

Web Archive Research and the role of (digital) academic libraries

Thomas Risse
Goethe University Frankfurt
University Library J. C. Senckenberg

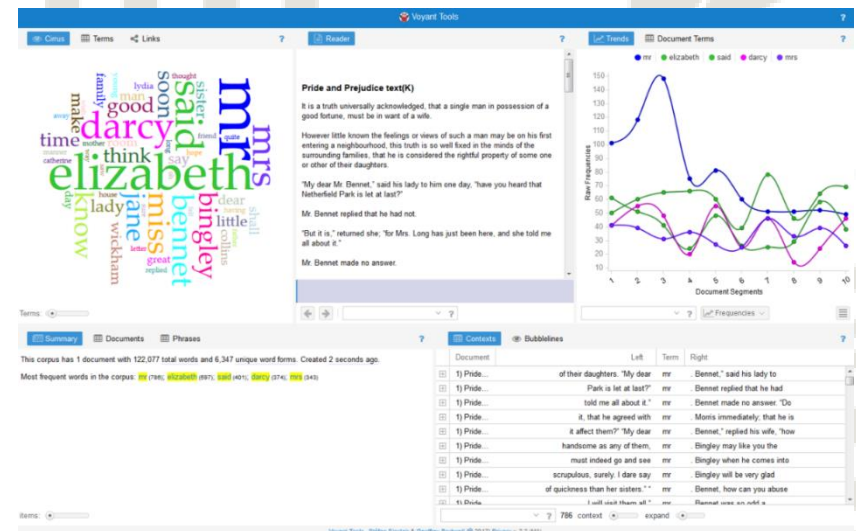
RESAW@Porto2018, 13. September 2018

Different Research Fields

- Digital Humanities
 - Historians (long term)
 - Social Science (more short term)
 - Cultural anthropology
 - ...
- Law
- Computer Science and related areas
 - Information Retrieval
 - Temporal Analytics
 - Natural Language Processing
 - ...

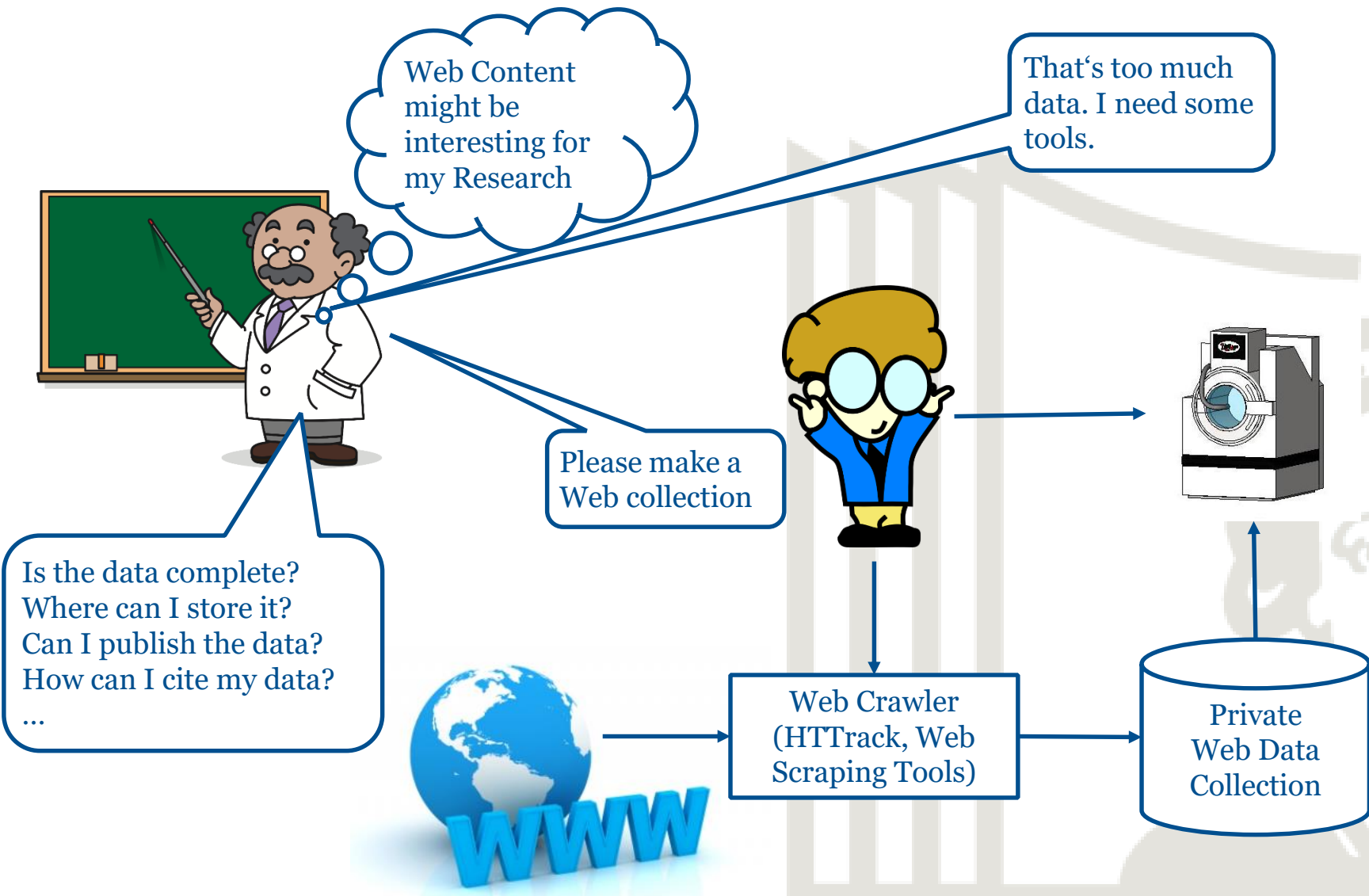
Different Requirements

- Selection strategy, Time frames
- Provenance, Validation
- Persistence
- Analysis, Visualizations
- ...



Thomas Risse, Elena Demidova and Gerhard Gossen. What Do You Want to Collect from the Web?
In: *Proc. of the Building Web Observatories Workshop (BWOW) 2014*. 2014.

A Typical Point of Departure



Sources of Web Content used in Research

Oh, there are
other sources!!!

- Archive.org
 - Broad in scope
 - URL and fulltext access
- ArchiveIT
 - Collections curated by libraries
 - URL and fulltext (inside collections)
- National Libraries
 - Broad in scope, different selection strategies
 - Limited access with some exceptions (e.g. Portugal)
- Organizational Archives, e.g. political parties
 - Focused collections
 - Limited access due to copyright restrictions



Doing Research with Web Archives

"I'm new to Web Archives. Where should I start?"

Typical Questions

- Where to find relevant content?
- What can be done with the content?
- How to handle the amount of data?
- Which tools can be used?
- What are regulatory limitations?
- ...

But there is more to know when working with Web Archives

- WAs differ from traditional libraries and archives
- WAs are never complete or fully coherent
- There are different versions of a page
- WAs are not a full copy of the Web (or Web Site)
- Different WAs have different selection strategies

Scholars need to get a good understanding of the sources, content properties and tools

→ Currently it's their own task



Academic Libraries could Help

- Many researchers are used to work with (local) academic libraries
- Libraries have a long tradition and experience in working with content of various kinds
- Academic libraries are already offering various courses and services e.g. reference management, research data management, digital humanities
- Provide direct support in finding literature
- Libraries might have specialists in various areas
 - Special collection focus since many years
 - Knowledgeable librarians who curate the collections

How does it help with Web Archives?

Guiding the Users

- Basic knowledge on Web Archives is independent of the research field
- Academic Libraries could offer courses on Web Archive usage
 - General Introduction
 - Major tools
- Publish general guidelines
- Faculties provide complementary specific knowledge

In collaboration with the computing center

- Offer tools and services



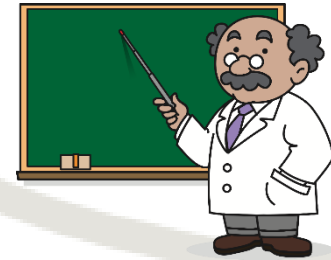
Providing Access to Web Archive Content

From the researchers point of view

- Academic libraries would be a logical source/starting point to access Web archives
- Similar to publications and research data

Some Major Challenges

- Resources
 - Academic libraries have limited (or no) resources for doing Web archiving
 - Staff
 - Computation & Storage Resources
- Content Selection
 - What should be collected?
 - What are the researchers needs?
- Regulation



CHALLENGE

Regulation: Copyright and Privacy

- Regulations differ among the countries and states
- Typically Academic Libraries are not allowed to archive Web content and to provide access
 - There is no regulation that enables this function
 - Germany: Some state libraries which are also academic libraries are allowed
- National libraries are allowed to harvest
 - Harvest national domains with different selection strategies
 - Provide limited access
 - Maybe 70 years after the death of author of a page
But huge amounts of orphan works
 - Maybe if the content is marked as open access
 - Other countries are more open: Portugal, US
 - Access is broad, fulltext search is not satisfying for focused research questions
 - Curated collections only exist for major topics (s. ArchiveIT)
 - Specialists in research areas are missing



Let's dream a bit ...

... of the nearly ideal library and research world

- Academic libraries would be allowed to harvest and archive the Web
- And could provide unlimited access

Web Harvesting

- What should a library harvest?
 - Everything? → Impossible
 - Selected Topics
 - On request? → Would only address research questions on current Web content
 - Topical areas of the Library → Would be a nice complement
- Access to content
 - Search among all collections might not be necessary
 - Fulltext search and browsing within collections should be provided
 - Any advanced search is welcome

Resources

- Libraries have only limited personal and monetary resources
- Most innovative services are financed by 3rd parties
- Sustainability is not guaranteed but necessary for Web archiving

... but it was only a nice dream

... the Regulation will not change
... the Resources will not increase



What do we have?

Academic Libraries

- Could offer courses on Web Archive usage
- Have specialists in certain areas
- No mandate to harvest or to provide Web archive content

National libraries / Archive.org

- Allowed to collect and to provide access
- Have the resources and experience for harvesting and Web archiving
- Technology for retrieval
- No specialists in specific research fields

→ Let's join forces to provide academic related access

Providing a Window to Web Archives

The real legal problem of access: The storage location of the data

Countries with strict regulations

- It is forbidden is to deliver archived content that is physically stored in this country

But

- It is allowed to store URLs to archived content in any country
- It is allowed to show users the archived URLs
- It is allowed is to open a frame in the browser that loads the archived URL

Approach:

- Technology: Micro Archives + Viewer
- Organizational Structure at the library

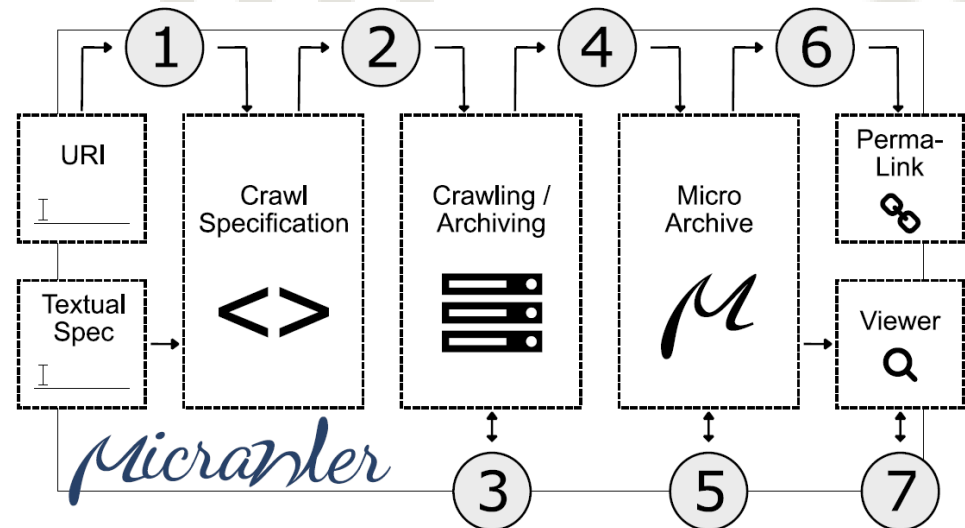
Micro Archives

Micro Archives consists of

- Crawl Specification: URLs (to the Web or a Web Archive)
- Meta-Information
 - Labels and Comments for URLs
 - Semantic descriptions
 - Type

MiCrawler

- Creates an instance of the Micro Archive
- Provides an unique identifier (e.g. DOI, URN)



Helge Holzmann and Mila Runnwerth. 2018. Micro Archives as Rich Digital Object Representations. In Proceedings of the 10th ACM Conference on Web Science (WebSci '18). ACM, New York, NY, USA, 353-357.

Organizational Aspect

e.g. Specialized Information Services

- German Research Infrastructure Activities funded by German Research Association
- Build upon special collections und existing knowledge in libraries
- Provide researchers a comprehensive view on their topic
 - Aggregates library catalogues from all over Germany/International
 - Provide free access to licensed content (collections, databases)
 - Provide community services, e.g. blogs, researcher directories, tool boxes
- For example: University Library Frankfurt
 - Comparative literature, Linguistics, German Studies, African Studies, Jewish Studies, Performing Arts, Biodiversity
- Idea: Semi Curated Web Archive Windows

iliss[africa](http://iliss-africa.org)

avldigital.de
Fachinformationsdienst
Allgemeine und Vergleichende Literaturwissenschaft

BIOfid.de

 www.performing-arts.eu
Fachinformationsdienst für darstellende Kunst

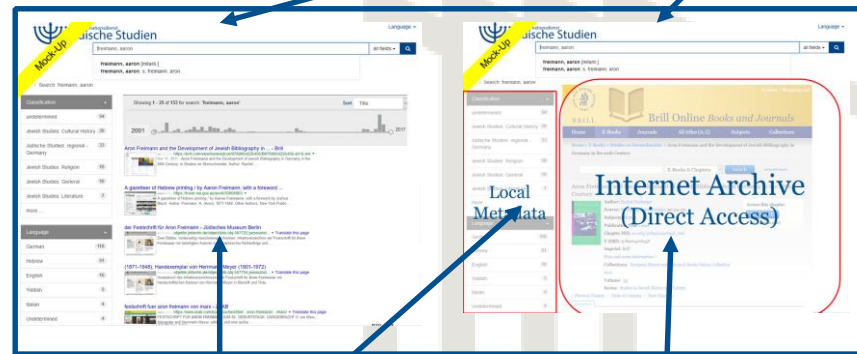
 Fachinformationsdienst
Jüdische Studien

Lin|gu|is|tik
PORTAL FÜR
SPRACHWISSENSCHAFT

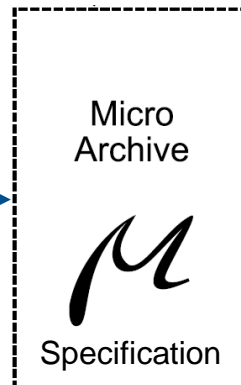
Semi Curated Web Archive Windows



Portal



Curates
Micro Archive



freimann, aaron

all fields ▾



freimann, aaron [mitarb.]

freimann, aaron: s. freimann, aron

🏠 / Search: freimann, aaron

Classification ▾

undetermined 54

Jewish Studies: Cultural History 38

Jüdische Studien: regional - Germany 33

Jewish Studies: Religion 18

Jewish Studies: General 10

Jewish Studies: Literature 7

more ...

Language ▾

German 118

Hebrew 51

English 10

Yiddish 5

Italian 4

Undetermined 4

Showing 1 - 25 of 153 for search: 'freimann, aaron'

Sort

Title ▾

1896



2017

1 [Die Abtheilung der israelitischen Ritualgegenstände im städtischen historischen Museum zu Frankfurt am Main](#)

Freimann, Aron

[Frankfurt a.M.] : [Druck des Frankfurter Intelligenz-Blattes], 1900

book



2 [Die Abtheilung der israelitischen Ritualgegenstände im städtischen historischen Museum zu Frankfurt am Main. \[Druck des Frankfurter Intelligenz-Blattes\]](#)

[Elektronische Ressource]

Freimann, Aron

[Online-Ausg.]

2012

ebook



3 [Annalen der hebräischen Druckerei in Wilhermsdorf](#)

Freimann, Aron

Berlin : Itzkowski, 1903

book



4 [Annalen der hebräischen Druckerei in Wilhermsdorf \[Elektronische Ressource\]](#)

Freimann, Aron

2010

ebook



5 [Annalen der hebräischen Druckerei in Wilhermsdorf](#)

Freimann, Aron





Mock-Up

freimann, aaron

all fields ▾



freimann, aaron [mitarb.]

freimann, aaron: s. freimann, aron

/ Search: freimann, aaron

Classification ▾

undetermined 54

Jewish Studies: Cultural History 38

Jüdische Studien: regional - Germany 33

Jewish Studies: Religion 18

Jewish Studies: General 10

Jewish Studies: Literature 7

more ...

Language ▾

German 118

Hebrew 51

English 10

Yiddish 5

Italian 4

Undetermined 4

Showing 1 - 25 of 153 for search: 'freimann, aaron'

Sort

Title ▾

2001

2017

Aron Freimann and the Development of Jewish Bibliography in ... - Brill



<https://brill.com/view/book/edcoll/9789004226456/B9789004226456-s016.xml> ▾

Nov 11, 2011 - Aron Freimann and the Development of Jewish Bibliography in Germany in the 20th Century. in Studies on Steinschneider. Author: Rachel ...

A gazetteer of Hebrew printing / by Aaron Freimann, with a foreword ...



<https://trove.nla.gov.au/work/20806903> ▾

A gazetteer of Hebrew printing / by Aaron Freimann, with a foreword by Joshua Bloch. Author. Freimann, A. (Aron), 1871-1948. Other Authors. New York Public ...

der Festschrift für Aron Freimann - Jüdisches Museum Berlin



objekte.jmberlin.de/object/jmb-obj-587722.jsessionid... ▾ [Translate this page](#)

Zwei Blätter, vorderseitig maschinengeschrieben. Inhaltsverzeichnis der Festschrift für Aron Freimann mit beteiligten Autoren in alphabetischer Reihenfolge und ...

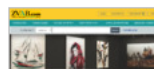
(1871-1948), Handexemplar von Hermann Meyer (1901-1972)



objekte.jmberlin.de/object/jmb-obj-587704.jsessionid... ▾ [Translate this page](#)

Vorabdruck des Inhaltsverzeichnisses der Festschrift für Aron Freimann mit handschriftlichen Notizen von Hermann Meyer in Bleistift und Tinte.

festschrift fuer aron freimann von marx - ZVAB



<https://www.zvab.com/buch-suchen/titel/...aron-freimann/.../marx/> ▾ [Translate this page](#)

FESTSCHRIFT FÜR ARON FREIMANN ZUM 60. GEBURTSTAGE: DARGEBRACHT V. von Marx, Alexander and Hermann Meyer, editors. und eine große ...



Mock-Up

all fields ▾

freimann, aaron [mitarb.]
freimann, aaron: s. freimann, aron

/ Search: freimann, aaron

Classification	
undetermined	54
Jewish Studies: Cultural History	38
Jüdische Studien: regional - Germany	33
Jewish Studies: Religion	18
Jewish Studies: General	10
Jewish Studies: Literature	7
more ...	
Language	
German	118
Hebrew	51
English	10
Yiddish	5
Italian	4
Undetermined	4

Local
Metadata



Brill Online *Books and Journals*

Home | E-Books | Journals | All titles (A-Z) | Subjects | Collections

Home > E-Books > Studies on Steinschneider > Aron Freimann and the Development of Jewish Bibliography in Germany in the 20th Century



Broad Archive (e.g. Internet Archive) Direct Access in an iframe

Author: Rachel Heuser
Source: Studies on Steinschneider, pp 319-337
Subjects: History
Publication Year: 2011
E-ISBN: 9789004226456
Imprint: Brill
Print and series information

Collections: European History and Culture E-Books Online, Collection 2012
Volume: 33
Series: Studies in Jewish History and Culture

« Previous Chapter | Table of Contents | Next Chapter »

Overview

Buy 25,00 €
+ Tax (if applicable)

Access this chapter

Share

Advanced search

Conclusions

Academic Libraries should play an more important role

- Close contact to the users
- Knowledge about selected areas

Experienced in training users in using library services

- Offer training courses on Web archive usage

Resources and legal regulations are hindering

- Harvesting by libraries overall not realistic
- Could provide windows into large scale Web archives
- Curation of content for specific areas are possible
 - Requires more IT support to ease the task
- Requires more collaboration among libraries and Web archives





Dr. Thomas Risse
Head of Electronic Services
Tel. +49 69 798 39 905
Email: t.risse@ub.uni-frankfurt.de

Universitätsbibliothek J. C. Senckenberg
Bockenheimer Landstraße 134 - 138
60325 Frankfurt am Main
<http://www.ub.uni-frankfurt.de>