



Introduction to Digital Humanities

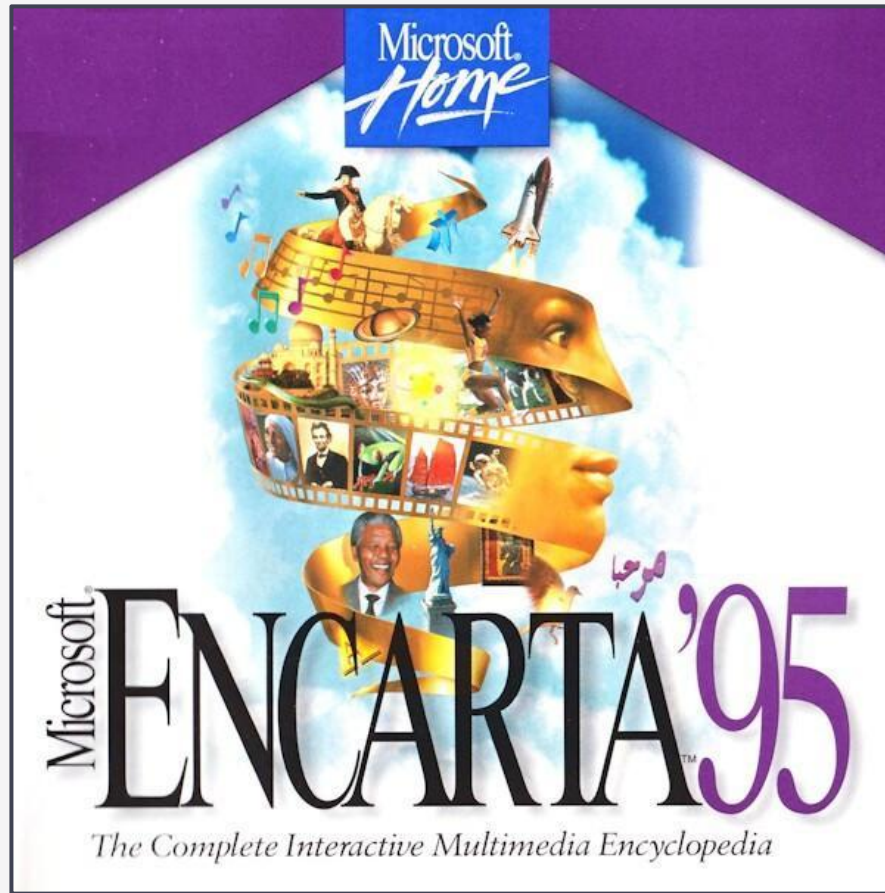
Week 2.2 – Data's narratives

database <-> narrative

database <-> narrative

> Lev Manovich (1999), *Database as symbolic form*

- database = a new symbolic form of the 20th century
- new form of cultural expression
- distinct from the narrative
- ... although it will sometimes present it as such



Microsoft Encarta '95 cd-rom case

person ID	call ID	call time	location	call to
#41	#5345	00:01:30	40.3° N, 74.6° W	home improvement store
#41	#5346	00:05:43	40.2° N, 74.8° W	locksmith
#41	#5347	00:16:51	41.1° N 74.3° W	hydroponics dealer
#41	#5348	00:04:09	41.1° N 74.3° W	head shop

after Pomerantz (2015, p. 2)



René Magritte, *La Trahison des Images* (LACMA, 1929)

understanding metadata

understanding metadata

> types of metadata

- **Descriptive** – description of the object
e.g. title, author, keywords, etc.
- **Administrative** – information to help manage a resource
e.g. when and how it was created, file type, and other technical information
- **Structural** – indicates how compound objects are put together
e.g. how pages are ordered to form chapters
- **Preservation** – information needed to archive and preserve a resource
e.g. do we need to emulate an application or operating system to interact with a digital file?
- **Use** – information about how the resource has been used
e.g. number of downloads the file has received, on what dates, data about the users etc.

understanding metadata

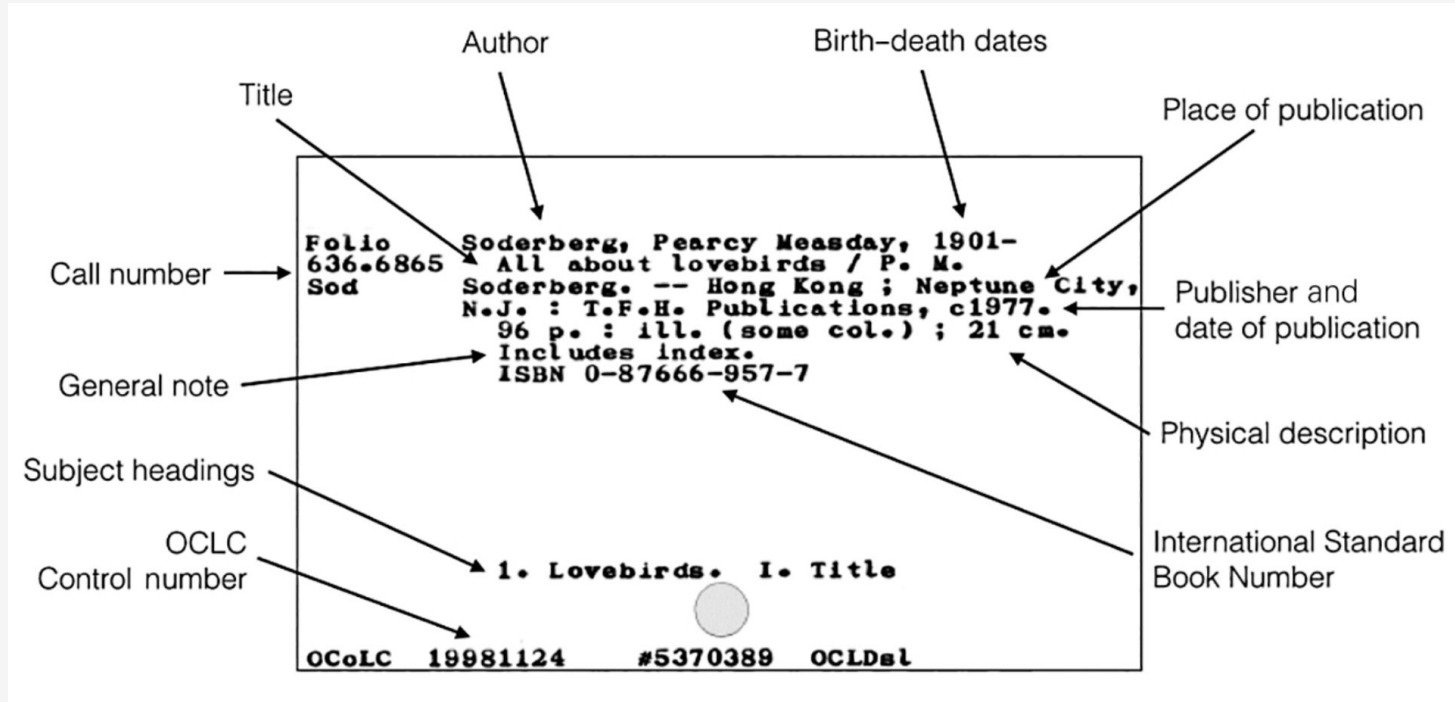


Figure 1 in Pomerantz (2015, p. 9)

understanding metadata

Table 1

Title	Author	Date of publication	Subject	Call number	Pages
<i>Intellectual Property Strategy</i>	Palfrey, John	2012	Intellectual property—Management	HD53 .P35 2012	172
<i>Open Access</i>	Suber, Peter	2012	Open access publishing	Z286.O63 S83 2012	242
<i>Memes in Digital Culture</i>	Shifman, Limor	2014	Social evolution. Memes. Culture diffusion. Internet—Social aspects. Memetics.	HM626 .S55 2014	200

metadata standards

metadata standards

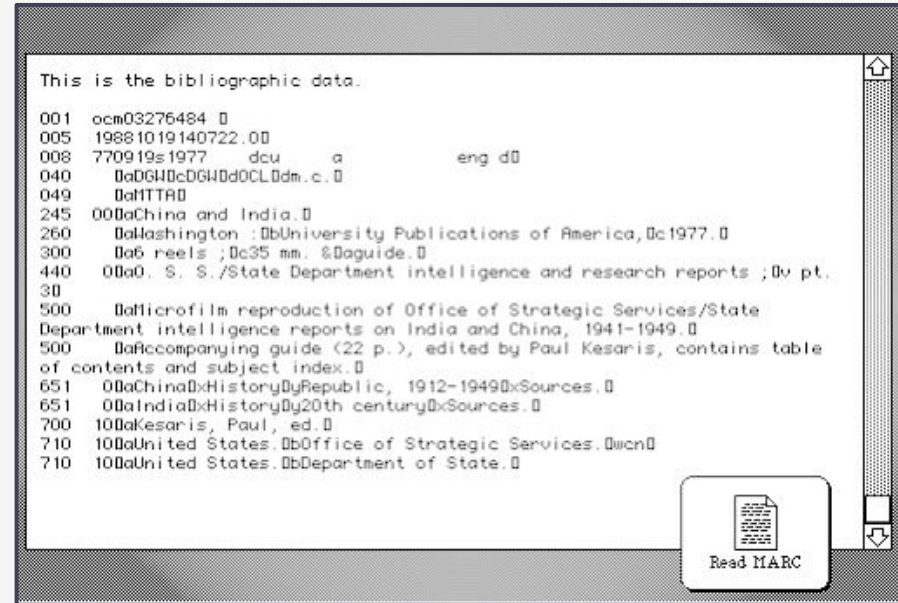
> Metadata standards are **essential** but also often **highly contested**

- They embody **value judgements** (implicitly or explicitly)

e.g. Do all works, for instance, have “creators” or “authors”?

- **MARC** (Machine-Readable Cataloging), cornerstone in library cataloging!

- innovative +50 years ago; what about today?



This is the bibliographic data.

```
001  ocm03276484  0
005  19881019140722.00
008  770919s1977    dcu    a                eng d0
040  0a0GM0c0GM0d0CL0dm.c.0
049  0a1TTAD
245  000aChina and India.0
260  0aWashington :0bUniversity Publications of America,0c1977.0
300  0a6 reels ;0c35 mm. 8.0aguide.0
440  00a0. S. S./State Department intelligence and research reports ;0v pt.
    30
500  0aMicrofilm reproduction of Office of Strategic Services/State
    Department intelligence reports on India and China, 1941-1949.0
500  0aicompanying guide (22 p.), edited by Paul Kesaris, contains table
    of contents and subject index.0
651  00aChina0xHistory0yRepublic, 1912-19490xSources.0
651  00aIndia0xHistory0y20th century0xSources.0
700  100aKesaris, Paul, ed.0
710  100aUnited States.0bOffice of Strategic Services.0wcn0
710  100aUnited States.0bDepartment of State.0
```

Read MARC

metadata standards

> Problems with MARC

- **Technological limitations** – not well-suited to modern data models like those used in contemporary web technologies
- **Complexity and costly** – complex and requires specialized knowledge to use and interpret
- **Limited flexibility** – difficult to accommodate diverse types of resources, such as multimedia, and community-generated content
- **Multilingualism** – MARC struggles with non-Latin scripts
- ...

metadata standards



LJ INFOTECH **DIGITAL** LIBRARIES BY ROY TENNANT

MARC Must Die

WHEN MARC was created, the Beatles were a hot new group and those of us alive at the time wore really embarrassing clothes and hairstyles. Computers were so large, complex, and expensive that it was ludicrous to think that you would one day have one in your home, let alone hold one in the palm of your hand. Although age by itself is not necessarily a sign of technological obsolescence (how much has the wooden pencil

For more on the wired library, see the **netConnect** supplement mailed with the January, April 15, July, and October 15 issues of *LJ*

irregularities of MARC are beginning to create problems for catalogers and users," says David Flanders in the article "Applying XML to the Bibliographic Description." He continues, "MARC

the book should not be buried in a text string in a subfield of the title field. A more egregious example is the ambiguous encoding of respective parts of a personal name (last name, first name, etc.).

Extensibility and language

Migrating our catalogs from printed cards to computers was a massive job, now largely completed. A number of libraries are now enriching those

humanities metadata

humanities data

- **Contributor** – "An entity responsible for making contributions to the resource".
- **Coverage** – "The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant".
- **Creator** – "An entity primarily responsible for making the resource".
- **Date** – "A point or period of time associated with an event in the lifecycle of the resource".
- **Description** – "An account of the resource".
- **Format** – "The file format, physical medium, or dimensions of the resource".
- **Identifier** – "An unambiguous reference to the resource within a given context".
- **Language** – "A language of the resource".
- **Publisher** – "An entity responsible for making the resource available".
- **Relation** – "A related resource".
- **Rights** – "Information about rights held in and over the resource".
- **Source** – "A related resource from which the described resource is derived".
- **Subject** – "The topic of the resource".
- **Title** – "A name given to the resource".
- **Type** – "The nature or genre of the resource".



DublinCore

classification

classification

> thinking through the way what will be described but also how something will be described is intellectual work and

> e.g. an archive of a queer performance

- what fields in a metadata scheme allow that identity to be expressed?
- what types of performance need to be identified?
- what (controlled) vocabulary is used?
- who defines that vocabulary?

classification

> in short: **who speaks in the metadata?**

- no classification system is value neutral, objective, or self-evident
- all classification systems bear within them the ideological imprint of their production

> e.g. the **Library of Congress** has 26 divisions (see [here](#)):

- roman alphabet is used for structure
- **E & F**: history of the Americas
- **D**: history of the rest of the world
- naval science, military have dedicated letters assigned

classification

> in short: **who speaks in the metadata?**

- no classification system is value neutral, objective, or self-evident
- all classification systems bear within them the ideological imprint of their production

> e.g. the **Dewey Decimal System:**

- **100:** Philosophy & Psychology
- ...
- **700:** Arts & Recreation

Datasheets for Datasets

~ a new narrative?

postcards' narrative

postcards' narrative

- > Next week: visit to Princeton Special Collections @ Mudd Library
 - Princeton Historical Postcard Collection
 - When does a pile of postcards become a story?



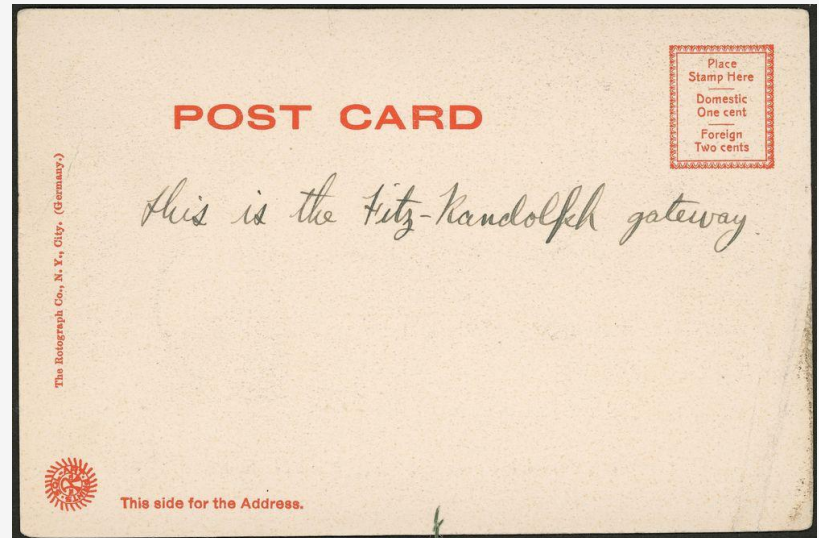
postcards' narrative

- > Next week: visit to Princeton Special Collections @ Mudd Library
 - Princeton Historical Postcard Collection
 - When does a pile of postcards become a story?

?	?	?	?	?
?	?	?	?	?
?	?	?	?	?
?	?	?	?	?
?	?	?	?	?

postcards' narrative

- > Next week: visit to Princeton Special Collections @ Mudd Library
 - Princeton Historical Postcard Collection
 - When does a pile of postcards become a story?



next session

for next session

Pre-Class Exercise

Explore the [Princeton University Historical Postcard Collection](#) and **select a postcard** that you find interesting. Analyze it using **one** of the following guiding questions:

1. If available, find the same postcard on a different online platform. Note differences in color, detail, or cropping. What do these variations suggest about digitizing analog materials?
2. Assess the quality of the digital images. Are there details potentially lost or misrepresented in the digital version?
3. Look at the metadata provided alongside the digital surrogate. How does this information affect your understanding of the postcard?
4. Reflect on your experience navigating the digital collection. How does the digital interface impact your exploration and understanding?
5. Hypothesize about how viewing the postcards digitally might differ from viewing them in person.
6. What do you think the digitization process for this collection was like? What challenges might have been encountered? What choices were made?

Post your reflection in the [#postcards](#) channel on Slack **before 9:00 AM on the day of our class.**

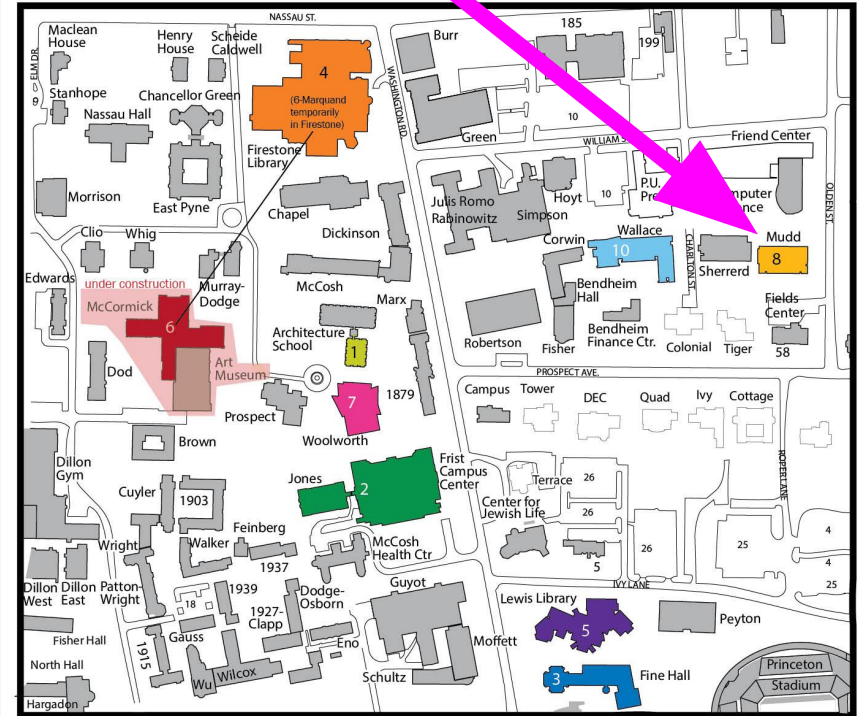
for next session

Pre-Class Reflection

- Listen to [this podcast episode](#), in which Dot Porter (Curator of Digital Research Services at UPenn's Schoenberg Institute for Manuscript Studies) is being interviewed by Stewart Varner (Price Lab at UPenn).
- [Manžuch, Zinaida. "Ethical Issues in Digitization of Cultural Heritage." 2017.](#)
- [Kropf, Evyn. "Will That Surrogate Do? \[...\]" 2016.](#)

Post your reflection in the [#reflections](#) channel on Slack [before 9:00 AM on the day of our class.](#)

next session **MEET AT MUDD**



references

Drucker, Johanna. *The Digital Humanities Coursebook. An Introduction to Digital Methods for Research and Scholarship*. First edition, Routledge/Taylor & Francis, 2021.

Gebru, Timnit, et al. "Datasheets for Datasets." *Communications of the ACM*, vol. 64, no. 12, Dec. 2021, pp. 86–92.

Hoffman, Gretchen. "How Are Cookbooks Classified in Libraries? An Examination of LCSH and LCC." *Proceedings from North American Symposium on Knowledge Organization*, vol. 4, no. 1, 2013, pp. 100–11.

Manovich, Lev. "Database as Symbolic Form." *Convergence: The International Journal of Research into New Media Technologies*, vol. 5, no. 2, June 1999, pp. 80–99.

Pomerantz, Jeffrey. "Introduction." *Metadata*, The MIT Press, 2015, pp. 1–18.

Tennant, Roy. "MARC Must Die." *Library Journal*, vol. 127, no. 17, 2002, p. 26.