

Software, Hard Facts and the Quality of Interpretation – Big Data in Historical Research

Workshop

"Digital Humanities und die Auswertung digitaler Unterlagen", Federal Archives Bern, 12.6.2013

Patrick Kammerer

Forschungsstelle für Sozial- und Wirtschaftsgeschichte, Universität Zürich

Associaion "Histoire & Informatique"

Overview - Contents

Big Data in Historical Research - Focus on Archives

- 1) Hard Facts - Historical Statistics
- 2) Software - Google NgramViewer
- 3) Qualities of Interpretation - Conclusion

Wie verstehen wir grosse Datenmengen als historische Archivquellen unter digitalen bzw. neuen medialen Bedingungen?

Big Data - Hard Facts [Historical Statistics]

Historical Statistics of Switzerland Online,

Patrick Kammerer, Margrit Müller, Jakob Tanner, Ulrich Woitek (Eds.). [<http://www.fsw.uzh.ch/histstat>, accessed 2.1.2013].

Webbased statistical database, more than 30'000 timeseries; covering the modern federal state

Development branch [<http://www.fsw.uzh.ch/pka>, accessed 2.1.2013].

historical statistics of switzerland online

[overview] [search] [upload] [documentation] [home] [readme] [message board]

for graphical support activate the [timeline] language [0] [1] [-]

Geld und Kredit

O.1a.	Prägung von Schweizermünzen 1850-1938 (in Tausend Stück) -	→	📄
O.2a.	Notenumlauf der früheren Emissionsbanken 1826-1910 -	→	📄
O.3.	Schätzung der Geldmenge in der Schweiz am Jahresende 1851-1910 (in Millionen Franken) -	→	📄
O.4.	Geldmengen am Jahresende 1907-1954 (in Millionen Franken) -	→	📄
O.5a.	Geldmengen am Jahresende 1950-1983 und im Jahresmittel 1979-1990 (in Millionen Franken) -	→	📄
O.6.	Giroverkehr der Schweizerischen Nationalbank 1907-1972 (in Millionen Franken) -	→	📄
O.7.	Bilanzposten am Jahresende und Ertragsrechnung der Schweizerischen Nationalbank 1907-1986 -	→	📄
O.8.	Kantonalbanken, Bodenkreditbanken, andere Lokalbanken, Sparkassen und Grossbanken: Anzahl und Bilanzsumme 1800-1995 -	→	📄
O.9a.	Kantonalbanken, Bodenkreditbanken, andere Lokalbanken und Sparkassen: -	→	📄
O.10.	Kantonalbanken, Bodenkreditbanken, andere Lokalbanken, Sparkassen, Grossbanken und Nationalbank: Wertschöpfung 1850-1913 -	→	📄
O.11.	Banken und Finanzgesellschaften nach Grössenklassen 1935-1985 (ausgewählte Jahre) -	→	📄



Literatur:

Alain Desrosières: Die Politik der grossen Zahlen, eine Geschichte der statistischen Denkweise. Berlin 2005. -> Entstehungskontext und Verwendungskontext von Statistiken und statistischen Konzepten

Gerd Gigerenzer, Zeno Swijtink, Theodore Porter, Lorraine Daston, John Beatty, Lorenz Krüger: Das Reich des Zufalls. Wissen zwischen Wahrscheinlichkeiten, Häufigkeiten und Unschärfen. Berlin 1999 (Original: The Empire of Chance: How Probability Changed Science and Everyday Life),

Hans Ulrich Jost ; Carlo Malaguerra: Geschichte und Gegenwart der öffentlichen Statistik in der Schweiz, Zürich 2012.

Heiner Ritzmann-Blickenstorfer (Hg.): Historische Statistik der Schweiz / Statistique historique de la Suisse / Historical Statistics of Switzerland; unter der Leitung von Hansjörg Siegenthaler, Zürich 1996. (5,2 kg)



Geschichte der Statistik - Paradigmenwechsel seit der Mitte der 1990er Jahre

Big Data: Entstehungs- und
Verwendungszusammenhang von Statistiken und
statistischen Konzepten

Wichtige Begriffe für die Forschungsperspektiven der
Geschichte der Statistik und der historischen Statistik:
Entstehungs- und Verwendungszusammenhang von
Statistiken und statistischen Konzepten (Derosières,
Daston/Giegerenzer)

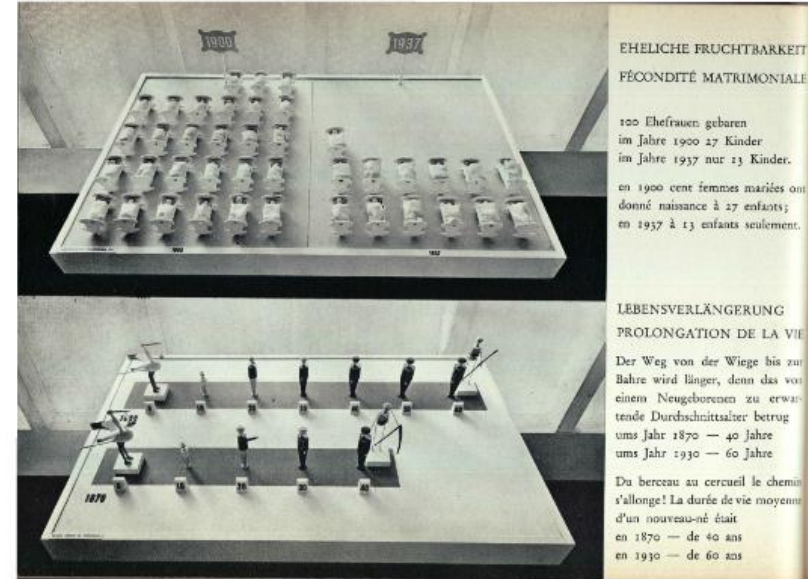
Beispiel zum Verwendungskontext von historisch-
statistischen Daten:

Geburtenstatistik an der Landesausstellung 1939

Geburtenstatistik an der Landesausstellung 1939



Die Sondernummern der Zürcher Illustrierten zur Schweizerischen Landesausstellung 1939, Heft 1 (Nr. 27), Zürich 1939, S. 838.



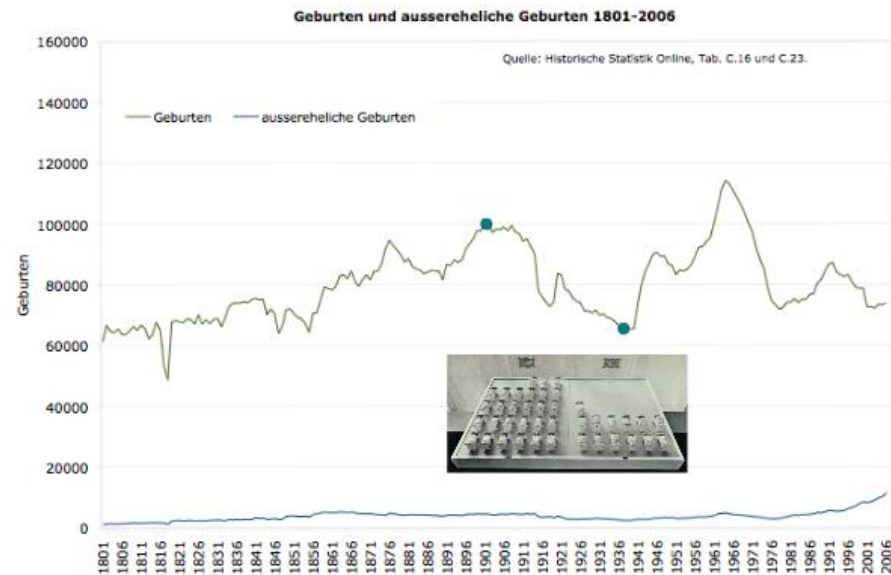
Heimat und Volk: Schweizerische Landesausstellung 1939 Zürich, von Hans Hofmann, Chefarchitekt der Schweizerischen Landesausstellung 1939 mit 141 Photos von M. Wolgensinger, Zürich 1939. S. 35..

Geburtenstatistik an der Landesausstellung 1939



Heimat und Volk: Schweizerische Landesausstellung 1939 Zürich, von Hans Hofmann, Chefarchitekt der Schweizerischen Landesausstellung 1939 mit 141 Photos von M. Wolgensinger, Zürich 1939. S. 35.

Geburtenstatistik an der Landesausstellung 1939



[http://www.fsw.uzh.ch/pka/home/projects/histstat/nls_development/overview.php]

Thesen Fragen zu Big Data und "Hard Facts"

- 1) Archivierung von Entstehungs- und Verwendungskontexten als Herausforderung bei der Archivierung von "Big Data" in (zukünftigen) digitalen Archiven
- 2) Welche Strukturen als Verbindung von technischen Lösungen und Wissensbeständen sind zukunftsweisend ?



Google Books Ngram Viewer

[books.google.com/ngrams]

Google Books Ngram Viewer
[\[http://books.google.com/ngrams\]](http://books.google.com/ngrams)

In 500 Billion Words, New Window on Culture

"With little fanfare, Google has made a mammoth database culled from nearly 5.2 million digitized books available to the public for free downloads and online searches, opening a new landscape of possibilities for research and education in the humanities. "

"Their study, to be published in the journal Science on Friday, offers a tantalizing taste of the rich buffet of research opportunities now open to literature, history and other liberal arts professors who may have previously avoided quantitative analysis. Science is taking the unusual step of making the paper available online to nonsubscribers."

[New York Times, December 16, 2010.]

Version 1.0 released in December 2010

- Based on 5,2 million books scanned by Google Inc. up to 2008.
- Paper in "Science": Quantitative Analysis of Culture Using Millions of Digitized Books

Version 2.0 Released in October 2012

- 8,116,746 books (8,1 million)
- syntax tagging
- improved ocr
- html 5 interface

Jean-Baptiste Michel, Erez Lieberman Aiden, Steven Pinker et al.: Quantitative Analysis of Culture Using Millions of Digitized Books, Science 331, 176 (2011).

Google Books Ngram Viewer
[\[http://books.google.com/ngrams\]](http://books.google.com/ngrams)



Google Books Ngram Viewer

[books.google.com/ngrams]

Google Books Ngram Viewer

[<http://books.google.com/ngrams>]

In 500 Billion Words, New Window on Culture

"With little fanfare, Google has made a mammoth database culled from nearly 5.2 million digitized books available to the public for free downloads and online searches, opening a new landscape of possibilities for research and education in the humanities. "

"Their study, to be published in the journal Science on Friday, offers a tantalizing taste of the rich buffet of research opportunities now open to literature, history and other liberal arts professors who may have previously avoided quantitative analysis. Science is taking the unusual step of making the paper available online to nonsubscribers."

[New York Times, December 16, 2010.]

Version 1.0 released in December 2010

- Based on 5,2 million books scanned by Google Inc. up to 2008.

- Paper in "Science": Quantitative Analysis of Culture Using Millions of Digitized Books

Version 2.0 Released in October 2012

- 8,116,746 books (8,1 million)

- syntax tagging

- improved ocr

- html 5 interface

Jean-Baptiste Michel, Erez Lieberman Aiden, Steven Pinker et al.: Quantitative Analysis of Culture Using Millions of Digitized Books, Science 331, 176 (2011).

Google Books Ngram Viewer

[<http://books.google.com/ngrams>]



Ngram Viewer als Tool für Historikerinnen und Historiker

- Breite Diskussion auf Blogs: Good or bad? Nutzen und Nachteil
- Noch kaum Diskussion in historischen Zeitschriften und Sammelbänden (Ausnahmen: z.B. Sarasin 2012)
- Persönliche Beobachtung: Tools, wie der NgramViewer, werden von Studierenden relativ häufig benutzt

Fragen zu Big Data und "Software"

- 1) Wie würde man den Ngram Viewer archivieren?
(Rohdaten, User Interface bzw. Software, Benutzerdaten, etc.)
- 2) Sind "Tools" wie der Ngram Viewer eine neue Quellengattung bzw. eine neue Form von Archivgut?
- 3) Wie kann Software archiviert werden? Gehen Archive zukünftig in diese Richtung?





gource, software version control visualization.
visualization of svc-data.
<http://code.google.com/p/gource/> [2.2.2013].

SVC of openoffice.org october 2000- january 2009

