

Forschungsdatenmanagement

Fachhochschule für den öffentlichen Dienst,
SS 2017

Michael Franke, Max Planck Digital Library

Literaturempfehlungen

- Büttner, Stephan; Hobohm, Hans-Christoph; Müller, Lars (2011): Handbuch Forschungsdatenmanagement; <https://opus4.kobv.de/opus4-fhpotsdam/frontdoor/index/index/docId/208>
- Rat für Informationsinfrastrukturen (2016): Leistung durch Vielfalt; <http://www.rfii.de/?wpdmdl=1998>

Forschungsdaten – Situation normal, alles im Eimer



Data Sharing and Management Snafu in 3 Short Acts



NYU Health Sciences Library

Abonnieren 71

32.692

Wofür Forschungsdatenmanagement?

- Gute wissenschaftliche Praxis ([DFG](#))
- Nachnutzung der Daten
- Wissenschaftliches Renommee
 - z.B. Enhanced Publication
(<http://hdl.handle.net/11858/00-001M-0000-0029-B0C0-8>)

Struktur der Veranstaltung

Teil 1: Forschungsdaten

Teil 2: Forschungsdateninfrastruktur

Teil 3: Forschungsdatenmanagement in Bibliotheken

Teil 4: Weiterführende Überlegungen

Struktur der Veranstaltung

Teil 1: Forschungsdaten

Teil 2: Forschungsdateninfrastruktur

Teil 3: Forschungsdatenmanagement in Bibliotheken

Teil 4: Weiterführende Überlegungen

Beispiele für Forschungsdaten

Gravitationswellen

BICEP2

BICEP2 2014 Results Release

1. BICEP2 2014 Release Papers
2. BICEP2 2014 Release Data Products
3. BICEP2 2014 Release Figures from Papers
4. Previous Publications

[Videos for Technical Talk](#) \Rightarrow and [News Conference](#) \Rightarrow
[BICEP2 2014 Release Frequently Asked Questions](#) \Rightarrow
[BICEP2 2014 Release Image Gallery](#) \Rightarrow
[BICEP2 Public Web Pages and News Releases](#) \Rightarrow

1. BICEP2 2014 Release Papers

BICEP2 2014 I: Detection of B-mode Polarization at Degree Angular Scales
The BICEP2 Collaboration, 2014

[PDF](#) / [figures](#) / [arXiv](#) / [ADS](#)

BICEP2 2014 II: Experiment and Three-year Data Set

The BICEP2 Collaboration, 2014

[PDF](#) / [figures](#) / [arXiv](#) / [ADS](#)


2. BICEP2 2014 Release Data Products

B2_3yr_rlikelihood_20140314.txt	Text file containing the tabulated likelihood for the tensor-to-scalar ratio, r , computed using the "direct likelihood calculation" described in Section 9.3.1 of Barkats <i>et al.</i> and Section 11.1 of BICEP2 2014 I.
B2_3yr_bandpowers_20140314.txt	Text file containing bandpowers and statistical uncertainties, corresponding to Figure 2 of BICEP2 2014 I.
B2_3yr_bandpower_window_functions.tgz B2_3yr_bpwf_bin1_20140314.txt B2_3yr_bpwf_bin2_20140314.txt B2_3yr_bpwf_bin3_20140314.txt B2_3yr_bpwf_bin4_20140314.txt	Bandpower window functions for each l bin.

Beispiele für Forschungsdaten

Gravitationswellen

LIGO



LIGO Open Science Center
LIGO is operated by California Institute of Technology and Massachusetts Institute of Technology and supported by the U.S. National Science Foundation.

Getting Started

- Tutorials
- Data**
 - Events
 - Bulk Data
- Timelines
- My Sources
- Software
- GPS ↔ UTC
- About LIGO
- Data Analysis Projects
- Acknowledgement

Welcome to the LIGO Open Science Center

About LIGO
Get Started **with LIGO data**
Join the E-mail list for updates
For general information on LIGO, please visit ligo.org
If you have LSC credentials, you may go to the [development site](#)

Discoveries from the LIGO detectors!

released 2017 June 1:
Event of January 4, 2017: GW170104: total mass 50

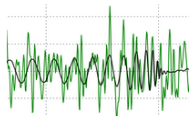
released 2016 June 15:
Event of December 26, 2015: GW151226: total mass 22

released 2016 June 15:
Candidate event of October 12, 2015: LVT151012: total mass 37

released 2016 Feb 11:
Event of September 14, 2015: GW150914: total mass 65

The LIGO Laboratory's [Data Management Plan](#) describes the scope and timing of LIGO data releases.

Jupyter notebook
See the new tutorial on signal processing with LIGO data, as a Jupyter (Python) notebook.
[Tutorial on Binary Black Hole Signals in LIGO Open Data](#)



Beispiele für Forschungsdaten

Gensequenzen

The screenshot shows the NCBI GenBank database entry for the Homo sapiens class II AP endonuclease (APE) gene, partial CDS. The entry includes a search bar, navigation links, and detailed gene information.

NCBI Resources | **How To** | **Sign in to NCBI**

Nucleotide | Nucleotide | Search | Limits | Advanced | Help

Display Settings: GenBank | Send:

Homo sapiens class II AP endonuclease (APE) gene, partial CDS

GenBank: M99703.1
[FASTA](#) | [Graphics](#)

[Go to:](#)

LOCUS HUMAPEB 3019 bp DNA linear PRI 31-DEC-1994
DEFINITION Homo sapiens class II AP endonuclease (APE) gene, partial CDS.
ACCESSION M99703
VERSION M99703.1 GI:178748
KEYWORDS 3' diesterase associated activity; DNA repair protein; DNA-binding; class II AP endonuclease; nuclear localized protein.
SOURCE Homo sapiens (human)
ORGANISM [Homo sapiens](#)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 3019)
AUTHORS Harrison,L., Ascione,G., Menninger,J.C., Ward,D.C. and Demple,B.
TITLE Human apurinic endonuclease gene (APE): structure and genomic mapping (chromosome 14q11.2-12)
JOURNAL Hum. Mol. Genet. 1 (9), 677-680 (1992)
PUBMED [1284593](#)

FEATURES

Location/Qualifiers	
source	1..3019 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606"
intron	596..778 /gene="APE" /number=1
gene	join(779..904,1115..1302,1869..2061) /gene="APE"
mRNA	join(779..904,1115..1302,1869..2061) /gene="APE"
exon	779..904 /gene="APE" /number=2
CDS	join(847..904,1115..1302,1869..2061)

Analyze this sequence
Run BLAST
Pick Primers
Highlight Sequence Features
Find in this Sequence

Articles about the APEX1 gene
Expression and prognostic significance of APE1/Ref-1 and NPM1 | [Am J Clin Pathol. 2014]
Dysregulation of human apurinic/apyrimidinic endonuclease 1 (APE1) | [Er J Ophthalmol. 2014]
PRP19 transforms into a sensor of RPA-ssDNA after DNA damage and drives ATR | [Mol Cell. 2014]
See all...

Reference sequence information
RefSeq alternative splicing
See 6 reference mRNA sequence splice variants for the APEX1 gene.

More about the APEX1 gene
Apurinic/apyrimidinic (AP) sites occur frequently in DNA molecules by spontaneous hydrolysis, by DNA damaging agents or by DNA glycosylases
...
Also Known As: APE, APE1, APEN, APEX, ...

Genbank

Beispiele für Forschungsdaten

Nahrungsmittelanalysen

USDA
Agricultural Research Service
United States Department of Agriculture

National Nutrient Database for Standard Reference
Release 26

[NDL Home](#) [Food Search](#) [Nutrients List](#) [Ground Beef Calculator](#) [Contact Us](#) [SR26 Documentation](#) [Help](#)

FOOD GROUP

Baby Foods (1)
Cereal Grains and Pasta (6)
Fast Foods (1)
Meals, Entrees, and Side Dishes (1)
Vegetables and Vegetable Products (17)

Enter one or more terms and click 'Go' for a new search

26 foods found. Click on a food name to view details

1 2 Next

NDB No.	Description	Food Group
11658	Spinach souffle	Vegetables and Vegetable Products
11457	Spinach, raw	Vegetables and Vegetable Products
11986	Malabar spinach, cooked	Vegetables and Vegetable Products
20126	Spaghetti, spinach, dry	Cereal Grains and Pasta
20127	Spaghetti, spinach, cooked	Cereal Grains and Pasta
11274	Mustard spinach, (tendergreen), raw	Vegetables and Vegetable Products
11276	New Zealand spinach, raw	Vegetables and Vegetable Products
03127	Babyfood, vegetables, spinach, creamed, strained	Baby Foods
11458	Spinach, cooked, boiled, drained, without salt	Vegetables and Vegetable Products
11463	Spinach, frozen, chopped or leaf, unprepared	Vegetables and Vegetable Products
11854	Spinach, cooked, boiled, drained, with salt	Vegetables and Vegetable Products
20095	Pasta, fresh-refrigerated, spinach, as purchased	Cereal Grains and Pasta
20096	Pasta, fresh-refrigerated, spinach, cooked	Cereal Grains and Pasta
20111	Noodles, egg, spinach, dry, enriched	Cereal Grains and Pasta
20112	Noodles, egg, spinach, cooked, enriched	Cereal Grains and Pasta
21500	KASHI Pizza, Mushroom Trio & Spinach	Fast Foods
11275	Mustard spinach, (tendergreen), cooked, boiled, drained, without salt	Vegetables and Vegetable Products
11277	New Zealand spinach, cooked, boiled, drained, without salt	Vegetables and Vegetable Products
11459	Spinach, canned, regular pack, solids and liquids	Vegetables and Vegetable Products
11461	Spinach, canned, regular pack, drained solids	Vegetables and Vegetable Products
11801	Mustard spinach, (tendergreen), cooked, boiled, drained, with salt	Vegetables and Vegetable Products
11802	New zealand spinach, cooked, boiled, drained, with salt	Vegetables and Vegetable Products
11855	Spinach, canned, no salt added, solids and liquids	Vegetables and Vegetable Products
11464	Spinach, frozen, chopped or leaf, cooked, boiled, drained, without salt	Vegetables and Vegetable Products
11856	Spinach, frozen, chopped or leaf, cooked, boiled, drained, with salt	Vegetables and Vegetable Products

USDA

Beispiele für Forschungsdaten

Tierbewegungen

Movebank

The screenshot displays the Movebank website interface. At the top, the logo "MOVEBANK For Animal Tracking Data" is visible. Below the navigation bar, there are sections for "User login" (with fields for Username and Password) and a "Search" section. The search results are sorted by "Animal Identifier" and list various studies, including "TBMU_Digges_Glchirat_2013" and "TBMUCOMU_GastonMontevocchioNWAatlantic". A map of Europe is shown on the right, with several green dots representing tracking data points across countries like Spain, France, Germany, and Italy. The map includes a search bar, zoom controls, and a "Link" button. At the bottom, there is a footer with the text "Movebank is coordinated by the Max Planck Institute for Ornithology, the North Carolina Museum of Natural Sciences, and the University of Konstanz. For questions or feedback, contact us at support@movebank.org".

Beispiele für Forschungsdaten

Gesichter/Mimik

The image shows a composite of two web pages. The top page is from the Max Planck Institute for Human Development, displaying a gallery of facial expressions with a map of Europe in the background. The bottom page is a Scopus search result for the paper 'FACES-a database of facial expressions in young, middle-aged, and older women and men: Development and validation' by Ebner N.C. et al. (2010). The Scopus page shows 227 documents have cited the paper, with a list of results including authors like Prada, M., Rodrigues, D., Garrido, M.V., Lopes, J., Lefevre, C.E., Saxton, T.K., Luckhardt, C., Kröger, A., Cholemery, H., Bender, S., Freitag, C.M., Cortes, D.S., Laukka, P., Lindahl, C., and Fischer, H.

Max Planck Institute for Human Development
HOME PICTURES ALBUMS
Sort Organize
Sorted by Emotion Person-ID Picture Set Sort
Sad, Happy
Go to page 1 of 2

Scopus Search Sources Alerts Lists Help SciVal Register
227 documents have cited:
FACES-a database of facial expressions in young, middle-aged, and older women and men: Development and validation
Ebner N.C., Riediger M., Lindenberg U.
(2010) Behavior Research Methods, 42 (1) , pp. 351-362.
Is cited by: Set feed
227 documents results for: Analyze search results
Search within results...
Refine results: Limit to Exclude
Year: 2017 (22), 2016 (58), 2015 (51), 2014 (34), 2013 (29), 2012 (20), 2011 (13)
Author name: Ebner, N.C. (18)
Prada, M., Rodrigues, D., Garrido, M.V., Lopes, J. 2017 Food
Lefevre, C.E., Saxton, T.K. 2017 Evol
Luckhardt, C., Kröger, A., Cholemery, H., Bender, S., Freitag, C.M. 2017 Journ
Cortes, D.S., Laukka, P., Lindahl, C., Fischer, H. 2017 PLoS

Definitionsversuche

- „Ganz allgemein gesprochen sind Forschungsdaten Daten, die im Forschungsprozess erzeugt, gesammelt oder zusammengestellt werden und auf deren Grundlage wissenschaftliche Hypothesen, Modelle oder Theorien gebildet werden“ (Uni Heidelberg).
- „Forschungsdaten stellen im weitesten Sinne Primärdaten, Sekundäranalysen, Visualisierungen, Modelle, Analysewerkzeuge, Objektsammlungen oder Produkte dar, die während des wissenschaftlichen Arbeitsprozesses erzeugt und benutzt werden“ (Johanna Vompras, Uni Bielefeld).
- *„Forschungsdaten sind Daten, die im Zuge wissenschaftlicher Vorhaben z.B. durch Digitalisierung, Quellenforschungen, Experimente, Messungen, Erhebungen oder Befragungen entstehen.“ (Allianz AG „Forschungsdaten“)*
- *„(...) researchers know within their projects what research data is“ (Leibniz-Gemeinschaft)*

Worum geht es nicht

- Physische Daten
- Wissenschaftliche Publikationen
- Daten über Wissenschaft

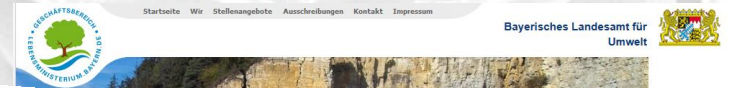
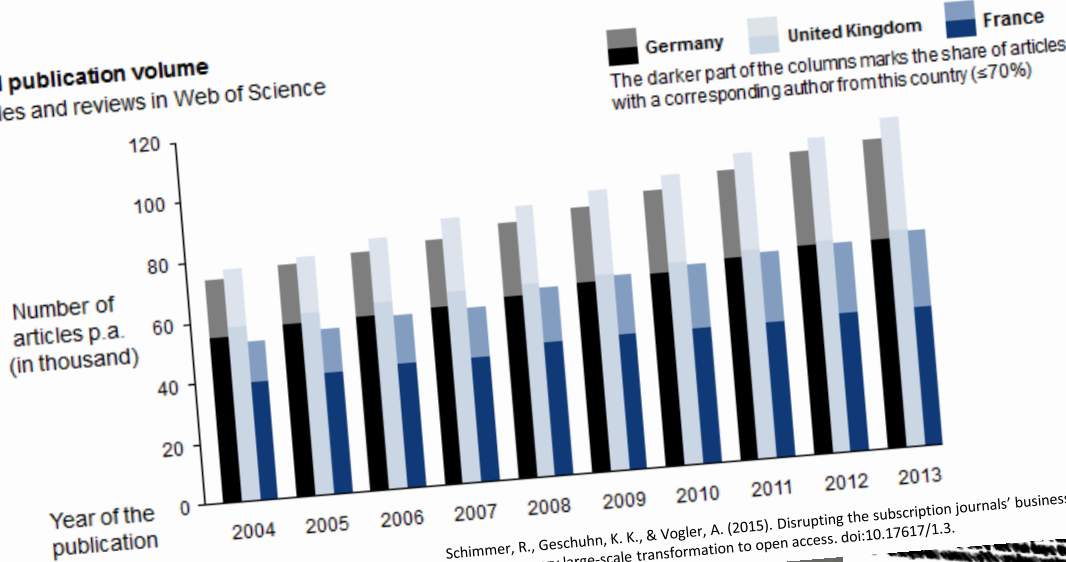


Figure 1: Publication volume with APC relevant share of selected European countries

Total publication volume
Articles and reviews in Web of Science



Schimmer, R., Geschuhn, K. K., & Vogler, A. (2015). Disrupting the subscription journals' business model for the necessary large-scale transformation to open access. doi:10.17617/1.3.

Klassifikation von Forschungsdaten

Nach dem Charakter der Daten

Texte (Quellen, Transkripte)

Bilder (Scans, Fotos, Mikroskopdaten, Teleskopdaten, Satellitendaten)

Multimedia (Audiodaten, Videodaten, 3D-Daten, 4D-Daten)

Zahlenreihen (Umfragedaten, Experimentaldaten, Sensormessreihen, Gensequenzen, Geodaten)

Born-digitals (Simulationsdaten, Algorithmen, Websites)

Qualitative Daten



Quantitative Daten



Klassifikation von Forschungsdaten

the APICS Consortium (Max Planck Institute for Evolutionary Anthropology). APICS supplemental files. <http://edmond.mpg.de/imeji/collection/4WKY1hHhw8iEuNQ4>

Nach dem Format der Daten

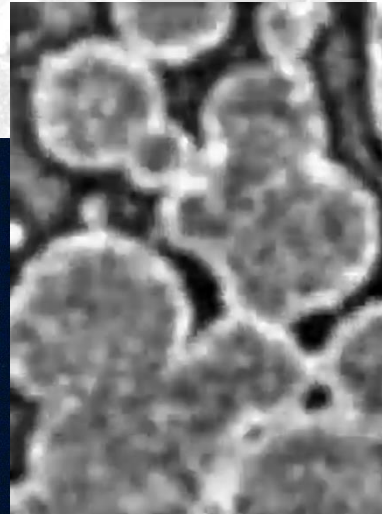
Texte (TXT, ODT, PDF, ...)

Bilder (TIFF, JPEG, ...)

Multimedia (MP4, Dicom, ...)

Zahlenreihen (CSV, XLSX, ...)

Born-digitals (netCDF, grib, HTML, ...)



```
HCCT-2010
Schmücke station
MAAP @ interstitial inlet
MPIC Mainz
Contact: Johannes Schneider, johannes.schneider@mpic.de
```

```
Data status: Final (1.3.2011)
(unreliable data points have been removed)
```

```
Date_and_Time
Black_Carbon_ng/m3
```

11.09.2010 18:30:05	144
11.09.2010 18:31:05	117
11.09.2010 18:32:05	157
11.09.2010 18:33:05	165
11.09.2010 18:34:05	198
11.09.2010 18:35:05	226
11.09.2010 18:36:05	230
11.09.2010 18:37:05	216
11.09.2010 18:38:05	232
11.09.2010 18:39:05	232
11.09.2010 18:40:05	249
11.09.2010 18:41:05	241
11.09.2010 18:42:05	449
11.09.2010 18:43:05	329
11.09.2010 18:44:05	252
11.09.2010 18:45:05	215
11.09.2010 18:46:05	119
11.09.2010 18:47:05	173
11.09.2010 18:48:05	215

```
netcdf tos_01_2001-2002 {
dimensions:
  lon = 180 ;
  lat = 170 ;
  time = UNLIMITED ; // (24 currently)
  bnds = 2 ;
variables:
  double lon(lon) ;
    lon:standard_name = "longitude" ;
    lon:long_name = "longitude" ;
    lon:units = "degrees_east" ;
    lon:axis = "X" ;
    lon:bounds = "lon_bnds" ;
    lon:original_units = "degrees_east" ;
  double lon_bnds(lon, bnds) ;
  double lat(lat) ;
    lat:standard_name = "latitude" ;
    lat:long_name = "latitude" ;
    lat:units = "degrees_north" ;
    lat:axis = "Y" ;
    lat:bounds = "lat_bnds" ;
    lat:original_units = "degrees_north" ;
  double lat_bnds(lat, bnds) ;
  double time(time) ;
    time:standard_name = "time" ;
    time:long_name = "time" ;
    time:units = "days since 2001-1-1" ;
    time:axis = "T" ;
    time:calendar = "360_day" ;
    time:bounds = "time_bnds" ;
    time:original_units = "seconds since 2001-1-1" ;
  double time_bnds(time, bnds) ;
  float tos(time, lat, lon) ;
    tos:standard_name = "sea surface temperature" ;
    tos:long_name = "Sea Surface Temperature" ;
    tos:units = "K" ;
```

Sea surface temperatures collected by PCMDI for use by the [IPCC](http://www.ipcc.org).
<https://www.unidata.ucar.edu/software/netcdf/examples/files.html>

235
256
230
252
238
244
215
196
162
188
216
197
213
243
239

<http://dx.doi.org/10.17617/3.i>

[dl.mpg.de/imeji/collection/KVTF6tbHX0y49NC](http://edmond.mpg.de/imeji/collection/KVTF6tbHX0y49NC)

Klassifikation von Forschungsdaten

Nach der Quelle der Daten
bzw. der Art der Forschung

Observational: data captured in real time that is usually unique and irreplaceable. For example, remote sensing data, survey data, field recordings, sample data.

Experimental: data captured from lab equipment that is often reproducible. For example, gene sequences, chromatograms, magnetic field data.

Models or simulation: data generated from test models where model and metadata may be more important than output data from the model. For example, climate models, economic models.

Derived or compiled: resulting from processing or combining 'raw' data. For example, text and data mining, compiled databases, 3D models.

Reference or canonical: a static or organic conglomeration or collection of datasets, probably published and curated. For example, gene sequence databanks, collection of letters.

(University of Bristol, Boston University)

Klassifikation von Forschungsdaten

Nach dem Grad der Aggregation/Verdichtung

Rohdaten: Als Rohdaten soll die Gesamtmenge aller Daten bezeichnet werden, die empirischer Wissenschaft als Forschungsgrundlage zur Verfügung stehen.

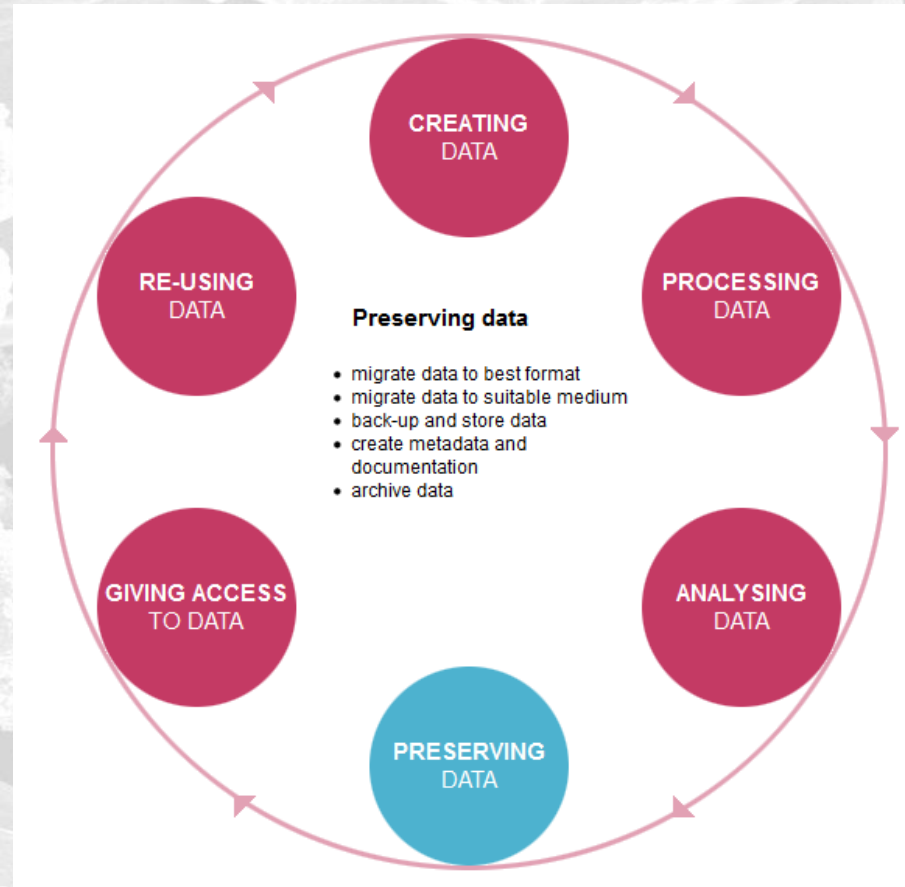
Primärdaten: Primärdaten seien diejenigen Daten, die als Teilmenge der Rohdaten tatsächlich zur Forschung herangezogen werden.

Sekundärdaten: Sekundärdaten seien Daten, die in Prozessschritten aus Primärdaten gewonnen wurden.

(Tertiärdaten: Als Tertiärdaten sollen Informationen bezeichnet werden, die nicht aus den Primärdaten herleitbar sind, aber im wissenschaftlichen Prozess anfallen.)

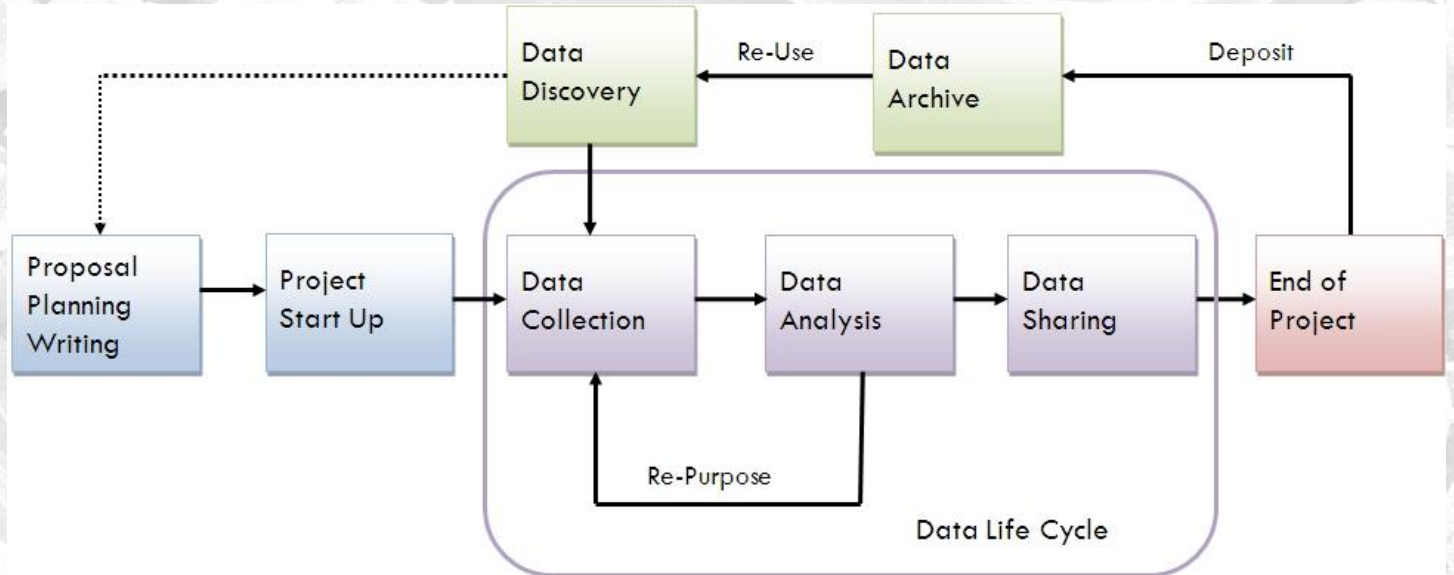
(Franke)

Forschungsdatenzyklus



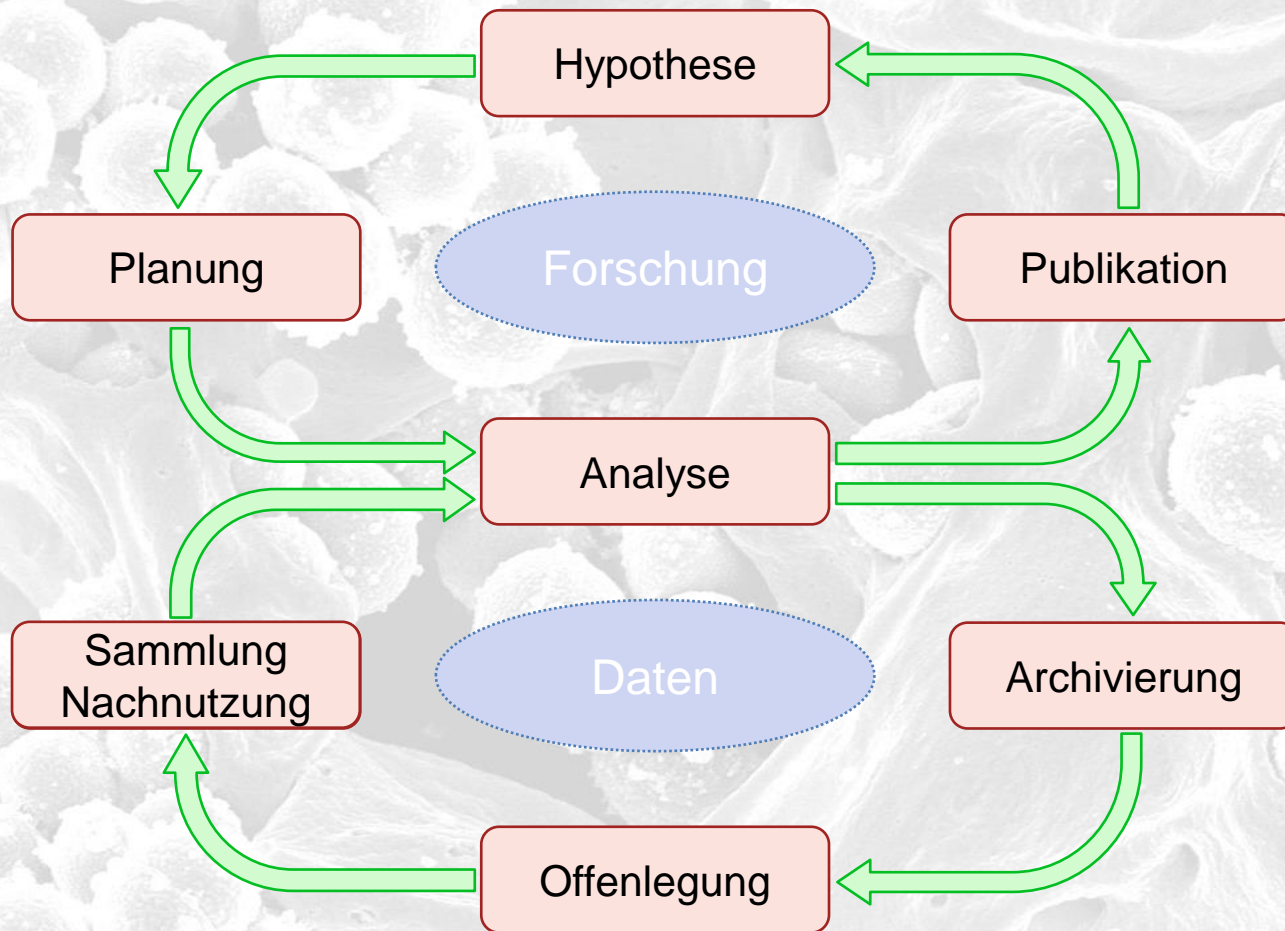
UK Data Archive

Forschungs-Datenzyklus



University of Virginia

Forschungs- und Datenzyklus



Struktur der Veranstaltung

Teil 1: Forschungsdaten

Teil 2: Forschungsdateninfrastruktur

Teil 3: Forschungsdatenmanagement in Bibliotheken

Teil 4: Weiterführende Überlegungen

ktur?

- z.B. Verkehrsinfrastruktur

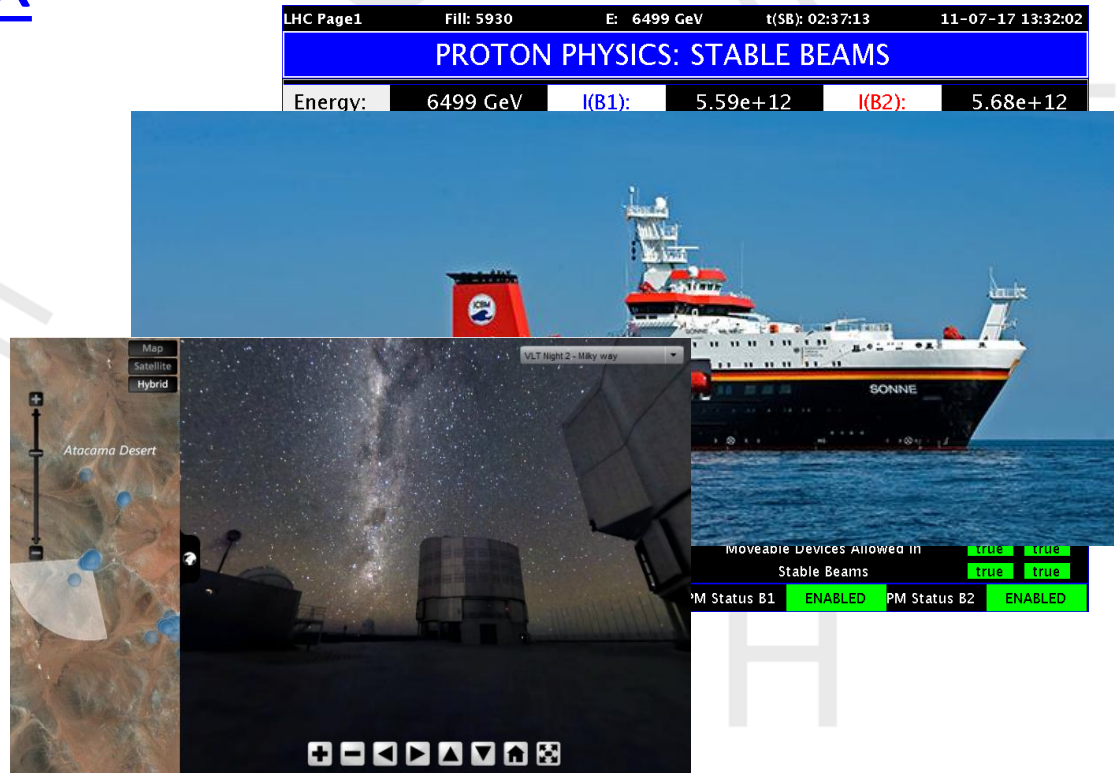
- 
- 
- Straßen, Brücken
 - Schienen, Wasserwege
 - Tankstellennetz, Verkehrsfunk, GPS
 - Straßenverkehrsordnung, Straßenwacht, ADAC
 - Räumdienst, Mautsystem, KFZ-Steuer
 - Autoindustrie, Verkehrsministerium, ÖPNV

Forschungsdateninfrastruktur

- Wiss. Großanlagen
- Metadaten, Standards, Identifikatoren
- Repositorien, Registries, Datenjournale
- Rechtsgrundlagen

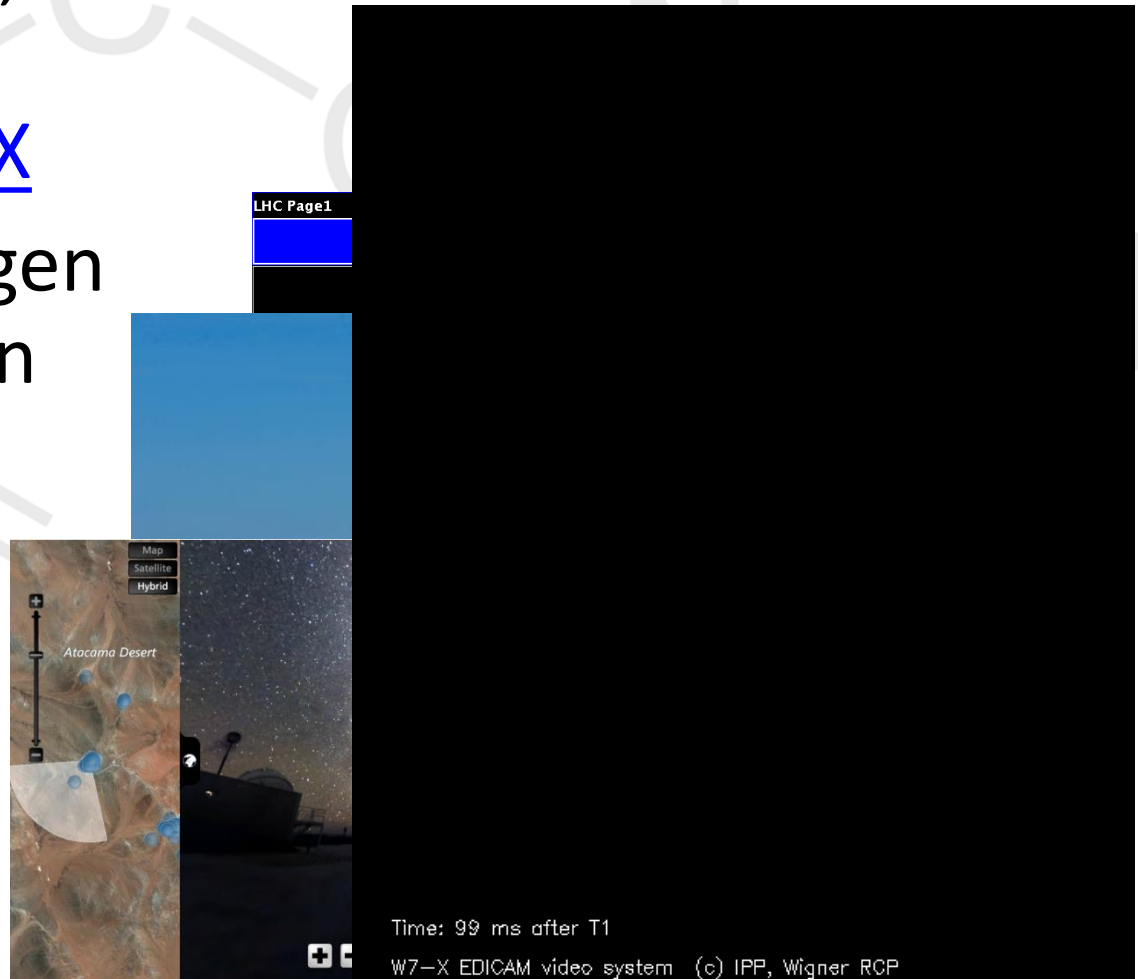
Wissenschaftliche Großanlagen

- z.B. [LHC](#), [Sonne](#),
[E-ELT](#), SKA
[Wendelstein 7-X](#)



Wissenschaftliche Großanlagen

- z.B. [LHC](#), [Sonne](#), [E-ELT](#), SKA
[Wendelstein 7-X](#)
- Erheben/Erzeugen
/Verteilen Daten
- Beschleunigen
Prozesse
- Zwingen zur
Kooperation



Metadaten

- [DC](#) – Dublin Core Metadata Initiative
- [RDF](#) – Resource Description Framework
([Beispiel](#))
- [SKOS](#) – Simple Knowledge Organization System
- ([Metadaten-Registries](#))

Standards

- Metadaten
- Daten-/Dateiformate
- Protokolle/Schnittstellen
- Lizenzen

[Pronom](#)

IANA

HTTP

REST

OAI-PMH

Z39.50

Creative Commons

Open Data Commons

Public Domain

Persistente Identifikation

- URL/URI

<http://zuse.zib.de/collection/wl3XoEDHO8v0ImCa/item/VqNgKUacPrIhqPKu>

- URN

urn:nbn:de:bsz:25-opus-14124

- Handle-System

hdl:11858/00-001M-0000-0019-D20F-6

- DOI

doi:10.1088/0004-637X/715/2/1453

- Researcher-ID

G-2512-2011

- ORCID

0000-0002-2661-8242

Forschungsdatenrepositorien I

Institutionell

Fachlich

Allgemein

[Open Data LMU](#)

[PsychData](#)

[DANS Easy](#)

[Dryad](#)

[Edinburgh DataShare](#)

[Pangaea](#)

[Zenodo](#)

[Edmond](#)

[Radar](#)

Forschungsdatenrepositorien II

Aufgaben

- Daten aufnehmen
- Langzeitarchivierung
- Daten auffindbar machen
- Daten exportieren
- Geschäftsmodell

Repositorien finden

- [Re3data](#)

Datenjournale

- Lebenswissenschaften
 - [Gigascience](#)
 - [Biodiversity Data Journal](#)
 - [Open Health Data](#)
- Geowissenschaften
 - [Earth System Science Data](#)
- Naturwissenschaften
 - [Scientific Data](#)

Rechtsgrundlagen

- Grundgesetz
- Urheberrecht
- Datenbankrecht
- Archivrecht
- Persönlichkeitsrechte
- Strafrecht
- Kriegswaffenkontrollgesetz
- Jugendschutzgesetz
- Kooperationsverbot

- Lizenzen für Forschungsdaten
- Einwilligungserklärungen
- Pre-Registration

Nationale Forschungsdateninfrastruktur

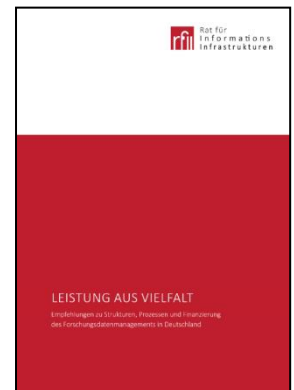
- „Der RfII schlägt daher die Gründung eines Verbundes vor, der die vorhandenen Kompetenzen bündelt und für eine Grundversorgung mit Speicherinfrastrukturen und Services sowie einen schnellen Kompetenztransfer im Wissenschaftssystem sorgt. Diese *Nationale Forschungsdateninfrastruktur* (NFDI) sollte die Form eines Netzwerkes haben, disziplinen- bzw. communityübergreifend angelegt sein und sowohl die existierenden großen Informationseinrichtungen als auch die nationale Ebene der ESFRI-Projekte und die Repositorien weiterer, in ihren Bedarfen hinreichend homogener Nutzergruppen einbinden.“
Rat für Informationsinfrastrukturen (2016)

- GWK



Nationale Forschungsdateninfrastruktur

- „Als funktionaler Verbund hat die NFDI folgende nationale, das Gesamtsystem betreffende Herausforderungen anzugehen:
 - Die Definition von übergreifenden Mindeststandards und von Qualitätsmanagement bei der Datenbeschreibung und -haltung;
 - die Entwicklung von generischen Datenanalyseverfahren und ihre Übertragung auf spezifische Daten und Fragestellungen;
 - Entwicklung, Aufbau und Bereitstellung von generischen Datenservices und Datenspeichern sowie Schnittstellen zur Anbindung verteilter lokaler Repositorien;
 - Aus-, Fort- und Weiterbildung.“ Rat für Informationsinfrastrukturen (2016)
- GWK



European Open Science Cloud

- Europäische Kommission

OPPORTUNITIES

EUROPEAN OPEN SCIENCE CLOUD
BRINGING TOGETHER CURRENT AND FUTURE DATA INFRASTRUCTURES

en and seamless
ces to analyse and
se research data

and technology

future
es,
industry and SMEs

Bring benefits to citizens

reduce the cost of big data storage
and high-performance analysis

€2 BN IN OVERALL HORIZON 2020 FUNDING TO THE EUROPEAN CLOUD INITIATIVE, WITH ESTIMATED ADDITIONAL PUBLIC AND PRIVATE INVESTMENT OF €4.7 BN REQUIRED TO FURTHER DEVELOP THE EUROPEAN DATA INFRASTRUCTURE.

0.2bn

Widening the user base to the public and private sectors



1bn

EU-wide Quantum technologies flagship

3.5 bn

Data Infrastructure

#DigitalSingleMarket

EUROPEAN CLOUD INITIATIVE
UNLOCKING THE POWER OF BIG DATA FOR OPEN SCIENCE

OPPORTUNITIES

Spurring new solutions in core areas such as research, transport, environment

Better public services such as diagnostics

Better science for complex problems

Commercial opportunities for innovative companies

Better value for taxpayers - cutting up data produced by projects funded by the Horizon 2020 research and innovation programme. (Protein, Healthcare, Transport data, Research)

EUROPEAN OPEN SCIENCE CLOUD
BRINGING TOGETHER CURRENT AND FUTURE DATA INFRASTRUCTURES

A trusted open environment for sharing scientific data

Open and seamless services to analyse and reuse research data

Linking data

Connecting across borders and scientific disciplines

Connecting scientists globally

Improving science

Long term and sustainable

EUROPEAN DATA INFRASTRUCTURE
UNLOCKING THE VALUE OF BIG DATA, DIGITAL BY DEFAULT

Facilitate access to and re-use of data for researchers, innovators and public sector

Work in collaboration with national and regional scientific and public data and computing centres

Reduce the cost of big data storage and high-performance analysis

1 HIGH PERFORMANCE COMPUTING (HPC)

Provide access to high performance computing resources to support the needs of big data

2 HPC COMPUTING CORE AND SOFTWARE INFRASTRUCTURE

Storage, dissemination, access and management of big data

European big data centre for high-impact cloud services

3 CONNECTIVITY

Interconnect high-speed networks and secure connectivity and big data storage for EU-wide HPC access

Integration of European public services networks

WIDENING ACCESS; BUILDING TRUST

Widening the user base to the public and private sectors

A safe, secure environment with privacy and data protection built on trusted data and intellectual property protected by highest standards of quality, reliability and confidentiality

WHO IS IT FOR?

1.7 million researchers

70 million professionals in science and technology

Open up in the future to public services, industry and SMEs

Bringing benefits to citizens

€2 BN IN OVERALL HORIZON 2020 FUNDING TO THE EUROPEAN CLOUD INITIATIVE, WITH ESTIMATED ADDITIONAL PUBLIC AND PRIVATE INVESTMENT OF €4.7 BN REQUIRED TO FURTHER DEVELOP THE EUROPEAN DATA INFRASTRUCTURE

0.2bn

Widening the user base to the public and private sectors



1bn

EU-wide Quantum technologies flagship

3.5 bn

Data Infrastructure

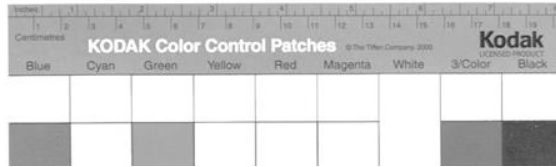
Struktur der Veranstaltung

Teil 1: Forschungsdaten

Teil 2: Forschungsdateninfrastruktur

Teil 3: Forschungsdatenmanagement in Bibliotheken

Teil 4: Weiterführende Überlegungen



Open Science: Bibliotheken spielen eine zentrale Rolle - sie wissen es nur nicht



106. Bibliothekartag

Medien — Menschen — Märkte

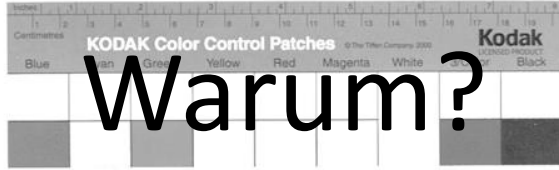
30.05.-02.06.2017 | Frankfurt am Main

Mai 2017

Klaus Tochtermann

ZBW Kiel/Hamburg





- Neue Technologien
- Neue Anforderungen
- Neue kommerzielle Dienste
- Wahrnehmung
- → Krise der Bibliotheken?

Deutsche Zentralbibliothek für Medizin - Leibniz-
Informationszentrum Lebenswissenschaften, Köln und
Bonn (ZB MED)

17.03.16: Stellungnahme des Senats einschließl. Anlagen
Download (PDF)

Demobündnis zur Rettung der ZB MED in Köln und Bonn
Leibniz Informationszentrum Lebenswissenschaften
eigen. Deutsche Zentralbibliothek für Medizin

Demos: 24.05.2016 - Ab 09:00 Uhr
vor der Gemeinsamen Wissenschaftskonferenz
Friedrich-Ebert-Allee 38, 53113 Bonn
Demo während der entscheidenden Sitzung!

ZB MED #KeepZBMED
JUSO UNIVERSITÄT ZU KÖLN
ASIA Universität zu Köln
campus:grün
PIRATEN HSG BONN

24.05.2016
KEEP ZB MED! - Demo zur Rettung während der GWK-Sitzung!

Öffentlich - Gastgeber: KEEP ZB MED

★ Interessiert + Zusagen ✉ Einladen

Dienstag, 24. Mai um 09:00 (UTC+02)
Nächste Woche: 12-22? Teilweise bewillt

Friedrich-Ebert-Allee 38, 53113 Bonn, Deutschland Karte anzeigen

GÄSTE		
53	15	564
interessiert	Zusagen	Eingeladen

FREUNDE EINLADEN

+ Freunde zu dieser Veranstaltung hinzufügen

Info Diskussion

Details

DEMOAUFRUF

Die Leibniz-Gemeinschaft hat der GWK die Einstellung der Finanzierung der ZB MED empfohlen, was einer Schließung der Deutschen Zentralbibliothek für Medizin gleichkommt. Von der ZB MED profitieren nicht nur die Medizin und Zahnmedizin in Köln, Bonn und ganz Deutschland, sondern Informations- und Bibliothekswissenschaftler sowie alle Life Sciences: Agrar-, Umwelt- & Ernährungsbiotechnologie, Pharmazie u.v.m.

Wir halten diese Entscheidung für grundlegend falsch und gehen den Studierenden und Forschenden die ZB MED als zentrale Informations- und Publikationszentrale verloren. Wir fordern die Leibniz-Gemeinschaft auf, die Finanzierung der ZB MED zu gewährleisten und die Schließung zu verhindern.

Mehr anzeigen

Bibliotheken
Weg mit den Büchern!

NZZ am Sonntag von Michael Furger · 7.2.2016, 01:00 Uhr

Das Internet mache Bibliotheken überflüssig, sagt der Chef der ETH-Bibliothek im Interview. Entweder sie räumen ihre Bücherbestände aus und erfinden sich neu – oder sie werden verschwinden.

f t w i n e 48 KOMMENTARE

MEISTGELESEN

Alexander von Plato: **Seltene Flugmanöver in den letzten Minuten**
Beitrag bewertet: 20.0.2016

Osterliche starke Rechtspopulisten: **Verführerische Scheinwelt**
Beitrag bewertet: 16.0.2016

Alexander von Plato: **MERKA Luftfahrtbehörde bestätigt Rauch an Bord**
vor 2 Stunden

NZZ am Sonntag: **Herr Ball, brauchen wir heute noch Bibliotheken?**

Rafael Ball: Nein, in ihrer bestmöglichen Form nicht. Bibliotheken machen sie bisher nichts anderes, als für die Menschen Inhalte zu sammeln. Dieses Konzept funktioniert heute nicht mehr, jetzt ist das Internet da. Wer Inhalte sucht, braucht keine Bibliothek mehr.

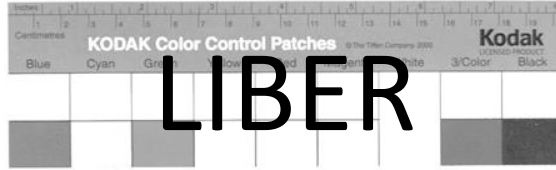
Das Internet macht Bibliotheken überflüssig?

Ja, mit der Vollbildung sind die öffentlichen Bibliotheken gekommen, mit dem Internet gehen sie wieder. Ist das ein Problem?

Vielleicht, weil Bibliotheken eben Kultur-Auftrag haben!

unicef

Allergien sind auf dem Vormarsch



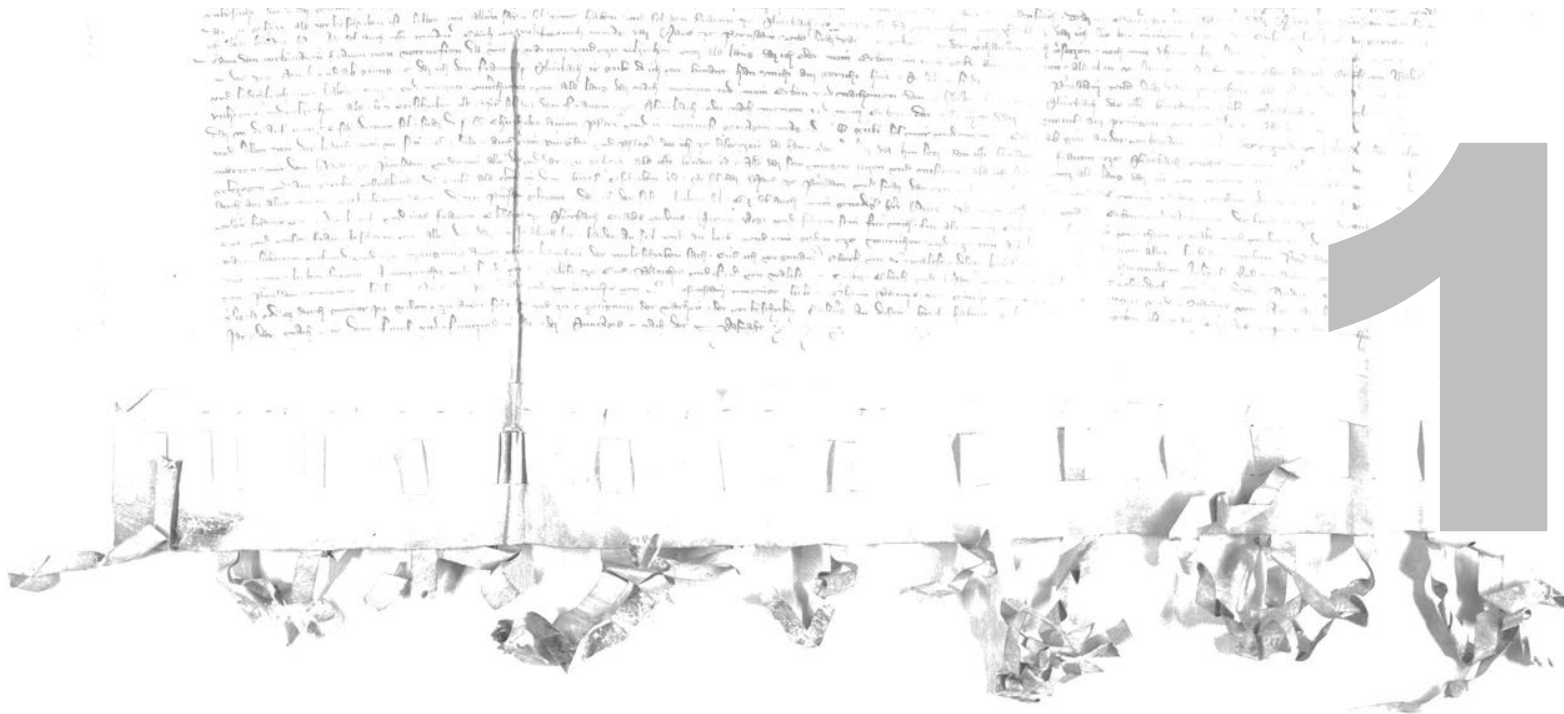
- [Ligue des bibliothèques européennes de recherche](#)
- 400 Forschungsbibliotheken
- 51 deutsche Bibliotheken
- München: UB TU, BSB, UB UniBwM



Ten recommendations for libraries to get started with research data management

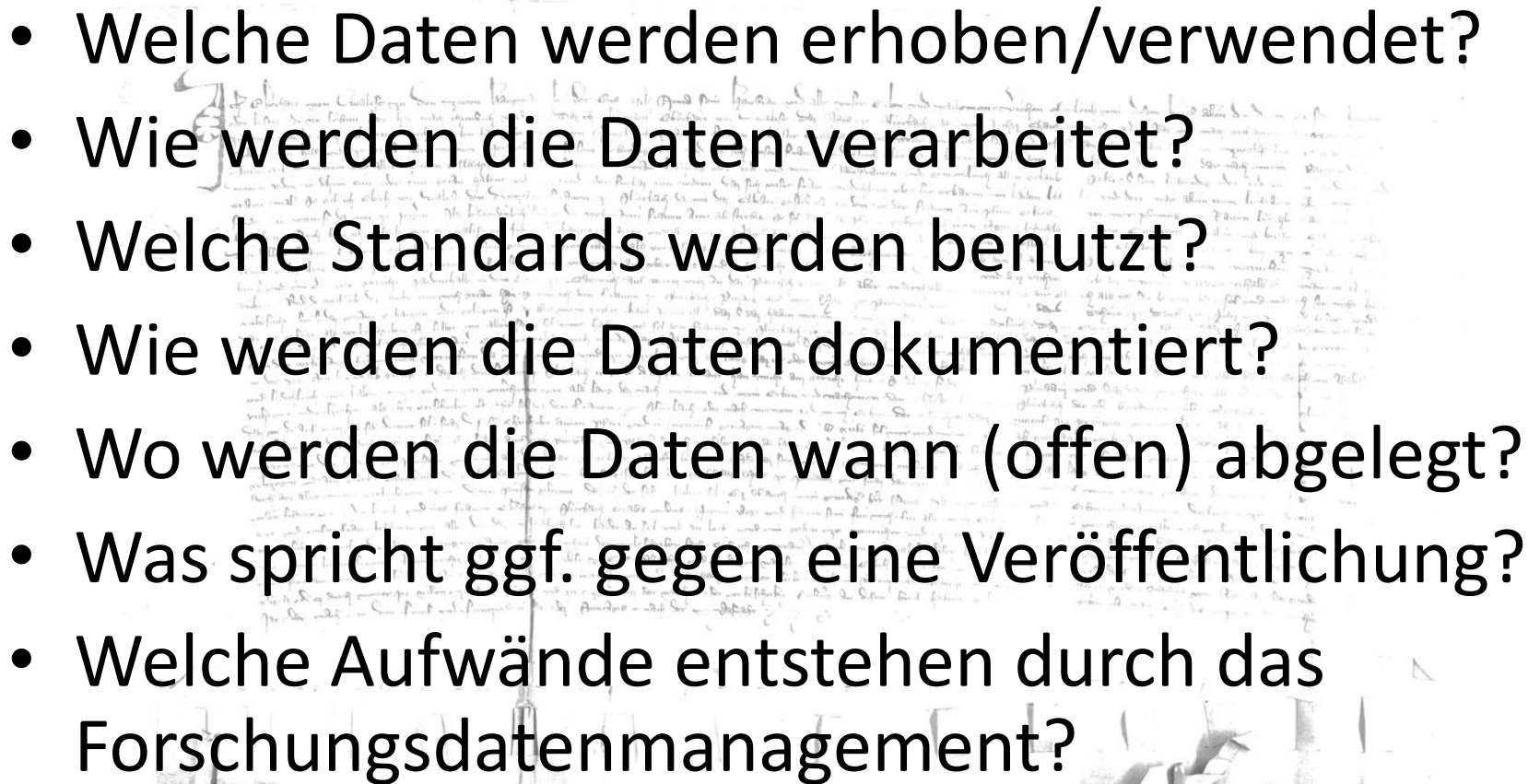
[Final report](#) of the LIBER working group on E-Science / Research Data Management, 2012

Offer research data management support, including data management plans for grant applications, intellectual property rights advice and information materials. Assist faculty with data management plans and the **integration of data management into the curriculum.**



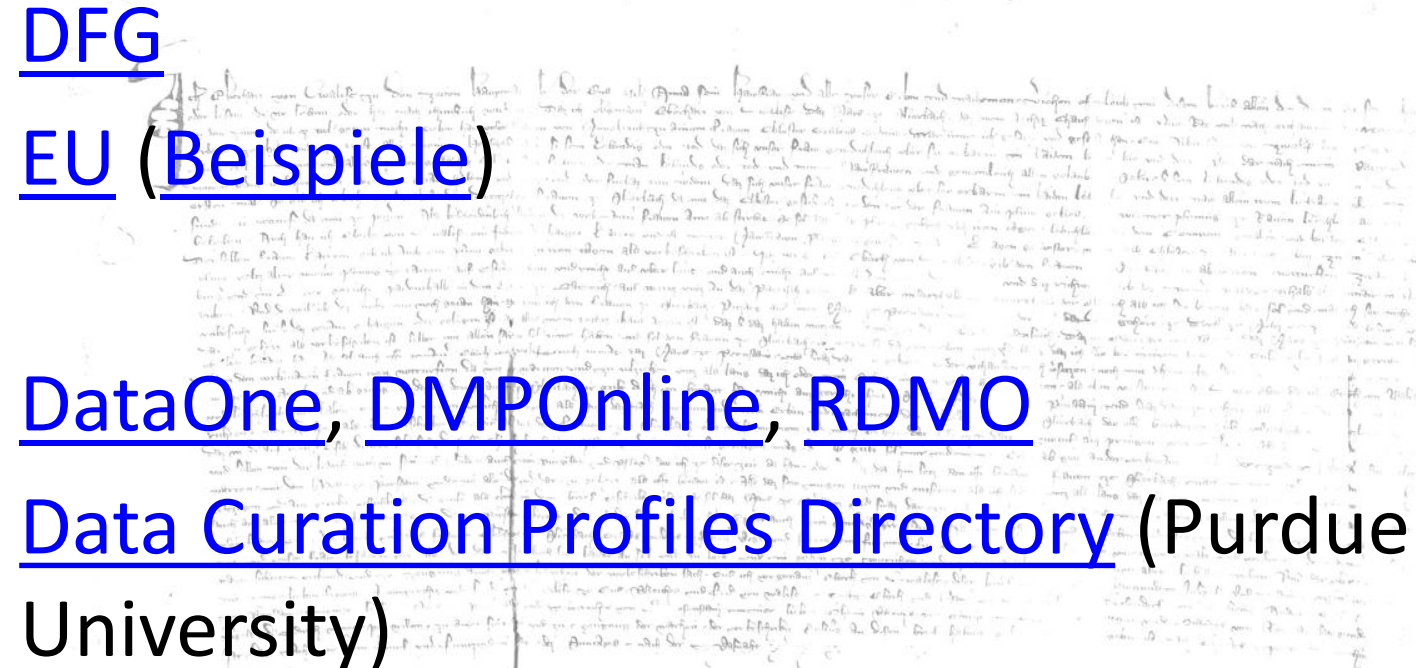


Datenmanagementpläne I

- Welche Daten werden erhoben/verwendet?
 - Wie werden die Daten verarbeitet?
 - Welche Standards werden benutzt?
 - Wie werden die Daten dokumentiert?
 - Wo werden die Daten wann (offen) abgelegt?
 - Was spricht ggf. gegen eine Veröffentlichung?
 - Welche Aufwände entstehen durch das Forschungsdatenmanagement?
- 

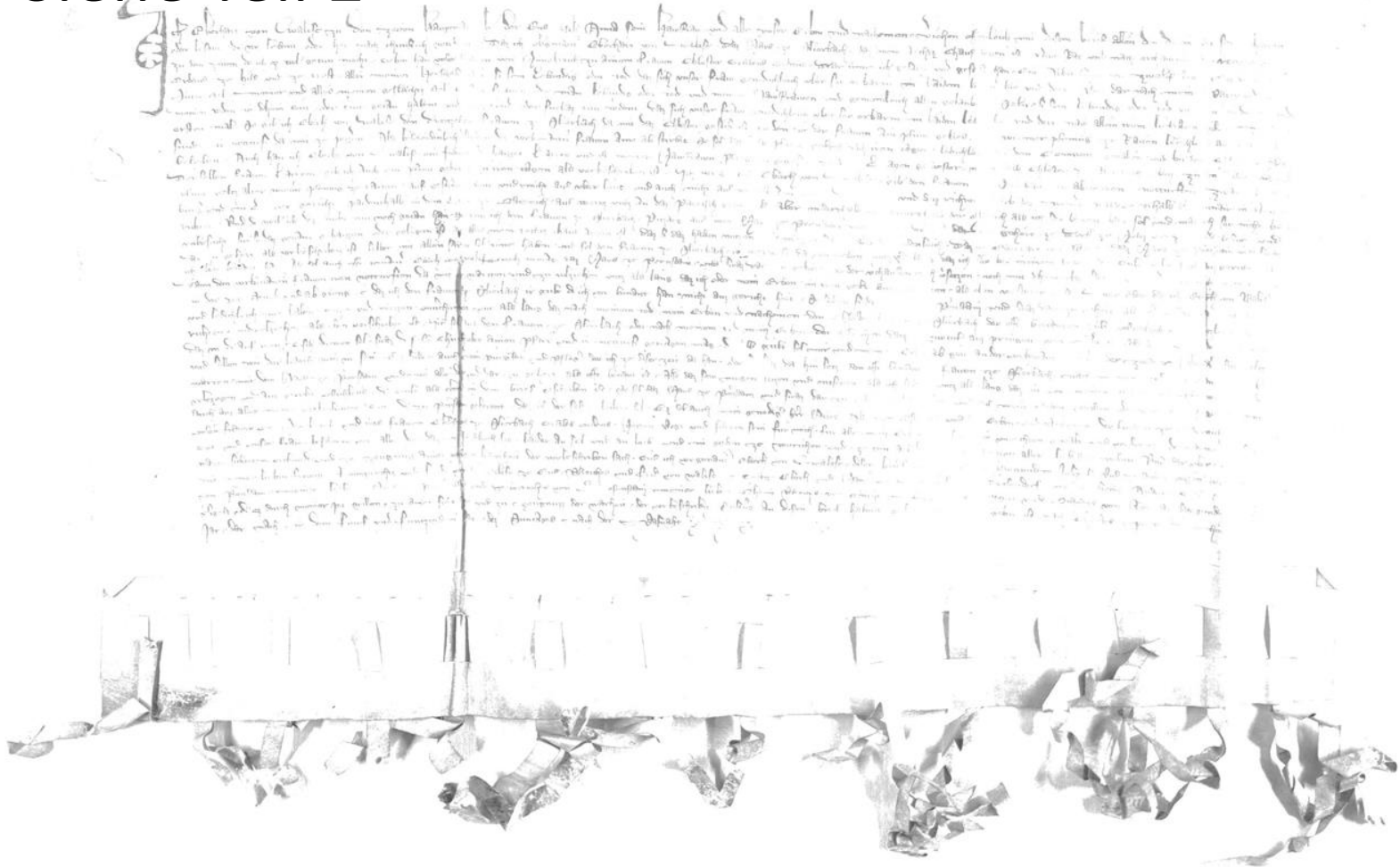


Datenmanagementpläne II

- [DFG](#)
 - [EU \(Beispiele\)](#)
 - [DataOne](#), [DMPOnline](#), [RDMO](#)
 - [Data Curation Profiles Directory](#) (Purdue University)
- 

Lizenzen für Forschungsdaten

- Siehe Teil 2



Kurse für Forschungsdatenmanagement

- Uni Bielefeld, Uni Hannover, Uni Regensburg,
Uni Heidelberg, FU Berlin, ...
- HföD
- TUM
- ~~LMU~~



Engage in the development of metadata and data standards and provide **metadata services** for research data.

- Beispiel [RDA](#), [DC](#)
- Beispiel [DDI](#)
- **Metadaten-Services:**
 - Transformationen
 - Kontrollierte Vokabulare

2

Create Data Librarian posts and develop professional staff skills for **data librarianship**.

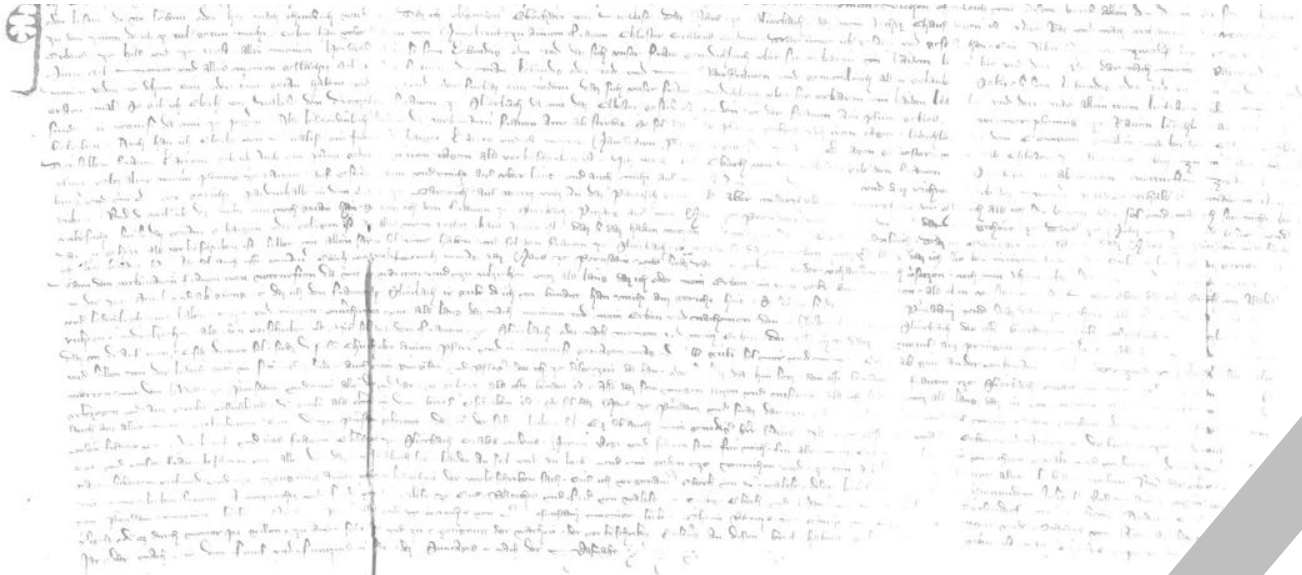
[Google](#)

Hapke, H. (2016) Data Librarian: Das moderne Berufsbild; [b.i.t. online](#)

3



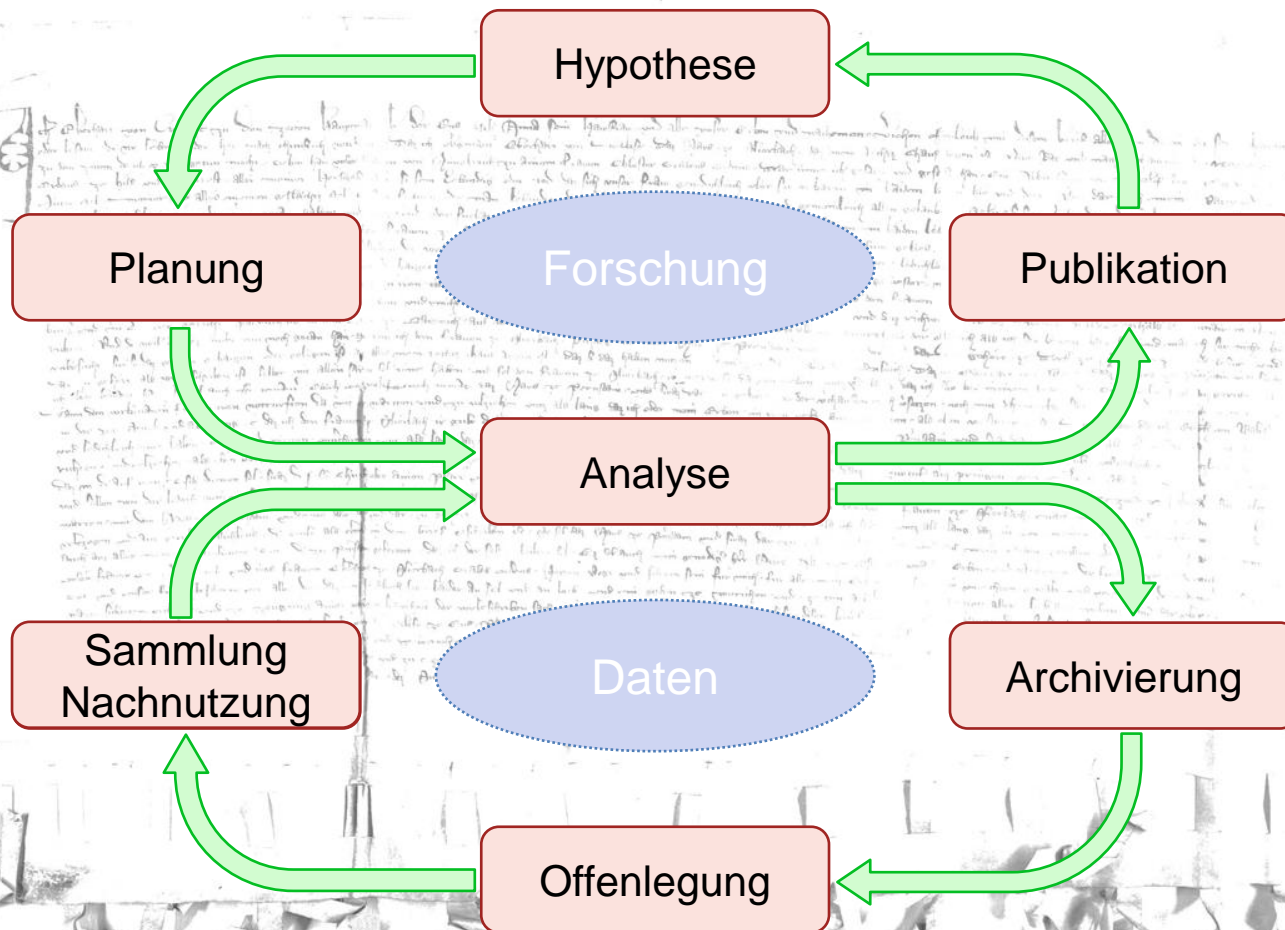
Actively participate in institutional research data policy development, including resource plans. Encourage and adopt open data policies where appropriate in the **research data life cycle**.



4



Forschungs- und Datenzyklus





(Offene) Fragen

- Who owns the data?
- What Requirements are Imposed By Others?
- Which Data Should Be Retained?
- For How Long Should Data Be Maintained?
- How Should Digital Data Be Preserved?
- Are there Ethical Considerations?
- How are Data Accessed?
- How Open Should the Data Be?
- How Will Costs Be Managed?
- What are the Alternatives to Local Data Management?

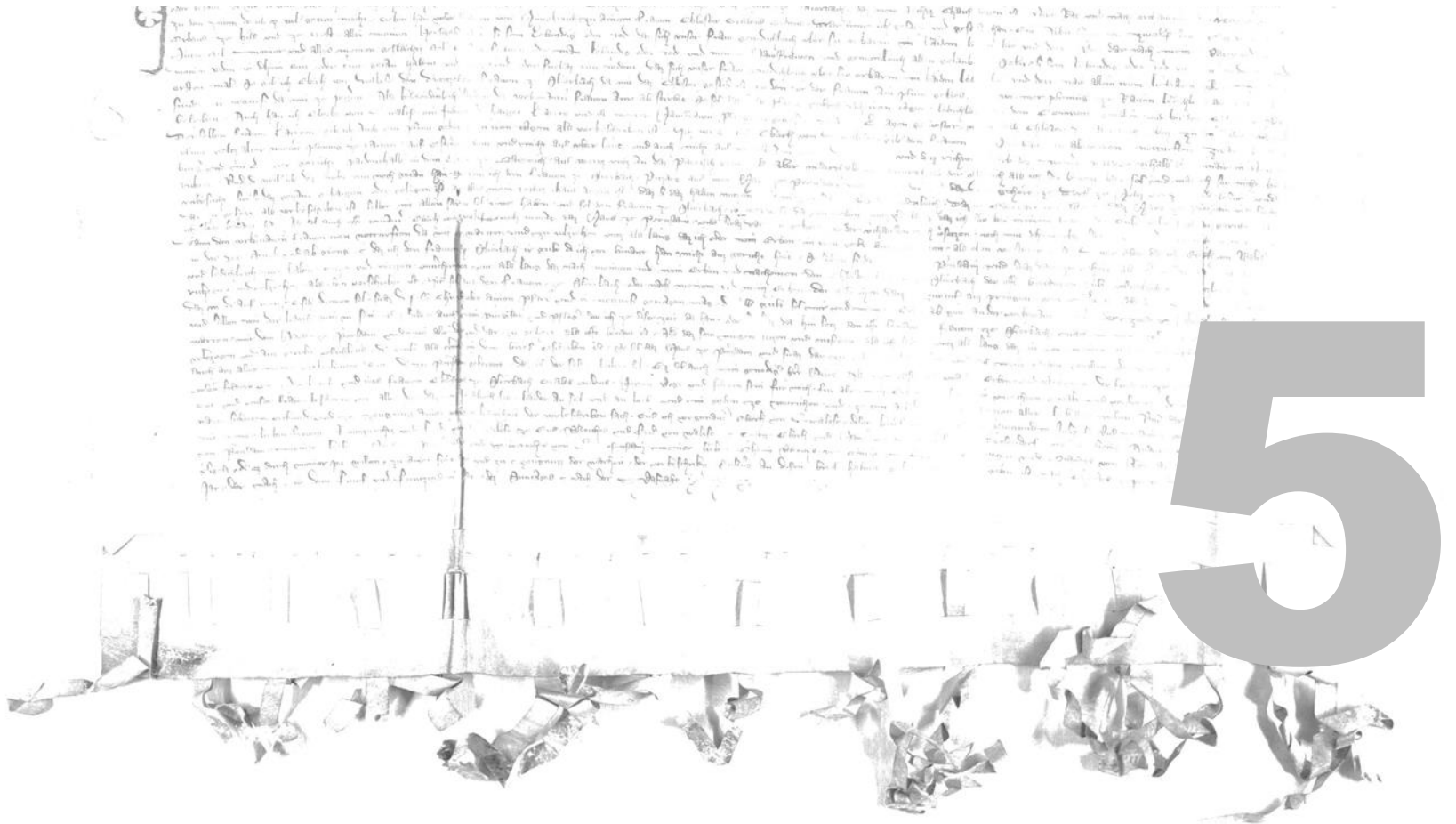


Forschungsdaten-Policies

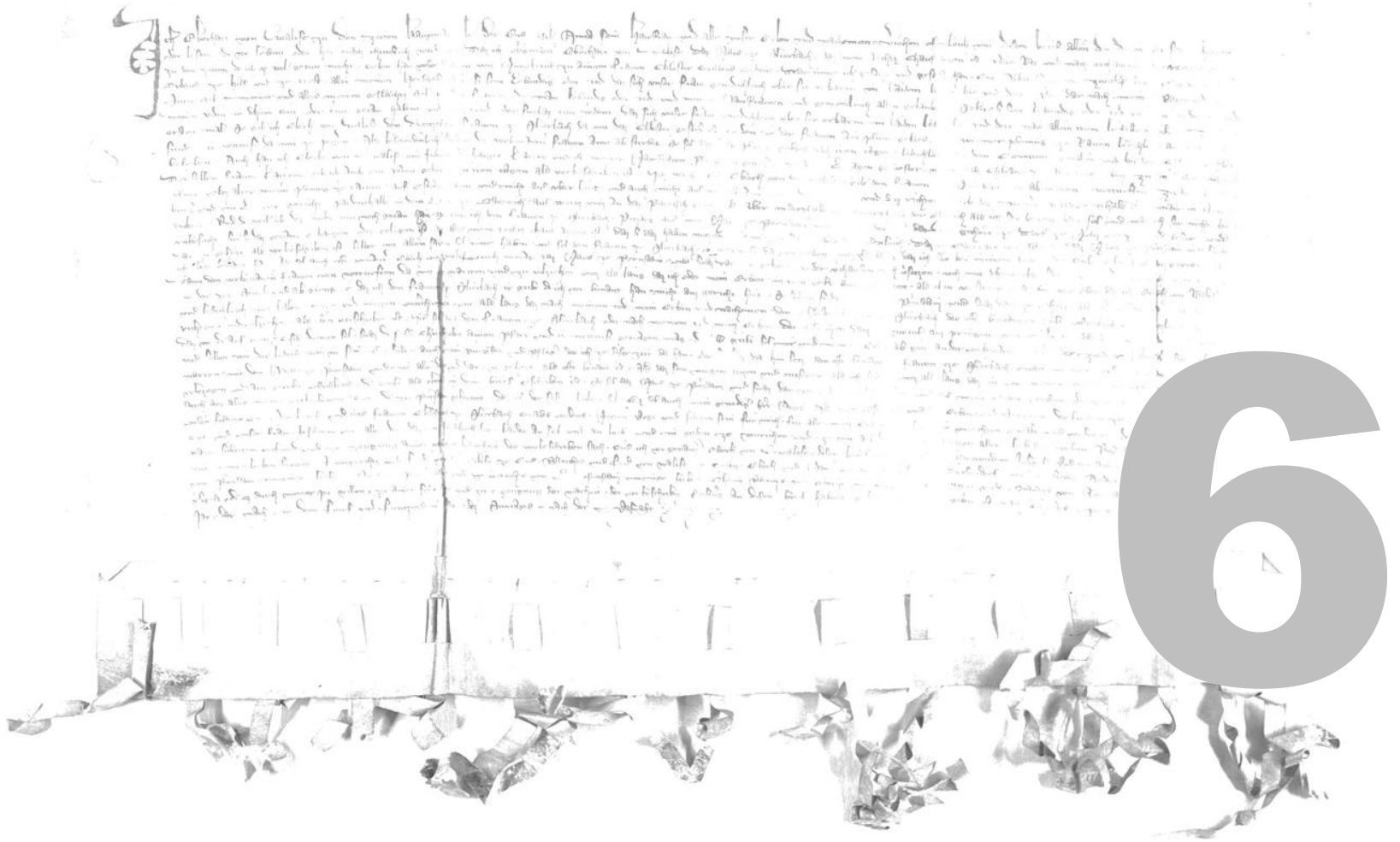
- [Universität Bielefeld](#) (2012)
- [HU Berlin](#) (2014)
- [Universität Kiel](#) (2015)
- [Universität Bayreuth](#) (2016)
- [Universität Erlangen-Nürnberg](#) (2016)



Liase and partner with researchers, research groups, data archives and data centers to foster an interoperable infrastructure for data access, discovery and data sharing.



Support the lifecycle for research data by providing services for storage, discovery and permanent access.

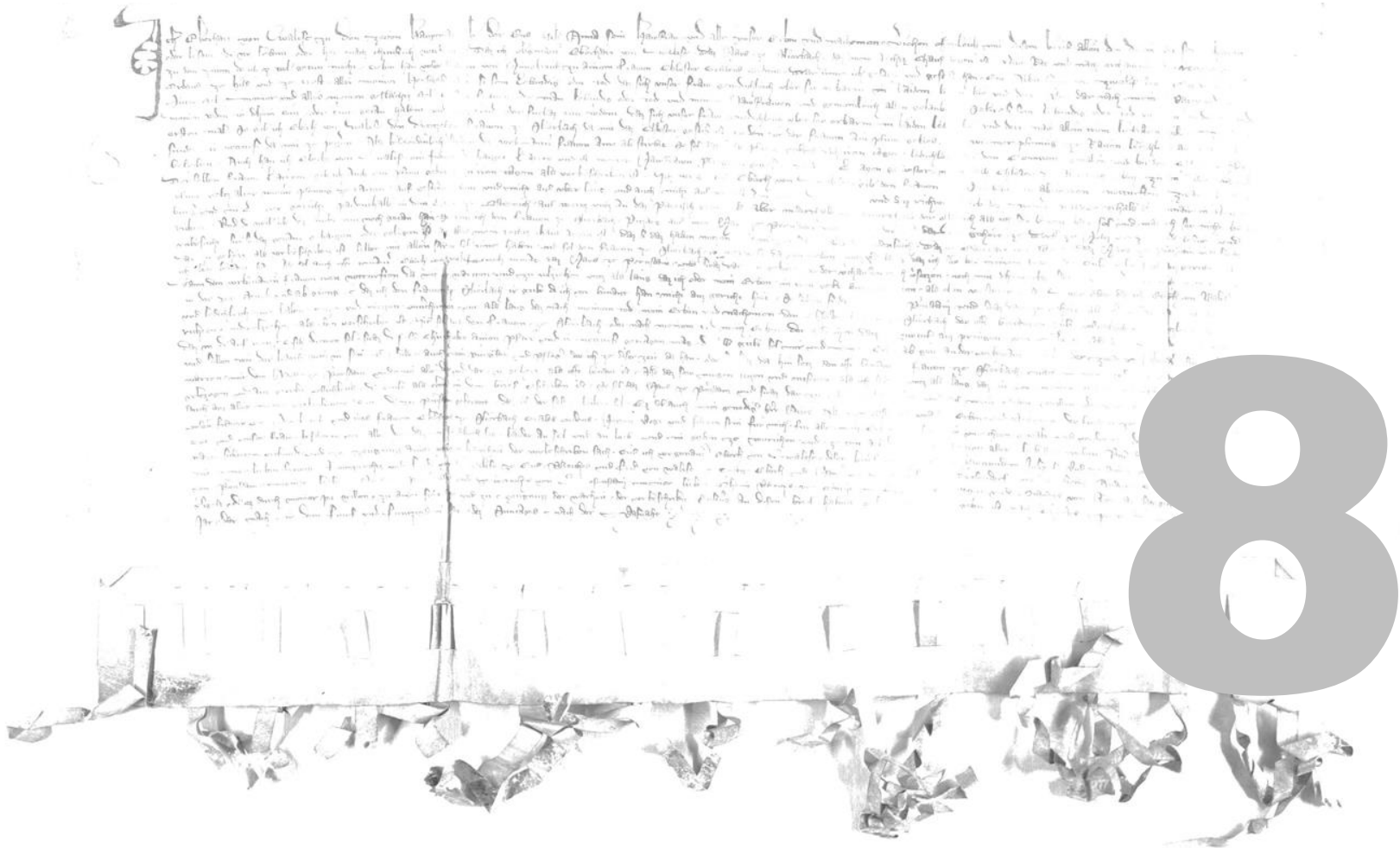


Promote research data citation by **applying persistent identifiers to research data.**

- URNs → [DNB](#) – Deutsche Nationalbibliothek
- Handles → [GWDG](#) – Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen
- DOIs
 - [TIB](#) – Technische Informationsbibliothek Hannover
 - [DaRa](#) – Registrierungsagentur für Sozial- und Wirtschaftsdaten



Provide an institutional Data Catalogue or Data Repository, depending on available infrastructure.



Get involved in subject specific data management practice.

The image shows a page of handwritten text in a historical script, likely a manuscript or ledger. The text is dense and fills most of the page. A large, decorative initial 'S' is visible at the top left. The handwriting is in a cursive style, characteristic of historical documents. The paper appears aged and slightly yellowed.

A large, semi-transparent grey number '9' is overlaid on the bottom right corner of the page. The number is bold and has a slight shadow, making it stand out against the background of the handwritten text.



Offer or mediate secure storage for dynamic and static research data in co-operation with institutional IT units and/or seek exploitation of **appropriate cloud services.**



Struktur der Veranstaltung

Teil 1: Forschungsdaten

Teil 2: Forschungsdateninfrastruktur

Teil 3: Forschungsdatenmanagement in Bibliotheken

Teil 4: Weiterführende Überlegungen

Wissenschaftler

- Erheben Daten
- Verwenden Daten
- Nutzen Daten nach
- Archivieren Daten?
- Beschreiben Daten?
- Veröffentlichen Daten?

There's no API	We will get too many enquiries	I don't mind, but someone else might	Thieves will use it
People may misinterpret the data	It's too complicated	Data Protection	Terrorists will use it
It's not very interesting	We might want to use it in a paper	It's too big	We'll get spam
Lawyers want a custom License	What if we want to sell it later	There's already a project to...	Poor Quality

"Open Data Excuse" Bingo, JISC

➔ Motivation?

Spieltheorie

Exkurs: Gefangenendilemma

„Wenn wir dichthalten kommen wir mit zwei Jahren davon“

„Wenn Sie auspacken und ihr Komplize schweigt, machen wir Sie zum Kronzeugen und nach einem Jahr sind Sie draußen“

„Wenn Sie leugnen und ihr Komplize packt aus, dann wandern Sie für vier Jahre ein“

„Wenn Sie und ihr Komplize reden, dann erhalten Sie ihre gerechte Strafe von drei Jahren“

Spieltheorie

Exkurs: Gefangenendilemma

		B	
		C _{ooperation}	D _{efection}
A	C	-2 -2	-4 -1
	D	-1 -4	-3 -3

Spieltheorie

Exkurs: Gefangenendilemma

B

		C _{ooperation}	D _{efection}
A	C	-2 -2	-4 -1
	D	-1 -4	-3 -3

Pareto-Optimum

Nash-Gleichgewicht

Wissenschaftler vs. Community

		Community	
		Open Data	Do Not Open Data
Wissenschaftler	O	4 4	-1 1
	N	5 3	0 0

Wissenschaftler vs. Community

		Community	
		Open Data	Do Not Open Data
Wissenschaftler	O	4 4	-1 1
	N	5 3	0 0

Wissenschaftler vs. Community

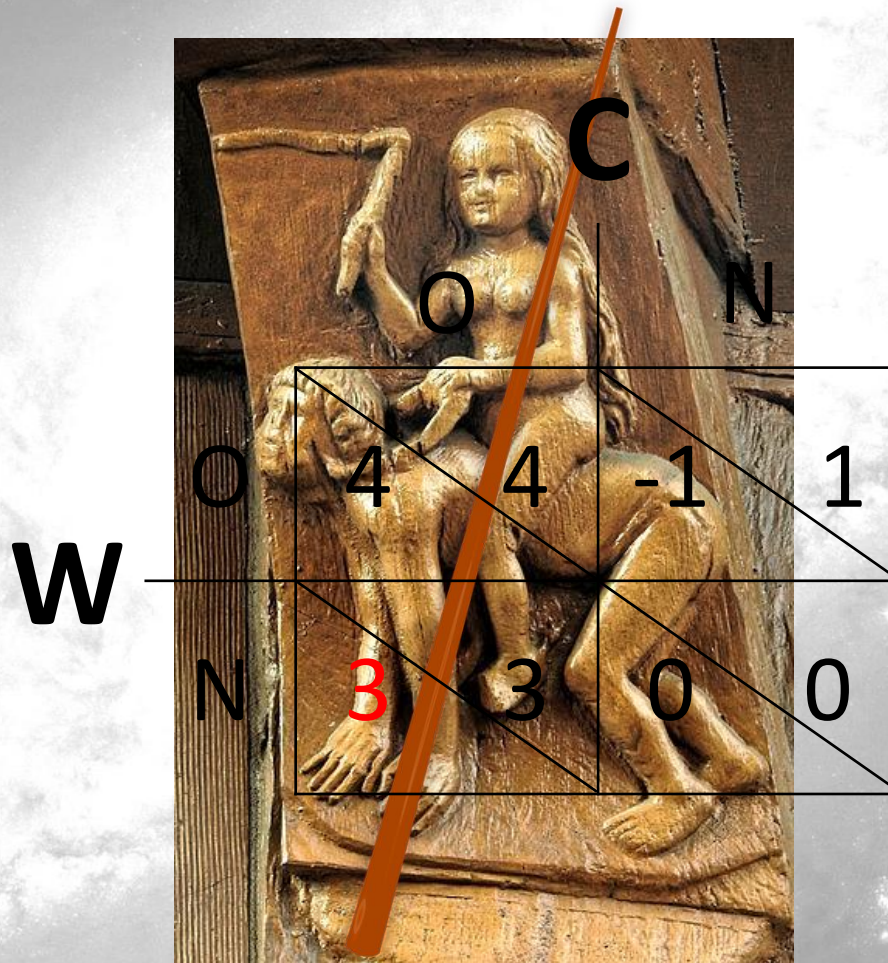
		Community	
		Open Data	Do Not Open Data
Wissenschaftler	O	4 4	-1 1
	N	5 3 →	0 0

The table is a 2x2 matrix representing a game between a Scientist (Wissenschaftler) and a Community. The Scientist's strategies are 'Open Data' (O) and 'Not Open Data' (N). The Community's strategies are 'Open Data' and 'Do Not Open Data'. The payoffs are shown as pairs of numbers (Scientist, Community). The cell (N, Open Data) is highlighted in light blue, and a red arrow points to the value 3.

Wissenschaftler vs. Community

		Community	
		Open Data	Do Not Open Data
Wissenschaftler	O	4 4	-1 1
	N	5 3	0 0

Möglichkeit 1: Sanktionen



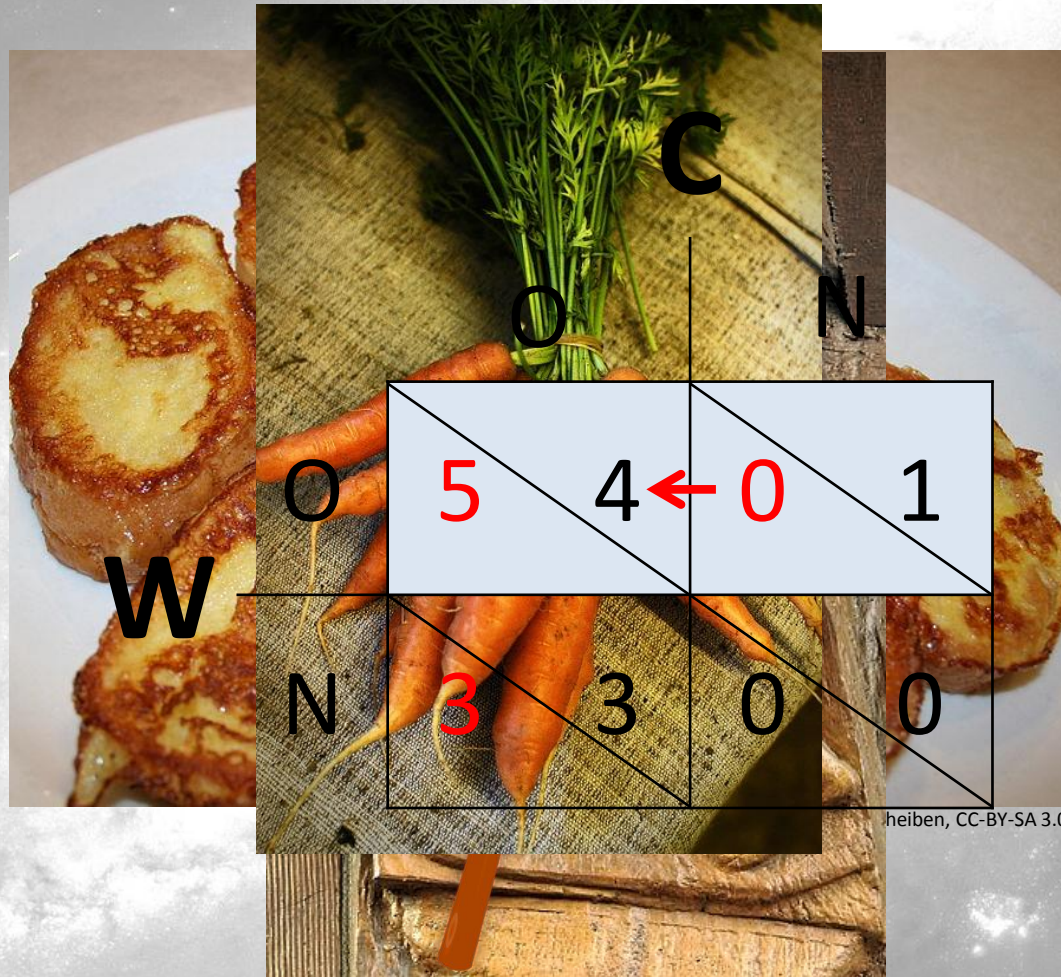
Holger Uwe Schmitt: Der Butterhanne-Comic, CC-BY-SA

Möglichkeit 2: Belohnung



Nettescheiben, CC-BY-SA 3.0

Möglichkeit 3: Kombination



heiben, CC-BY-SA 3.0

Holger Uwe Schmitt: Der Butterhane-Comic, CC-BY-SA

Open Science

- Transparency in experimental methodology, observation, and collection of data.
- Public availability and reusability of scientific data.
- Public accessibility and transparency of scientific communication.
- *Using web-based tools to facilitate scientific collaboration.*

([What, exactly, is Open Science?](#), Dan Gezelter, 2009)

Open Science

The conduction of science in a way that others can collaborate and contribute, where research data, lab notes and other research processes are freely available, with terms that allow reuse, redistribution and reproduction of the research.

([FOSTER](#))

Open Science

meint

- den öffentlichen Zugang zu Forschungsergebnissen in Form von Publikationen (**Open Access**)
- die freie Verfügbarkeit von Forschungsdaten/Rohdaten (**Open Research Data**)
- die Zugänglichkeit und Vernetzung von öffentlichen Datenbeständen (**Linked Open Data**)
- die Suche nach Alternativen zu den traditionellen Begutachtungsverfahren der Zeitschriften-Verlage, um eine höhere Transparenz in diesem Bereich zu gewährleisten (**Open Review**)
- die Entwicklung und den Einsatz quelloffener Software für die Wissenschaft (**Open Source**)
- die Einbeziehung nicht-professioneller Wissenschaftler in den wissenschaftlichen Prozess (**Citizen Science**)
- freien Zugang zu digitalen Lehrmaterial und (Aufzeichnungen von) Lehrveranstaltungen (**Open Education**)

Science 2.0

- Facebook, ResearchGate
 - Twitter
 - Altmetrics
 - Dropbox
 - Blogs
 - Wikis
- ➔ meistens kommerzielle Anbieter, kein soziales Dilemma

Big Data/Data-Driven Research

- [Definition](#) (Hasso Plattner)

- Induktiver Forschungsansatz



Kausalitätsproblem



Statistisches Problem

- Tipp: [Hans Rosling „The Joy of Stats“](#)

Open ~~Research~~ Data

- [Open Data Portal München](#)
- [Open Data Portal der Bundesregierung](#)
- [Open Data Portal der EU](#)
- <http://opendataenterprise.org/map/reports/May2016Report.pdf>

Forschungsdaten-Discovery

- [Research Data Australia](#)
- [Jisc Research Data Discovery Service \(Großbritannien\)](#)
- [Portage Network \(Kanada\)](#)
- [CESSDA \(Europa\)](#)
- Data Citation Index (Clarivate)

Archive

- Langzeitperspektive
- Kulturelles Erbe
- [DDB](#) – Deutsche Digitale Bibliothek
- [Europeana](#)
- [WDL](#) – World Digital Library

Wissenschaftliche Software

- Repositorien ?
- Metadaten ?
- Langzeitarchivierung ?
- Persistente Identifikation ?
- Standards ?

