HUMANITIES DATA WORKSHOP SERIES

THE CENTER FOR DIGITAL HUMANITIES @PRINCETON

Princeton Research Data Service

Structuring Humanities Data

December 1, 2021

Sierra Eckert | scecerkert@princeton.edu
Matthew Chandler | mchandler@princeton.edu

Overview of the Humanities Data Workshop Series

Workshop 1: What is “Data” in the Humanities?

- Oct. 6

Workshop 2: Finding Humanities Data

- Nov. 3

Workshop 3: Structuring Humanities Data

- Dec.1

<u>Spring 2022:</u>

Workshop 4: Cleaning Humanities Data
Workshop 5: Preserving Humanities Data
Workshop 6: Capstone on ethical considerations in working with humanities data

Today's Agenda

- Quick recap of the second workshop and the workshop series (done!)
- Overview of what structured data is (and why a humanities scholar might want to use it)
- "Tidy" data
- Making structured data: 3 case studies
- Ask Us Anything (Q&A)

I

What is “structured” data?

The situation

Most data that humanities scholars work with––whether texts, images, objects, or records––are messy.

An example

Suppose you're a historian working with a list of names written in columns in the ledgers of an early 20th century lending library. Your data might look like this:

1937 DE BEAUVOIR

Mlle Simone
71 rue de Rennes

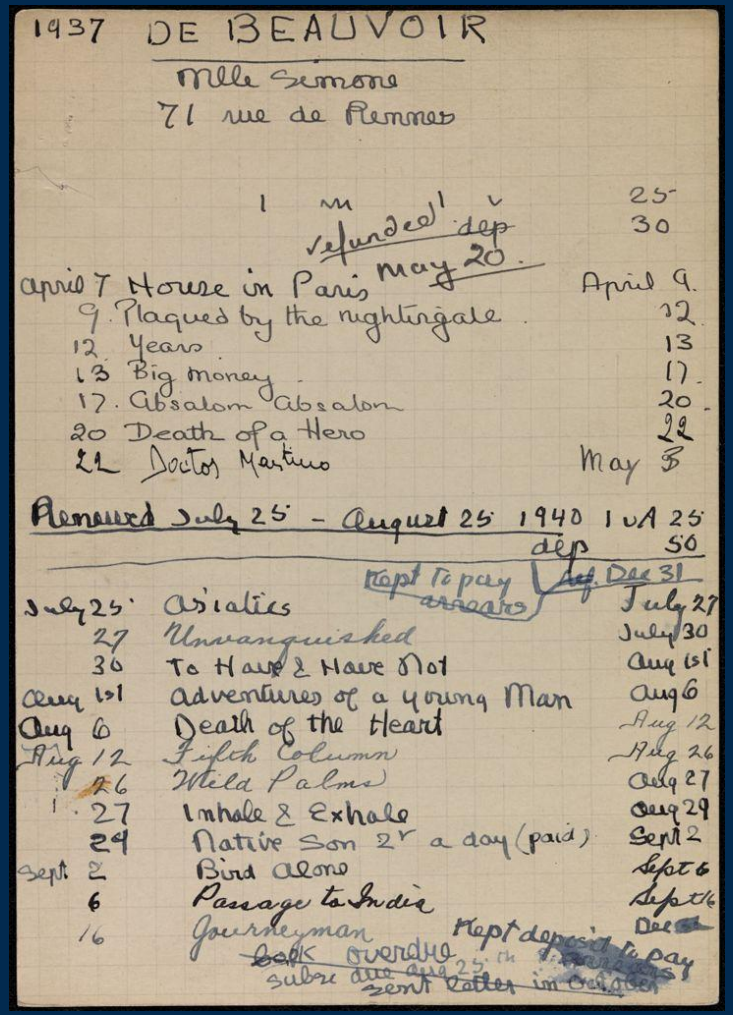
	1 m	25
	refunded dep	30
April 7	House in Paris May 20.	April 9.
9	Plagued by the nightingale	12
12	Years	13
13	Big money	17
17	Absalom Absalon	20
20	Death of a Hero	22
22	Doctor Martino	May 8

Renewed July 25 - August 25 1940 1 vA 25
dep 50

	Kept to pay arrears	~~Aug~~ Dec 31
July 25	Asiatics	July 27
27	Unvanquished	July 30
30	To Have & Have Not	Aug 1st
Aug 1st	Adventures of a Young Man	Aug 6
Aug 6	Death of the Heart	Aug 12
Aug 12	Fifth Column	Aug 26
26	Wild Palms	Aug 27
27	Inhale & Exhale	Aug 29
29	Native Son 2r a day (paid)	Sept 2
Sept 2	Bird Alone	Sept 6
6	Passage to India	Sept 16
16	Journeyman	Dec

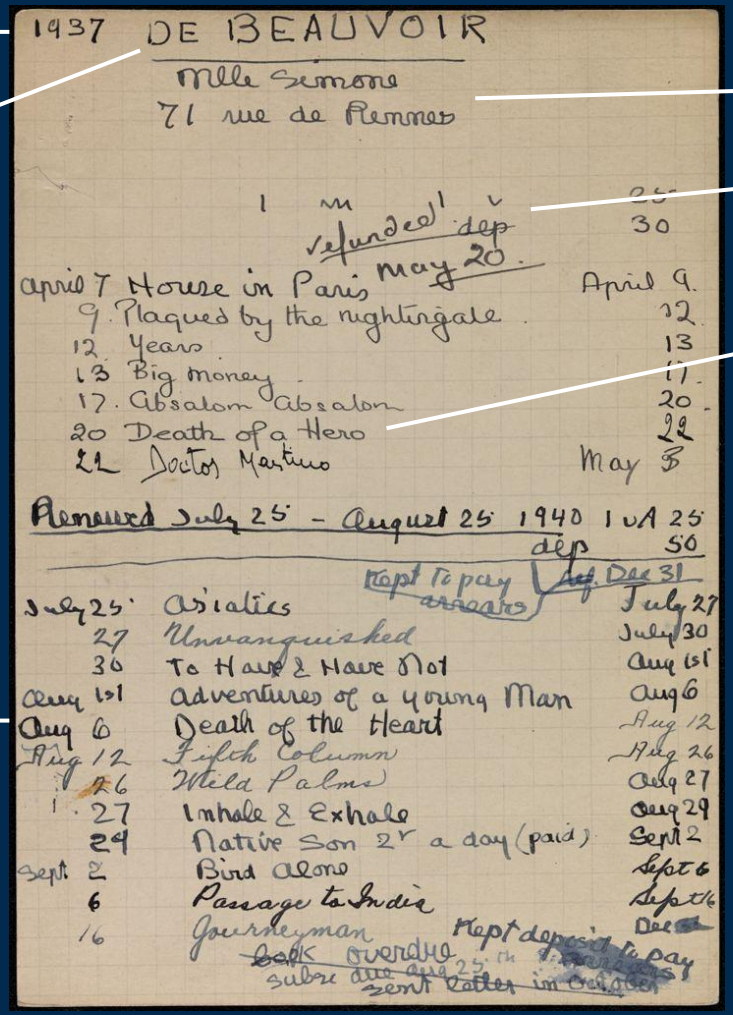
Kept deposit to pay
~~Book overdue~~ ~~sub due Aug 25th renewed~~ ~~sent letter in October~~

https://shakespeareandco.princeton.edu/members/beauvoir-simone-de/cards/



This lending library card for Simone de Beauvoir at Sylvia Beech's lending library is readable for human eyes, but not legible for a machine.

https://shakespeareandco.princeton.edu/members/beauvoir-simone-de/cards/



Date

Member Name

Address

Note

Title Borrowed

Membership Dates

Borrow date

Return date

https://shakespeareandco.princeton.edu/sources/cards/

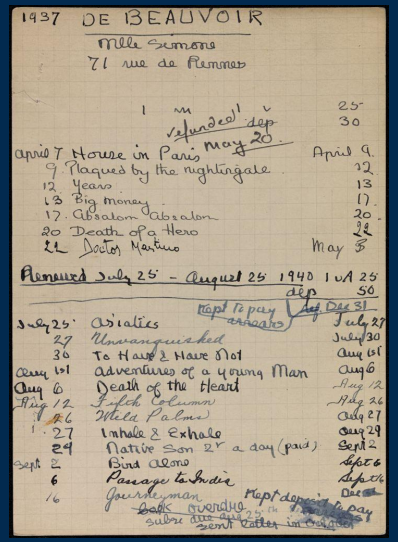
If you wanted to retrieve and compare any of this information for all of the lending library card records, you'd need to *structure* it.

event_type	start_date	end_date	member_uris	member_names	borrow_statu	borrow_duration_	item_uri	item_title	item_authors	item_year
Borrow	4/7/37	4/9/37	https://shake	Simone de Beauvoir	Returned	2	https://shake	The House in Paris	Bowen, Elizabeth	1935
Borrow	4/9/37	4/12/37	https://shake	Simone de Beauvoir	Returned	3	https://shake	Plagued by the Nightingale	Boyle, Kay	1931
Borrow	4/12/37	4/13/37	https://shake	Simone de Beauvoir	Returned	1	https://shake	The Years	Woolf, Virginia	1937
Borrow	4/17/37	4/20/37	https://shake	Simone de Beauvoir	Returned	3	https://shake	Absalom, Absalom!	Faulkner, William	1936
Borrow	4/20/37	4/22/37	https://shake	Simone de Beauvoir	Returned	2	https://shake	Death of a Hero	Aldington, Richar	1929
Borrow	4/22/37	5/3/37	https://shake	Simone de Beauvoir	Returned	11	https://shake	Doctor Martino and Other St	Faulkner, William	1934
Borrow	7/25/40	7/27/40	https://shake	Simone de Beauvoir	Returned	2	https://shake	The Asiatics	Prokosch, Frederi	1935
Borrow	7/27/40	7/30/40	https://shake	Simone de Beauvoir	Returned	3	https://shake	The Unvanquished	Faulkner, William	1938
Borrow	7/30/40	8/1/40	https://shake	Simone de Beauvoir	Returned	2	https://shake	To Have and Have Not	Hemingway, Erne	1937
Borrow	8/1/40	8/6/40	https://shake	Simone de Beauvoir	Returned	5	https://shake	Adventures of a Young Man	Dos Passos, John	1939

event_type	start_date	end_date	member_uris	member_names	borrow_statu	borrow_duration_	item_uri	item_title	item_authors	item_year
Borrow	4/7/37	4/9/37	https://shake	Simone de Beauvoir	Returned	2	https://shake	The House in Paris	Bowen, Elizabeth	1935
Borrow	4/9/37	4/12/37	https://shake	Simone de Beauvoir	Returned	3	https://shake	Plagued by the Nightingale	Boyle, Kay	1931
Borrow	4/12/37	4/13/37	https://shake	Simone de Beauvoir	Returned	1	https://shake	The Years	Woolf, Virginia	1937
Borrow	4/17/37	4/20/37	https://shake	Simone de Beauvoir	Returned	3	https://shake	Absalom, Absalom!	Faulkner, William	1936
Borrow	4/20/37	4/22/37	https://shake	Simone de Beauvoir	Returned	2	https://shake	Death of a Hero	Aldington, Richar	1929
Borrow	4/22/37	5/3/37	https://shake	Simone de Beauvoir	Returned	11	https://shake	Doctor Martino and Other St	Faulkner, William	1934
Borrow	7/25/40	7/27/40	https://shake	Simone de Beauvoir	Returned	2	https://shake	The Asiatics	Prokosch, Frederi	1935
Borrow	7/27/40	7/30/40	https://shake	Simone de Beauvoir	Returned	3	https://shake	The Unvanquished	Faulkner, William	1938
Borrow	7/30/40	8/1/40	https://shake	Simone de Beauvoir	Returned	2	https://shake	To Have and Have Not	Hemingway, Erne	1937
Borrow	8/1/40	8/6/40	https://shake	Simone de Beauvoir	Returned	5	https://shake	Adventures of a Young Man	Dos Passos, John	1939

Structuring humanities data =
Deciding how to make your data legible for a machine

Human-readable vs Machine-actionable


1937 DE BEAUVOIR
Mlle Simone
71 rue de Rennes

1 m refunded dep 25
30
April 7 House in Paris May 20 April 9
9 Plagued by the nightingale 12
12 Years 13
13 Big money 17
17 Absalom Absalom 20
20 Death of a Hero 22
22 Doctor Martino May 5
Remained July 25 – August 25 1940 1 vol 25
dep 50
Kept to pay arrears
July 25 Asiatics July 27
27 Unvanquished July 30
30 To Have & Have Not Aug 6
Aug 6 Adventures of a young Man Aug 6
Aug 6 Death of the Heart Aug 12
Aug 12 Fifth Column Aug 26
26 Mila Palma Aug 27
27 Inhale & Exhale Aug 29
29 Native Son 2¢ a day (paid) Sept 2
Sept 2 Bird alone Sept 6
6 Passage to India Sept 16
16 Journeyman Kept dep to pay
Balls OVERDUE

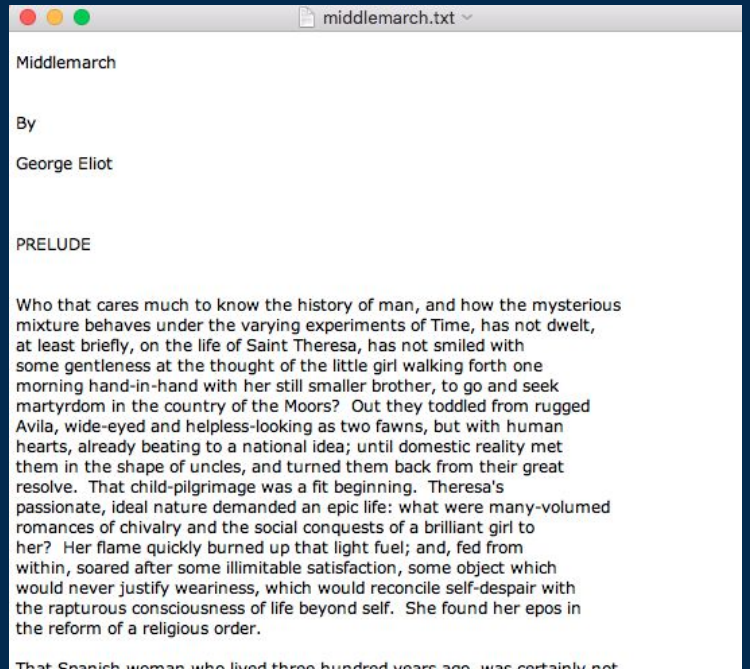
event_type	start_date	end_date	member_uris	member_names	borrow_statu	borrow_duration_	item_uri	item_title	item_authors	item_year
Borrow	4/7/37	4/9/37	https://shake	Simone de Beauvoir	Returned	2	https://shake	The House in Paris	Bowen, Elizabeth	1935
Borrow	4/9/37	4/12/37	https://shake	Simone de Beauvoir	Returned	3	https://shake	Plagued by the Nightingale	Boyle, Kay	1931
Borrow	4/12/37	4/13/37	https://shake	Simone de Beauvoir	Returned	1	https://shake	The Years	Woolf, Virginia	1937
Borrow	4/17/37	4/20/37	https://shake	Simone de Beauvoir	Returned	3	https://shake	Absalom, Absalom!	Faulkner, William	1936
Borrow	4/20/37	4/22/37	https://shake	Simone de Beauvoir	Returned	2	https://shake	Death of a Hero	Aldington, Richar	1929
Borrow	4/22/37	5/3/37	https://shake	Simone de Beauvoir	Returned	11	https://shake	Doctor Martino and Other St	Faulkner, William	1934
Borrow	7/25/40	7/27/40	https://shake	Simone de Beauvoir	Returned	2	https://shake	The Asiatics	Prokosch, Frederi	1935
Borrow	7/27/40	7/30/40	https://shake	Simone de Beauvoir	Returned	3	https://shake	The Unvanquished	Faulkner, William	1938
Borrow	7/30/40	8/1/40	https://shake	Simone de Beauvoir	Returned	2	https://shake	To Have and Have Not	Hemingway, Erne	1937
Borrow	8/1/40	8/6/40	https://shake	Simone de Beauvoir	Returned	5	https://shake	Adventures of a Young Man	Dos Passos, John	1939

Machine-actionable

Unstructured data

typically data where the the boundaries of individual items, the relations between items, and the meaning of items, are mostly implicit.

Example: plain text files


middlemarch.txt

Middlemarch

By

George Eliot

PRELUDE

Who that cares much to know the history of man, and how the mysterious mixture behaves under the varying experiments of Time, has not dwelt, at least briefly, on the life of Saint Theresa, has not smiled with some gentleness at the thought of the little girl walking forth one morning hand-in-hand with her still smaller brother, to go and seek martyrdom in the country of the Moors? Out they toddled from rugged Avila, wide-eyed and helpless-looking as two fawns, but with human hearts, already beating to a national idea; until domestic reality met them in the shape of uncles, and turned them back from their great resolve. That child-pilgrimage was a fit beginning. Theresa's passionate, ideal nature demanded an epic life: what were many-volumed romances of chivalry and the social conquests of a brilliant girl to her? Her flame quickly burned up that light fuel; and, fed from within, soared after some illimitable satisfaction, some object which would never justify weariness, which would reconcile self-despair with the rapturous consciousness of life beyond self. She found her epos in the reform of a religious order.

That Spanish woman who lived three hundred years ago, was certainly not

Machine-actionable

Structured data

typically tabular data: data held in a database in which all key/value pairs have identifiers and clear relations (each piece of data is separated by standard markers) and which follow an explicit data model.

Example: tabular data (spreadsheets, CSVs)

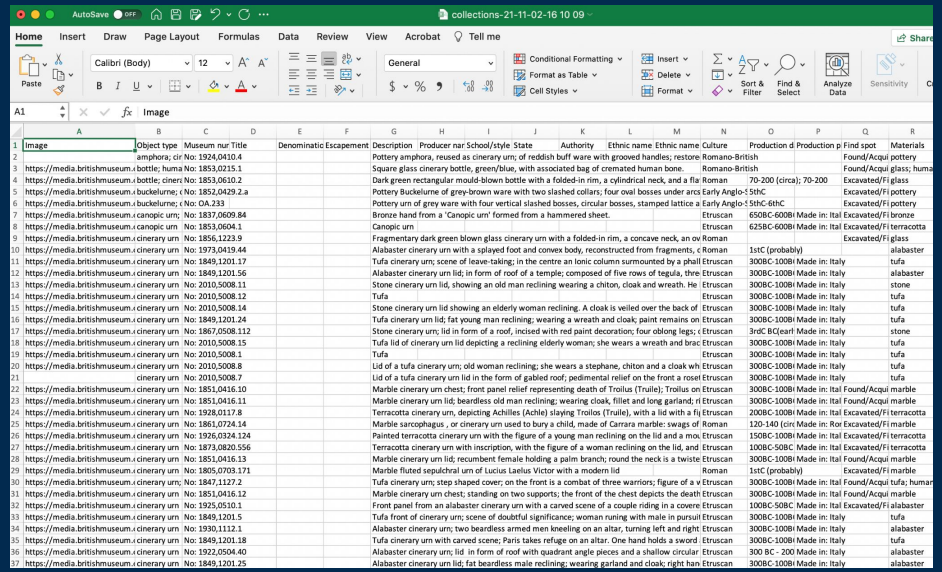


Image	Object type	Museum nur	Title	Denominatic	Escapement	Description	Producer nar	School/style	State	Authority	Ethnic name	Ethnic name	Culture	Production d	Production p	Find spot	Materials
	amphora; cir	No: 1924,0410.4				Pottery amphora, reused as cinerary urn; of reddish buff ware with grooved handles; restore							Romano-British			Found/Acqui	pottery
https://media.britishmuseum.c	bottle; huma	No: 1853,0215.1				Square glass cinerary bottle, green/blue, with associated bag of cremated human bone.							Romano-British			Found/Acqui	glass; huma
https://media.britishmuseum.c	bottle; cinera	No: 1853,0610.2				Dark green rectangular mould-blown bottle with a folded-in rim, a cylindrical neck, and a fla							Roman	70-200 (circa); 70-200		Excavated/Fi	glass
https://media.britishmuseum.c	buckelurne; c	No: 1852,0429.2.a				Pottery Buckelurne of grey-brown ware with two slashed collars; four oval bosses under arcs							Early Anglo-S	5thC		Excavated/Fi	pottery
https://media.britishmuseum.c	buckelurne; c	No: OA.233				Pottery urn of grey ware with four vertical slashed bosses, circular bosses, stamped lattice a							Early Anglo-S	5thC-6thC		Excavated/Fi	pottery
https://media.britishmuseum.c	canopic urn;	No: 1837,0609.84				Bronze hand from a 'Canopic urn' formed from a hammered sheet.							Etruscan	650BC-600B	Made in: Ital	Excavated/Fi	bronze
https://media.britishmuseum.c	canopic urn	No: 1853,0604.1				Canopic urn							Etruscan	625BC-600B	Made in: Ital	Excavated/Fi	terracotta
https://media.britishmuseum.c	cinerary urn	No: 1856,1223.9				Fragmentary dark green blown glass cinerary urn with a folded-in rim, a concave neck, an ov							Roman			Excavated/Fi	glass
https://media.britishmuseum.c	cinerary urn	No: 1973,0419.44				Alabaster cinerary urn with a splayed foot and convex body, reconstructed from fragments, c							Roman	1stC (probably)			alabaster
https://media.britishmuseum.c	cinerary urn	No: 1849,1201.17				Tufa cinerary urn; scene of leave-taking; in the centre an Ionic column surmounted by a phall							Etruscan	300BC-100B	Made in: Italy		tufa
https://media.britishmuseum.c	cinerary urn	No: 1849,1201.56				Alabaster cinerary urn lid; in form of roof of a temple; composed of five rows of tegula, thre							Etruscan	300BC-100B	Made in: Italy		alabaster
https://media.britishmuseum.c	cinerary urn	No: 2010,5008.11				Stone cinerary urn lid, showing an old man reclining wearing a chiton, cloak and wreath. He							Etruscan	300BC-100B	Made in: Italy		stone
https://media.britishmuseum.c	cinerary urn	No: 2010,5008.12				Tufa							Etruscan	300BC-100B	Made in: Italy		tufa
https://media.britishmuseum.c	cinerary urn	No: 2010,5008.14				Stone cinerary urn lid showing an elderly woman reclining. A cloak is veiled over the back of							Etruscan	300BC-100B	Made in: Italy		tufa
https://media.britishmuseum.c	cinerary urn	No: 1849,1201.24				Tufa cinerary urn lid; fat young man reclining; wearing a wreath and cloak; paint remains on							Etruscan	300BC-100B	Made in: Italy		tufa
https://media.britishmuseum.c	cinerary urn	No: 1867,0508.112				Stone cinerary urn; lid in form of a roof, incised with red paint decoration; four oblong legs; c							Etruscan	3rdC BC(earl	Made in: Italy		stone
https://media.britishmuseum.c	cinerary urn	No: 2010,5008.15				Tufa lid of cinerary urn lid depicting a reclining elderly woman; she wears a wreath and brac							Etruscan	300BC-100B	Made in: Italy		tufa
https://media.britishmuseum.c	cinerary urn	No: 2010,5008.1				Tufa							Etruscan	300BC-100B	Made in: Italy		tufa
https://media.britishmuseum.c	cinerary urn	No: 2010,5008.8				Lid of a tufa cinerary urn; old woman reclining; she wears a stephane, chiton and a cloak wh							Etruscan	300BC-100B	Made in: Italy		tufa
	cinerary urn	No: 2010,5008.7				Lid of a tufa cinerary urn lid in the form of gabled roof; pedimental relief on the front a rosel							Etruscan	300BC-100B	Made in: Italy		tufa
https://media.britishmuseum.c	cinerary urn	No: 1851,0416.10				Marble cinerary urn chest; front panel relief representing death of Troilus (Truile); Troilus on							Etruscan	300BC-100B	Made in: Ital	Found/Acqui	marble
https://media.britishmuseum.c	cinerary urn	No: 1851,0416.11				Marble cinerary urn lid; beardless old man reclining; wearing cloak, fillet and long garland; ri							Etruscan	300BC-100B	Made in: Ital	Found/Acqui	marble
https://media.britishmuseum.c	cinerary urn	No: 1928,0117.8				Terracotta cinerary urn, depicting Achilles (Achle) slaying Troilos (Truile), with a lid with a fi							Etruscan	200BC-100B	Made in: Ital	Excavated/Fi	terracotta
https://media.britishmuseum.c	cinerary urn	No: 1861,0724.14				Marble sarcophagus , or cinerary urn used to bury a child, made of Carrara marble: swags of							Roman	120-140 (circ	Made in: Ror	Excavated/Fi	marble
https://media.britishmuseum.c	cinerary urn	No: 1926,0324.124				Painted terracotta cinerary urn with the figure of a young man reclining on the lid and a mou							Etruscan	150BC-100B	Made in: Ital	Excavated/Fi	terracotta
https://media.britishmuseum.c	cinerary urn	No: 1873,0820.556				Terracotta cinerary urn with inscription, with the figure of a woman reclining on the lid, and							Etruscan	100BC-50BC	Made in: Ital	Excavated/Fi	terracotta
https://media.britishmuseum.c	cinerary urn	No: 1851,0416.13				Marble cinerary urn lid; recumbent female holding a palm branch; round the neck is a twiste							Etruscan	300BC-100B	Made in: Ital	Found/Acqui	marble
https://media.britishmuseum.c	cinerary urn	No: 1805,0703.171				Marble fluted sepulchral urn of Lucius Laelus Victor with a modern lid							Roman	1stC (probably)		Excavated/Fi	marble
https://media.britishmuseum.c	cinerary urn;	No: 1847,1127.2				Tufa cinerary urn; step shaped cover; on the front is a combat of three warriors; figure of a v							Etruscan	300BC-100B	Made in: Ital	Found/Acqui	tufa; human
https://media.britishmuseum.c	cinerary urn	No: 1851,0416.12				Marble cinerary urn chest; standing on two supports; the front of the chest depicts the death							Etruscan	300BC-100B	Made in: Ital	Found/Acqui	marble
https://media.britishmuseum.c	cinerary urn	No: 1925,0510.1				Front panel from an alabaster cinerary urn with a carved scene of a couple riding in a covere							Etruscan	100BC-50BC	Made in: Ital	Excavated/Fi	alabaster
https://media.britishmuseum.c	cinerary urn	No: 1849,1201.5				Tufa front of cinerary urn; scene of doubtful significance; woman runing with male in pursuit							Etruscan	300BC-100B	Made in: Italy		tufa
https://media.britishmuseum.c	cinerary urn	No: 1930,1112.1				Alabaster cinerary urn; two beardless armed men kneeling on an altar, turning left and right							Etruscan	300BC-100B	Made in: Italy		alabaster
https://media.britishmuseum.c	cinerary urn	No: 1849,1201.18				Tufa cinerary urn with carved scene; Paris takes refuge on an altar. One hand holds a sword							Etruscan	300BC-100B	Made in: Italy		tufa
https://media.britishmuseum.c	cinerary urn	No: 1922,0504.40				Alabaster cinerary urn; lid in form of roof with quadrant angle pieces and a shallow circular							Etruscan	300 BC - 200	Made in: Italy		alabaster
https://media.britishmuseum.c	cinerary urn	No: 1849,1201.25				Alabaster cinerary urn lid; fat beardless male reclining; wearing garland and cloak; right han							Etruscan	300BC-100B	Made in: Italy		alabaster

Machine-actionable

Semi-structured data

typically *not* organized in tables, this form uses tags or other markup to annotate semantic units and arrange them in a hierarchy according to a schema

Example: eXtensible Markup Language (XML) or TEI

A Simple XML Record Example

```
<?xml version="1.0" encoding="UTF-8"?>
<work type="play">
  <workName>The Tempest</workName>
  <writtenBy>
      <playwright>
         <playwrightName>William Shakespeare</playwrightName>
         <bornInPlace>Stratford Upon Avon</bornInPlace>
      </playwright>
   </writtenBy>
</work>
```

`<work>`, `<playwright>`, and `<bornInPlace>` are examples of XML elements
Stratford Upon Avon is an example of an element's value
`type="play"` is an example of at attribute name (type) and value (play)

Machine-actionable

Semi-structured data

typically *not* organized in tables, this form uses tags or other markup to annotate semantic units and arrange them in a hierarchy according to a schema

Example: eXtensible Markup Language (XML) or TEI

Used in archival and museum metadata (DACS, EAD, MARC)

A Simple XML Record Example

```
<?xml version="1.0" encoding="UTF-8"?>
<work type="play">
  <workName>The Tempest</workName>
  <writtenBy>
      <playwright>
         <playwrightName>William Shakespeare</playwrightName>
         <bornInPlace>Stratford Upon Avon</bornInPlace>
      </playwright>
   </writtenBy>
</work>
```

`<work>`, `<playwright>`, and `<bornInPlace>` are examples of XML elements
Stratford Upon Avon is an example of an element's value
`type="play"` is an example of at attribute name (type) and value (play)

Why use structured/semi-structured data?

- Makes it possible to use software to search for more meaningful and granular patterns
- Allows you to cross-reference other datasets
- Can allow for fast look-up or information retrieval
- Structured data is what allows for more abstract forms of exploratory analysis, data visualization or presentation

II

“Tidy” humanities data

What is “tidy” data?

- One variable per column
- One observation per row
- Consistent data types
- Unique identifiers for each object
- Clear names for your fields
- *If you find you’re trying to fit too much information into a column or row, use another table and cross-reference!*

Matthew Lincoln, “Tidy Humanities Data.” Carnegie Mellon, May 30, 2019.

Hadley Wickham, “Tidy Data.” *Journal of Statistical Software* 59 (September 12, 2014): 1–23.

Example of rich but messy data

Collection	Shelfmark	Document Type	People Mentioned	Locations Mentioned	Date in Document	Description (from PGP)
BODL	Bodl. MS Heb. c 72/14	Letter in Hebrew and Italian	Eliyya b. Elyaqim; Moshe b. Ye	Alexandria | Crete	5254, Hebrew calendar	Letter. In Hebrew. Eliyya b. Elyaqim, in Crete,
CUL	T-S Ar.29.105	Charity letter in Ladino	R. Jacob	Cairo; Jerusalem	5270?, Hebrew calendar	Letter in Ladino. Dating: possibly 16th century. T
CUL	T-S 13J7.30	Commercial letter in Ladino	Avraham Mayo; Yaʿaqov b. Ḥa	Cairo; Alexandria; Istanbul	5284, Hebrew calendar	Letter in Ladino from the merchant Avraham M

Example of rich but messy data

Collection	Shelfmark	Document Type	People Mentioned	Locations Mentioned	Date in Document	Description (from PGP)
BODL	Bodl. MS Heb. c 72/14	Letter in Hebrew and Italian	Eliyya b. Elyaqim; Moshe b. Ye	Alexandria | Crete	5254, Hebrew calendar	Letter. In Hebrew. Eliyya b. Elyaqim, in Crete,
CUL	T-S Ar.29.105	Charity letter in Ladino	R. Jacob	Cairo; Jerusalem	5270?, Hebrew calendar	Letter in Ladino. Dating: possibly 16th century. T
CUL	T-S 13J7.30	Commercial letter in Ladino	Avraham Mayo; Yaʿaqov b. Ḥa	Cairo; Alexandria; Istanbul	5284, Hebrew calendar	Letter in Ladino from the merchant Avraham M

- Document type includes genre *and* language

Example of rich but messy data

Collection	Shelfmark	Document Type	People Mentioned	Locations Mentioned	Date in Document	Description (from PGP)
BODL	Bodl. MS Heb. c 72/14	Letter in Hebrew and Italian	Eliyya b. Elyaqim; Moshe b. Ye	Alexandria | Crete	5254, Hebrew calendar	Letter. In Hebrew. Eliyya b. Elyaqim, in Crete,
CUL	T-S Ar.29.105	Charity letter in Ladino	R. Jacob	Cairo; Jerusalem	5270?, Hebrew calendar	Letter in Ladino. Dating: possibly 16th century. T
CUL	T-S 13J7.30	Commercial letter in Ladino	Avraham Mayo; Yaʿaqov b. Ḥa	Cairo; Alexandria; Istanbul	5284, Hebrew calendar	Letter in Ladino from the merchant Avraham M

- Document type includes genre *and* language
- Dates are not in standard numbers (ie "?") and not in Gregorian calendar format (and includes both numerical and textual data)

Example of rich but messy data

Collection	Shelfmark	Document Type	People Mentioned	Locations Mentioned	Date in Document	Description (from PGP)
BODL	Bodl. MS Heb. c 72/14	Letter in Hebrew and Italian	Eliyya b. Elyaqim; Moshe b. Ye	Alexandria | Crete	52[illegible]54, Hebrew calendar	Letter. In Hebrew. Eliyya b. Elyaqim, in Crete,
CUL	T-S Ar.29.105	Charity letter in Ladino	R. Jacob	Cairo; Jerusalem	52[illegible]70?, Hebrew calendar	Letter in Ladino. Dating: possibly 16th century. T
CUL	T-S 13J7.30	Commercial letter in Ladino	Avraham Mayo; Ya'aqov b. Ḥa	Cairo; Alexandria; Istanbul	52[illegible]84, Hebrew calendar	Letter in Ladino from the merchant Avraham M

- Document type includes genre *and* language
- Dates are not in standard numbers (ie "?") and not in Gregorian calendar format (and includes both numerical and textual data)
- Multiple values in the "Locations Mentioned" fields are separated by different kinds of delimiters (e.g. |, ;)

Example of “tidier” data

Collection	Shelfmark	Document Type	People Mentioned	Locations Mentioned	Language(s	Date in Document	Calendar	Converted Date	Date Uncertainty	Description (from PGP)
BODL	Bodl. MS Heb. c 72/1	Letter	Eliyya b. Elyaqim; Moshe b. Yehuda; Avraham Talmid; Yos	Alexandria; Crete	Hebrew; Ita an	5254	Hebrew	1484	date known	Letter. In Hebrew. Eliyya b. Elyaqim, in Crete,
CUL	T-S Ar.29.105	Charity Letter	R. Jacob	Cairo; Jerusalem	Ladino	5270	Hebrew	1500	date approx.	Letter in Ladino. Dating: possibly 16th century. T
CUL	T-S 13J7.30	Commercial Letter	Avraham Mayo; Ya'aqov b. Ḥabīb; Rahamin; Sulaymān	Cairo; Alexandria; Istanbul	Ladino	5284	Hebrew	1514	date known	Letter in Ladino from the merchant Avraham M

- Dates are now broken into separate numerical and qualitative fields and include converted Gregorian dates

Example of “tidier” data

Collection	Shelfmark	Document Type	eople Mentioned	Locations Mentioned	Language(s)	ate in Document	Calendar	Converted Date	Date Uncertainty	Description (from PGP)
BODL	Bodl. MS Heb. c 72,1	Letter	liyya b. Elyaqim; Moshe b. Yehuda; Avraham Talmid; Yos	Alexandria; Crete	Hebrew; Italian	5254	Hebrew	1484	date known	Letter. In Hebrew. Eliyya b. Elyaqim, in Crete,
CUL	T-S Ar.29.105	Charity Letter	L. Jacob	Cairo; Jerusalem	Ladino	5270	Hebrew	1500	date approx.	Letter in Ladino. Dating: possibly 16th century. T
CUL	T-S 13J7.30	Commercial Letter	vraham Mayo; Ya'aqov b. Ḥabīb; Rahamin; Sulaymān	Cairo; Alexandria; Istanbul	Ladino	5284	Hebrew	1514	date known	Letter in Ladino from the merchant Avraham M

- Dates are now broken into separate numerical and qualitative fields and include converted Gregorian dates
- Genre (document type) and document language(s) are now in separate columns

Example of “tidier” data

Collection	Shelfmark	Document Type	People Mentioned	Locations Mentioned	Language(s)	Date in Document	Calendar	Converted Date	Date Uncertainty	Description (from PGP)
BODL	Bodl. MS Heb. c 72/1	Letter	Eliyya b. Elyaqim; Moshe b. Yehuda; Avraham Talmid; Y	Alexandria; Crete	Hebrew; Italian	5254	Hebrew	1484	date known	Letter. In Hebrew. Eliyya b. Elyaqim, in Crete,
CUL	T-S Ar.29.105	Charity Letter	R. Jacob	Cairo; Jerusalem	Ladino	5270	Hebrew	1500	date approx.	Letter in Ladino. Dating: possibly 16th century. T
CUL	T-S 13J7.30	Commercial Letter	Avraham Mayo; Ya'aqov b. Ḥabīb; Rahamin; Sulaymān	Cairo; Alexandria; Istanbul	Ladino	5284	Hebrew	1514	date known	Letter in Ladino from the merchant Avraham M

- Dates are now broken into separate numerical and qualitative fields and include converted Gregorian dates
- Genre (document type) and document language(s) are now in separate columns
- Multiple values are separated by a standard delimiter (;)

Dataset excerpted from Grace Masback, "Romance Language Documents in the PGP" (2021): https://docs.google.com/spreadsheets/d/14J6ofY1crEpiOWDHXLxo812G_jFxiaXO-LaMDvkKrWg/edit#gid=1692110779

Controlled vocabularies

Controlled vocabulary = standardized language and terms

Often used to describe genres, formats, types, roles, or relationships

Examples include subject headings (like those used by the Library of Congress), as well as any predetermined set of vocabularies you've determined.

Controlled vocabularies

Use to link data to other datasets

Ex:

- Linking datasets internally
- Linking to other standards that other scholars/professionals might use (“Linked Open Data”)
  - https://vocab.getty.edu
  - https://www.w3.org/TR/rdf11-concepts/
  - https://programminghistorian.org/en/lessons/intro-to-linked-data

I

Making tidy data involves archival & editorial decisions

Some tips:

- Choose categories that matter for YOUR research question
- Record what you know about what you don't know
- Make a vocabulary for recording uncertainties:
  - missing, illegible, approximated
- Add a column for context!
  - Ex: you might have a column called `**date**` and a column called `**date uncertainty**`, where you might record the level of uncertainty
- Include a general "notes" or “descriptive notes” column
- Create "data dictionaries" explaining what your fields (your column labels) mean and how they were derived

Try to avoid:

- Using different ways to indicate precision
- Being imprecise about uncertainty in your data
  - Ex: including a column called `**uncertainty**` rather than specifying *what* is uncertain
- Aiming for an exhaustively detailed dataset: no dataset will be able to capture everything about the objects it describes

IV

Making Structured Data: 3 Case Studies

Comparative Literature

Let’s say you wanted to compare multiple translations (and multiple editions) of the first lines of Hamlet’s “to be or not to be” soliloquy...

To be or not to be - multilingual

21 Total Versions | Line numbers | Bibliographic panel | Critical introduction

Bibliographic Information

To be or not to be - multilingual

Original Source

See witness list

Witness List

- **Witness HQ1:** First Quarto 1603
- **Witness HQ2:** Second Quarto 1604/5
- **Witness HaF:** First Folio 1623
- **Witness TS:** T. J. B. Spencer 1980
- **Witness MW:** B. W. Mowat and P. Werstine 1992
- **Witness BQ1:** D. Bevington ISE Q1
- **Witness BQ2:** D. Bevington ISE Q2
- **Witness BF:** D. Bevington ISE F1
- **Witness BE:** D. Bevington ISE Editor's
- **Witness LFM:** Luis Fernández de Moratín 1798
- **Witness GM73:** Guillermo Macpherson 1873
- **Witness GM79:** Guillermo Macpherson 1879
- **Witness GM82:** Guillermo Macpherson 1882
- **Witness GM85:** Guillermo Macpherson 1885
- **Witness FVH2:** François-Victor Hugo (Second) 1865
- **Witness FVH1:** François-Victor Hugo (Premier) 1865
- **Witness FG:** François Guizot 1864
- **Witness CMW:** C. M. Wieland 1762-66
- **Witness ST:** A. W. Schlegel und L. Tieck 1825-33
- **Witness CB:** Carlo Busconi 1852

Version HQ1: First Quarto 1603

1710 To be, or not to be, I there's the point,
1714 To Die, to sleepe,

Version HQ2: Second Quarto 1604/5

1710 To be, or not to be, that is the question,
1711 Whether tis nobler in the minde to suffer
1712 The slings and arrowes of outragious fortune,
1713 Or to take Armes against a sea of troubles,
1714 And by opposing, end them, to die to sleepe

Version GM73: Guillermo Macpherson 1873

1710 Ser o dejar de ser: he ahí el problema.
1711 Si es con arreglo á la razón, mas noble
1712 Sufrir los golpes, los agudos dardos
De atroz fortuna, ó terminar la lucha
1713 Armas haciendo contra un mar de penas.
1714 Morir; dormir;

Version FVH2: François-Victor Hugo (Second) 1865

1710 Être, ou ne pas être, c'est là la question.
1711 — Y a-t-il plus de noblesse d'âme à subir
1712 — la fronde et les flèches de la fortune
outrageante,
1713 — ou bien à s'armer contre une mer de
douleurs
1714 — et à l'arrêter par une révolte ? Mourir…
dormir,

https://www.uv.es/shaxpere/vm/samples/To_be_multiling2.html

4. William Shakespeare: *Titus Andronicus*

The following example is a fragment (the front matter, and scene 2 of the second act) of William Shakespeare's *Titus Andronicus*, encoded and made available by the Perseus Digital Library.

The text of the play is preceded by front matter, consisting of a character list, and a prologue. The
is encoded as a **<castList>** structure within a **<div1>** container in the **<front>** part. The cast list c
<castItem> elements, listing the roles (**<role>**) with their description (**<roleDesc>**). Each role is id
the **@xml:id** attribute. Three named groups of characters are grouped into **<castGroup>** elements;
nameless group of minor characters is listed as **<castItem type="list">**. Notice, how in the latter
both **<role>** and **<roleDesc>** are used a bit indiscriminate at first sight (e.g., both

```
<roleDesc>Romans</roleDesc>
```

and

```
<role xml:id="tit-11">Goths and Romans</role>
```

occur). On second sight, however, **<role>** appears to be used for all speaking characters, who are f
identified with an **@xml:id** attribute. The front matter is concluded with a prologue (**<prologue>**) co
28 lines spoken by the Chorus. The cast list is succeeded by a general description of the setting in
action takes place, in the **<set>** element.

The body of the play (**<body>**) consists of 14 scenes, grouped into 5 acts. Acts are encoded in **<div**
in which the scenes occur as **<div2>** elements. Each speech is marked with **<sp>**, containing the in
the speaker as it occurs in the source text (**<speaker>**), as well as a formal indication (using the **@w**
attribute). Stage instructions are encoded inside **<stage>**. The speeches are encoded as verse lines (**<l>**)
Notice, however, how logical lines (**<l>**) are distinguished from typographic lines: the latter are explicitly
encoded with the **<lb>** element, occurring inside **<l>**.

```
<body>
  <!— ... —>
  <div1 n="2" type="act" org="uniform" sample="complete">
    <head>ACT II</head>
    <!— ... —>
    <div2 n="2" type="scene" org="uniform" sample="complete">
      <head>SCENE II</head>
      <stage type="setting"> A forest near Rome. Horns and cry of hounds heard. </stage>
      <lb n="699" ed="F1"/>
      <stage type="entrance">Enter TITUS ANDRONICUS, <lb n="700" ed="F1"/>with Hunters, etc., MARCUS, LUCIUS,
      QUINTUS, and MARTIUS.</stage>
      <lb ed="G"/>
      <lb n="701" ed="F1"/>
      <sp who="#tit-25">
        <speaker>Tit.</speaker>
        <l>The hunt is up, the morn is bright and grey, <lb ed="G"/> <lb n="702" ed="F1"/></l>
        <l>The fields are fragrant and the woods are green: <lb ed="G"/> <lb n="703" ed="F1"/></l>
        <l>Uncouple here and let us make a bay <lb ed="G"/> <lb n="704" ed="F1"/></l>
        <l>And wake the emperor and his lovely bride <lb ed="G"/> <lb n="705" ed="F1"/></l>
        <l>And rouse the prince and ring a hunter's peal, <lb ed="G"/> <lb n="706" ed="F1"/></l>
        <l>That all the court may echo with the noise. <lb ed="G"/> <lb n="707" ed="F1"/></l>
        <l>Sons, let it be your charge, as it is ours, <lb ed="G"/> <lb n="708" ed="F1"/></l>
        <l>To attend the emperor's person carefully: <lb ed="G"/> <lb n="709" ed="F1"/></l>
        <l>I have been troubled in my sleep this night, <lb ed="G"/> <lb n="710" ed="F1"/></l>
        <l>But dawning day new comfort hath inspired. <lb n="711" ed="F1"/> <stage>A cry of hounds, and horns
        winded in a peal.</stage> <lb n="712" ed="F1"/> <lb n="713" ed="F1"/> <stage type="entrance">Enter S
        ATURNINUS, TAMORA, BASSIANUS, LAVINIA, DEMETRIUS, CHIRON, <lb n="714" ed="F1"/>and Attendants.
        </stage> <lb n="11" ed="G"/> <lb n="715" ed="F1"/></l>
        <l>Many good morrows to your majesty; <lb ed="G"/> <lb n="716" ed="F1"/></l>
        <l>Madam, to you as many and as good: <lb ed="G"/> <lb n="717" ed="F1"/></l>
        <l>I promised your grace a hunter's peal. <lb ed="G"/> <lb n="718" ed="F1"/></l>
      </sp>
      <sp who="#tit-22">
        <speaker>Sat.</speaker>
        <l>And you have rung it lustily, my lord: <lb ed="G"/> <lb n="719" ed="F1"/></l>
```

https://teibyexample.org/examples/TBED05v00.htm#shakespeare

4. William Shakes

The following example is a
Titus Andronicus, encoded

The text of the play is prec
is encoded as a **<castList**
<castItem> elements, listi
the **@xml:id** attribute. Thre
nameless group of minor c
both **<role>** and **<roleDes**

```
<roleDesc>Romans</roleDesc
```

and

```
<role xml:id="tit-11">Goth
```

occur). On second sight, h
identified with an **@xml:id**
28 lines spoken by the Cho
action takes place, in the **<**

The body of the play (**<bod**
in which the scenes occur a
the speaker as it occurs in
attribute). Stage instructio
Notice, however, how logic
encoded with the **<lb>** element, occurring inside **<l>**.

```
<body>
  <!-- ... -->
  <div1 n="2" type="act" org="uniform" sample="complete">
    <head>ACT II</head>
    <!-- ... -->
    <div2 n="2" type="scene" org="uniform" sample="complete">
      <head>SCENE II</head>
      <stage type="setting"> A forest near Rome. Horns and cry of hounds heard. </stage>
      <lb n="699" ed="F1"/>
      <stage type="entrance">Enter TITUS ANDRONICUS, <lb n="700" ed="F1"/>with Hunters, etc., MARCUS, LUCIUS,
      QUINTUS, and MARTIUS.</stage>
      <lb ed="G"/>
      <lb n="701" ed="F1"/>
      <sp who="#tit-25">
        <speaker>Tit.</speaker>
        <l>The hunt is up, the morn is bright and grey, <lb ed="G"/> <lb n="702" ed="F1"/></l>
        <l>The fields are fragrant and the woods are green: <lb ed="G"/> <lb n="703" ed="F1"/></l>
        <l>Uncouple here and let us make a bay <lb ed="G"/> <lb n="704" ed="F1"/></l>
        <l>And wake the emperor and his lovely bride <lb ed="G"/> <lb n="705" ed="F1"/></l>
        <l>And rouse the prince and ring a hunter's peal, <lb ed="G"/> <lb n="706" ed="F1"/></l>
        <l>That all the court may echo with the noise. <lb ed="G"/> <lb n="707" ed="F1"/></l>
        <l>Sons, let it be your charge, as it is ours, <lb ed="G"/> <lb n="708" ed="F1"/></l>
        <l>To attend the emperor's person carefully: <lb ed="G"/> <lb n="709" ed="F1"/></l>
        <l>I have been troubled in my sleep this night, <lb ed="G"/> <lb n="710" ed="F1"/></l>
        <l>But dawning day new comfort hath inspired. <lb n="711" ed="F1"/> <stage>A cry of hounds, and horns
        winded in a peal.</stage> <lb n="712" ed="F1"/> <lb n="713" ed="F1"/> <stage type="entrance">Enter S
        ATURNINUS, TAMORA, BASSIANUS, LAVINIA, DEMETRIUS, CHIRON, <lb n="714" ed="F1"/>and Attendants.
        </stage> <lb n="11" ed="G"/> <lb n="715" ed="F1"/></l>
        <l>Many good morrows to your majesty; <lb ed="G"/> <lb n="716" ed="F1"/></l>
        <l>Madam, to you as many and as good: <lb ed="G"/> <lb n="717" ed="F1"/></l>
        <l>I promised your grace a hunter's peal. <lb ed="G"/> <lb n="718" ed="F1"/></l>
      </sp>
      <sp who="#tit-22">
        <speaker>Sat.</speaker>
        <l>And you have rung it lustily, my lord: <lb ed="G"/> <lb n="719" ed="F1"/></l>
```

https://teibyexample.org/examples/TBED05v00.htm#shakespeare

Art History

Let’s say you’re interested in studying the artists and individuals represented in the holdings of Princeton paintings and sculptures held in the Princeton Portraits, Campus Collection….

PRINCETON UNIVERSITY
ART MUSEUM

VISIT COLLECTIONS EXHIBITIONS SUPPORT LEARN ABOUT

Campus Collections

The Campus Collections include dozens of sculptures, paintings, monuments, and memorials important to the University’s rich history and its singular traditions. One segment of the Campus Collections, the Princeton Portraits, is a collection of more than 600 paintings and sculptures, predominantly likenesses of important Princetonians as well as non-portrait works that relate to the history of the University. The Princeton Portraits span centuries and encompass enormously varying styles and subjects, including Charles Willson Peale’s landmark portrait of *George Washington at the Battle of Princeton* (1783), commissioned with funds given by Washington himself. Other highlights include a remarkable sequence of dinosaur paintings by the Victorian artist and naturalist Benjamin Waterhouse Hawkins, commissioned in 1876 by then University President James McCosh as a progressive response to Darwin’s theories.

Perhaps the most celebrated works in the Campus Collections belong to the John B. Putnam Jr. Memorial Collection of outdoor sculpture, distributed across the University campus. Comprising 22 sculptures by master artists of the 20th and 21st centuries, the Putnam Collection is the result of a generous gift of funds in the 1960s by an anonymous donor as a memorial to a Princeton alumnus killed in World War II. Commonly regarded as one of the greatest single collections of public art, the Putnam Collection can be thought of as a plein-air lesson in art history by many of the great masters of the modern canon. Highlights include Sir Henry Moore’s cast bronze organic abstraction *Oval with*

objectid	objectnumber	title	displaydate	datebegin	dateend	displaymaker	medium	classificatior	department
44861	PP1	William III, King of England, Prince o	ca. 1672,Äì89	1672	1689	Caspar Netscher, Dutch, 1639,Äì1684 | Person depicted: William III, Prince of Orange & Nassau &	Oil on canvas	Paintings	European Painting and Scu
44862	PP10	Gilbert Tennent (1703,Äì1764), Trustee	(1747,Äì64)	1796	1842	Jacob Eichholtz, American, 1776,Äì1842 | after Gustavus Hesselius, American, 1682,Äì1755	Oil on canvas	Paintings	American Art
44862	PP10	Gilbert Tennent (1703,Äì1764), Trustee	(1747,Äì64)	1796	1842	Jacob Eichholtz, American, 1776,Äì1842 | after Gustavus Hesselius, American, 1682,Äì1755	Oil on canvas	Paintings	American Art
44857	PP100	Laurence Hutton (1843-1904)	1892	1892	1892	Dora Wheeler Keith, American, 1857,Äì1940	Oil on canvas	Paintings	American Art
44863	PP101	Portrait of Laurence Hutton	1894	1894	1894	Dora Wheeler Keith, American, 1857,Äì1940 | Person depicted: Laurence Hutton, American, 1843,/	Pastel	Drawings	Prints and Drawings
44867	PP103	Christian Frederick Gauss (1878-195	1946	1946	1946	Henry Ernest Schnakenberg, American, 1892,Äì1970	Oil on canvas	Paintings	American Art
44874	PP104	Robert Kilburn Root (1877-1950)	1945	1945	1945	Ernest Ludwig Ipsen, American, 1869,Äì1951	Oil on canvas	Paintings	American Art
44875	PP105	Henry Norris Russell, Class of 1897 (	1929	1929	1929	Samuel Morrow Palmer, American, 1876,Äì1936	Oil on canvas	Paintings	American Art
44876	PP106	William John Sinclair (1877-1935)	1938	1938	1938	Robert Bruce Horsfall, American, 1869,Äì1948	Oil on canvas	Paintings	American Art
45070	PP107	Augustus Trowbridge (1870-1934)	1934	1934	1934	Enrico Licari	Bronze	Sculpture	European Painting and Scu
45071	PP108	Edwin Grant Conklin (1863-1952)	1938	1938	1938	John Young-Hunter, American, 1874,Äì1955	Oil on canvas	Paintings	American Art
45072	PP109	Radcliffe Heermance (1882-1958)	1946	1946	1946	Edmund Minor Archer, American, 1904-,Äì1986	Oil on canvas	Paintings	American Art
45073	PP11	Samuel Finley (1715,Äì1766), Presid	1870	1870	1870	Charles Walker Lind, American, active 1850,Äì1875, died 1882 | after John Hesselius, American, 1	Oil on canvas	Paintings	American Art
45074	PP110	Frank Albert Fetter (1863-1949)	1946	1946	1946	Joe Brown, American, 1909,Äì1985	Bronze	Sculpture	American Art
45075	PP111	Alfred Goldsborough Mayor (1868-1922)		1922	1923	Harriet Hyatt Mayor, American, 1868,Äì1960	Composition plac	Reliefs	American Art
45076	PP112	Henry Lane Eno (1871-1928)		1891	1934	Unknown	Oil on canvas	Paintings	American Art
45078	PP113	Arthur Maurice Greene, Dean of the	1946	1946	1946	Maurice Molarsky, American, 1885,Äì1950	Oil on canvas	Paintings	American Art
45081	PP116	Samuel Woodruff (1700-1768)	1897	1897	1897	Clara Wokeman	Oil on canvas	Paintings	American Art

Data collected from PUAM REST API. See:
https://github.com/Princeton-University-Art-Museum/puam-api-docs

pportraits.csv

objectid	objectnumber	title	displaydate	datebegin	dateend	displaymaker	medium	classificatior	department
44861	PP1	William III, King of England, Prince o	ca. 1672,Äì89	1672	1689	Caspar Netscher, Dutch, 1639,Äì1684 | Person depicted: William III, Prince of Orange & Nassau &	Oil on canvas	Paintings	European Painting and Scu
44862	PP10	Gilbert Tennent (1703,Äì1764), Trustee (1747,Äì64)		1796	1842	Jacob Eichholtz, American, 1776,Äì1842 | after Gustavus Hesselius, American, 1682,Äì1755	Oil on canvas	Paintings	American Art
44862	PP10	Gilbert Tennent (1703,Äì1764), Trustee (1747,Äì64)		1796	1842	Jacob Eichholtz, American, 1776,Äì1842 | after Gustavus Hesselius, American, 1682,Äì1755	Oil on canvas	Paintings	American Art
44857	PP100	Laurence Hutton (1843-1904)	1892	1892	1892	Dora Wheeler Keith, American, 1857,Äì1940	Oil on canvas	Paintings	American Art
44863	PP101	Portrait of Laurence Hutton	1894	1894	1894	Dora Wheeler Keith, American, 1857,Äì1940 | Person depicted: Laurence Hutton, American, 1843,/	Pastel	Drawings	Prints and Drawings
44867	PP103	Christian Frederick Gauss (1878-195	1946	1946	1946	Henry Ernest Schnakenberg, American, 1892,Äì1970	Oil on canvas	Paintings	American Art
44874	PP104	Robert Kilburn Root (1877-1950)	1945	1945	1945	Ernest Ludwig Ipsen, American, 1869,Äì1951	Oil on canvas	Paintings	American Art
44875	PP105	Henry Norris Russell, Class of 1897 (	1929	1929	1929	Samuel Morrow Palmer, American, 1876,Äì1936	Oil on canvas	Paintings	American Art
44876	PP106	William John Sinclair (1877-1935)	1938	1938	1938	Robert Bruce Horsfall, American, 1869,Äì1948	Oil on canvas	Paintings	American Art
45070	PP107	Augustus Trowbridge (1870-1934)	1934	1934	1934	Enrico Licari	Bronze	Sculpture	European Painting and Scu
45071	PP108	Edwin Grant Conklin (1863-1952)	1938	1938	1938	John Young-Hunter, American, 1874,Äì1955	Oil on canvas	Paintings	American Art
45072	PP109	Radcliffe Heermance (1882-1958)	1946	1946	1946	Edmund Minor Archer, American, 1904-,Äì1986	Oil on canvas	Paintings	American Art
45073	PP11	Samuel Finley (1715,Äì1766), Presid	1870	1870	1870	Charles Walker Lind, American, active 1850,Äì1875, died 1882 | after John Hesselius, American, 1	Oil on canvas	Paintings	American Art
45074	PP110	Frank Albert Fetter (1863-1949)	1946	1946	1946	Joe Brown, American, 1909,Äì1985	Bronze	Sculpture	American Art
45075	PP111	Alfred Goldsborough Mayor (1868-1922)		1922	1923	Harriet Hyatt Mayor, American, 1868,Äì1960	Composition plac	Reliefs	American Art
45076	PP112	Henry Lane Eno (1871-1928)		1891	1934	Unknown	Oil on canvas	Paintings	American Art
45078	PP113	Arthur Maurice Greene, Dean of the	1946	1946	1946	Maurice Molarsky, American, 1885,Äì1950	Oil on canvas	Paintings	American Art
45081	PP116	Samuel Woodruff (1700-1768)	1897	1897	1897	Clara Wokeman	Oil on canvas	Paintings	American Art

Because our data is structured, we could sort, say, to examine all works where the medium is “bronze” by sorting by tag:

Data collected from PUAM REST API. See:
https://github.com/Princeton-University-Art-Museum/puam-api-docs

objectid	objectnumber	title	displaydate	datebegin	dateend	displaymaker	medium	classificatior	department
45070	PP107	Augustus Trowbridge (1870-1934)	1934	1934	1934	Enrico Licari	Bronze	Sculpture	European Painti
45074	PP110	Frank Albert Fetter (1863-1949)	1946	1946	1946	Joe Brown, American, 1909,Äì1985	Bronze	Sculpture	American Art
45091	PP121	Elias Boudinot (1740-1821)		1881	1902	William Ordway Partridge, American, 1861,Äì1930	Bronze	Reliefs	American Art
44695	PP146	John Insley Blair (1802-1899)	1883	1883	1883	Olin Levi Warner, American, 1844,Äì1896 | Gorham Manufacturing Co., American, established in 1	Bronze	Sculpture	American Art
44695	PP146	John Insley Blair (1802-1899)	1883	1883	1883	Olin Levi Warner, American, 1844,Äì1896 | Gorham Manufacturing Co., American, established in 1	Bronze	Sculpture	American Art
44699	PP149	Nathaniel Norris Halsted (1816-1884	1891	1891	1891	Unknown | Henry-Bonnard Bronze Company, American, founded 1890's	Bronze	Reliefs	American Art
45158	PP158	[Stephen] Grover Cleveland (1837-19	1937	1937	1937	Rudolph Evans, American, 1878,Äì1960	Bronze	Sculpture	American Art
45161	PP161	William Cooper Procter, Class of 188	1923	1923	1923	Robert Tait McKenzie, American, born Canada, 1867,Äì1938	Bronze	Reliefs	American Art
45175	PP172	James Madison, Class of 1771 (1751	1892	1892	1892	John Flanagan, American, 1865,Äì1952	Bronze	Reliefs	American Art
45209	PP196	Huntington Wolcott Jackson, Class of	1911	1911	1911	Henry Hering, American, 1874,Äì1949	Bronze	Reliefs	American Art
45213	PP199	Charles Ira Young, Class of 1883 (1862-1912)		1891	1915	Theodore Spicer-Simson, British, 1871,Äì1959	Bronze	Reliefs	American Art
45225	PP206	William Logan MacCoy, Jr., Class of 1	1945	1945	1945	Joe Brown, American, 1909,Äì1985	Bronze	Reliefs	American Art
45203	PP21	John Witherspoon (1723-1794)	1876	1876	1876	Unknown	Bronze	Medals and I	American Art
45229	PP211	Henry Pomeroy Davison (1867-1922)	1920	1920	1920	Jo Davidson, American, 1883,Äì1952	Bronze	Sculpture	American Art

objectid	objectnumber	title	displaydate	datebegin	dateend	displaymaker	medium	classificatior	department
44861	PP1	William III, King of England, Prince o	ca. 1672,Äì89	1672	1689	Caspar Netscher, Dutch, 1639,Äì1684 | Person depicted: William III, Prince of Orange & Nassau &	Oil on canvas	Paintings	European Painting and Scu
44862	PP10	Gilbert Tennent (1703,Äì1764), Trustee	(1747,Äì64)	1796	1842	Jacob Eichholtz, American, 1776,Äì1842 | after Gustavus Hesselius, American, 1682,Äì1755	Oil on canvas	Paintings	American Art
44862	PP10	Gilbert Tennent (1703,Äì1764), Trustee	(1747,Äì64)	1796	1842	Jacob Eichholtz, American, 1776,Äì1842 | after Gustavus Hesselius, American, 1682,Äì1755	Oil on canvas	Paintings	American Art
44857	PP100	Laurence Hutton (1843-1904)	1892	1892	1892	Dora Wheeler Keith, American, 1857,Äì1940	Oil on canvas	Paintings	American Art
44863	PP101	Portrait of Laurence Hutton	1894	1894	1894	Dora Wheeler Keith, American, 1857,Äì1940 | Person depicted: Laurence Hutton, American, 1843,/	Pastel	Drawings	Prints and Drawings
44867	PP103	Christian Frederick Gauss (1878-195	1946	1946	1946	Henry Ernest Schnakenberg, American, 1892,Äì1970	Oil on canvas	Paintings	American Art
44874	PP104	Robert Kilburn Root (1877-1950)	1945	1945	1945	Ernest Ludwig Ipsen, American, 1869,Äì1951	Oil on canvas	Paintings	American Art
44875	PP105	Henry Norris Russell, Class of 1897 (	1929	1929	1929	Samuel Morrow Palmer, American, 1876,Äì1936	Oil on canvas	Paintings	American Art
44876	PP106	William John Sinclair (1877-1935)	1938	1938	1938	Robert Bruce Horsfall, American, 1869,Äì1948	Oil on canvas	Paintings	American Art
45070	PP107	Augustus Trowbridge (1870-1934)	1934	1934	1934	Enrico Licari	Bronze	Sculpture	European Painting and Scu
45071	PP108	Edwin Grant Conklin (1863-1952)	1938	1938	1938	John Young-Hunter, American, 1874,Äì1955	Oil on canvas	Paintings	American Art
45072	PP109	Radcliffe Heermance (1882-1958)	1946	1946	1946	Edmund Minor Archer, American, 1904-,Äì1986	Oil on canvas	Paintings	American Art
45073	PP11	Samuel Finley (1715,Äì1766), Presid	1870	1870	1870	Charles Walker Lind, American, active 1850,Äì1875, died 1882 | after John Hesselius, American, 1	Oil on canvas	Paintings	American Art
45074	PP110	Frank Albert Fetter (1863-1949)	1946	1946	1946	Joe Brown, American, 1909,Äì1985	Bronze	Sculpture	American Art
45075	PP111	Alfred Goldsborough Mayor (1868-1922)		1922	1923	Harriet Hyatt Mayor, American, 1868,Äì1960	Composition plac	Reliefs	American Art
45076	PP112	Henry Lane Eno (1871-1928)		1891	1934	Unknown	Oil on canvas	Paintings	American Art
45078	PP113	Arthur Maurice Greene, Dean of the	1946	1946	1946	Maurice Molarsky, American, 1885,Äì1950	Oil on canvas	Paintings	American Art
45081	PP116	Samuel Woodruff (1700-1768)	1897	1897	1897	Clara Wokeman	Oil on canvas	Paintings	American Art

How else might we make this dataset more “tidy”?

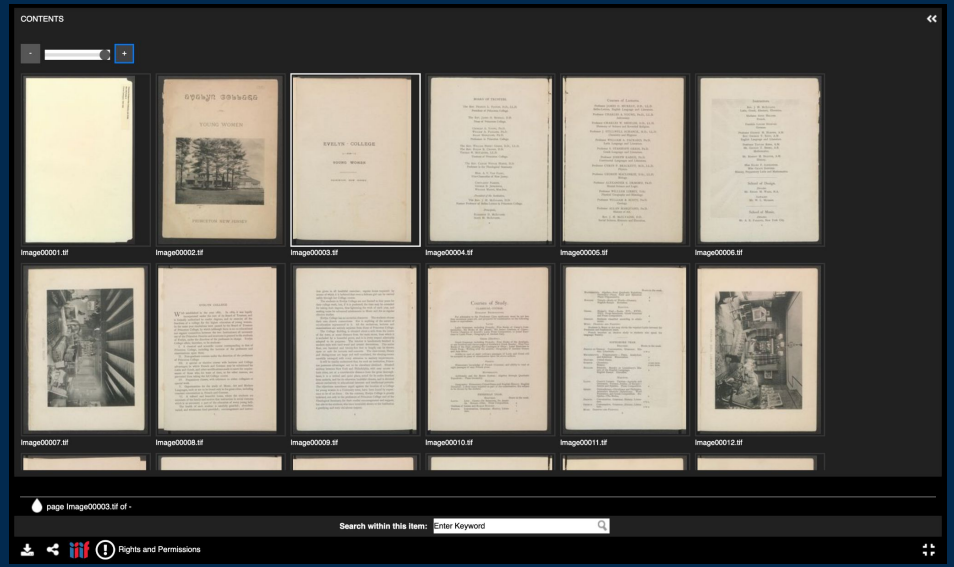
History

Let’s say you want to create a dataset drawing from a complicated primary source like the Evelyn College Course Catalogue (1890-1897).

Link to the facsimile of the catalogue: https://bit.ly/2ZrhSJk

Let’s say you want to create a dataset drawing from a complicated primary source like the Evelyn College Course Catalogue (1890-1897).

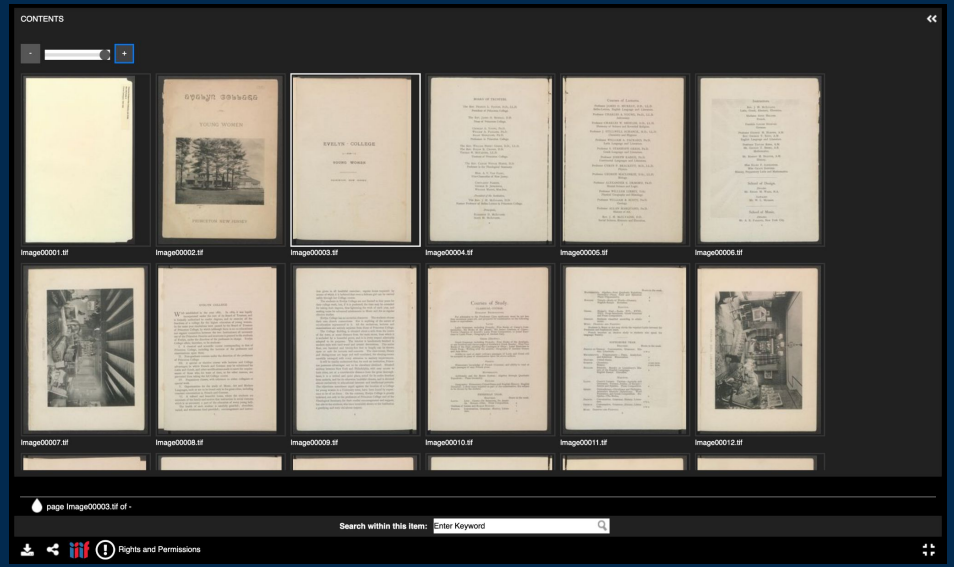
Link to the facsimile of the catalogue: https://bit.ly/2ZrhSJk



Let’s say you want to create a dataset drawing from a complicated primary source like the Evelyn College Course Catalogue (1890-1897).

Link to the facsimile of the catalogue: https://bit.ly/2ZrhSJk

How would you transform this primary source into structured (or semi-structured) data?



Making tidy data involves archival & editorial decisions

V

Ask Us Anything (Q & A)

THANK YOU!

Join us on **Feb. 2** for

Workshop 4: Cleaning Humanities Data

sceckert@princeton.edu

References:

Berry, Dorothy. "Digitizing and Enhancing Description Across Collections to Make African American Materials More Discoverable on Umbra Search African American History." *The Design for Diversity Learning Toolkit* (blog), August 2, 2018.

Blaney, Jonathan. " ntroduction to the Principles of Linked Open Data." *Programming Historian*, May 7, 2017.

Koeser, Rebecca Sutton. "Is a Spreadsheet a Database?" The Center for Digital Humanities at Princeton, February 11, 2021.

———. "Old Project, New Database — a Milestone for the Princeton Geniza Project." The Center for Digital Humanities at Princeton, July 1, 2021.

Lincoln, Matthew. "Tidy Humanities Data." Carnegie Mellon, May 30, 2019.

Lincoln, Matthew D. "Best Practices for Using Google Sheets in Your Data Project." *Matthew Lincoln, PhD* (blog), March 26, 2018.

Merry, Mark. "Designing Databases for Historical Research: Historical Research Handbook." School of Advanced Study, University of London..

Sanderson, Robert. "RDF: Resource Description Failures and Linked Data Letdowns." *Journal of Digital Humanities* 2, no. 3 (Summer 2013).

Schöch, Christof. "Big? Smart? Clean? Messy? Data in the Humanities." *Journal of Digital Humanities* 2, no. 3 (Summer 2013).

Wickham, Hadley. "Tidy Data." *Journal of Statistical Software* 59 (September 12, 2014): 1–23. https://doi.org/10.18637/jss.v059.i10.

Ziegler, S.L. "Open Data in Cultural Heritage Institutions: Can We Be Better Than Data Brokers?" *Digital Humanities Quarterly* 014, no. 2 (June 19, 2020).