is named raw and contains all of our data in its raw state before we make any changes to it. Our second branch is named normalized and includes all of the changes we make to normalize and clean our data to get it ready for reliable analysis. Finally, our last branch is named projects and is used to store the projects that use our various data sets. Projects often require using multiple data sets and our projects branch includes a folder for each project and, subsequently, each project folder includes the transformed data sets in individual folders.

All folders within all branches of our repository include metadata about the data sets to properly document both the
<table>
<thead>
<tr>
<th>File Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>collection-expenditures.csv</td>
<td>A comma-delimited (CSV) file that contains data related to collection expenditures. It also includes an Excel file with the original formulas so that the data can be checked if needed. The files include the following columns:</td>
</tr>
<tr>
<td>collection-expenditures-percent-change.csv</td>
<td></td>
</tr>
<tr>
<td>collection-expenditures-percent-change-minus-byu.csv</td>
<td></td>
</tr>
<tr>
<td>collection-expenditures.xlsx</td>
<td></td>
</tr>
</tbody>
</table>
ARL Statistics: [https://www.arlstatistics.org/analytics](https://www.arlstatistics.org/analytics)
Description

collection-expenditures

Data is a comma-delimited (CSV), UTF-8 plain text file. It also includes the Excel file with the original formulas so that data can be checked if needed. The files include the following columns:

1. Line (institution name)
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>inan</td>
<td>$542,508.00</td>
<td>$577,778.00</td>
<td>$676,916.00</td>
<td>$770,157.00</td>
<td>$905,914.00</td>
<td>$932,378.00</td>
<td>$2,021,860.00</td>
</tr>
<tr>
<td>2</td>
<td>ALABAMA</td>
<td>$1,523,200.00</td>
<td>$1,369,383.00</td>
<td>$2,065,080.00</td>
<td>$2,017,639.00</td>
<td>$2,166,605.00</td>
<td>$2,594,383.00</td>
<td>$2,276,681.00</td>
</tr>
<tr>
<td>3</td>
<td>ALBERTA</td>
<td>$1,210,350.00</td>
<td>$1,335,504.00</td>
<td>$1,416,157.00</td>
<td>$1,744,566.00</td>
<td>$1,964,024.00</td>
<td>$2,481,478.00</td>
<td>$3,195,570.00</td>
</tr>
<tr>
<td>4</td>
<td>ARIZONA</td>
<td>$882,357.00</td>
<td>$981,390.00</td>
<td>$1,362,171.00</td>
<td>$1,193,360.00</td>
<td>$1,430,587.00</td>
<td>$1,635,074.00</td>
<td>$1,911,764.00</td>
</tr>
<tr>
<td>5</td>
<td>ARIZONA STATE</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>6</td>
<td>AUBURN</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>7</td>
<td>BOSTON</td>
<td>$947,563.00</td>
<td>$884,053.00</td>
<td>$865,902.00</td>
<td>$1,030,807.00</td>
<td>$1,034,293.00</td>
<td>$1,514,014.00</td>
<td>$1,514,162.00</td>
</tr>
<tr>
<td>8</td>
<td>BOSTON COLLEGE</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>9</td>
<td>BOSTON PUBLIC</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>10</td>
<td>BRIGHAM YOUNG</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>11</td>
<td>BRITISH COLUMBIA</td>
<td>$1,513,856.00</td>
<td>$1,502,317.00</td>
<td>$1,741,021.00</td>
<td>$1,954,121.00</td>
<td>$2,473,368.00</td>
<td>$2,722,613.00</td>
<td>$2,457,842.00</td>
</tr>
<tr>
<td>12</td>
<td>BROWN</td>
<td>$655,366.00</td>
<td>$736,596.00</td>
<td>$864,248.00</td>
<td>$717,001.00</td>
<td>$837,880.00</td>
<td>$1,029,616.00</td>
<td>$1,136,000.00</td>
</tr>
<tr>
<td>13</td>
<td>CALGARY</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Collection development data management at BYU

For the purposes of managing our collection development data Brigham Young University, we use three branches for our data that we treat independently and will never merge. We have a central repository for all of our data with three branches. The first is named raw and contains all of our data in its raw state before we make any changes to it. Our second branch is named normalized and includes all of the changes we make to normalize and clean our data to get it ready for reliable analysis. Finally, our last branch is named projects and is used to store the projects that use our various data sets.

Projects often require using multiple data sets and our projects branch includes a folder for each project and, subsequently, each project folder includes the transformed data sets in individual folders.

All folders within all branches of our repository include metadata about the data sets to properly document both the
<table>
<thead>
<tr>
<th>Dataset Name</th>
<th>Description</th>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>Print Circulation</td>
<td>Circulation of print materials</td>
<td>CSV</td>
</tr>
<tr>
<td>In-house Print Circulation</td>
<td>Circulation of in-house materials</td>
<td>CSV</td>
</tr>
<tr>
<td>BYU Undergraduate Program Data</td>
<td>Information about BYU Undergraduate programs</td>
<td>TSV</td>
</tr>
<tr>
<td>BYU Graduate Program Data</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Collection Development
Datasets to help with collection development at the Harold B. Lee Library.
13 Datasets
What are Groups?

You can use CKAN Groups to create and manage collections of datasets. This could be to catalogue datasets for a particular project or team, or on a particular theme, or as a very simple way to help people find and search your own published datasets.

3 groups found

- **Budget**
  - Collection development budget data

- **University**
  - Data about Brigham Young University

- **Usage**
  - Collection development usage data
<table>
<thead>
<tr>
<th>Dataset Title</th>
<th>Description</th>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>Print Circulation</td>
<td>Circulation of print materials</td>
<td>CSV</td>
</tr>
<tr>
<td>In-house Print Circulation</td>
<td>Circulation of in-house materials</td>
<td>CSV</td>
</tr>
<tr>
<td>BYU Undergraduate Program Data</td>
<td>Information about BYU Undergraduate programs</td>
<td>TSV</td>
</tr>
<tr>
<td>BYU Graduate Program Data</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
BYU Class Schedule

Enrollment data for BYU courses.

- 2016-1: Winter 2016 Semester
- 2016-3: Spring 2016 Term
- 2016-4: Summer 2016 Term
- 2016-5: Fall 2016 Semester

Data and Resources

- 2013-5 (TSV)
- 2014-1 (TSV)
- 2014-3 (TSV)
- 2014-4 (TSV)
- 2014-5 (TSV)
- 2015-1 (TSV)
- 2015-3 (TSV)
- 2015-4 (TSV)
Byu class schedule

Enrollment data for Byu courses.

- 2016-1: Winter 2016 Semester
- 2016-3: Spring 2016 Term
- 2016-4: Summer 2016 Term
- 2016-8: Fall 2016 Semester

Data and resources

2013-5 (TSV)
2014-1 (TSV)
2014-3 (TSV)
2014-4 (TSV)
2014-5 (TSV)
2015-1 (TSV)
2015-3 (TSV)
2015-4 (TSV)
BYU Class Schedule

Enrollment data for BYU courses.

- 2016-1: Winter 2016 Semester
- 2016-3: Spring 2016 Term
- 2016-4: Summer 2016 Term
- 2016-8: Fall 2016 Semester

Data and Resources

- 2013-5 (TSV)
- 2014-1 (TSV)
- 2014-3 (TSV)
- 2014-4 (TSV)
- 2014-5 (TSV)
- 2015-1 (TSV)
- 2015-3 (TSV)
- 2015-4 (TSV)
- 2016-4 (TSV)
### 2013-5 (TSV)

**URL:** [http://data.jaredhowland.com/dataset/9fbe0de2-2d90-4385-9143-a970740c4f5/resource/e9ea5be2-b537-4717-8a12-27bed8e8e05f...](http://data.jaredhowland.com/dataset/9fbe0de2-2d90-4385-9143-a970740c4f5/resource/e9ea5be2-b537-4717-8a12-27bed8e8e05f...)

**From the dataset abstract:**
-

**Source:** BYU Class Schedule

**Add Filter**

- **Graph**

|  |  |  |  |  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 2659 | 6 | A | HTG | S | 100 | 1 | DAY | America... |
| 2 | 2659 | 6 | A | HTG | S | 100 | 2 | DAY | America... |
| 3 | 2659 | 6 | A | HTG | S | 100 | 3 | DAY | America... |
| 4 | 2659 | 6 | A | HTG | S | 100 | 4 | DAY | America... |
| 5 | 2659 | 6 | A | HTG | S | 100 | 5 | DAY | America... |
| 6 | 2659 | 6 | A | HTG | S | 100 | 6 | DAY | America... |
| 7 | 2659 | 6 | A | HTG | S | 100 | 7 | DAY | America... |
| 8 | 7499 records | | | | | | | | |
| 9 | 152 | 191 | 12 | AEROS | S | 110 | 2 | DAY | America... |
| 10 | 9 | | | | | | | | |
THANK YOU
Thank you!