

Who cares about yesterday's news?

Use cases and requirements for newspaper digitization

*IFLA International News Media Conference
Hamburg, 20-22 April 2016*

Clemens Neudecker
Staatsbibliothek zu Berlin
Europeana Newspapers
[@cneudecker](https://twitter.com/cneudecker)



**Staatsbibliothek
zu Berlin**

Preußischer Kulturbesitz



Europeana
newspapers

Topics

- Current state of newspaper digitization
 - European Newspapers Survey
 - ICON Comparative Analysis
- Exemplary use cases
 - Digital Humanities / Text Mining
 - Creative Industries / Apps
 - Industry / Family History
- Requirements and best practices

Europeana Newspapers Survey

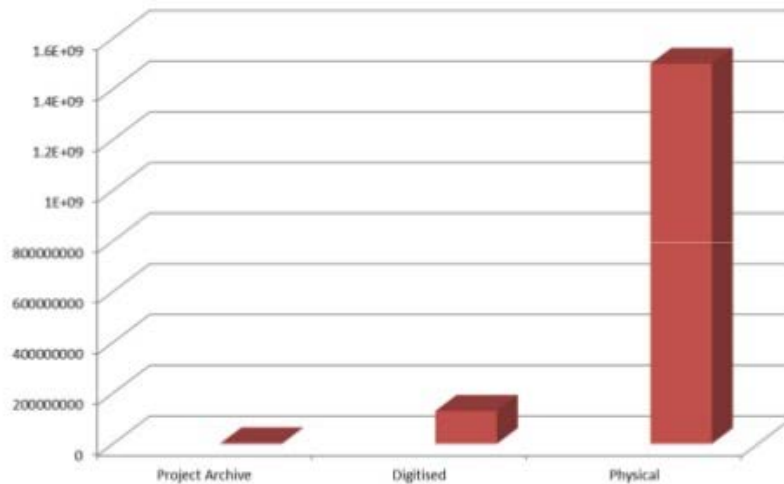
- [Europeana Newspapers survey](#) (2012):
47 respondents from European libraries
- Most EU countries have (national/major) newspaper digitization programmes in place
- Approx. 130,000,000 pages already digitized
- 87% of respondents offer access to their newspaper collection free-of-charge

ICON Comparative Analysis

- ICON [Comparative Analysis](#) (2015)
- (Awareness of) newspaper digitization mostly limited to Western countries (US-UK-EU)
- The vast majority of digital newspapers have been produced from microfilm / cost-efficiency
- Estimated 30,000 titles digitized in US-UK-EU, approximately 45,000 titles worldwide
→ Lack of material other than English

Representation of Absence

Quantities of newspapers – a) in project b) digitised in total c) in physical libraries

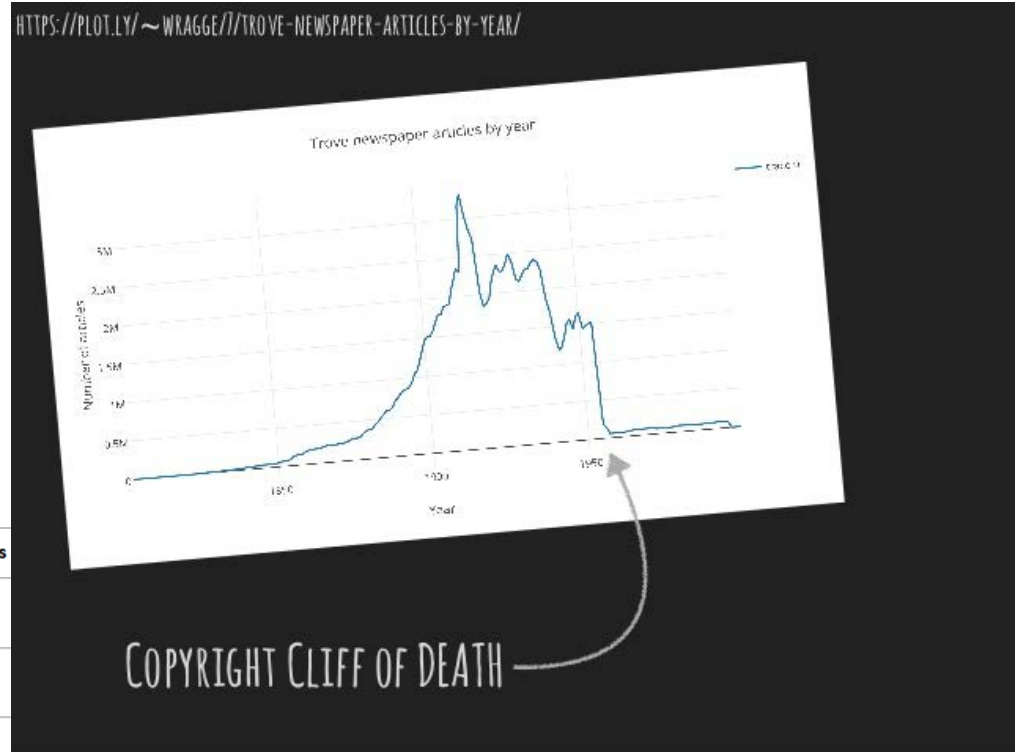
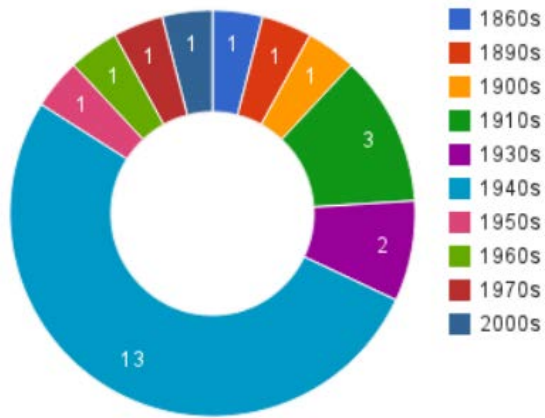


Source: European Newspaper Survey Report
<http://www.europeana-newspapers.eu/wp-content/uploads/2012/04/D4.1-Europeana-newspapers-survey-report.pdf>

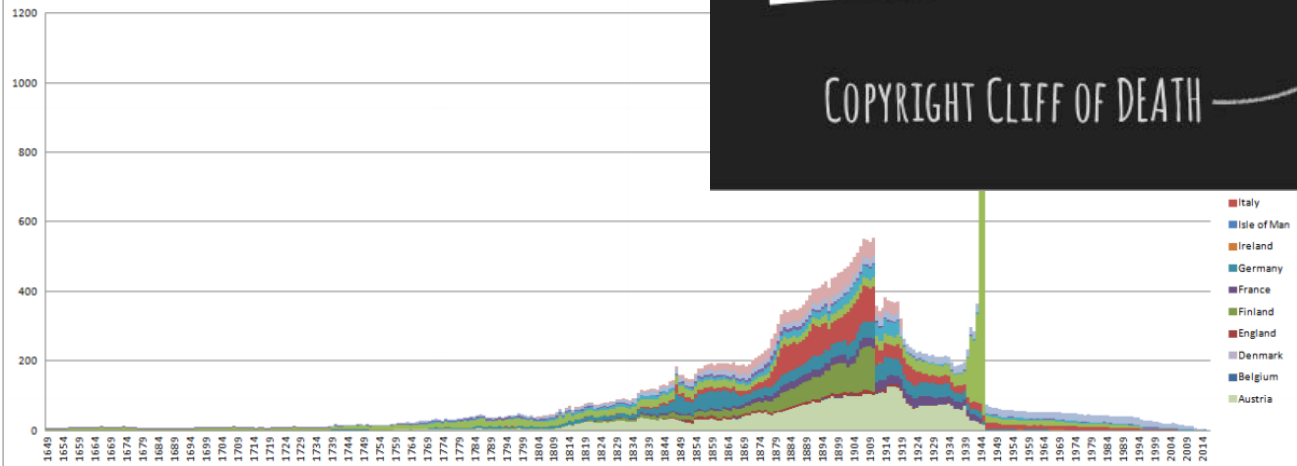
- Scale of what is still left to digitize is mindboggling ...only about 0,001% done in Europe

„Copyright cliff of death“

Cut off date for presentation of digitised newspapers



European Newspapers - Open Access



Use cases

Example use cases: 1

- Digital Humanities / Text & Data Mining
 - Broad interest in societal, cultural developments
 - Newspapers cover „daily life“, events that do not make it into the history textbooks
 - OCR/full-text almost always a requirement
 - For text mining, large quantities of data can be more important than the quality of the OCR
 - Prefer API or bulk download over search & browse
 - See also <http://www.europeana-newspapers.eu/category/interviews-with-researchers/>

viraltexts.org

Network of "Viral Text" Sharing, 1836-1899

This graph illustrates connections among pre-Civil-War newspapers from the Library of Congress' [Chronicling America newspaper archive](#) and Making of America magazine archive. The circles (nodes) in the graph represent individual newspapers, while the lines between them (edges) represent shared texts discovered by researchers working on the [Viral Texts project](#), sponsored by Northeastern University's [NULab for Texts, Maps, and Networks](#). For more information about the project, visit [viraltexts.org](#).

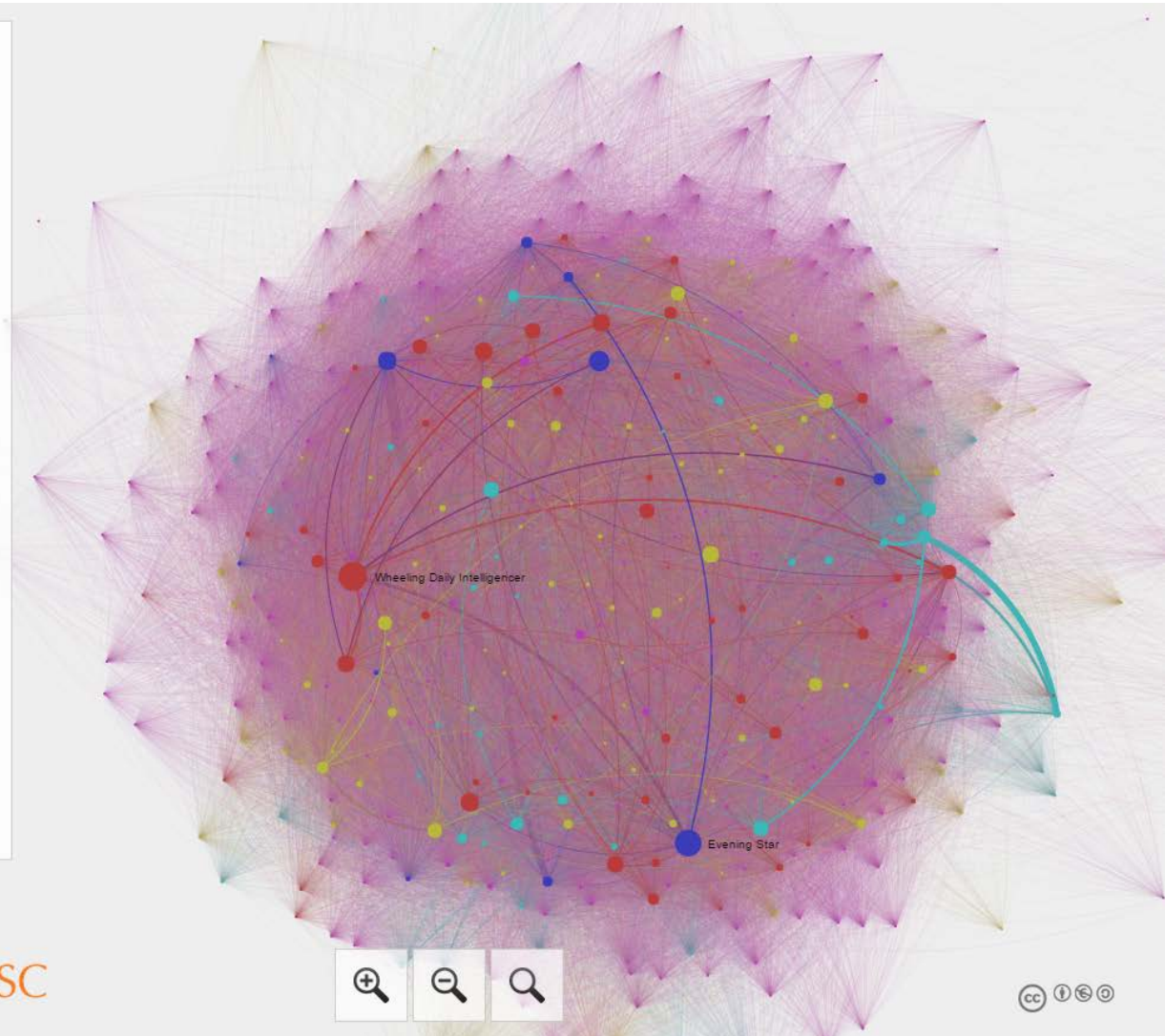
i More about this visualisation

Legend:

- Newspapers or Magazines
- "Viral Texts" shared Between Publications. A thicker line indicates more shared reprints
- Communities of periodicals that frequently share reprinted texts

Search:

Group Selector:



Elegant Elephant

SURE to be a great favorite with the children, this elephant will be a durable toy.

Materials: 5 balls grey Peacock Kooba; 2 small buttons; 1 pr. No. 12 knitting needles; yarn or suitable lining; thread of black wool.


This toy is worked in double-wool and garter-stitch.

TRUNK (Make 2)
Cast on 4 sts.
K 2 times.
3rd Row: K 2 tog., k to last st., k twice into last st.
K 3 rows.
Rep. last 4 rows once.
11th Row: K to last st., k twice into last st.
12th Row: Knit.
Rep. last 2 rows.
15th Row: K twice into last st., k to end.
16th Row: Knit.
17th Row: K twice into last st., k to last 2 sts., k 2 tog.
K 3 rows.
21st Row: K twice into last st., k to end.

ELEGANT ELEPHANT

22nd Row: Knit.
23rd Row: K twice into last st., k to last 2 sts., k 2 tog.
24th Row: Knit.
25th Row: K twice into last st., k to end.
26th Row: Knit.
Rep. last 2 rows 3 times.
Leave on a spare needle.

BODY
Front Leg: Cast on 14 sts.
K 2 rows.
3rd Row: K to last 2 sts., k 2 tog.
2 tog.
4th Row: Knit.
Rep. 3rd row.
K 19 rows.
25th Row: K to last st., k twice into last st.
K 3 rows.
Rep. last 4 rows. Leave on spare needle.
Back Leg: Cast on 12 sts.
K 2 rows.
3rd Row: K to last 2 sts., k 2 tog.
2 tog.
4th Row: Knit.
Rep. 3rd row.
K 21 rows.
27th Row: K, inc. 1 st. each end of row.
28th Row: Knit.
29th Row: K to last st., k twice into last st.
30th Row: Knit.
31st Row: K twice into last st., k to end.
Cast on 10 sts., pick up sts. on spare needle for front leg, k to last st., k twice into last st.
32nd Row: K to last st., k twice into last st.
33rd Row: K twice into last st., k to end.
34th Row: Knit.
Rep. last 2 rows.
37th Row: K to last st., k twice into last st.
38th Row: Knit.
Rep. last 2 rows 3 times.
45th Row: K to end, pick up sts. on spare needle for trunk, k to end.
46th Row: K 2 tog., k to end.
K 5 rows.
50th Row: K 2 tog., k to end.
K 5 rows.
56th Row: K 2 tog., k to end.
Rep. last row.
58th Row: Knit.
59th Row: K to last 2 sts., k 2 tog.
60th Row: Knit.
61st Row: K 2 tog. each end.
62nd Row: Knit.
63rd Row: K to last 2 sts., k 2 tog.
Rep. last 2 rows 3 times.
70th Row: Knit.
71st Row: K 2 tog., k to end.
Rep. last row twice.
74th Row: K 2 tog., k to end.
75th Row: Cast off 8 sts., k to end.
Rep. last 2 rows twice.
80th Row: K 2 tog. each end.
Rep. last row twice.




Example use cases: 2

- Creative industries / Apps
 - Unfamiliar but intriguing uses
 - Potential to reach out to novel audiences
 - Not necessarily commercial interest
 - Almost exclusively require API
 - Serendipity effect
 - Tracing the use:
Trove:
<http://trovespace.webfactional.com/traces/>
NDNP: <http://www.loc.gov/ndnp/extras/#reuse>

hierwashetnieuws.nl

KB Koninklijke Bibliotheek
Nationale bibliotheek van Nederland



Hier was het nieuws

Wat is hier ooit gebeurd? Op deze vraag geeft de app Hier was het nieuws je een antwoord. De app toont – overal in Nederland – historische krantenartikelen over de locatie waar jij je bevindt. Wil je weten wat er vroeger bij jou in de buurt gebeurde? Download dan de gratis app voor het nieuws van toen. Laat je verrassen door historische gebeurtenissen in kranten van vroeger. Vanaf begin december is de app ook beschikbaar via Google Play (Android).



Over de app

De app Hier was het nieuws van de Koninklijke Bibliotheek is gebaseerd op de website Historische Kranten. Op deze website zijn miljoenen gedigitaliseerde krantenpagina's uit de periode 1618-1995 voor iedereen gratis toegankelijk. De app toont de krantenartikelen met gebruik van GPS.



Wat kun je met de app?



Bekijk de promo en kom er in 30 seconden achter wat je met de app kunt doen!

Example use cases: 3

- Commercial / Family History
 - Aim to identify individuals within articles, obituaries
 - Benefit greatly from Named Entity Recognition
 - Huge volunteer base for crowd-sourcing

familysearch.com

The screenshot displays the FamilySearch website interface. At the top left is the FamilySearch logo. To the right are navigation links: Family Tree, Memories, Search, and Indexing. Below the logo is a navigation bar with icons and labels for TREE, PERSON, FIND, LISTS, and WHAT'S NEW. A secondary bar shows 'VIEW: TRADITIONAL' and navigation controls (directional arrows, zoom in/out, and print). The main area shows a family tree structure with several individual profiles highlighted in white boxes. Each profile includes a name, birth-death dates, and a FamilySearch ID.

FamilySearch Family Tree Memories Search Indexing

TREE PERSON FIND LISTS WHAT'S NEW

VIEW: TRADITIONAL

Frederick Walton Seaver
1876-1942 • KL8K-54D

Alma Bessie Richmond
1882-1962 • LH5L-ZG1

Frederick Walton Seaver
1911-1983 • LH5L-4P9

Betty Virginia Carringer
1919-2002 • LH5L-HC6

Lyle Lawrence Carringer
1891-1976 • LH5P-RMT

Emily Kemp Auble
1899-1977 • LH5P-R9F

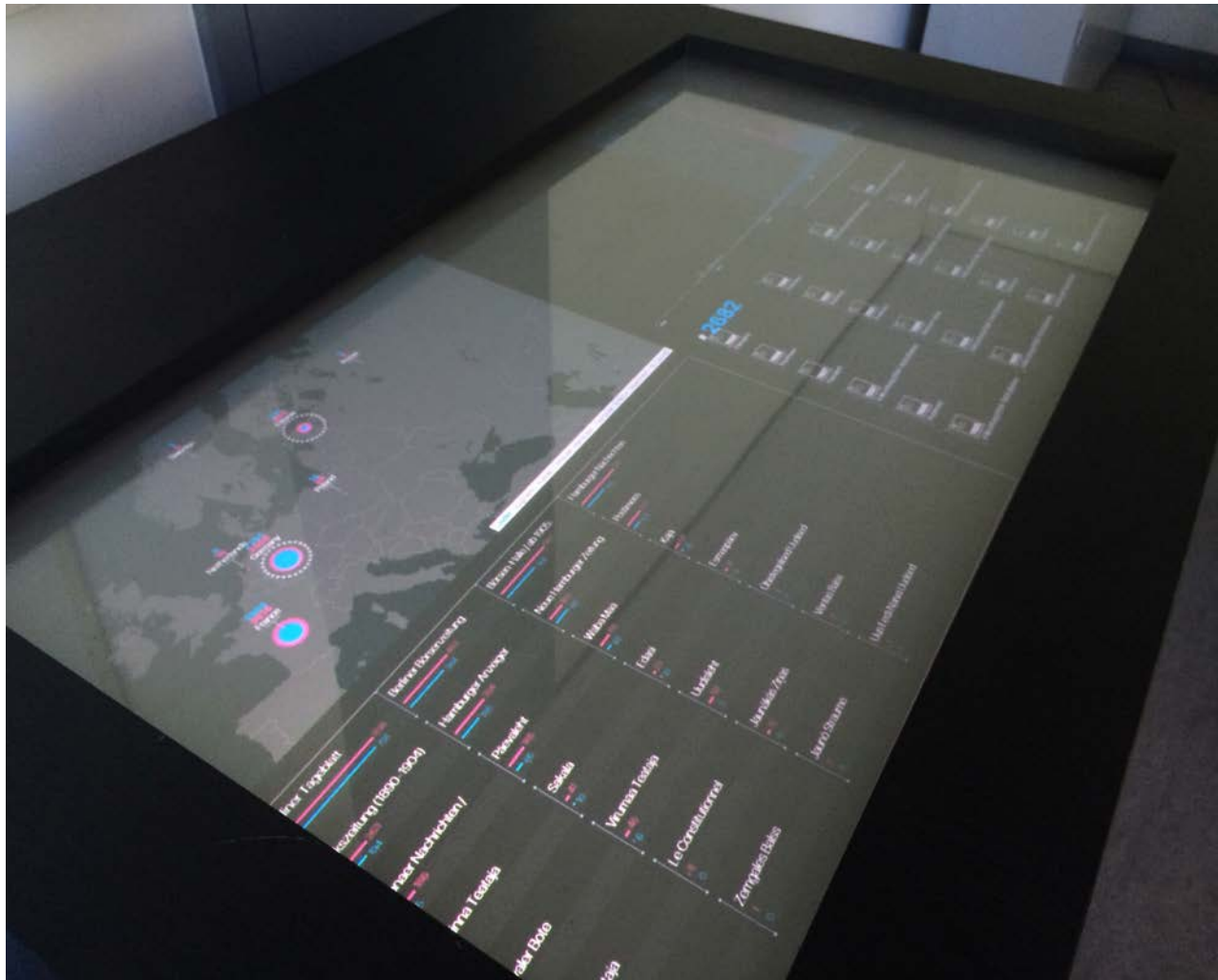
Randall Jeffrey Seaver
1943-Living • LH5N-K6D

Linda Joan Leland
1942-Living • LFDW-T5L

Severt Oliver Leland
1878-1940 • KCRW-F67

Amelia Anna Brocke
1884-1975 • KCNL-S3T

Interactive Newspaper Desk



Summary: Requirements

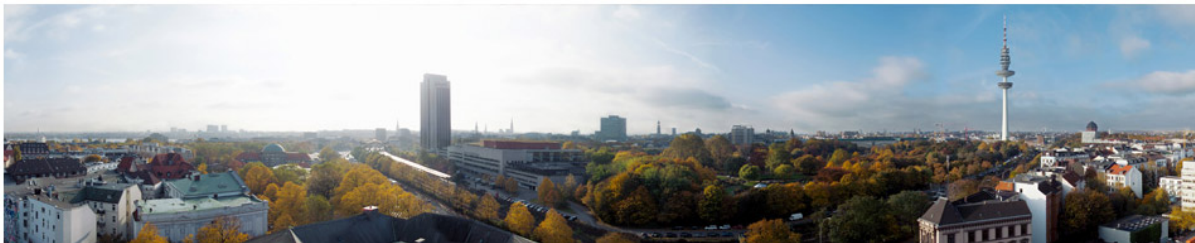
- Interest in digital newspapers is as diverse as the newspaper content
- OCR is nearly always a must-have
- NER can enhance some use cases greatly
- Access should be as open as possible
- APIs provide a means for third parties to create additional outreach and exposure

Summary: Best Practices

- Make available a critical mass through cost-efficient microfilm digitization
- Always provide OCR and be transparent about the quality
- Open access to the content is not a threat but can help create unforeseeable exposure and added value through creative reuse
- Work with the public!

„The coolest thing to do with
your data will be thought of
by someone else“

Jo Walsh & Rufus Pollock:
The Many Minds Principle



Thank you for your attention!

Questions?

Clemens Neudecker
Staatsbibliothek zu Berlin
Europeana Newspapers
[@cneudecker](https://twitter.com/cneudecker)



**Staatsbibliothek
zu Berlin**

Preußischer Kulturbesitz



Europeana
newspapers