

**Persistence**

Mint ARKs as PIDs  
Launch at n2t

**Preservation**

OCR to XML-TEI  
Transform into SKOS



**Performance**

Import vocabulary into HIVE

# Project Pipeline: Preservation, Persistence and Performance

Enabling HIVE for Accessing and Using Historical Vocabularies

<https://cci.drexel.edu/mrc/>



@all\_metadata



DREXEL UNIVERSITY

# Metadata Research Center

*College of Computing & Informatics*

Jane Greenberg (jg3243@drexel.edu)

Christopher Rauch (cr625@drexel.edu)

Mat Kelly (mkelly@drexel.edu)

# Preservation

---

Preservation of historical texts is contributing to an ever increasing body of recorded knowledge.



# Preservation

## Action 1: **Access, Use and Reuse**

- Ensure that data is accessible to both designated users and re-users, on a day-to-day basis.

## Action 2: **Transform**

- Create new data from the original.

# Persistence

- Digital renderings of historical texts have facilitated much broader access to their intellectual content but require context.

THE LEON LEVY  
Dead Sea  
Scrolls  
DIGITAL LIBRARY

[HOME](#)  
› [FEATURED SCROLLS](#)  
[EXPLORE THE ARCHIVE](#)  
[LEARN ABOUT THE SCROLLS](#)  
[ABOUT THE PROJECT](#)

Israel Antiquities  
Authority

Accessibility • Help | עברית • العربية • Deutsch • Русский

Featured  
Scrolls

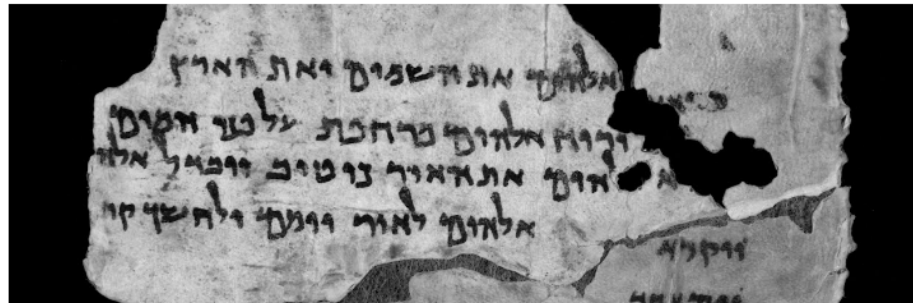
.....  
GENESIS  
.....  
THE TEN  
COMMANDMENTS  
.....  
PSALMS  
.....  
PALEOLEVITICUS  
.....  
TEFILLIN  
.....  
MINOR PROPHETS  
.....  
COMMUNITY RULE  
.....  
ENOCH  
.....  
APOCRYPHON OF DANIEL  
.....  
THE BOOK OF WAR  
.....  
LEGAL PAPYRUS

The discovery of the first Dead Sea Scrolls in a remote Judean Desert cave in 1947 is widely considered the greatest archaeological event of the twentieth century. Bedouin treasure hunters and archaeologists ultimately found the remains of hundreds of ancient scrolls. These fragile pieces of parchment and papyrus, including the oldest existing copies of the Hebrew Bible, were preserved for two thousand years by the hot, dry desert climate and the darkness of the caves where they were placed. The scrolls provide an unprecedented picture of the diverse religious beliefs of ancient Judaism, and of daily life during the turbulent Second Temple period when Jesus lived and preached.

.....

Biblical Scrolls

Fragments of every book of the Hebrew Bible (except the Book of Esther) were found in the Qumran caves, the most famous of the Dead Sea Scrolls sites. Remarkably, some of these ancient copies are identical to the traditional text of the Hebrew Bible that is used today. Other copies preserve differences in the text, which was in the process of standardisation.



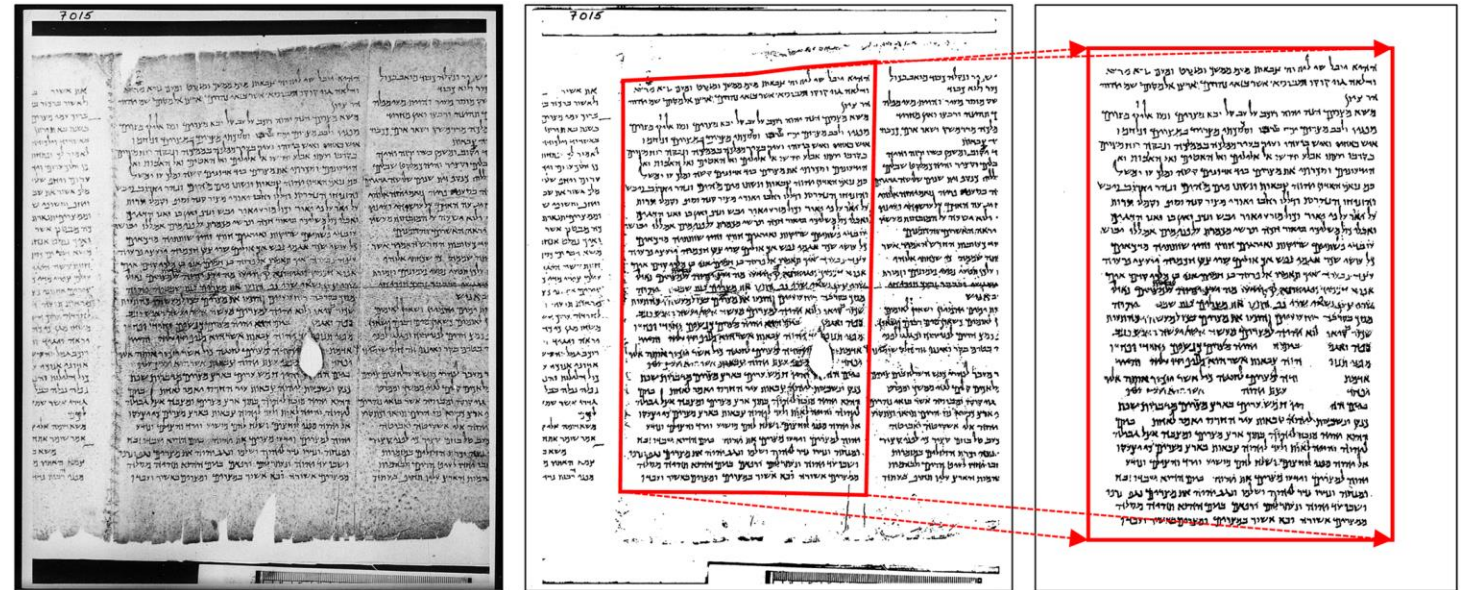
Project Pipeline: Preservation, Persistence and Performance

5



# Performance

- Processing digital representations of historic text into machine readable and actionable formats supports the goal of preservation by engaging a wider community and enabling access.



Popović M, Dhali MA, Schomaker L (2021) Artificial intelligence based writer identification generates new evidence for the unknown scribes of the Dead Sea Scrolls. PLoS ONE 16(4)

# Archival Resource Key (ARK) Identifiers

- Have a decentralized infrastructure
- Are stable and portable
- Allow inflections to facilitate services
- Can be carried in a URI

# Historical Vocabularies

- If digitized historical documents are to be properly understood in their temporal context, it is also necessary to include historical vocabularies in the preservation process that align with the textual resources.



# LCSH 1910

**Subject headings used in the dictionary catalogues ... 1st ed.:v.1 1909/1914.**

**About this Item**

748 page scans  
[Catalog Record](#)  
[Text-Only View](#)

**Rights**  
[Public Domain](#), [Google-digitized](#).

**Download**

**Search in This Text**

☐ Open results in a new window

**Jump to Section**

**Get This Item**

**Collections**

**Share**

**Version**  
2021-02-10 16:52 UTC  
OwnerID: 13510798902640113-5 / Seq: 1  
[About the version](#)

#1

#2

#3

#4

#5 (p.1)

#6 (p.1)

#7 (p.2)

#8 (p.2)

#9 (p.3)

→|←

1 / 748

VIEW

+

-

⏮

⏪

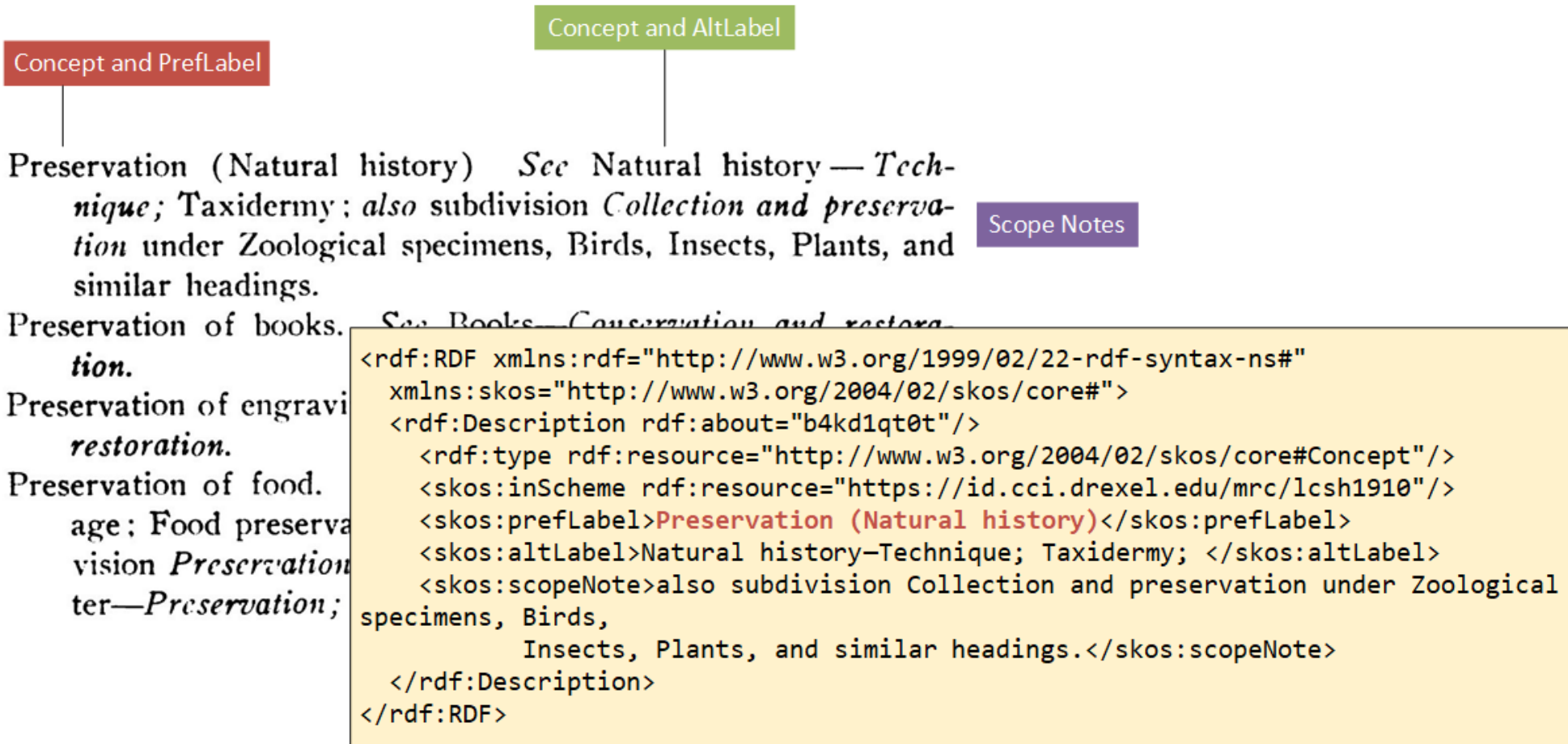
⏩

⏭

🔍

?

# SKOS enabled vocabularies



# Concept Extraction



Helping Interdisciplinary  
Vocabulary Engineering

Vocabularies Search Index

HIVE automatically extracts concepts from a file, or URL, using selected vocabularies.

1 Select vocabularies

- |   |                                       |   |
|---|---------------------------------------|---|
| <input type="checkbox"/> AGROVOC          | <input type="checkbox"/> Asthma       | <input type="checkbox"/> Cardiology                 |
| <input type="checkbox"/> Chambers         | <input type="checkbox"/> ChambersTest | <input type="checkbox"/> Diabetes                   |
| <input type="checkbox"/> FASTgeo          | <input type="checkbox"/> FASTpersonal | <input type="checkbox"/> FASTtopical                |
| <input type="checkbox"/> Gastroenterology | <input type="checkbox"/> LCSH         | <input checked="" type="checkbox"/> <b>LCSH1910</b> |
| <input type="checkbox"/> MeSH             | <input type="checkbox"/> Metals       | <input type="checkbox"/> Oncology                   |
| <input type="checkbox"/> Pediatrics       | <input type="checkbox"/> RADLEX       | <input type="checkbox"/> ROO                        |
| <input type="checkbox"/> Respiratory      | <input type="checkbox"/> SAREF        | <input type="checkbox"/> UAT                        |
| <input type="checkbox"/> USGS             |                                       |   |

2 Enter a URL, or select a file, to index

URL

or

[Browse for file](#)

EthicsandReligion.txt

3 Select indexing filters (optional)

[Index](#)

Cloud View List View

Rank Order Alpha Order

**LCSH1910**

[Intuitionism](#) [God](#)

[Utilitarianism](#) [Moralities](#)

[Morals](#) [Rations](#) [Rationalism](#) [Religion](#) [Religions](#) [Christianity](#) [Devil](#) [Insanity](#)

[Insane](#) [Church](#) [Churches](#) [Ethics, Practical](#)



# Conclusion

Historical documents are being transformed to digital formats to enable analytics. Temporally aligned vocabularies have not been addressed on the same level. We have sought to address this deficit by transforming historical vocabularies from their printed or electronic forms and prepare them for analysis and encoding as linked data.

Thanks to Peter M. Logan, John Kunze, Joan P. Boone, and Sam Grabus  
Supported by NEH Grant #HAA-261228-18