

Digitized and Digital Primary Sources

Challenges and Opportunities

Martin Schaller

proposed plan

- ▶ brief talk on technical background
- ▶ discussion
- ▶ try resources by yourself
- ▶ feedback



Technical Background

▶ Scanning



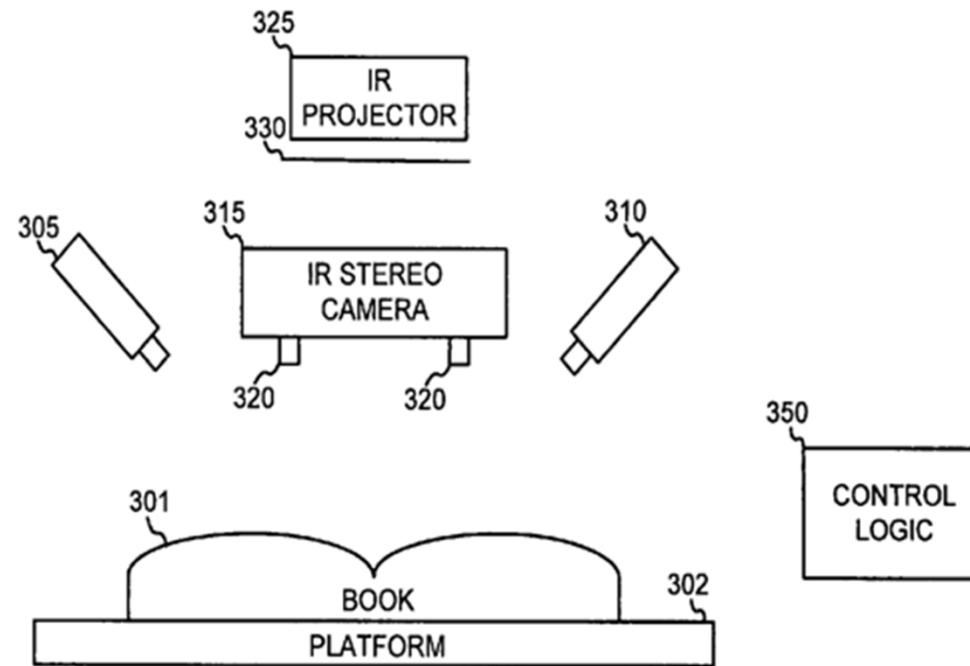
Scanners at work at the Internet Archive
Source: <http://www.nytimes.com/2007/10/22/technology/22library.html?>



Scan robots
<http://www.alfasoft.ro/>

Technical Background

▶ Scanning

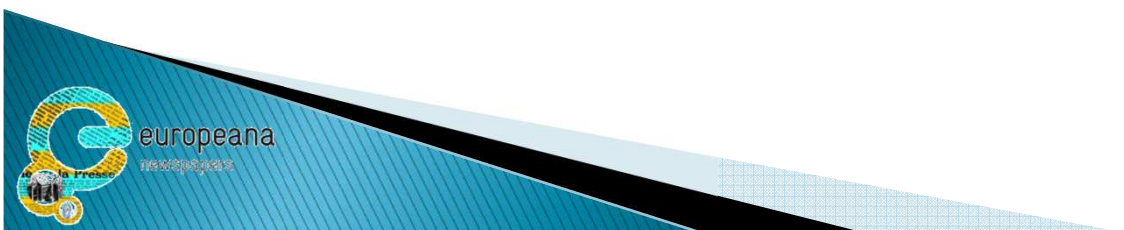


Google's Patent for scanning books

Source: <http://www.cnet.com/news/patent-reveals-googles-book-scanning-advantage/>

Technical Background

- ▶ Optical Character Recognition (OCR)
- ▶ Optical Layout Recognition (OLR)
- ▶ Named Entity Recognition (NER)



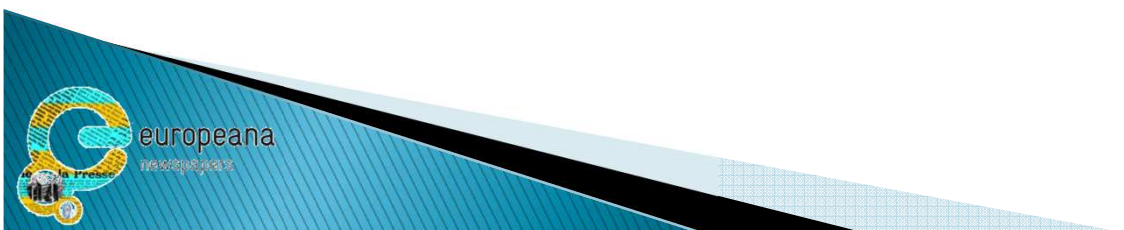
Digitized Sources – Some limitations

- ▶ decentralized initiatives (national, regional, focused on a topic, or even on one title)
- ▶ free of charge vs fees
- ▶ different portals, no federated search



Digitized Sources – Some limitations

- ▶ Europeana
- ▶ research.europeana.eu
- ▶ DPLA
- ▶ Google Books, Internet Archive, Project Gutenberg



Europeana Newspapers Project



project partners



associated partners



networking partners



Europeana Newspapers Project

- ▶ 11 million pages full-text searchable
+ meta data of 8 million pages

ANNO
*Historische Zeitungen
und Zeitschriften*

13 million pages



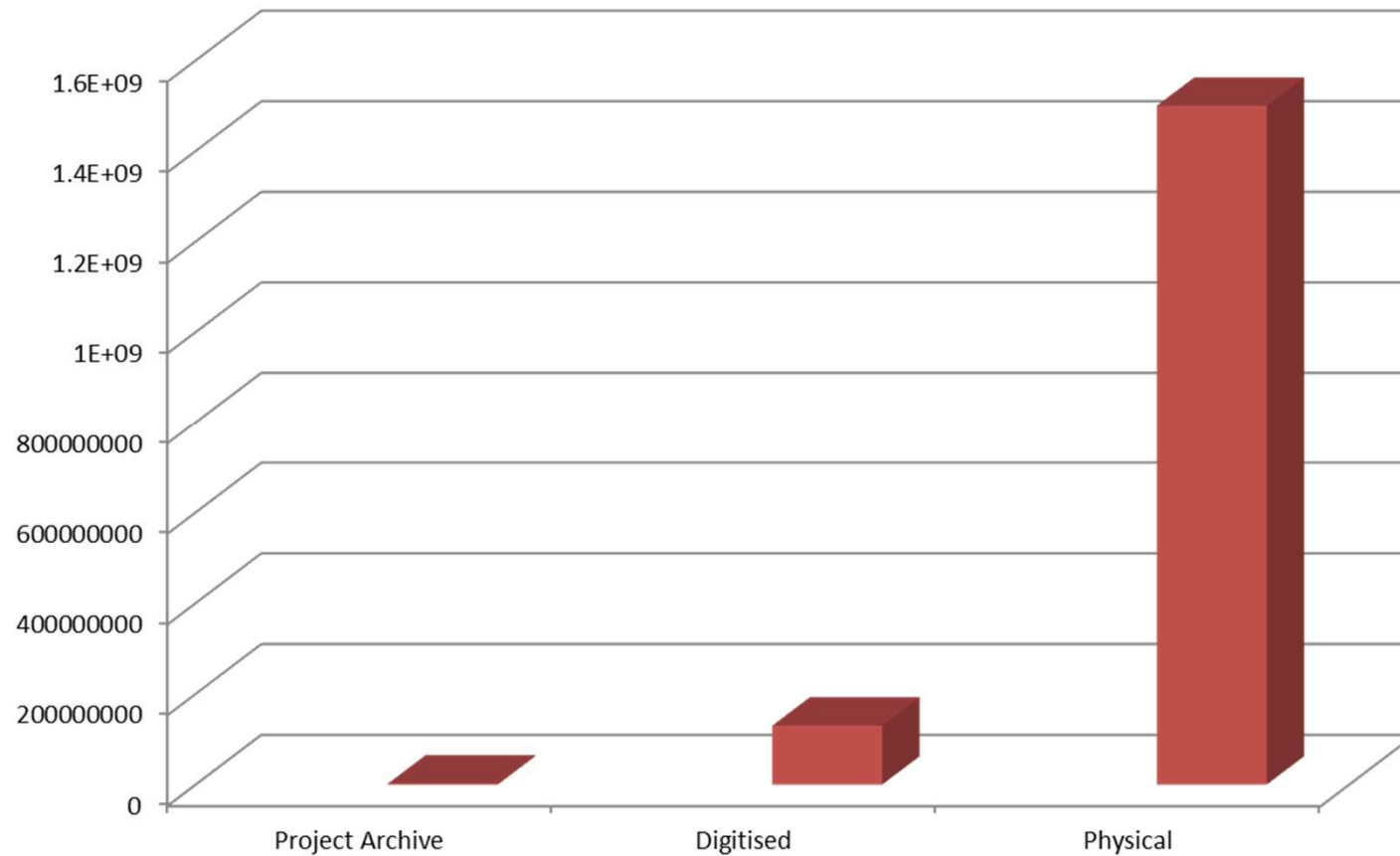
9 million pages

CHRONICLING AMERICA
Historic American Newspapers

8 million pages



Digitized Sources – Some limitations

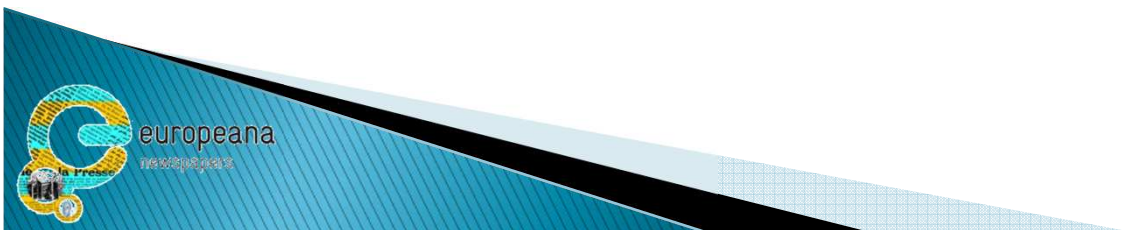


European Newspapers Survey Report

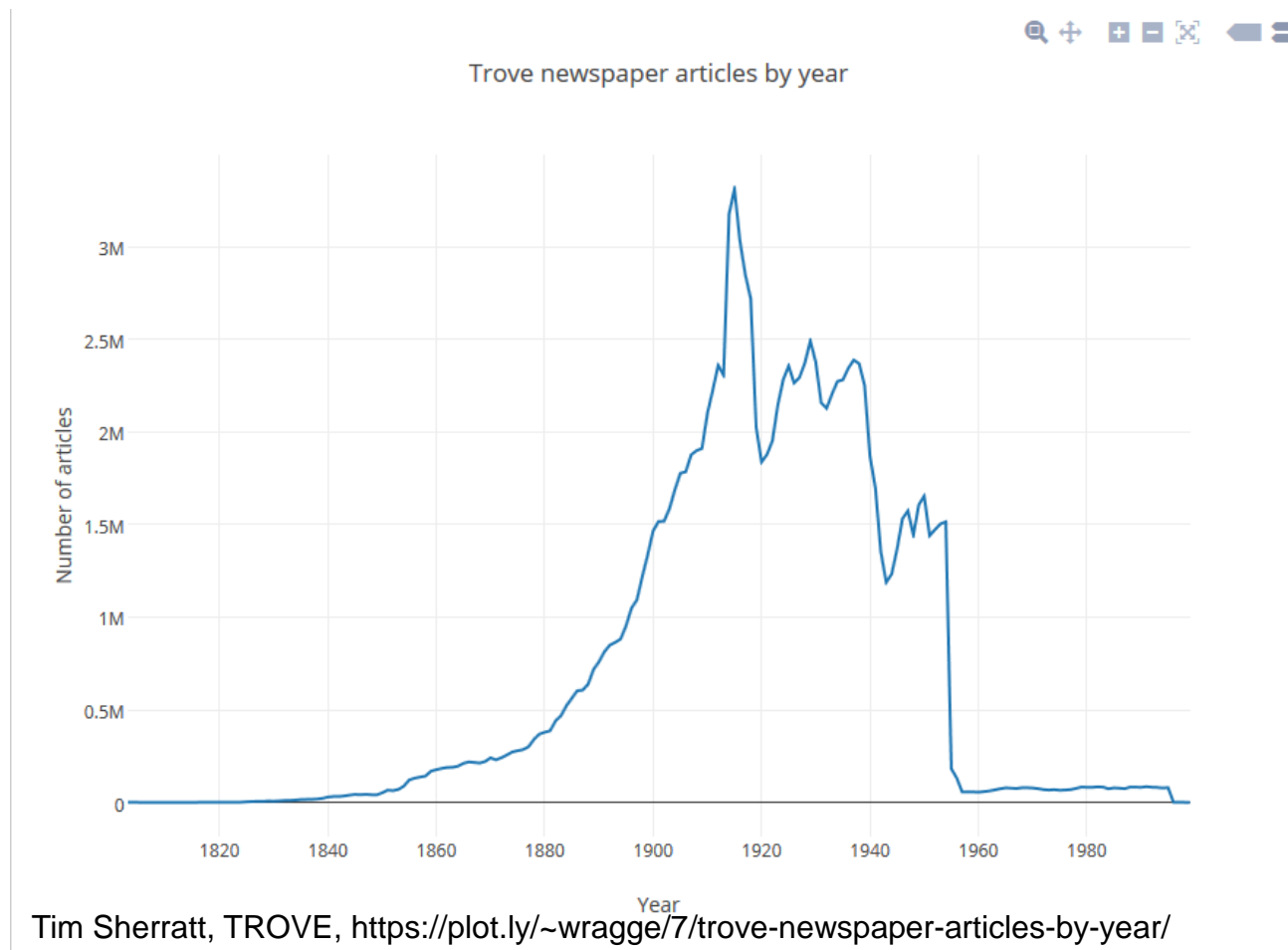
<http://www.europeana-newspapers.eu/wp-content/uploads/2012/04/D4.1-Europeana-newspapers-survey-report.pdf>

Digitized Sources – Some limitations

- ▶ „Copyright cliff of death“
- ▶ different national laws

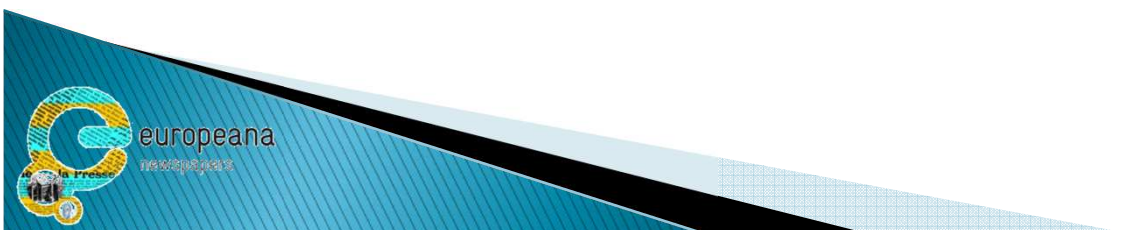


Digitized Sources – Some limitations



Digitized Sources – Some limitations

- ▶ OCR quality
- ▶ Fraktur vs Antiqua



Digitized Sources



Wiener Zeitung

← 25 June 1809 →

← 2 of 4 pages →

[See Original at Library...](#)

CONTENT

sie in Marsch zu setze», »nd alle Truppe
an die Gränzn rücken zn laffen. Die gan
Monarchie wir unter dei. Waffe.l. 'Seit
langer Zeit führte Oesterreich zu Konstar
nopol g^gen Frankreich Krieg; es bewirkt
die Wiederannäherung der Turkey und Eng
lands-; es erklärte sich endlich ganz fte?
offen. ,
Indessen beobachtete das Wiener - Kabi
noch das tiefste Stillfchweigen. Seit Sen
mag von Fontainebleau hatte es weder e
Beschwerde noch ei» Gesuch angebrach
Ew. Majestät hatten sich zu beschweren
die Ermordung Ihrer Couriere in Croatien;
über Französischen Officieren zu Trief z
fügte Beleidigungen; über an einigen Ihre
Italienischen Unterthanen verübte Gewalt
thätigkeiten. Sie erwarteten geduldig die
Ausgleichung dieser Beeinträchtigungen,
am s. März **) der Herr Gesandte von Oe
sterreich mir die Anzeige machte, der Ka
sein Herr, habe Befehl gaegeben, seine Tr
pen auf den Kriegsfuß zu fetzen. Er gab
Ursache dieser Maßregel an, den aus Ä
dolid an die Fürsten des Rheinbundes ge
teten Erlaß, einige Artikel in den Zeitunge
und endlich die Zurückkunft Ew. Majestät
nach Paris. Ew. Majestät befehlen mir
durch eine Note antworten, worin ich mi

sie in Marsch zu setze», »nd alle Truppe
an die Gränzn rücken zu lassen. Die ganze
Monarchie war unter den Waffen. Seit
langer Zeit führte Oesterreich zu Konstanti-
nopol gegen Frankreich Krieg; es bewirkte
die Wiederannäherung der Turkey und Eng-
lands-; es erklärte sich endlich ganz frey und
offen.
Indessen beobachtete das Wiener - Kabinet
noch das tiefste Stillfchweigen. Seit dem Ver-
trag von Fontainebleau hatte es weder eine
Beschwerde noch ein Gesuch angebracht *).
Ew. Majestät hatten sich zu beschweren über
die Ermordung Ihrer Couriere in Croatien;
über Französischen Officieren zu Triest zuge-
fügte Beleidigungen; über an einigen Ihrer
Italienischen Unterthanen verübte Gewalt-
thätigkeiten. Sie erwarteten geduldig die
Ausgleichung dieser Beeinträchtigungen, als
am 2. März **) der Herr Gesandte von Oe-
sterreich mir die Anzeige machte, der Kaiser,
sein Herr, habe Befehl gegeben, seine Trup-
pen auf den Kriegsfuß zu setzen. Er gab als
Ursache dieser Maßregel an, den aus Walla-
dolid an die Fürsten des Rheinbundes gerichteten
Erlaß, einige Artikel in den Zeitungen,

sie in Marsch zu setze», »nd alle
Truppe»
an die Gränzn rücken zn laffen. Die gan
Monarchie wir unter dei. Waffe.l. 'Seit
langer Zeit führte Oesterreich zu
Konstanti-
nopol g^gen Frankreich Krieg; es
bewirkte
die Wiederannäherung der Turkey
und Eng
lands-; es erklärte sich endlich ganz
fte? und
offen. ,

Digitized Sources

KB

Koninklijke Bibliotheek

[See Original at Library...](#)

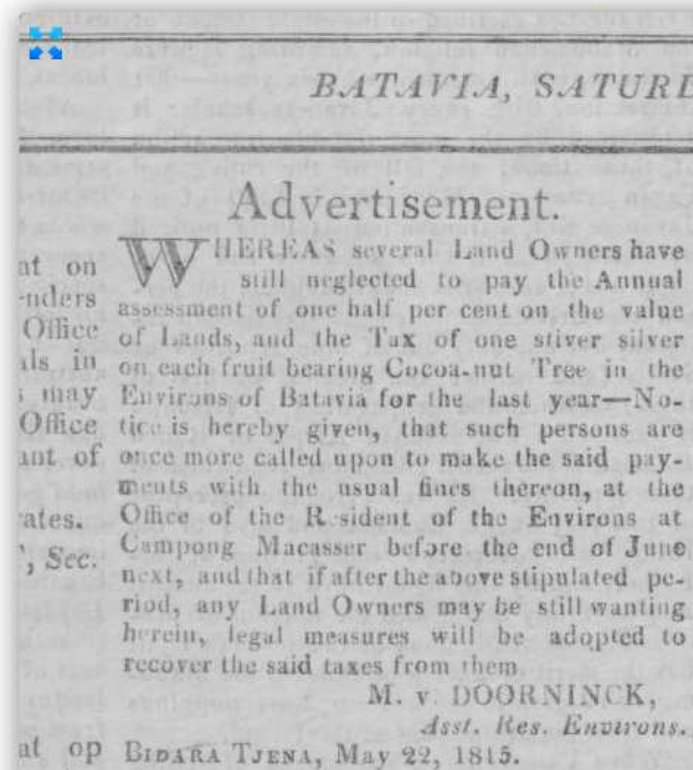
CONTENT

Advertisement.
WHEREAS several Land Owners have
suil neglected to pay the Annual
assessment of one half per cent on the v
of Lands, and the Tax of one sliver silver
on each fruit bearing Cocoa-nut Tree in tl
Environs of Batavia for (he last year —No
ti< ■ is hereby given, that such persons
at ee more called upon to make the said
ir uts with the usual tims thereon, at the
Office of the R sident of thj Environs at
Campong Macasser before the end of Jit
next, and that if after the aDove stipulate
riod, any Land Owners may be still want
h rein, legal measures will be adopted to
recover the said taxes from them
M. v DOORNINCK,
Asst. lies. Environs.
Bidara Tjena, May 1815.
Advertentie.
N ADEM AAL vershelde Land Eigenare
als nog nalatig zvu gebleven in het ~
betalen van het een h• If per cent op de c
taxeerde waarde der Landeryen, en de b
ting van een stuiver silver op de vrugt dia
gende Klapper Boonin in de Ommelande

Java government gazette

← 02 June 1815 →

1 of 8 pages → →|



Advertisement.
WHEREAS several Land Owners
have
suil neglected to pay the Annual
assessment of one half per cent
on the value
of Lands, and the Tax of one
sliver silver
on each fruit bearing Cocoa-nut
Tree in the
Environs of Batavia for (he last
year —No-
ti< ■ is hereby given, that such
persons ara
at ee more called upon to make
the said pay
ir uts with the usual tims

Digitized Sources

[See Original at Library...](#)

CONTENT

Xo^omsio vxecmo.
Zars, e Kor^o intern«, konksmiöa, Loe». ßns^oüo, v;elo, Losino, Largo l.ni??o, Lern«, Kikigno, Mslpsß», e 8. Lsssisn. (Vons, ^ÄNon il Ävni»> Lrevilsvizui, ?on- tsclür», Vvrvlie, ?leisbv>t>urfß, kogli^os, Xsravs, vrscevÄ«, L,rue, e ketor^osae. Kssan/ e, Ksänvin, e Lliuds. Äsi-sveeckl» coi suoi 8cogli,^snil ?ilppo o Oiscoi», -kvi>ette, liervme, o ?s- costisne. Uelaäil, kiergu^olie, 8csräs, ^spvntello, Isto, v l'reinucls eoi rispettivi 8cogli. Vglisn, l^oul^ooran ponentsle, ksvsz», 8e- »trugn, c 8virins!?. (ltre, küellsvs, ». Luteniis, l^oucorsn si- Google, Lsle, e Lu^oii^o.xa. 86reig?, La8w, DvdvopvAlisnz, Nsvisne, Uorlisnv, Dasmsr», L^ocon, VerAsä»,Lso Ponentsle, e<1 i 8eog1i säiacenti. 8sls, ^akäsrg», ^Sglavs, Isnle inoornsts, 8man, l»ues, Ksvs grsnile, e Ksvs pio- ool», 8anro, e<1 i 8coßli »üjacenti. Krilkignü, vrsrove, L»xsvs, 8e1ine. ?on-

ACQUISTO VECCHIO.

1	Zara.	Zara, e Borgo interno, Pontamica, Boca gnazzo, Dielo, Cosino, Borgo Erizzo Cerno, Bibigne, Malpaga, e s. Cassian
2	Nona.	Nona, Zaton di Nona, Brevilacqua, Pontadura, Verchè, Fleisemburg, Poglizza Zarava, Dračevaz, Crue, e Peterzane.
3	Rasanze.	Rasanze, Radovin, e Gliuba.
4	Zaravecchia.	Zaravecchia coi suoi Scogli, santi Filippo e Giacomo, Torrette, Kercine, e Pa costiane.
5	Melada.	Melada, Berguglie, Scarda, Zapontello, Isto, e Premuda coi rispettivi Scogli.
6	Uglian.	Uglian, Lucoran ponentale, Revagn, Se- strugn, e Svirinaz.
7	Cale.	Oltre, Pogliana, s. Eufemia, Lucoran si- rocale, Cale, e Cuçlizza.
8	Pasman.	Sdrelaz, Bagno, Dobropogliana, Neviane, Morliane, Pasman, Tcon, Vergada, Eso ponentale, ed i Scogli adjacenti.
9	Sale.	Sale, Labdara, Zaglava, Isole incoronate, Sman, Luca, Rava grande, e Rava pic- cola, Sauro, ed i Scogli adjacenti.
10	Bribigno.	Bribigno, Dragove, Boxava, Seline. Pon- tebianche, Verona, Poglie, e Scogli ad- jacenti, Eso sirocale.

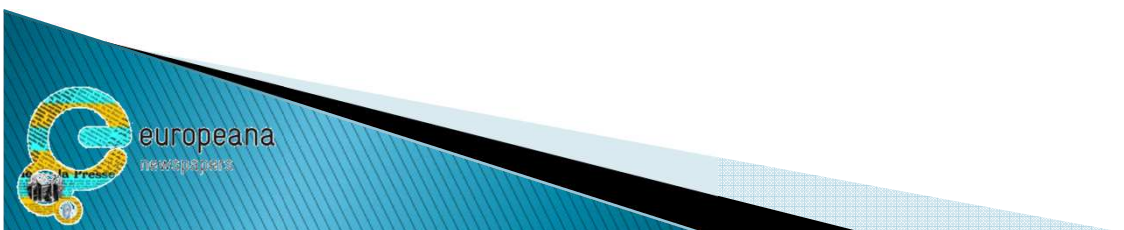
LOTTO No 7

1

Digital Sources

- ▶ Internet Archive
2012: 10 petabyte = 10.000 terabyte

- ▶ British Library's 2014 domain crawl:
july 2014
1,3 gigabyte viruses



Digital Sources

INTERNET ARCHIVE
Wayback Machine

http://www.univie.ac.at/ Go

1,908 captures
21 Dez 98 - 18 Nov 14

NOV DEZ FEB
1995 21 1996 1998

UNIVERSITÄT WIEN

WILLKOMMEN IM INFO-SERVER DER UNIVERSITÄT WIEN! KLICKEN SIE LINKS OBEN FÜR EINE ÜBERSICHT ÜBER DIE ORGANISATIONSTRUKTUR UND EINRICHTUNGEN DER UNIVERSITÄT

FORSCHUNG **LEHRE**

 [Institute und sonstige Einrichtungen](#)

 [Personal](#)

 [Vorlesungsverzeichnis](#)

 [DOnKey](#) Wissenschaftsdokumentation der Universität Wien

 [das EDV-Zentrum](#)

 lokale und internationale [Bibliotheken](#)

 [Unet](#) - [internet for students](#)

 die [HochschülerInnenschaft](#)

 Suchsysteme und Verzeichnisse von WWW-Servern:
[Weltweite Suche](#) ; Suche in [Österreich](#) sowie in [Usenet-Archiven](#)

 **DIE** UNIVERSITÄT
THE UNIVERSITY • LA UNIVERSIDAD
die Zeitung der Universität Wien

 **SCHRÖDINGER**
Thema: Kinder

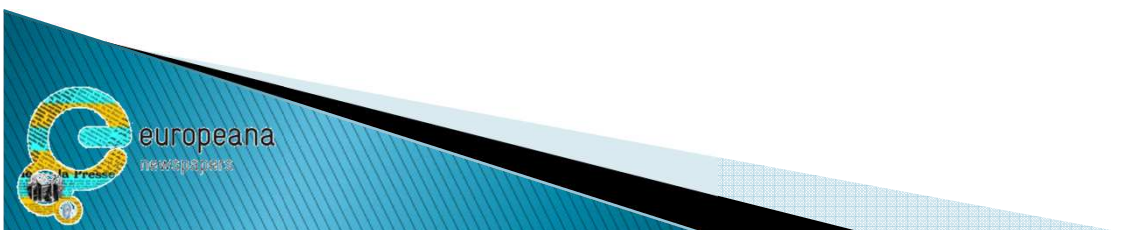
 **UND**
News Digest

 **EVENTOS**
Veranstaltungskalender



Research Infrastructures

- ▶ DARIAH (Digital Research Infrastructure for the Arts and Humanities)
- ▶ CLARIN (Common Language Resources and Technology Infrastructure)



Research Infrastructures

▶ Mission Statement DARIAH

DARIAH...aims to enhance and support digitally-enabled research and teaching across the humanities and arts...By working with communities of practice, DARIAH-EU will bring together individual state-of-the-art digital Arts and Humanities activities across Europe.

It will preserve, provide access to and disseminate research that stems from these collaborations and ensure that best practices, methodological and technical standards are followed.

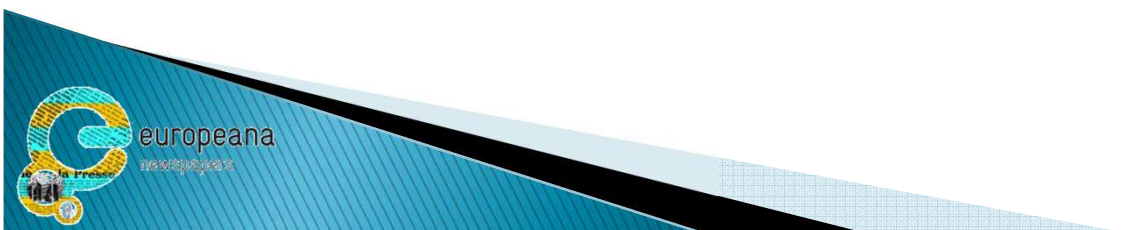
The DARIAH-EU infrastructure will be a connected network of tools, information, people and methodologies for investigating, exploring and supporting research across the broad spectrum of the Digital Humanities.

Source: <https://dariah.eu/about/mission.html>



Digitized and Digital Sources

- ▶ Emmanuel Le Roy Ladurie (1968):
„The historian of tomorrow will be a programmer or he will be nothing.“
quoted in L. Stone, ‚The Revival of Narrative‘, in *Past and Present* (1979), pp.13.



Digitized and Digital Sources

- ▶ <http://www.theeuropeanlibrary.org/tel4/newspapers> (European newspaper corpus)
- ▶ <http://anno.onb.ac.at/> (example of national initiative Austria)
- ▶ <http://gallica.bnf.fr/?lang=DE> (example of national initiative France)
- ▶ <http://trove.nla.gov.au/> (Australia „benchmark“)
- ▶ <http://www.eluxemburgensia.lu/> (example for completeness)
- ▶ <http://dp.la/> (USA)
- ▶ <http://archive.org/web/> (Webarchive)



Digitized and Digital Sources

- ▶ Roy Rosenzweig (2003):
„Historians, in fact, may be facing a fundamental paradigm shift from a culture of scarcity to a culture of abundance.“

Roy Rosenzweig, ‚Scarcity or Abundance? Preserving the Past in a Digital Era‘ in *The American Historical Review* (2003), p 735-762



Many thanks
for your attention!

