

Daten teilen?

Zur Bedeutung von Forschungsplattformen
für die Digital Humanities

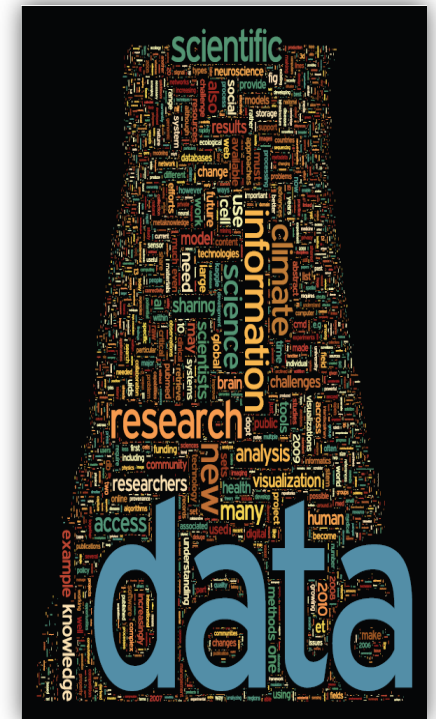
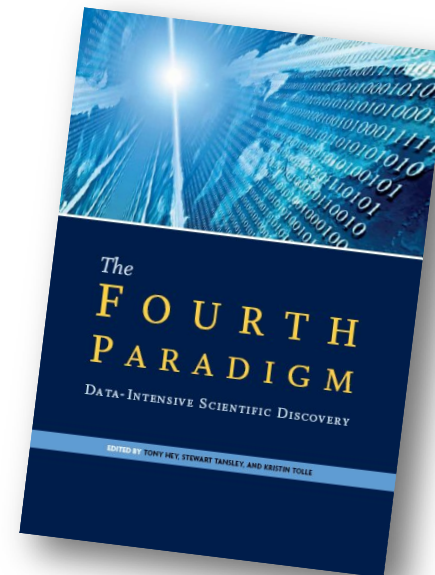
Gerhard Lauer

Viertes Paradigma: Datgetriebene Wissenschaften

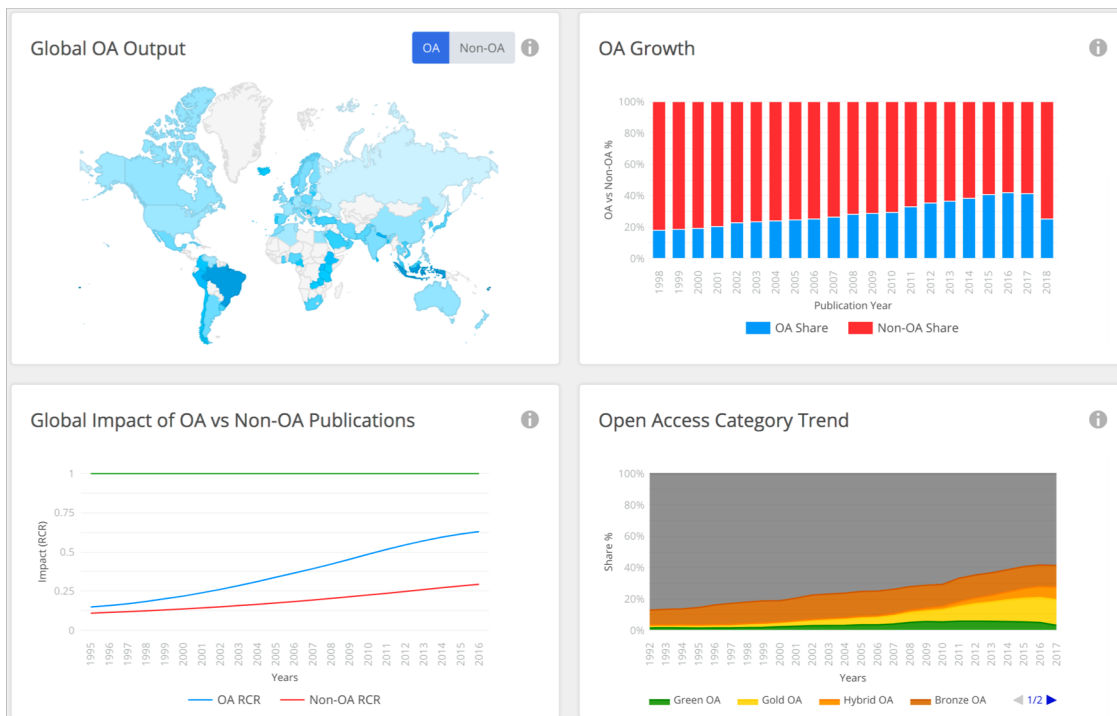
Tony Hey (2009), *The Fourth Paradigm*

Chris Anderson (2008), *The End of Theory*

Science Special (2011), *Dealing with Data*



Die Wirklichkeit des Datenteilens: Open Access Journals



OA Output by Countries

Country	OA Pubs	OA Share	Avg RCR(OA)
1 United States	4.55M	24%	0.90
2 United Kingdom	1.52M	27%	0.78
3 Japan	1.41M	33%	0.32
4 China	983K	26%	0.99
5 Germany	894K	24%	0.94
6 Canada	678K	25%	0.83
7 France	650K	25%	0.85
8 Brazil	596K	55%	0.36
9 Italy	522K	27%	0.86
10 India	480K	29%	0.64
11 Australia	478K	26%	0.91
12 Spain	433K	31%	0.85
13 South Korea	388K	34%	0.59
14 Netherlands	377K	33%	1.06
15 Switzerland	289K	30%	1.11
16 Sweden	253K	28%	0.93

<https://www.wizard.ai/dashboards/open-access>

Lösungen



cOAlition S



cOAlition S
Making
Open Access
a reality
by 2020

A DECLARATION OF COMMITMENT
BY PUBLIC RESEARCH FUNDERS

<http://scieur.org/coalition-s>

Login Area

News

Parliament did not sufficiently improve the far too restrictive proposal for TDM in the new Copyright Directive in today's plenary vote... [read more](#)

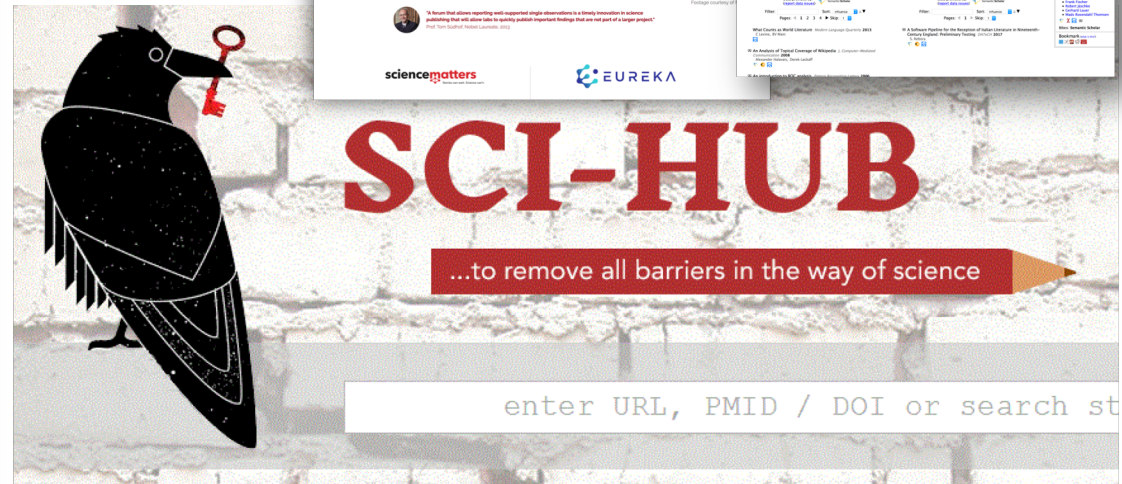
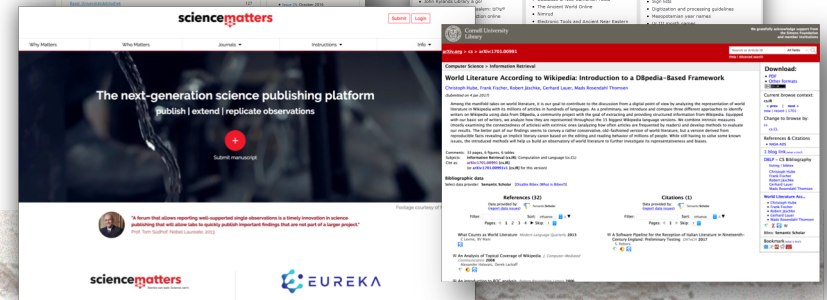
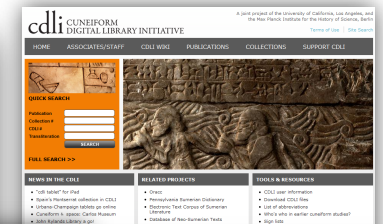
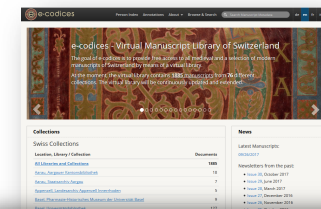
(04/09/18) Science without publication paywalls: cOAlition S for the realisation of full and immediate Open Access... [read more](#)

Documents

Science Without Publication Paywalls: cOAlition S for the Realisation of Full and Immediate Open Access... [read more](#)

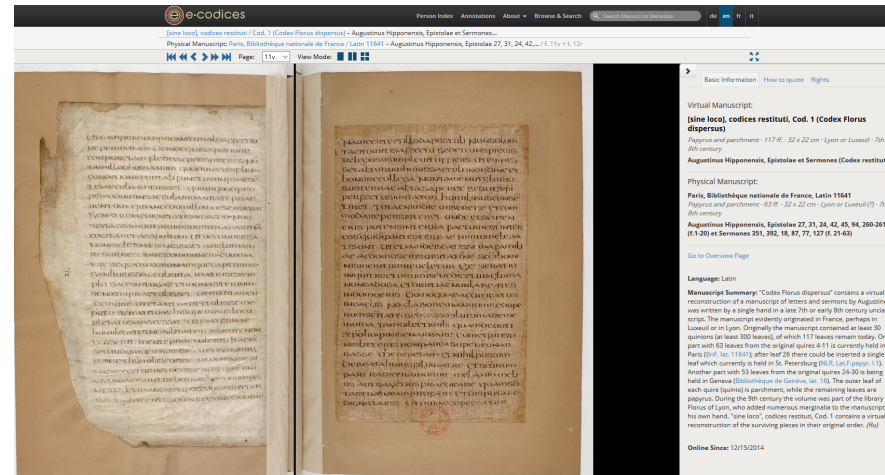
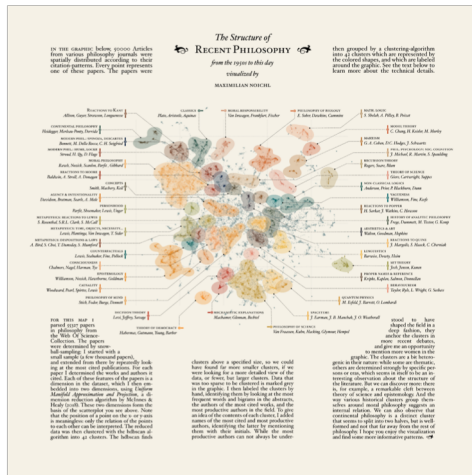
What is cOAlition S?

On 4 September 2018, a group of national research funding organisations, with the support of the European Commission and the European Research Council (ERC), announced the launch of cOAlition S, an initiative to make full and immediate Open Access to research publications a reality. It is built around Plan S, which consists of one target and 10 principles.



Daten: Retro-Digitalisate der Kultur

Formatierte und standardisierte (Meta-)Daten
(XML, rdf, TEI, IIIF, Dublin Core, METS/MODS, LOD, etc. ...)

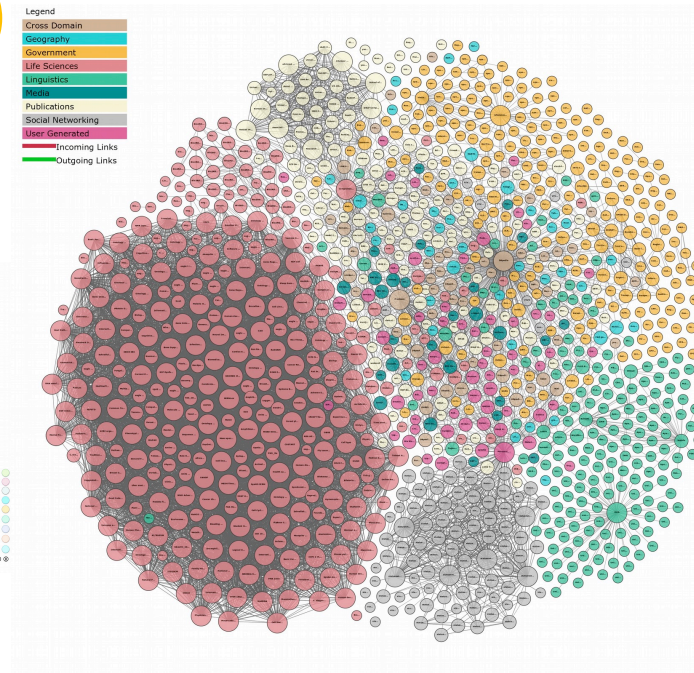
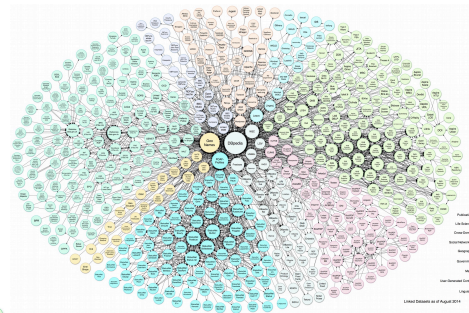
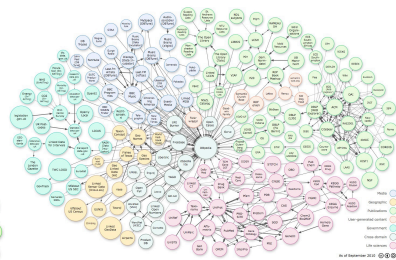
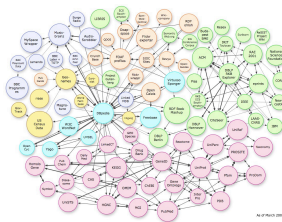
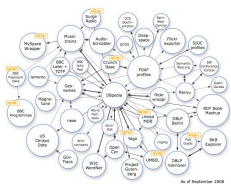


The screenshot shows a digital library interface with two panels, 'Mappe #1' and 'Mappe #2'. The 'Mappe #1' panel displays detailed metadata for a specimen:

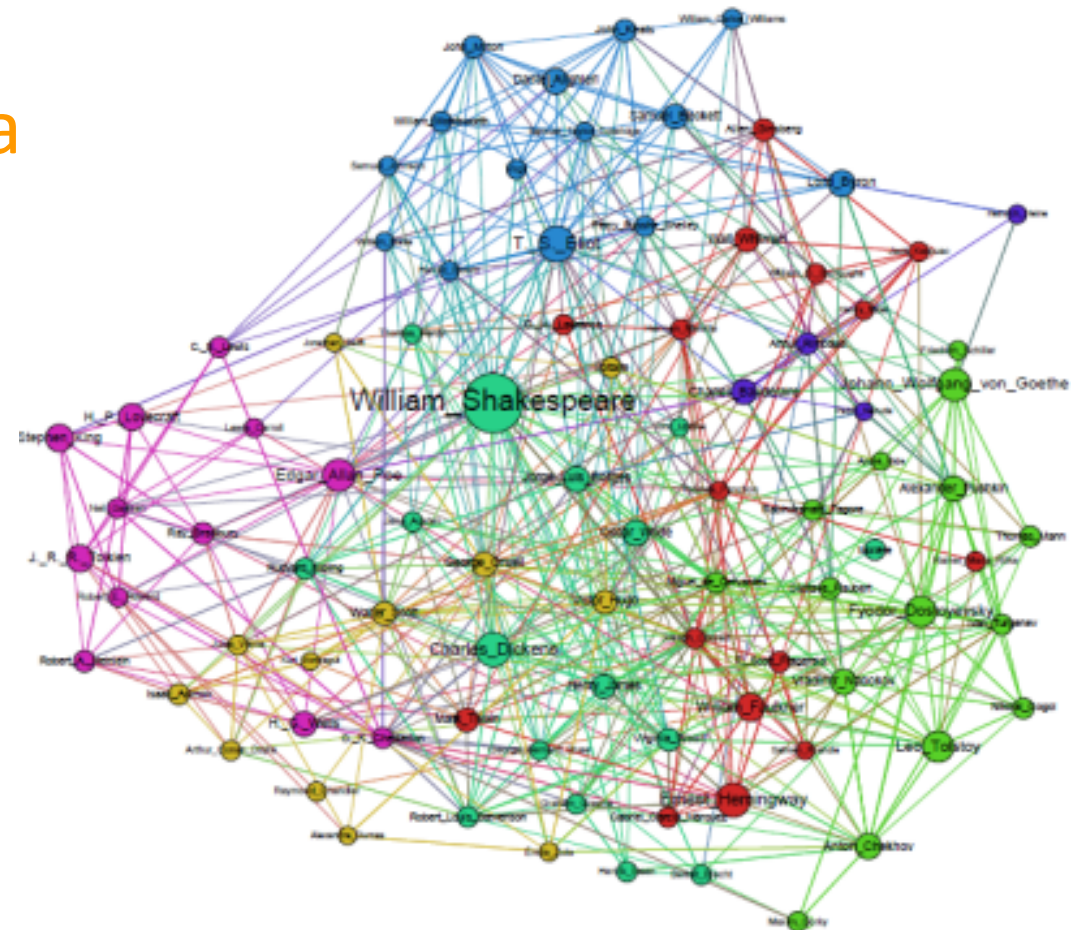
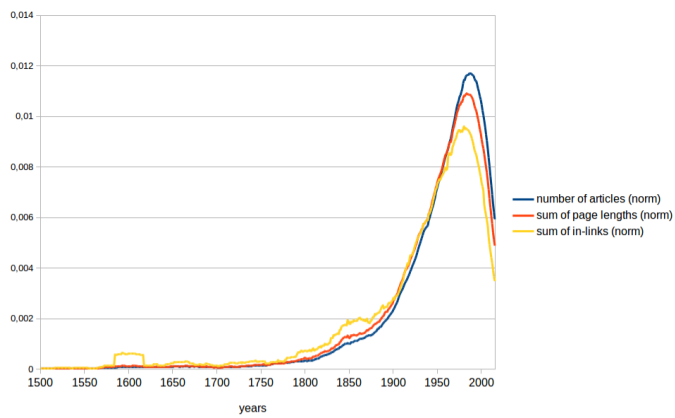
- Objectbezeichnung:** Mensch (Georgien) [Schädel]
- Fachbezeichnung:** Anatomie, Anthropologie
- Originalbezeichnung(en):** Georgianer, Femme Georgianae
- Invartnummer:** AG 546
- Sammlung:** Georg-August-Universität Göttingen, Medizinische Fakultät, Zentrum Anatomie (Blumenbachsche Schädelammlung)
- Anzahl der Objekte:** 1
- Allgemeine Bezeichnung:** Mensch(en)
- Fachbezeichnung:** <Homo sapiens-civ (Mammalia Homidae: Homo)
- Abkürzbezeichnung:** racem
- Fundort:** Moskau, Moskva autonomous city - Rossiya republic - Rossya - Asia - World (Getty-Id: 101207)
- Fundortkoordinaten:** 55° 45' 00"N, 037° 42' 00"E
- Erhalten von:** Hübnerbrant, Johann Konrad (1747-1831) (GZSL:61)
- Erwerbzeitpunkt:** 1793-05-29
- Bibliographische Verweise:**
 - *Originalverke Blumenbachs sind mit der Bibliographienumm (Roske 2010) versehen. Referenzen für die weiteren Werke finden sich unter Sekundärliteratur

Fast growing semantic web (Linked Open Data)

2008-2017

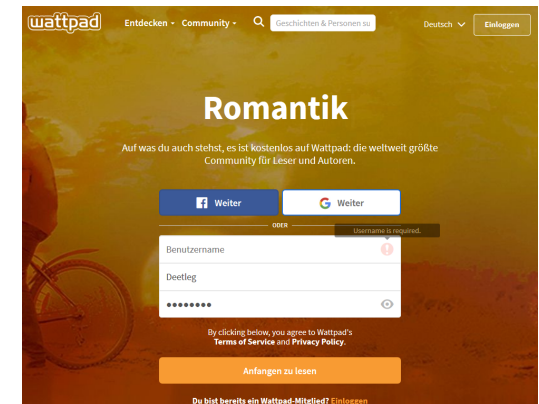
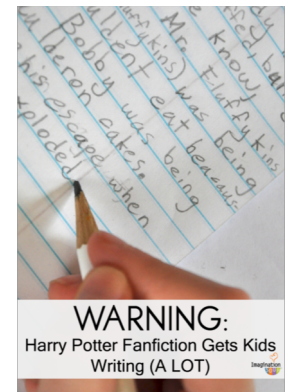


Weltliteratur nach Wikipedia



Hube, Fischer, Lauer et al. (2017)

Daten: Real-time Daten der Kultur



Scraping der Daten

Table 1. Sample of the Wattpad corpus. The column "Reply" indicates whether the comment is a reply to the previous one.

Book	Chapter	Paragraph	Username	Date	Comment	Reply
1	Pride and Prejudice (1813)	Chapter I It is a truth universally acknowledged, that a single man in possession of a good fortune, must be in want of a wife.	[REDACTED]	Jan 16, 2018	I read it once. I don't know how people take the effort to read historical stories. I actually don't understand 99 percent of stuff here	FALSE
2	[REDACTED]	Jan 16, 2018	I tried to read Sherlock Holmes, but I didn't finish it because it was too hard for me to understand. I hope this book is not like that... I heard good things about this book sooooo let's see.	FALSE
3	[REDACTED]	Jan 15, 2018	The Lizzie Bennett Diaries ❤️	FALSE
4	[REDACTED]	Jan 17, 2018	same yes 🍷	TRUE
5	[REDACTED]	Jan 08, 2018	Is this book going to be difficult to understand because I can already tell from this paragraph??	FALSE

Chapter I

Dedicated to TheLaughingLady

2.54M ★ 22.8K 9.54K

It is a truth universally acknowledged, that a single man in possession of a good fortune, must be in want of a wife.

However little known the feelings or views of such a man may be on his first entering a neighbourhood, this truth is so well fixed in the minds of the surrounding families, that he is considered the rightful property of some one or other of their daughters.

"My dear Mr. Bennet," said his lady to him one day, "have you heard that Netherfield Park is let at last?"

Mr. Bennet replied that he had not.

342 Comments

However little known the feelings or views of such a man may be on his first entering a neighbourhood, this truth is so well fixed in the minds of the surrounding families, that he is considered the rightful property of some one or other of their daughters.

[REDACTED]

Even without knowing what the man is like, the surrounding families wants to claim him already. Women didn't have any other means of supports other than their husband, so finding a

View entire message

Aug 28, 2017 at 3:22 am

2 replies

[REDACTED]

I need to read this again 🍷

Aug 28, 2017 at 7:58 am

1 reply

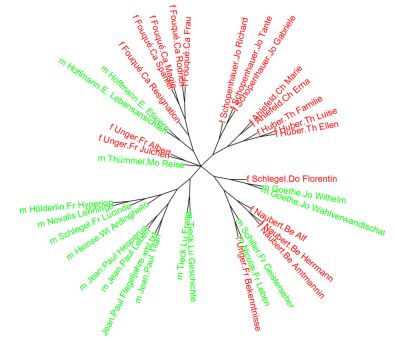
[REDACTED]

[REDACTED]

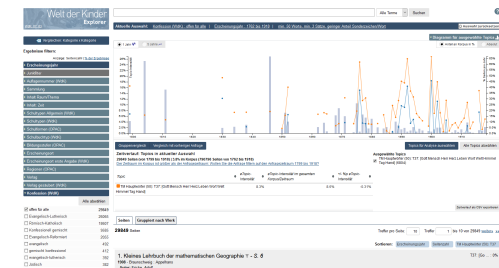
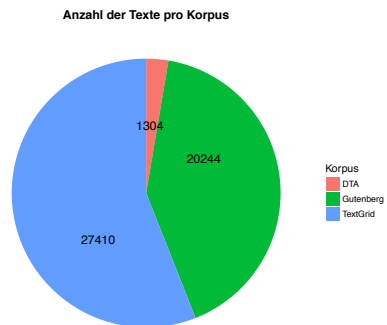
Post a comment...

Korpusbildung (Stichproben) ermöglichen

Ablegen und Nachnutzen von strukturierten Forschungsdaten

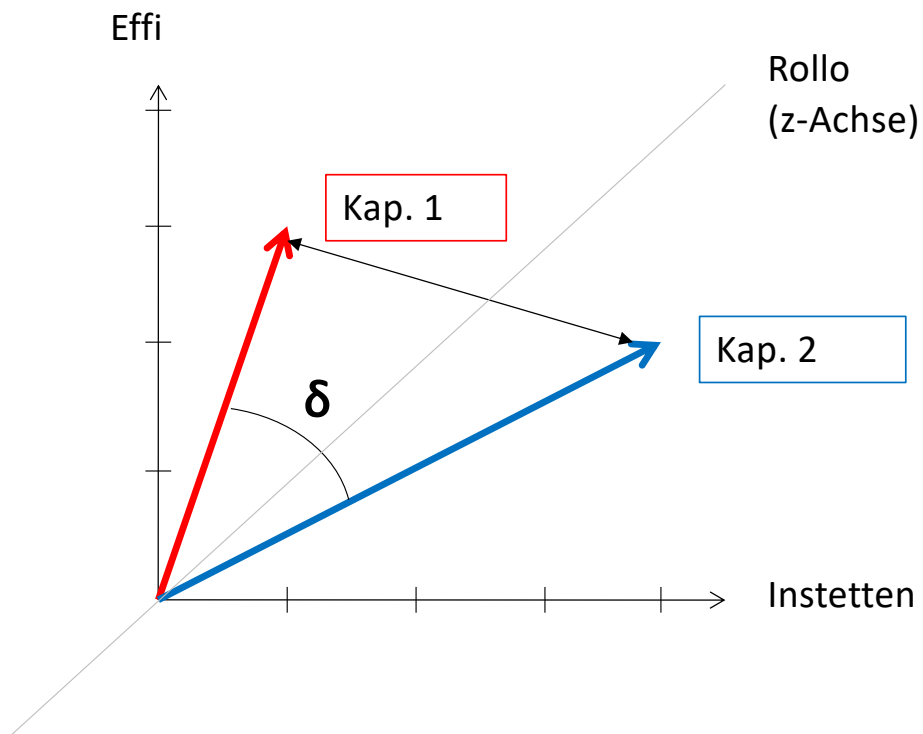


100-2000 MFW Culled @ 50%
Pronouns deleted Classic Delta distance Consensus 0.5




Jannidis & Lauer, 2014
 Hube, Fischer, & Lauer, 2017
<http://wdk.gei.de/>
<https://kolimo.uni-goettingen.de>

Modellierung durch Vektorisierung



	Effi	Instetten	Rollo
Kap. 1	3	1	
Kap. 2	2	4	



Multidimensionale Analysen
(e.g. for Machine Learning)

Datengetriebene Geisteswissenschaften

Wechsel von
hypothesengeleitete Wissenschaften und
explorativer, datengetriebener Generierung von Hypothesen

Tognini-Bonelli, 2001

Forschungsplattformen: DARIAH

- Technische Infrastruktur für das Teilen von Daten und Werkzeugen (VCC 1)
- Unterstützung bei der Nutzung von Forschungsdaten (VCC 2)
- Management des Lebenszyklus von Daten (VCC 3)
- Verbreiterung der Zusammenarbeit (VCC 4)



Daten teilen in den Geisteswissenschaften

- Big Data als Forschungsgrundlage
- Wachsende Rolle von unüberwachten, datengeleiteten Methoden
- Nachhaltige Forschungsumgebungen von Gitlab bis zu Datenrepositories wie Zenodo
- FAIR-Prinzipien und das geltende Recht



Goethe, Faust, GSA 25W 1372