Jared Howland

Collection Development Coordinator, Brigham Young University

Rebecca L. Lubas

Associate Dean, Claremont Colleges Library

ARC17 #OCLCARC17

OCLC

Rebecca L. Lubas

Associate Dean, Claremont Colleges
Library

OCLC

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

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
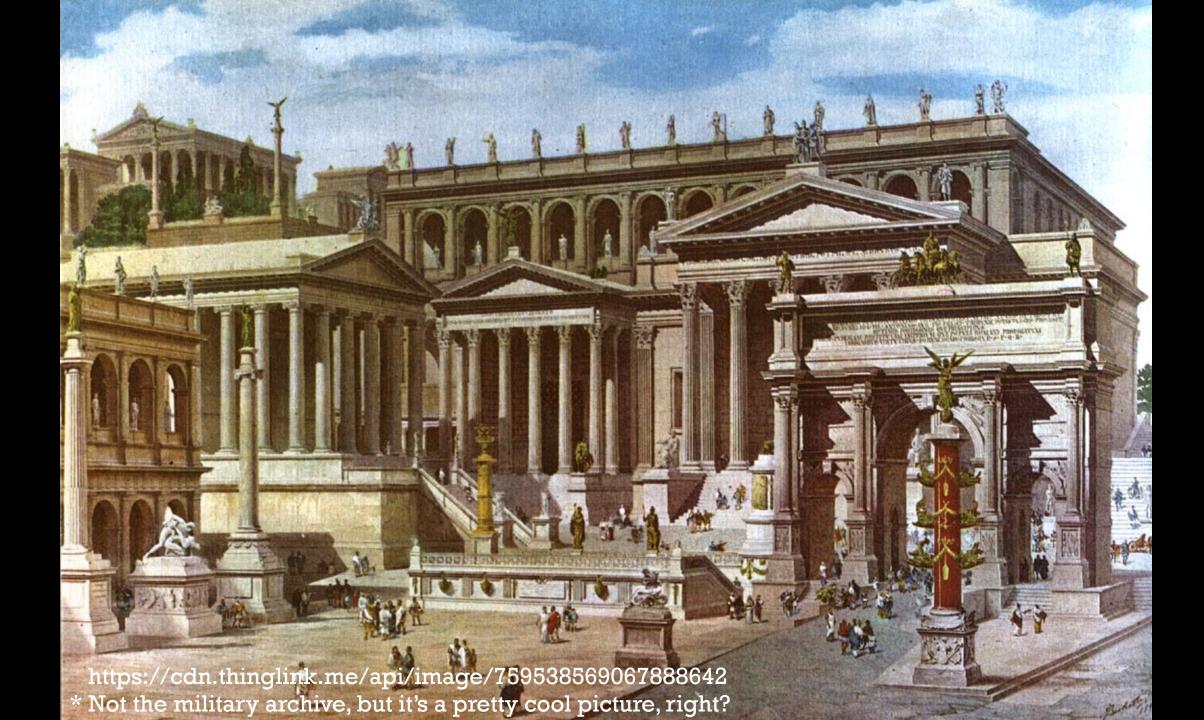
Collection Development Coordinator, Brigham Young University
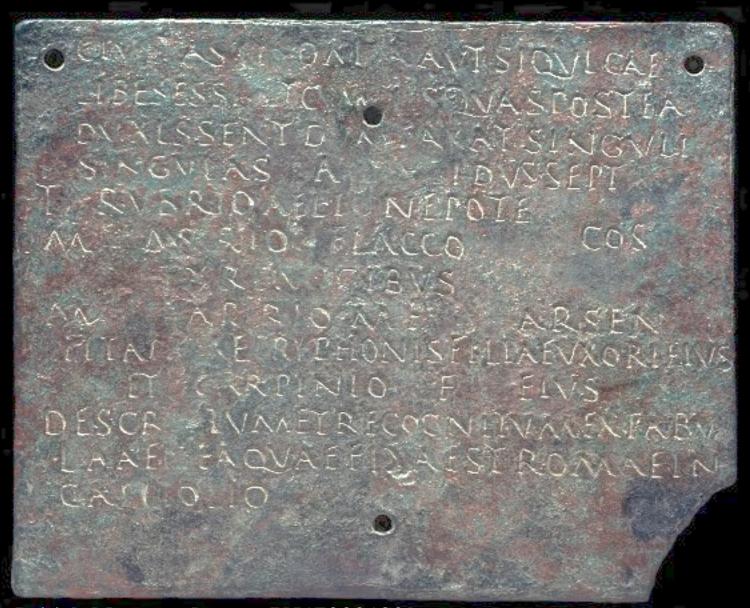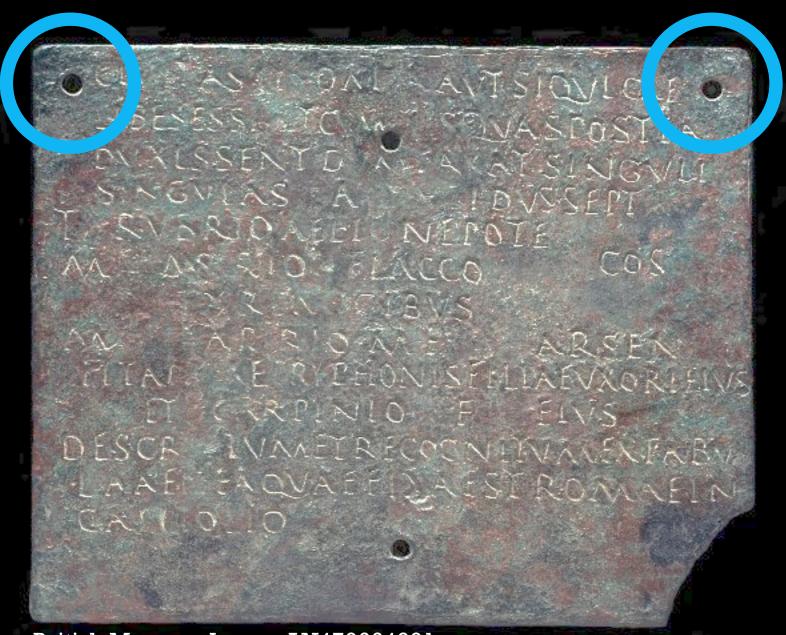
Small Data Management

Jared Howland
Collection Development Coordinator
Brigham Young University

https://cdn.thinglink.me/api/image/759538569067888642
* Not the military archive, but it's a pretty cool picture, right?

```
CAVEAS...QAE...AVT SIQVI CAE...
...BENESS...ICV...I...QVAS POSTEA
DVXISSENT D...A...A...T SINGVLI
SINGVLAS A......IDVS SEPT
...SVERIO A...IO NEPOTE
M...ASERIO FLACCO          COS
...R...TIBVS
M...AR SIO A...           ARGEN
...ITIA...E TRYPHONIS FILIAE VXOR EIVS
ET C...RPINIO F EIVS
DESCR...IVM ET RECOGN...TVM EX F...B...
L AAE...A QVAE...D A EST ROMA IN...
C...OLIO
```

British Museum Image AN470664001

... AS ... OAL ... AVT SIQVI Q ...
... BENES ... TCS ... ... QVAS POSTEA
DVXISSENT DA ... A ... T SINGVLI
SINGVLAS A ... IDVS SEPT
... EVSRIO ... FL NEPOTE
M ... ASERIO FLACCO   COS
... RI ... TIBVS
M ... AR ... IO ... E ... ARSEN
... TIA ... ET RYPHONIS FILIAE VXORI EIVS
ET C ... RPINIO F EIVS
DESCR ... IVM ET RECOGNIT VM EX TABV ...
LA AE ... EAQVAE FIXA EST ROMA EIN ...
CAPITOLIO

British Museum Image AN470664001

British Museum Image AN470664001

TI·CLAVDI   MENANDRI

A·FVLVI   IVSTI

TI·IVLI   VRBANI

L·PVILI   DAPHNI

L·NONI   VICTORIS

Q·LOLLI   FEST

L·PVLI   ANTI

British Museum Image AN378334001

CALCEAS...ONI...AVT SICVI CAE
HBENSS...ICO...S QVAS POSTEA
DVXISSENT D...A...ET SINGVLI
SINGVLAS A...IDVS SEPT
...SVBRIO...EIO NEPOTE
M ASERIO FLACCO COS
...RENTIBVS
M ARRIO...ARGEN
...TIAE TRYPHONIS FILIAE VXOR EIVS
ET CERPINIO F EIVS
DESCR...IVM ET RECOGNITVM EX FB...
LAAE...AQVAE EID AEST ROMAE IN...
CAPITOLIO

Big and Small Data

BIG DATA[1]
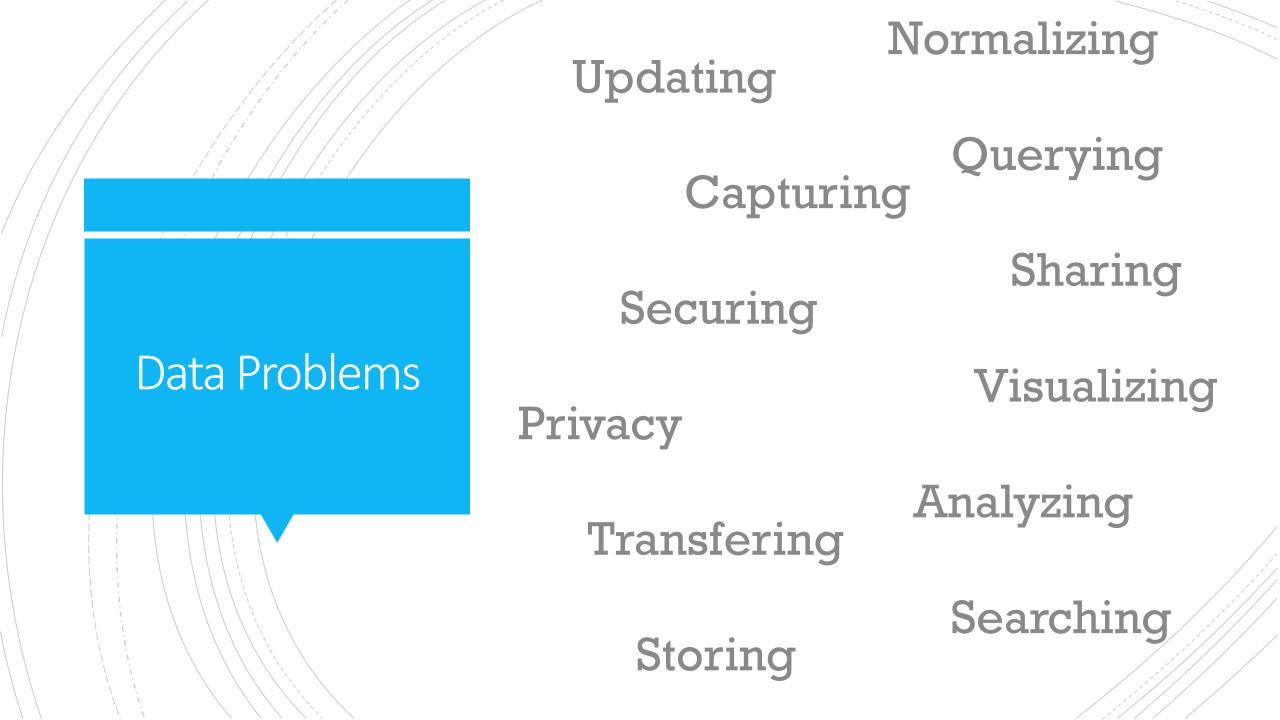
[D]ata sets that are so large or complex that traditional data processing application software is inadequate to deal with them

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

Normalizing

Updating

Querying

Capturing

Sharing

Securing

Visualizing

Privacy

Analyzing

Transfering

Searching

Storing

Data Management and Master Data Management

DATA MANAGEMENT[1]

Data governance	Reference & master data management
Data architecture, analysis, & design	Data warehousing & business intelligence management
Database management	Document, record, & content management
Data security management	Metadata management
Data quality management	Contact data management

MASTER DATA MANAGEMENT (MDM)[2]
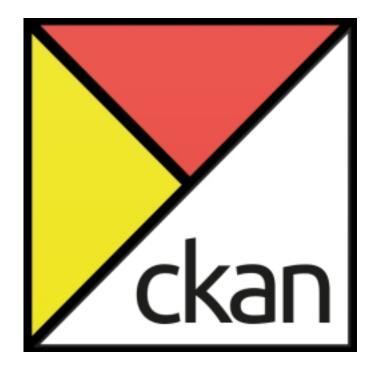
[C]omprises the processes, governance, policies, standards, and tools that consistently define and manage the critical data of an organization to provide a single point of reference

[1] https://en.wikipedia.org/wiki/Data_management
[2] https://en.wikipedia.org/wiki/Master_data_management

GitHub

https://github.com

Pros	Cons
1. Free https://education.github.com	1. No data set search tools
2. Desktop tools	2. More training required for new users
3. Version control	3. Difficult to organize
4. Easy to update data sets	
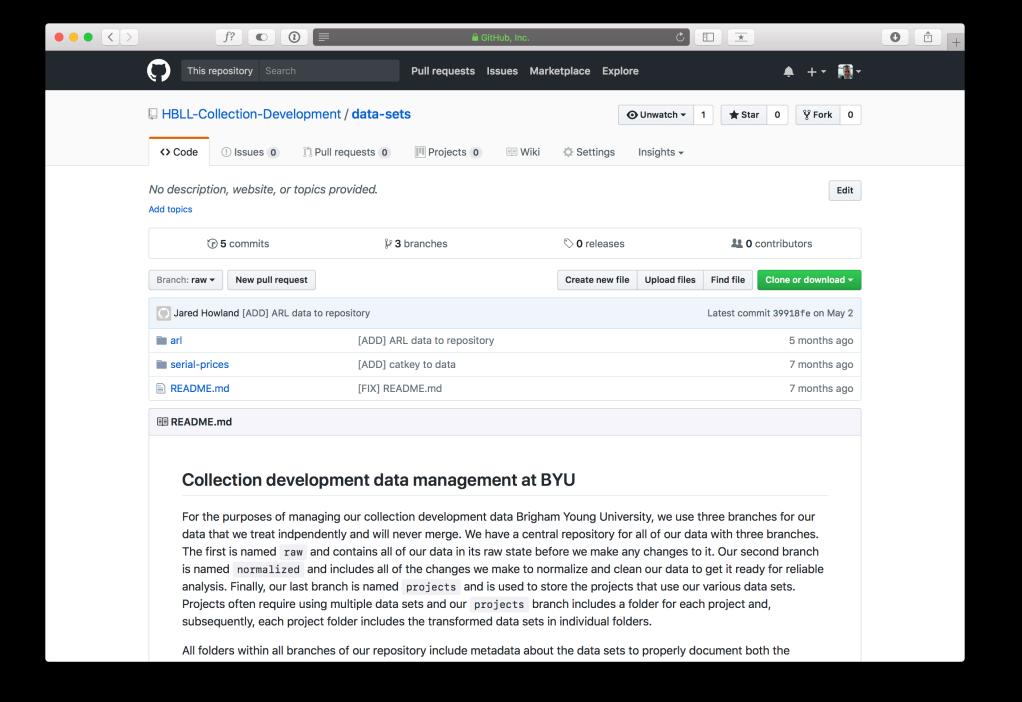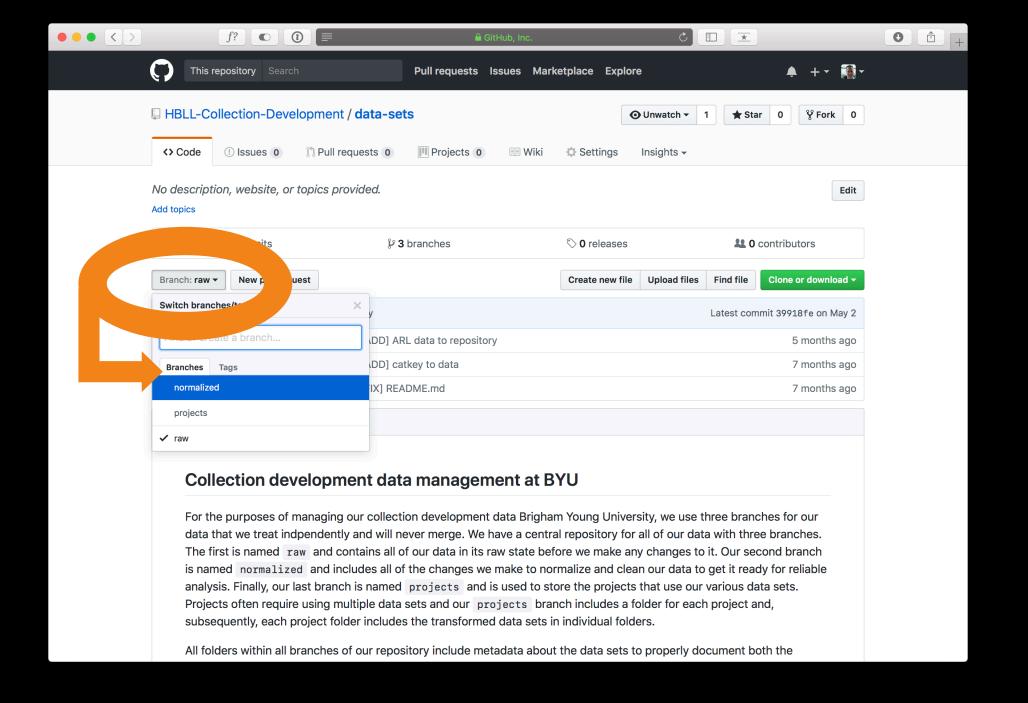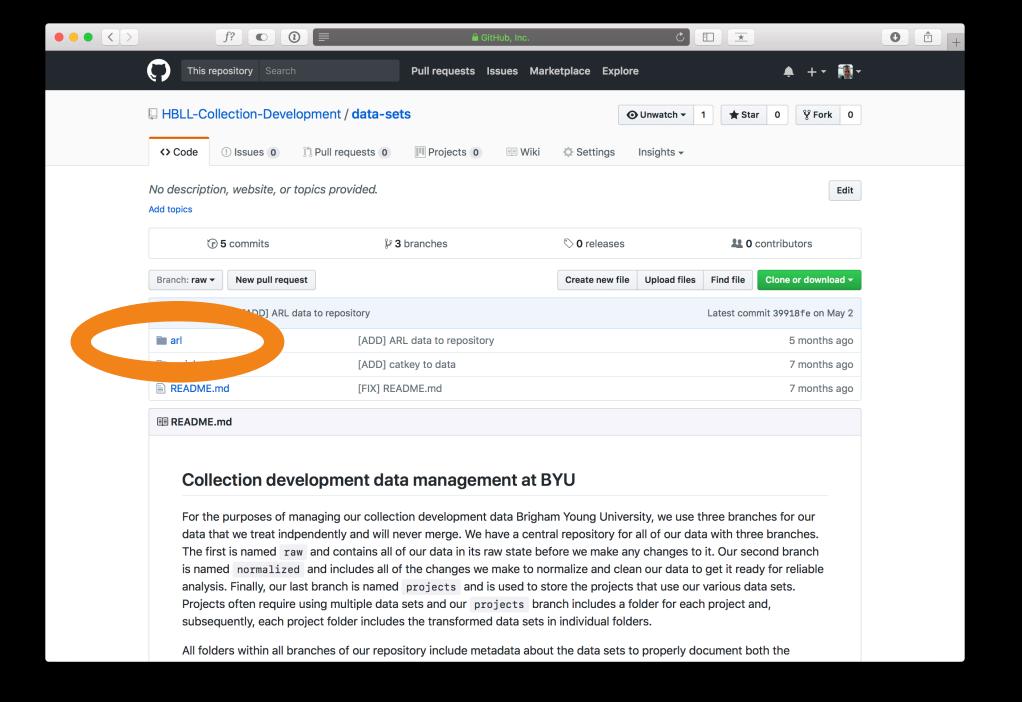
https://ckan.org
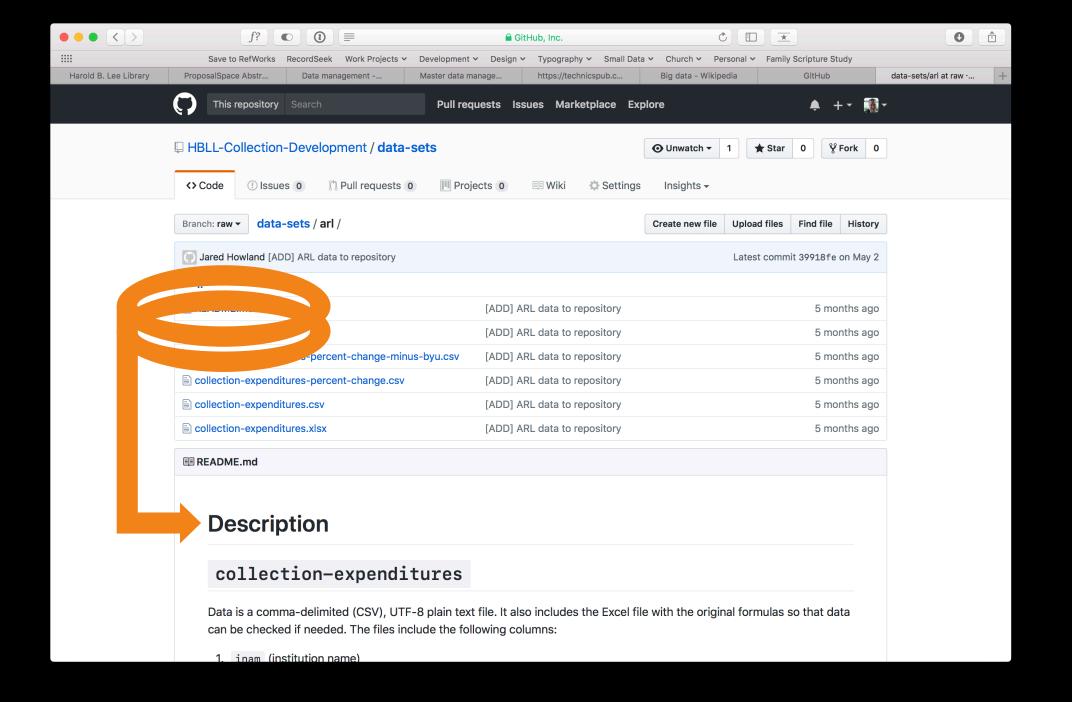
Pros	Cons
1. Open source	1. Need technical expertise to set up or pay someone to host
2. Searchable data sets	
3. Rudimentary visualizations	
4. Generated and customizable metadata per data set and/or per file	2. No version control

https://github.com/HBLL-Collection-Development/data-sets

📖 HBLL-Collection-Development / **data-sets**

👁 Unwatch ▾   1   ★ Star   0   ⑂ Fork   0

<> Code   ⓘ Issues 0   ⑂ Pull requests 0   ▤ Projects 0   ▤ Wiki   ⚙ Settings   Insights ▾

No description, website, or topics provided.

Edit

Add topics

⊙ **5** commits   ⑂ **3** branches   ⬫ **0** releases   👥 **0** contributors

Branch: raw ▾   New pull request   Create new file   Upload files   Find file   Clone or download ▾

👤 Jared Howland [ADD] ARL data to repository   Latest commit 39918fe on May 2

📁 arl	[ADD] ARL data to repository	5 months ago
📁 serial-prices	[ADD] catkey to data	7 months ago
📄 README.md	[FIX] README.md	7 months ago

📖 **README.md**

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 indpendently 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

This repository   Search   Pull requests   Issues   Marketplace   Explore

HBLL-Collection-Development / **data-sets**

Unwatch ▾   1   ★ Star   0   Fork   0

<> Code   ⚠ Issues 0   ⇡ Pull requests 0   ▦ Projects 0   📖 Wiki   ⚙ Settings   Insights ▾

No description, website, or topics provided.

Edit

Add topics

⎇ **3 branches**   🏷 **0 releases**   👥 **0 contributors**

Branch: raw ▾   New p...uest

Create new file   Upload files   Find file   Clone or download ▾

Switch branches/t...   ✕

Latest commit 39918fe on May 2

[find or create a branch...]

Branches   Tags

normalized   DD] ARL data to repository   5 months ago

projects   DD] catkey to data   7 months ago

✓ raw   IX] README.md   7 months ago

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 indpendently 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

This repository   Search   Pull requests   Issues   Marketplace   Explore

HBLL-Collection-Development / **data-sets**

👁 Unwatch ▾   1   ★ Star   0   ⑂ Fork   0

‹› Code   ⓘ Issues **0**   ⓘ Pull requests **0**   Projects **0**   Wiki   ⚙ Settings   Insights ▾

No description, website, or topics provided.

Edit

Add topics

⊙ **5** commits   ⑂ **3** branches   ⬚ **0** releases   👥 **0** contributors

Branch: raw ▾   New pull request

Create new file   Upload files   Find file   Clone or download ▾

[ADD] ARL data to repository   Latest commit 39918fe on May 2

📁 arl   [ADD] ARL data to repository   5 months ago

[ADD] catkey to data   7 months ago

📄 README.md   [FIX] README.md   7 months ago

📖 **README.md**

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 indpendently 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

This repository   Search   Pull requests   Issues   Marketplace   Explore   🔔   ＋▾   👤▾

📖 HBLL-Collection-Development / **data-sets**

👁 Unwatch ▾ 1   ★ Star 0   ⑂ Fork 0

<> Code   ⚠ Issues **0**   ⑂ Pull requests **0**   ▥ Projects **0**   📖 Wiki   ⚙ Settings   Insights ▾

Branch: raw ▾   **data-sets** / arl /   Create new file   Upload files   Find file   History

👤 Jared Howland [ADD] ARL data to repository   Latest commit 39918fe on May 2

[ADD] ARL data to repository   5 months ago

[ADD] ARL data to repository   5 months ago

…-percent-change-minus-byu.csv   [ADD] ARL data to repository   5 months ago

📄 collection-expenditures-percent-change.csv   [ADD] ARL data to repository   5 months ago

📄 collection-expenditures.csv   [ADD] ARL data to repository   5 months ago

📄 collection-expenditures.xlsx   [ADD] ARL data to repository   5 months ago

📖 **README.md**

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. `inam` (institution name)

Save to RefWorks   RecordSeek   Work Projects ⌄   Development ⌄   Design ⌄   Typography ⌄   Small Data ⌄   Church ⌄   Personal ⌄   Family Scripture Study

Harold B. Lee Library   ProposalSpace Abstr…   Data management -…   Master data manage…   https://technicspub.c…   Big data - Wikipedia   GitHub   data-sets/SOURCE.m…

This repository   Search

Pull requests   Issues   Marketplace   Explore

📖 HBLL-Collection-Development / **data-sets**

⊙ Unwatch ⌄   1   ★ Star   0   ⑂ Fork   0

<> Code   ⚠ Issues 0   Pull requests 0   Projects 0   Wiki   Settings   Insights ⌄

Branch: raw ⌄   **data-sets** / **arl** / SOURCE.md   Find file   Copy path

Jared Howland [ADD] ARL data to repository   39918fe on May 2

0 contributors

2 lines (1 sloc)   56 Bytes   Raw   Blame   History

ARL Statistics: https://www.arlstatistics.org/analytics

This repository   Search   Pull requests   Issues   Marketplace   Explore

📖 HBLL-Collection-Development / data-sets

⊙ Unwatch ⌄   1   ★ Star   0   ⑂ Fork   0

<> Code   ⊙ Issues 0   ⑂ Pull requests 0   ▦ Projects 0   ▤ Wiki   ⚙ Settings   Insights ⌄

Branch: raw ⌄   data-sets / arl /   Create new file   Upload files   Find file   History

Jared Howland [ADD] ARL data to repository   Latest commit 39918fe on May 2

..

📄 README.md	[ADD] ARL data to repository	5 months ago
📄 SOURCE.md	[ADD] ARL data to repository	5 months ago
📄 collection-expenditures-percent-change-minus-byu.csv	[ADD] ARL data to repository	5 months ago
📄 ...ection-expenditures-...change.csv	[ADD] ARL data to repository	5 months ago
📄 collection-expenditures.csv	[ADD] ARL data to repository	5 months ago
	[ADD] ARL data to repository	5 months ago

📖 README.md

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. `inam` (institution name)

This repository | Search   Pull requests   Issues   Marketplace   Explore

📖 **HBLL-Collection-Development** / **data-sets**

◉ Unwatch ∨ | 1   ★ Star | 0   ⑂ Fork | 0

<> Code   ⓘ Issues 0   ⑁ Pull requests 0   ▤ Projects 0   📖 Wiki   ⚙ Settings   Insights ∨

Branch: raw ∨   **data-sets** / **arl** / **collection-expenditures.csv**   Find file   Copy path

🧑 **Jared Howland** [ADD] ARL data to repository   39918fe on May 2

0 contributors

130 lines (130 sloc) | 87.6 KB   Raw   Blame   History   🖥 ✏ 🗑

🔍 Search this file...

	inam	1973-74	1974-75	1975-76	1976-77	1977-78	1978-79	1979-80	1980-81
1									
2	ALABAMA	$542,508.00	$577,778.00	$676,916.00	$770,157.00	$905,914.00	$932,375.00	$1,021,860.00	$1,327,485.
3	ALBERTA	$1,523,200.00	$1,369,363.00	$2,065,090.00	$2,017,639.00	$2,166,605.00	$2,594,383.00	$2,276,681.00	$1,816,159.
4	ARIZONA	$1,210,350.00	$1,335,504.00	$1,416,157.00	$1,744,566.00	$1,964,024.00	$2,481,478.00	$3,195,570.00	$3,674,784
5	ARIZONA STATE	$882,357.00	$981,390.00	$1,362,171.00	$1,193,360.00	$1,439,587.00	$1,635,074.00	$1,811,764.00	$1,983,121.
6	AUBURN	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
7	BOSTON	$947,563.00	$884,053.00	$865,902.00	$1,030,807.00	$1,034,293.00	$11,542,014.00	$1,514,162.00	$1,661,003.
8	BOSTON COLLEGE	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
9	BOSTON PUBLIC	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
10	BRIGHAM YOUNG	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
11	BRITISH COLUMBIA	$1,513,856.00	$1,502,317.00	$1,741,021.00	$1,954,121.00	$2,473,368.00	$2,722,613.00	$2,457,842.00	$2,726,692
12	BROWN	$655,366.00	$736,596.00	$864,248.00	$717,001.00	$837,880.00	$1,029,616.00	$1,136,000.00	$1,251,646.
13	CALGARY	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

This repository | Search | Pull requests | Issues | Marketplace | Explore

HBLL-Collection-Development / **data-sets**

Unwatch ▾ | 1 | ★ Star | 0 | Fork | 0

<> Code | ⓘ Issues 0 | ⫚ Pull requests 0 | ▤ Projects 0 | Wiki | ⚙ Settings | Insights ▾

No description, website, or topics provided.

Edit

Add topics

—its | ⑂ 3 branches | ◯ 0 releases | 0 contributors

Branch: **raw** ▾ | New p— uest

Create new file | Upload files | Find file | Clone or download ▾

Switch branches/t— ✕

Latest commit 39918fe on May 2

Find or create a branch...

ADD] ARL data to repository | 5 months ago

Branches | Tags

ADD] catkey to data | 7 months ago

normalized

IX] README.md | 7 months ago

projects

✓ raw

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 indpendently 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.

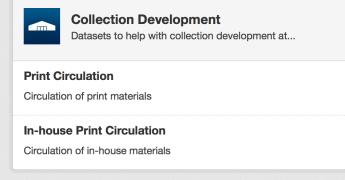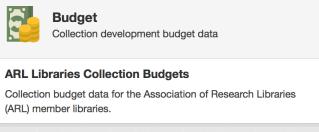All folders within all branches of our repository include metadata about the data sets to properly document both the

http://data.jaredhowland.com

BYU LIBRARY

Datasets   Organi   ns   Groups   A

Search

🏠 / **Organizations**

ℹ What are Organizations?

CKAN Organizations are used to create, manage and publish collections of datasets. Users can have different roles within an Organization, depending on their level of authorisation to create, edit and publish.

Search organizations...

1 organization found

Order by: Name Ascending

Collection Development

Datasets to help with collection development at the Harold B. Lee Library,...

13 Datasets

About BYU Library

CKAN API

CKAN Association

OPEN DATA

Powered by

✉ ckan

Language: English

Log in   Register

BYU LIBRARY   Datasets   Organizations   Groups   About   Search

🏠 / **Groups**

Search groups...   🔍

3 groups found

Order by: Name Ascending ▼

Budget
Collection development budget data

University
Data about Brigham Young University

Usage
Collection development usage data

About BYU Library

CKAN API

CKAN Association

OPEN DATA

Powered by
✉ ckan

Language: English ▼

BYU LIBRARY

Datasets   Organizations   Groups   About   Search

🏠 / **Datasets**

▼ **Organizations**

Collection Development (13)

▼ **Groups**

Usage (6)

Budget (1)

▼ **Tags**

counter (4)

alcts (1)

arl (1)

br1 (1)

br2 (1)

cpi (1)

db1 (1)

hepi (1)

jr1 (1)

lmpi (1)

▼ **Formats**

Search datasets...

13 datasets found

Order by: Relevance

Print Circulation

Circulation of print materials

`CSV`

In-house Print Circulation

Circulation of in-house materials

`CSV`

BYU Class Schedule

...all...cnt data for BYU...es. 2016-1: Winter 2016 Semester 2016-3: Spring 2016 Term 2016-4: Summer 2016 Term 2016-5: Fall 2016 Semester

`TSV`

BYU Undergraduate Program Data

Information about BYU Undergraduate programs

`TSV`

BYU Graduate Program Data

BYU LIBRARY

Datasets   Organizations   Groups   About

Search

⌂ / Organizations / Collection Development / **BYU Class Schedule**

BYU Class Schedule

Followers

0

🏢 **Organization**

Collection Development

Datasets to help with collection development at the Harold B. Lee Library, Brigham Young University

read more

↗ **Social**

Google+

Twitter

⊞ Dataset   👥 Groups   🕐 Activity Stream

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

DATA	**2013-5 (TSV)**	↗ Explore ▾
DATA	**2014-1 (TSV)**	↗ Explore ▾
DATA	**2014-3 (TSV)**	↗ Explore ▾
DATA	**2014-4 (TSV)**	↗ Explore ▾
DATA	**2014-5 (TSV)**	↗ Explore ▾
DATA	**2015-1 (TSV)**	↗ Explore ▾
DATA	**2015-3 (TSV)**	↗ Explore ▾
DATA	**2015-4 (TSV)**	↗ Explore ▾

2015-5 (TSV)

Explore

2016-1 (TSV)

Explore

2016-3 (TSV)

Explore

2016-4 (TSV)

Explore

2016-5 (TSV)

Explore

2017-1 (TSV)

Explore

2017-3 (TSV)

Explore

2017-4 (TSV)

Explore

License

Creative Commons CCZero

OPEN DATA

Additional Info

Field	Value
Source	http://saasta.byu.edu/noauth/classSchedule/
Author	BYU Registrar's Office
Maintainer	Jared Howland
Last Updated	July 17, 2017, 10:27 AM (UTC-06:00)
Created	July 14, 2017, 2:32 PM (UTC-06:00)

About BYU Library

CKAN API

CKAN Association

OPEN DATA

Powered by

ckan

Language: English

Log in   Register

BYU LIBRARY

Datasets   Organizations   Groups   About

Search

🏠 / Organizations / Collection Development / **BYU Class Schedule**

BYU Class Schedule

Followers

0

🏢 **Organization**

Collection Development

Datasets to help with collection development at the Harold B. Lee Library, Brigham Young University

read more

📤 **Social**

▪ Google+

▪ Twitter

⛓ Dataset   👥 Groups   🕐 Activity Stream

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)**   ↪ Explore ▾

📄 **2014-1 (TSV)**   ↪ Explore ▾

📄 **2014-3 (TSV)**   ↪ Explore ▾

📄 **2014-4 (TSV)**   ↪ Explore ▾

📄 **2014-5 (TSV)**   ↪ Explore ▾

📄 **2015-1 (TSV)**   ↪ Explore ▾

📄 **2015-3 (TSV)**   ↪ Explore ▾

📄 **2015-4 (TSV)**   ↪ Explore ▾

BYU LIBRARY

Datasets   Organizations   Groups   About

Search

🏠 / Organizations / Collection Development / BYU Class Schedule / 2013-5 (TSV)

2013-5 (TSV)

⬚ Go to resource   ⚗ Data API

URL: http://data.jaredhowland.com/dataset/9fbe0de2-2d90-4386-9143-ea970740c4f5/resource/e9ea9be2-b537-4717-8a12-27bed88e05f8…

From the dataset abstract

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

Source: BYU Class Schedule

⊞ Data Explorer

</> Embed

Add Fil…

		unknow…	course-…	course-…	course-…	course-…	section-…	unknown	course-t…	course-…	
Graph	7499 records	«	1 — 100	»	🔍 Search data …		Go »	Filters			
1	2859	6	A HTG	S	100		1		DAY	America…	
2	2859	6	A HTG	S	100		2		DAY	America…	
3	2859	6	A HTG	S	100		3		DAY	America…	
4	2859	6	A HTG	S	100		4		DAY	America…	
5	2859	6	A HTG	S	100		5		DAY	America…	
6	2859	6	A HTG	S	100		6		DAY	America…	
7	2859	6	A HTG	S	100		7		DAY	America…	
152	191	12	AEROS	S	110		2		DAY	Air Force…	
8	2859	6	A HTG	S	100		8		DAY	America…	
9	2859	6	A HTG	S	100		9		DAY	America…	

THANK YOU

Thank you!

ARC17 #OCLCARC17 OCLC