

DV-Systembetreuung in Bibliotheken

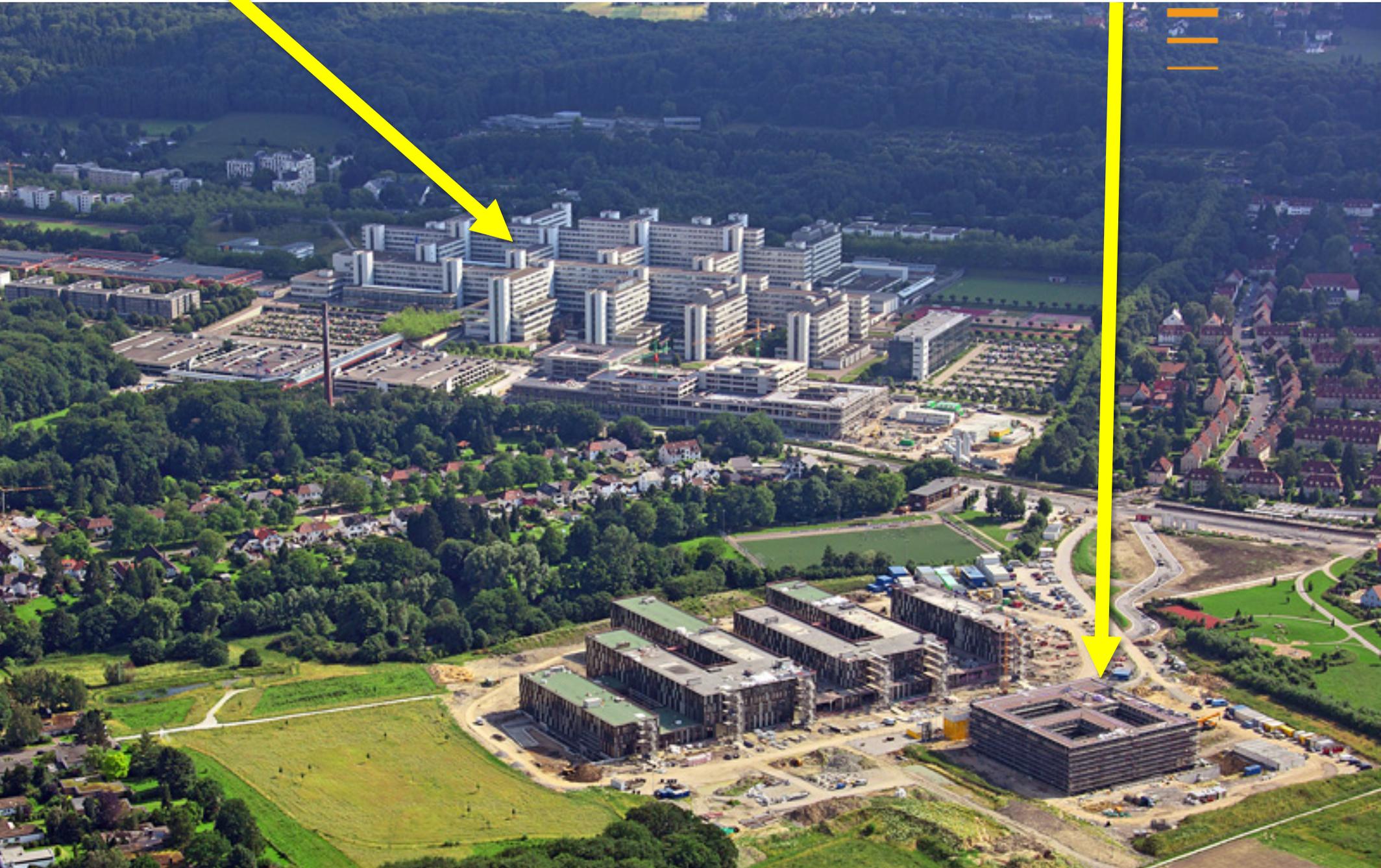
Einführung in das Forschungsdatenmanagement

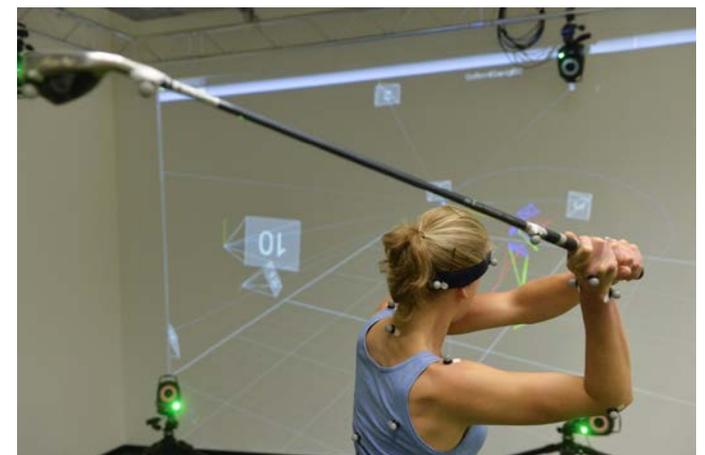
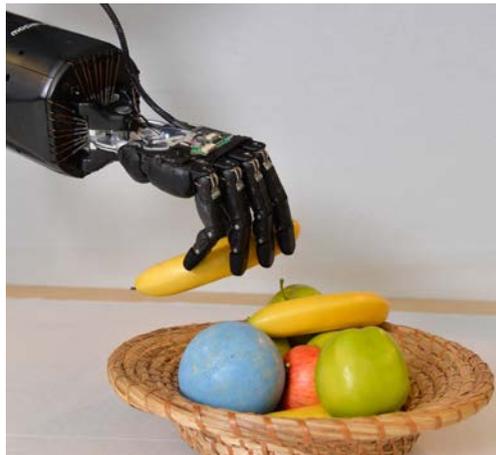
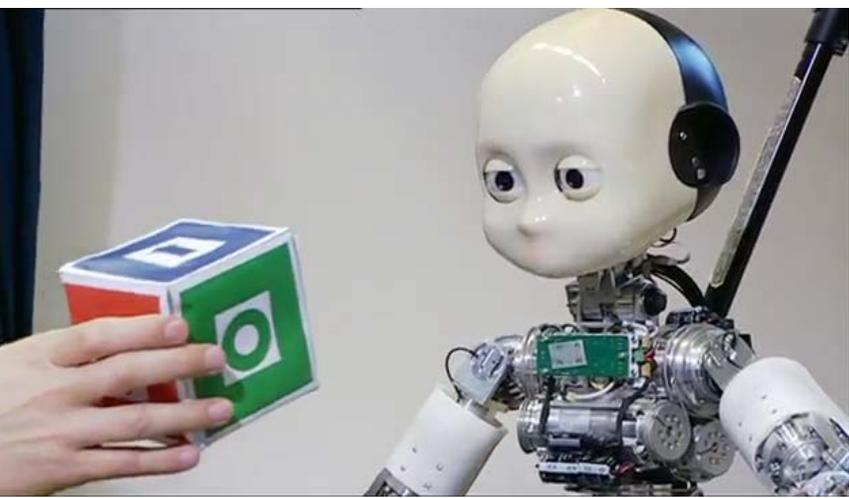
Do, 16.05.2018

Cord Wiljes
CITEC
Universität Bielefeld

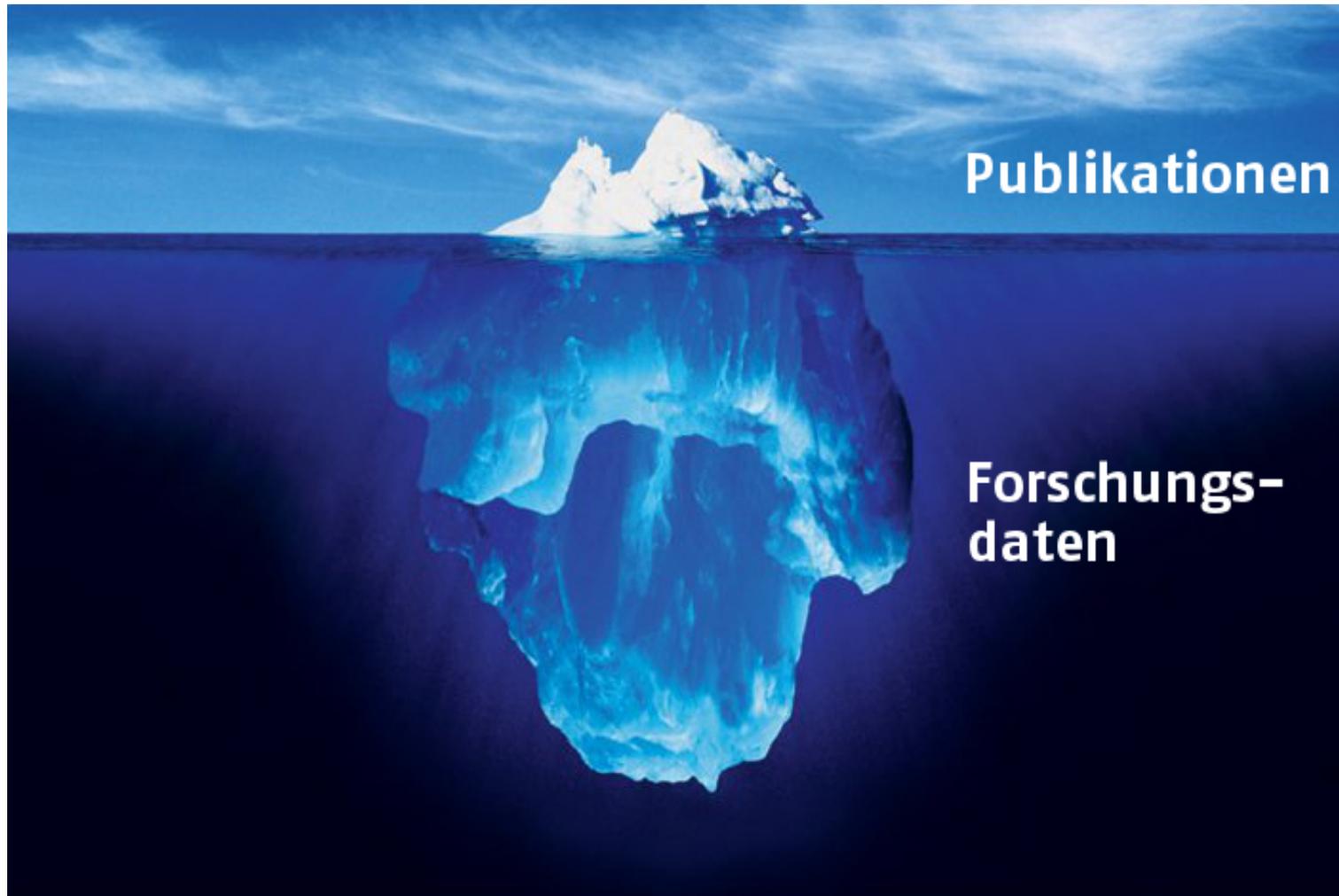
Universität Bielefeld

CITEC





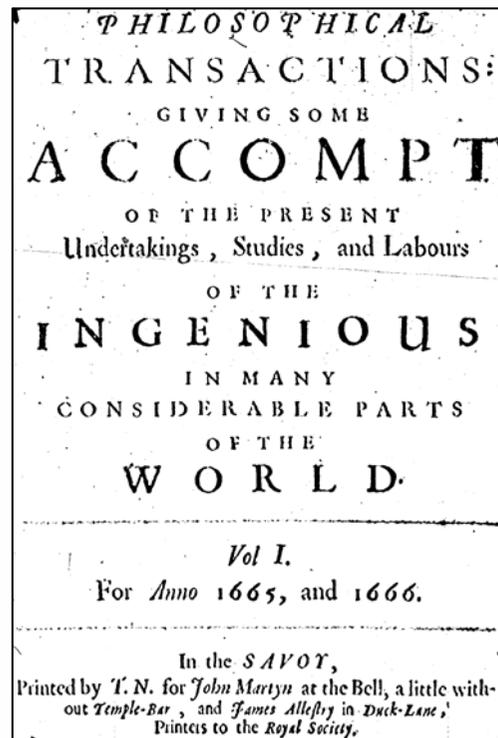
Die Spitze des Eisbergs



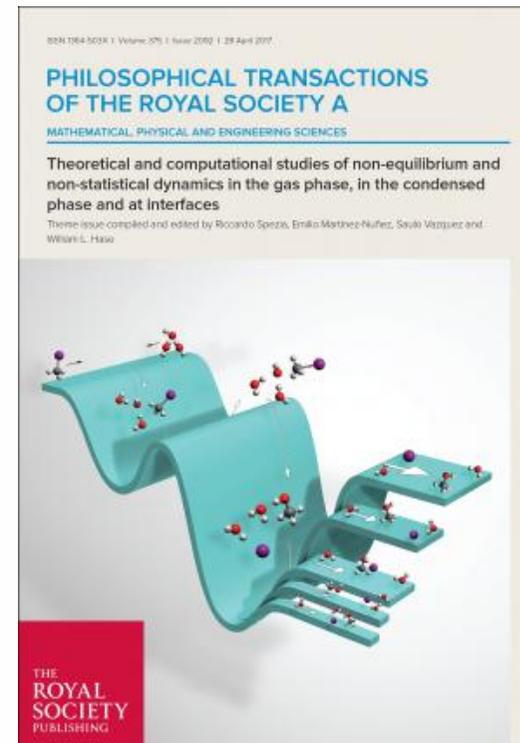
Publikationen

“Philosophical Transactions of the Royal Society”

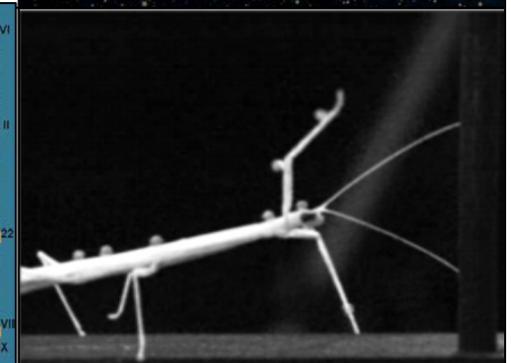
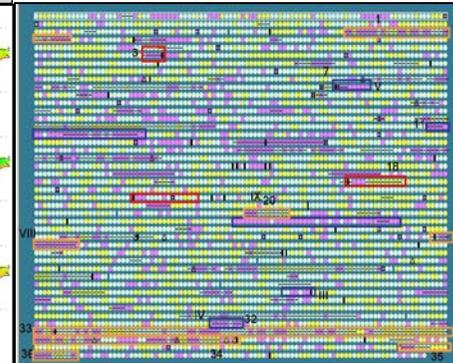
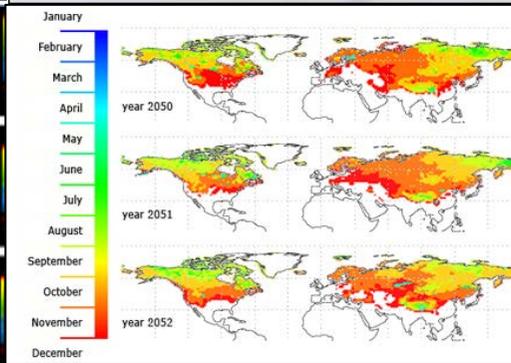
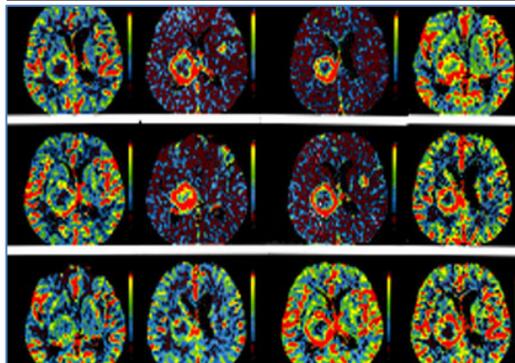
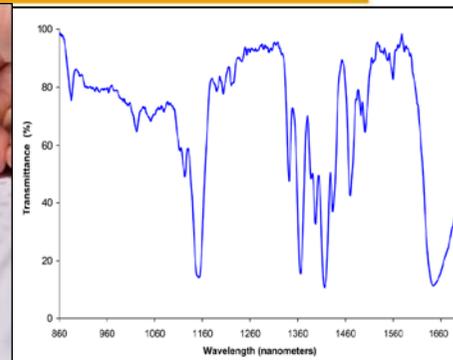
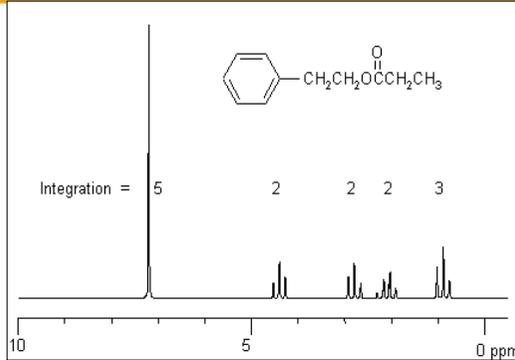
6. März 1665



12. Mai 2018

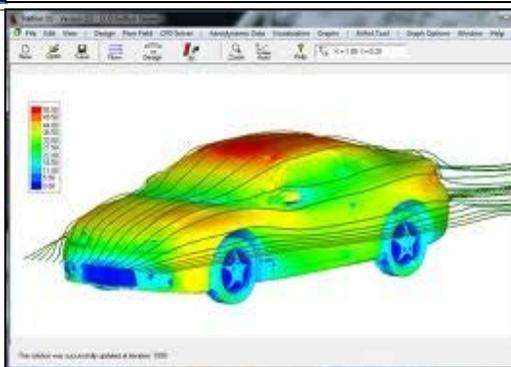


Forschungsdaten

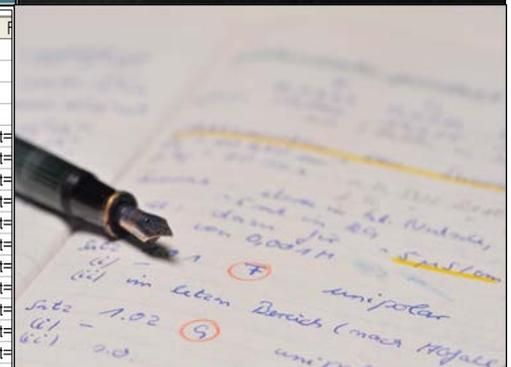


```

00221 Public Function GetChecksum(ByVal sentence As String) As String
00222 Dim Character As Char
00223 Dim Checksum As Integer
00224 For Each Character In sentence
00225     Select Case Character
00226     Case " $"
00227         *Ignore the dollar sign
00228     Case "*"
00229         *Stop processing before the asterisk
00230     Exit For
00231     Case Else
00232         *Is this the first value for the checksum?
00233         If Checksum = 0 Then
00234             *Yes. Set the checksum to the value
00235             Checksum = Convert.ToByte(Character)
00236         Else
00237             Checksum = Checksum Xor Convert.ToByte(Character)
00238         End If
00239     End Select
00240 Next
00241 Return Checksum.ToString("X2")
00242 End Function
    
```



	A	B	C	D	E	F
1						
2						
3	Q,1/A	SAXS	esd	Qdiff		
4	1/A	a.u.	a.u.			
5	0.022756	1107.6	8.586		<ldata><Q unit=	
6	0.023296	1038.9	7.6445	0.00054	<ldata><Q unit=	
7	0.023837	1071	7.919	0.000541	<ldata><Q unit=	
8	0.024377	1054.7	8.0684	0.00054	<ldata><Q unit=	
9	0.024917	1061.3	8.2971	0.00054	<ldata><Q unit=	
10	0.025457	1115.1	8.3305	0.00054	<ldata><Q unit=	
11	0.025998	1276.1	8.5378	0.000541	<ldata><Q unit=	
12	0.026538	1499.2	9.0048	0.00054	<ldata><Q unit=	
13	0.027078	1738.2	10.172	0.00054	<ldata><Q unit=	
14	0.027619	1802.5	10.335	0.000541	<ldata><Q unit=	
15	0.02816	1728.5	10.12	0.000541	<ldata><Q unit=	





Wie soll ich meine Daten dokumentieren?

Wie muss ich mit personenbezogenen Daten umgehen?

Wie kann ich mich vor Datenverlust schützen?

Wie kann ich meine Daten publizieren?

Wie stelle ich das Datenmanagement im Förderantrag dar?

Wie kann ich die Daten archivieren?

„Ökonomische Krisen durch Staatsverschuldung“

Ökonomische Krisen entstehen, wenn die Staatsverschuldung 90% des jährlichen Bruttoinlandsproduktes übersteigt



Kenneth Rogoff und Carmen Reinhart

Ökonomische Krisen durch Staatsverschuldung

~~Ökonomische Krisen entstehen, wenn die Staatsverschuldung 90% des jährlichen Bruttoinlandsproduktes übersteigt~~



Thomas Herndon

Reproduzierbarkeit wissenschaftlicher Forschung

IS THERE A REPRODUCIBILITY CRISIS?



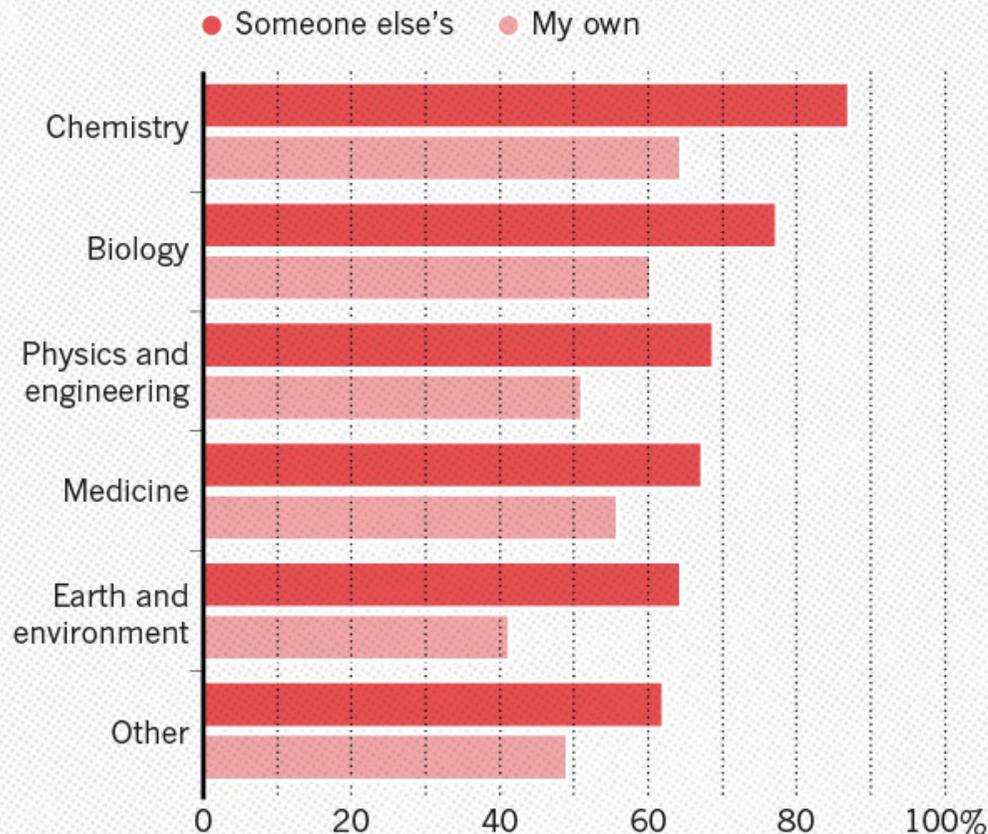
©nature

Nature 533, 452–454
(26 May 2016)

Reproduzierbarkeit wissenschaftlicher Forschung (2)

HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?

Most scientists have experienced failure to reproduce results.



Gute Wissenschaftliche Praxis

DFG: „**Sicherung guter wissenschaftlicher Praxis**“ (1998):

Empfehlung 7: Sicherung und Aufbewahrung von Primärdaten

Primärdaten als Grundlagen für Veröffentlichungen sollen auf haltbaren und gesicherten Trägern in der Institution, wo sie entstanden sind, zehn Jahre lang aufbewahrt werden.

→ <http://doi.org/10.1002/9783527679188.oth1>

Forschungsdatenmanagement in Deutschland

- ▶ 1998/2013 Deutsche Forschungsgemeinschaft (DFG): [Sicherung guter wissenschaftlicher Praxis](#)
- ▶ 2009 Deutsche Forschungsgemeinschaft: [Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten](#)
- ▶ 2010 Allianz der deutschen Wissenschaftsorganisationen [Grundsätze zum Umgang mit Forschungsdaten](#)
- ▶ 2014 Hochschulrektorenkonferenz (HRK) [Management von Forschungsdaten - eine zentrale strategische Herausforderung für Hochschulleitungen](#)
- ▶ 2015 Deutsche Forschungsgemeinschaft [Leitlinien zum Umgang mit Forschungsdaten](#)

Institutionelle Forschungsdaten-Policies

08.03.2013	CITEC, Universität Bielefeld
01.11.2013	Universität Bielefeld
01.07.2014	Georg-August-Universität Göttingen
08.07.2014	Humboldt-Universität zu Berlin
18.07.2014	Universität Heidelberg
21.09.2015	Robert Koch-Institut
14.07.2015	Christian-Albrechts-Universität Kiel
27.08.2015	Bergische Universität Wuppertal
16.12.2015	Technische Universität Darmstadt
08.03.2016	RWTH Aachen
31.03.2016	Deutsches GeoForschungsZentrum GFZ
14.06.2016	Helmholtz-Zentrum Berlin
17.10.2016	Karlsruher Institut für Technologie
08.11.2016	Universität Bayreuth
16.11.2016	Friedrich-Alexander-Universität Erlangen-Nürnberg
13.01.2017	Universität Kassel
30.03.2017	Universität Siegen
26.04.2017	Leibniz Universität Hannover
13.06.2017	Deutsches Institut für Erwachsenenbildung
14.06.2017	Westfälische Wilhelms-Universität Münster
25.07.2017	Universität Stuttgart

Quelle: http://www.forschungsdaten.org/index.php/Data_Policies#Institutionelle_Policies

CITEC Open Science Manifest

CITEC (Mai 2013):

"CITEC setzt sich nachdrücklich für das Ideal der Offenen Wissenschaft (Open Science) ein, dass „wissenschaftliche Erkenntnisse jeglicher Art offen geteilt werden sollen, so früh wie dies im Forschungsprozess praktikabel ist.“

→ <https://www.cit-ec.de/de/content/open-science-manifesto>

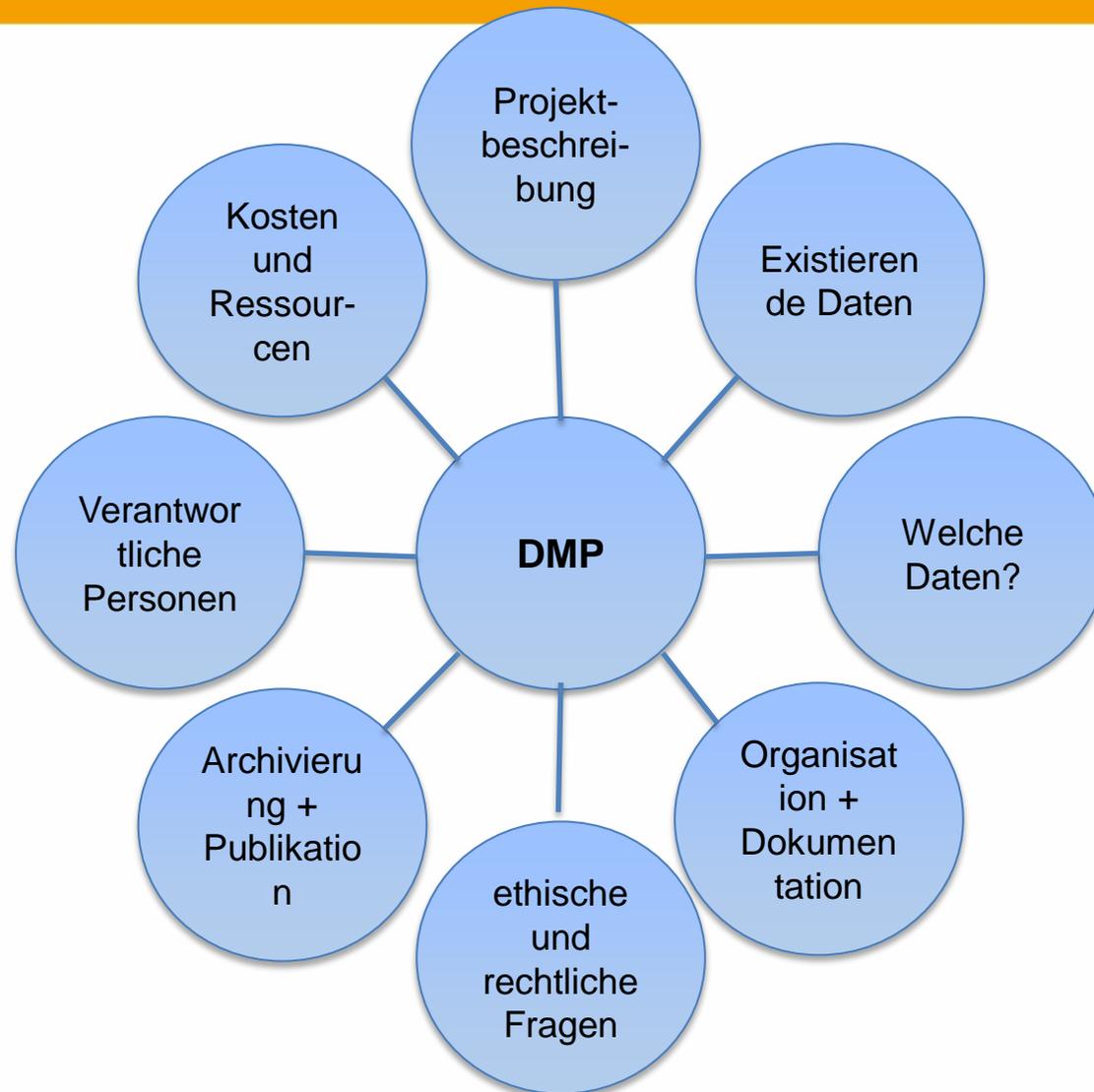
Warum sollten Forschungsdaten veröffentlicht werden?

- ▶ Reproduzierbarkeit von Forschung (Integrität der Forschung)
- ▶ Kostenersparnis (Effizienz der Forschung)
- ▶ “publication bias” vermeiden
- ▶ Daten für andere Forschungsfragen wiederverwenden
- ▶ globale Herausforderungen meistern

Forschungsdaten-Repositoryn

- ▶ Institutionell (Uni BI): <http://pub.uni-bielefeld.de>
- ▶ Catch-All: <http://www.zenodo.de>
- ▶ Disziplinär
- ▶ Liste: <https://www.re3data.org/>

Forschungsdatenmanagement-Pläne



Forschungsdatenmanagement-Services

- ▶ Beratung bei Fragen zu Management, Sicherung, Archivierung und Publikation von Daten
- ▶ Unterstützung bei der Erstellung eines individuellen Datenmanagementplanes
- ▶ Unterstützung beim Verfassen von Datenmanagement-Sektionen in Förderanträgen
- ▶ Verbindung zur Rechtsabteilung bei rechtlichen Fragen herstellen
- ▶ Fortbildungs- und Lehrangebote

Herausforderungen

- ▶ **Vielfältigkeit von Forschung**
- ▶ **Datenqualität**
 - ▶ Verständlichkeit
 - ▶ Interoperabilität / Standards
 - ▶ Nachnutzbarkeit
- ▶ **Datenschutz**
- ▶ **Honorierung + Finanzierung**

Was funktioniert?

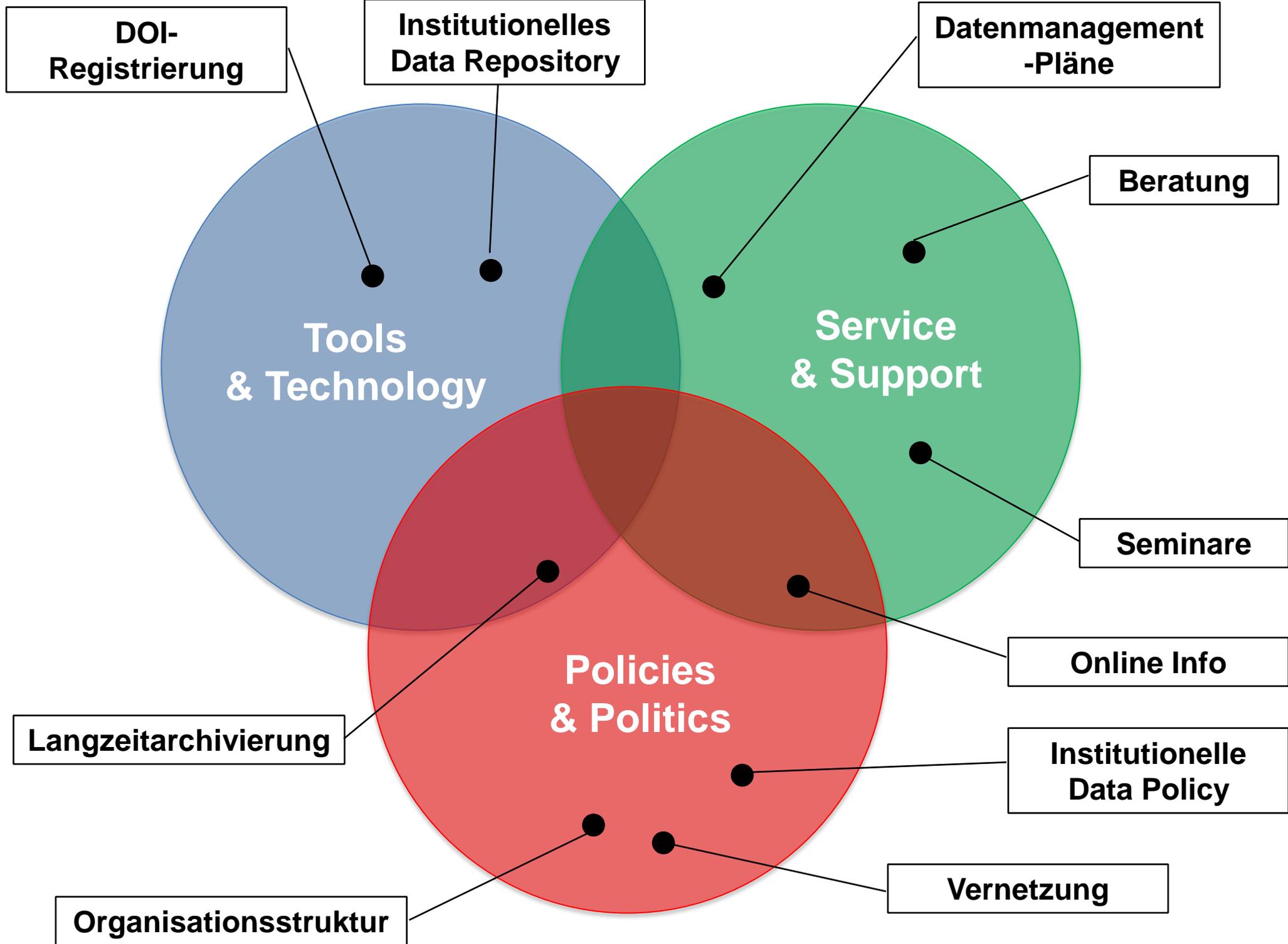
- ▶ Selbstverpflichtung der Institution („Forschungsdaten Policy“)
- ▶ Datenpublikationen mit persistenten Identifiern (DOI)
- ▶ individuelle Beratung
- ▶ Fortbildungsangebote (insbes. Vorträge in Gruppenseminaren)
- ▶ (Forschungsdatenmanagement-Pläne ... wenn sie konsequent geführt werden)

Wandel in der Rolle der Bibliotheken

- ▶ Bibliothekar/in → Informationsmanager/in
- ▶ neue Jobbilder: Datenmanager/in, Datenkurator/in, Data Steward...
- ▶ essenzielles Bindeglied zwischen Technischer Infrastruktur und Forschenden
- ▶ Beratung gewinnt zunehmend an Bedeutung

Vielen Dank!





Forschungsförderung: DFG

„Wenn aus Projektmitteln systematisch (Mess-)Daten erhoben werden, die für die Nachnutzung geeignet sind, legen Sie bitte dar, welche Maßnahmen ergriffen wurden bzw. während der Laufzeit des Projektes getroffen werden, um die Daten nachhaltig zu sichern und ggf. für eine erneute Nutzung bereit zu stellen.“

(DFG, 2010, http://www.dfg.de/download/formulare/1_02/1_02.pdf)

Forschungsförderung: H2020

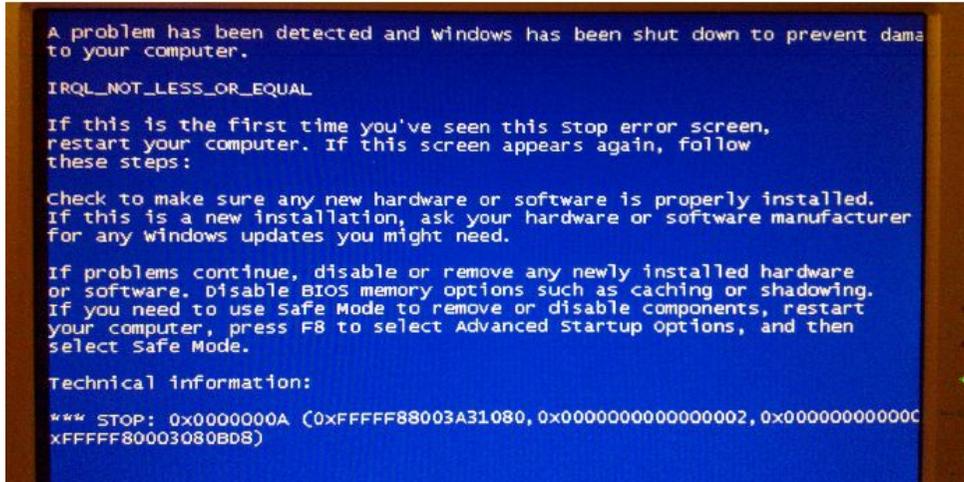
- ▶ Open Access Publikation für alle Calls verpflichtend
- ▶ „Open Data Pilot“ seit Jan 2017 für alle Calls verpflichtend (aber opt-out möglich)

Daten-Policies der wissenschaftlichen Journale

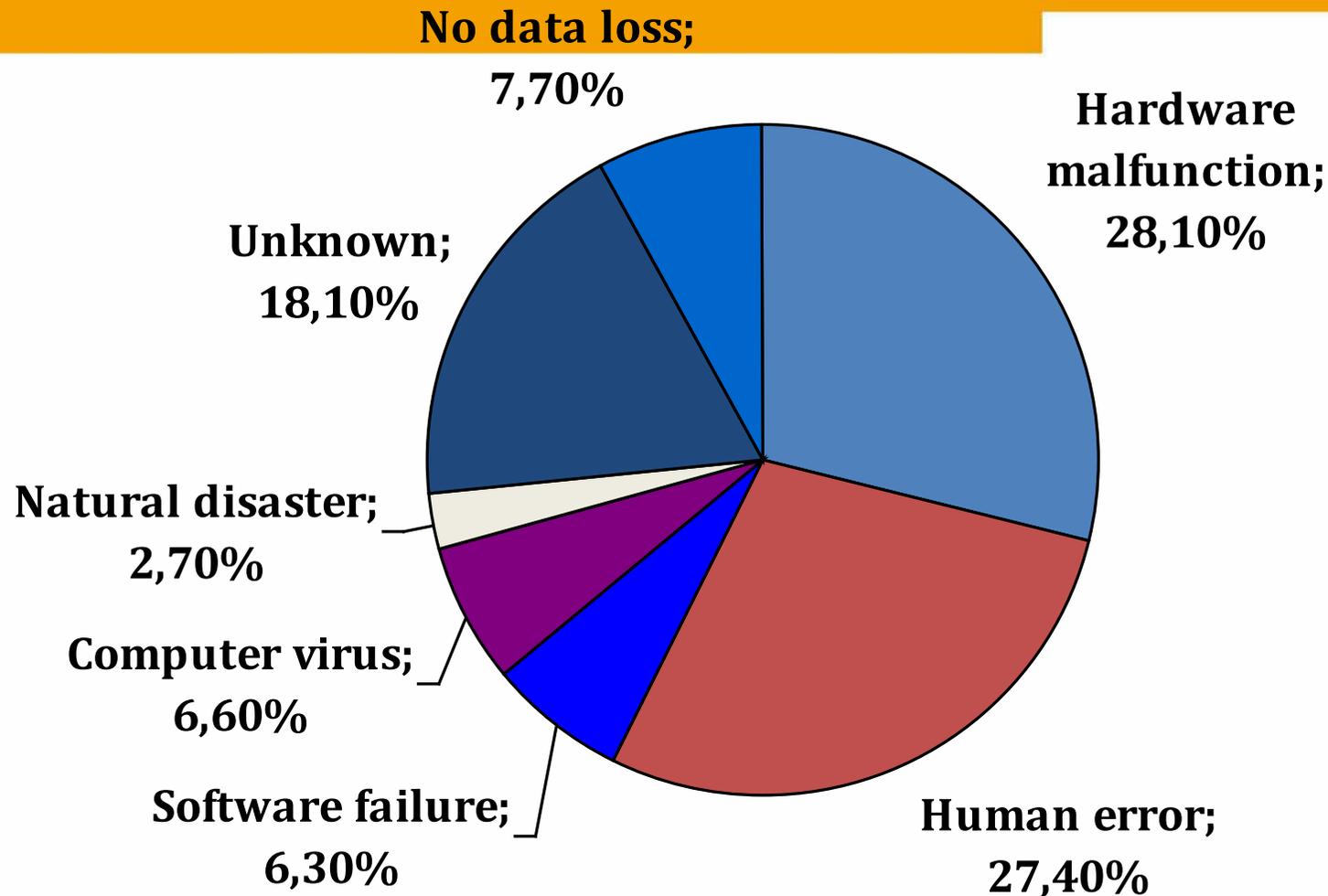
Mehr als 30 wichtige internationale Magazine und Forschungsinstitute haben sich auf einen neuen Codex der Reproduzierbarkeit geeinigt. Zu den Unterzeichnern zählen Fachjournale wie "Science", "Nature", das "British Medical Journal" oder "Cell". Das Ziel: Die Studienqualität zu verbessern und das Vertrauen in die Forschung wiederherstellen.

- ▶ Alle Datensätze, auf denen die Schlussfolgerungen einer Veröffentlichung beruhen, müssen von den Forschern zur Verfügung gestellt werden und auch nach der Publikation noch bereitliegen.

→ <http://www.spiegel.de/wissenschaft/mensch/qualitaet-in-der-forschung-weg-mit-dem-forschungsmuell-a-1001120.html>



Ursachen für Datenverlust



Source: Kroll Ontrack 2010,

→ <http://www.speicherguide.de/backup-recovery/datenrettung/datenverlust-ist-weiterhin-eine-realitaet-12793.aspx>

Erste Mondlandung am 21. Juli 1969

- ▶ Apollo 11 sendete während der Mondlandung einen Live Stream, der aus dem Slow-Scan Television (SSTV) Format konvertiert wurde
- ▶ Kopien der unbearbeiteten Apollo 11 SSTV wurden auf ca. 45 Datenbändern gespeichert
- ▶ Ein Team pensionierter NASA Mitarbeiter versuchte in den frühen 2000er Jahren die Bänder zu lokalisieren ...

... aber sie konnten sie nicht finden!

- ▶ Die Bändern wurden in frühen 1980er Jahren gelöscht und von der NASA wiederverwertet – weil während dieser Zeit ein akuter Mangel an Datenbändern herrschte.
- ▶ Das überlebende Filmmaterial besteht aus qualitativ geringer wertigen Fassungen, die von Fernsehstationen aufgezeichnet worden waren

https://en.wikipedia.org/wiki/Apollo_11_missing_tapes

DFG (2015)

DFG: "Leitlinien zum Umgang mit Forschungsdaten" (30. Sept. 2015)

„Soweit einer Veröffentlichung der Forschungsdaten aus einem DFG-geförderten Projekt Rechte Dritter (insbesondere Datenschutz, Urheberrecht) nicht entgegenstehen, **sollten** Forschungsdaten so zeitnah wie möglich verfügbar gemacht werden.“

http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf

H2020 Open Data Pilot: FAIR Prinzipien

Make research data **f**indable, **a**ccessible, **i**nteroperable and **r**eusable.

([Guidelines on FAIR Data Management in Horizon 2020](#))

DFG (2015)

DFG: "Leitlinien zum Umgang mit Forschungsdaten" (30. Sept. 2015)

"Die langfristige Sicherung und Bereitstellung der Forschungsdaten leistet einen Beitrag zur Nachvollziehbarkeit und Qualität der wissenschaftlichen Arbeit und eröffnet wichtige Anschlussmöglichkeiten für die weitere Forschung."

http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf

H2020

- ▶ Seit 1.1.2017 verpflichtend für alle Programme
- ▶ (begründeter) Opt-Out möglich
- ▶ innerhalb der ersten 6 Projektmonate
- ▶ Update während des Projektes und zum Ende

DFG

- ▶ bislang kein DMP erforderlich
- ▶ Aber „Angaben zum Umgang mit Forschungsdaten“ im Antrag
- ▶ Inhalte: [Leitlinie zum Umgang mit Forschungsdaten](#)

BMBF

- ▶ bislang nur bei einigen Programmen gefordert
- ▶ Ggf. Abgabe mit dem Antrag, keine Aktualisierung notwendig

Wozu ein Forschungsdatenmanagementplan?

- ▶ Anforderungen von Forschungsförderern erfüllen
- ▶ Nachweispflichten nachkommen → Überprüfbarkeit
- ▶ andere Forscher die Daten verstehen lassen → Kooperation
- ▶ erleichtert das Archivieren von Daten
- ▶ Vorbereitung für eine Datenpublikation
- ▶ Erhöhung der Datensicherheit → Schutz gegen Datenverlust
- ▶ Doppelaufwände vermeiden
- ▶ zukünftigen Nutzen der eigenen Arbeit sicherstellen













Aktivitäten

- ▶ Förderprogramme der Forschungsförderer
- ▶ Einrichtung von Forschungsdatenmanagement-Services an den Hochschulen
- ▶ Veranstaltungen der Research Data Alliance (<https://www.rd-alliance.org>)
- ▶ Digitale Hochschule NRW (<https://www.dh-nrw.de>)
- ▶ Aufbau einer Nationalen Forschungsdateninfrastruktur (NFDI)
- ▶ ...