Forschungsdaten in der digitalen Wissenschaft –

Umgang und Herausforderung

Sünje Dallmeier-Tiessen

CERN

Agenda

- Forschungsdaten: Wer, wie, was, wieso, weshalb, warum…?
- Der Forschungskreislauf und Forschungsdaten
- Rahmenbedingung der Bereitstellung von Forschungsdaten
- Umsetzungsbeispiele
- Bibliotheken: neue Aufgaben, neue Perspektiven
- Zusammenfassung

Digitale Wissenschaft heute: Mehr als Publikationen

Wer, wie,, was, wieso, weshalb, warum…

FORSCHUNGSDATEN

Warum

Nachvollziehbarkeit

Nachnutzbarkeit

… der wissenschaftlichen Ergebnisse

Langzeitarchivierung und Data Sharing

….ist gefragt von Forschungsförderern und der Gesellschaft

Empfehlung

EUROPEAN COMMISSION

Brussels, 17.7.2012
C(2012) 4890 final

COMMISSION RECOMMENDATION

of 17.7.2012

on access to and preservation of scientific information

{SWD(2012) 221 final}
{SWD(2012) 222 final}

http://ec.europa.eu/research/science-society/index.cfm?fuseaction=public.topic&id=1301

Empfehlung

Ensure that, as a result of these policies:

- research data that result from publicly funded research become publicly accessible, usable and re-usable through digital e-infrastructures. Concerns in particular in relation to privacy, trade secrets, national security, legitimate commercial interests and to intellectual property rights shall be duly taken into account. Any data, know-how and/or information whatever their form or nature which are held by private parties in a joint public/private partnership prior to the research action and have been identified as such shall not fall under such an obligation;

- datasets are made easily identifiable and can be linked to other datasets and publications through appropriate mechanisms, and additional information is provided to enable their proper evaluation and use;

http://ec.europa.eu/research/science-

Aspekte

Forschunsdaten sind:

Komplex

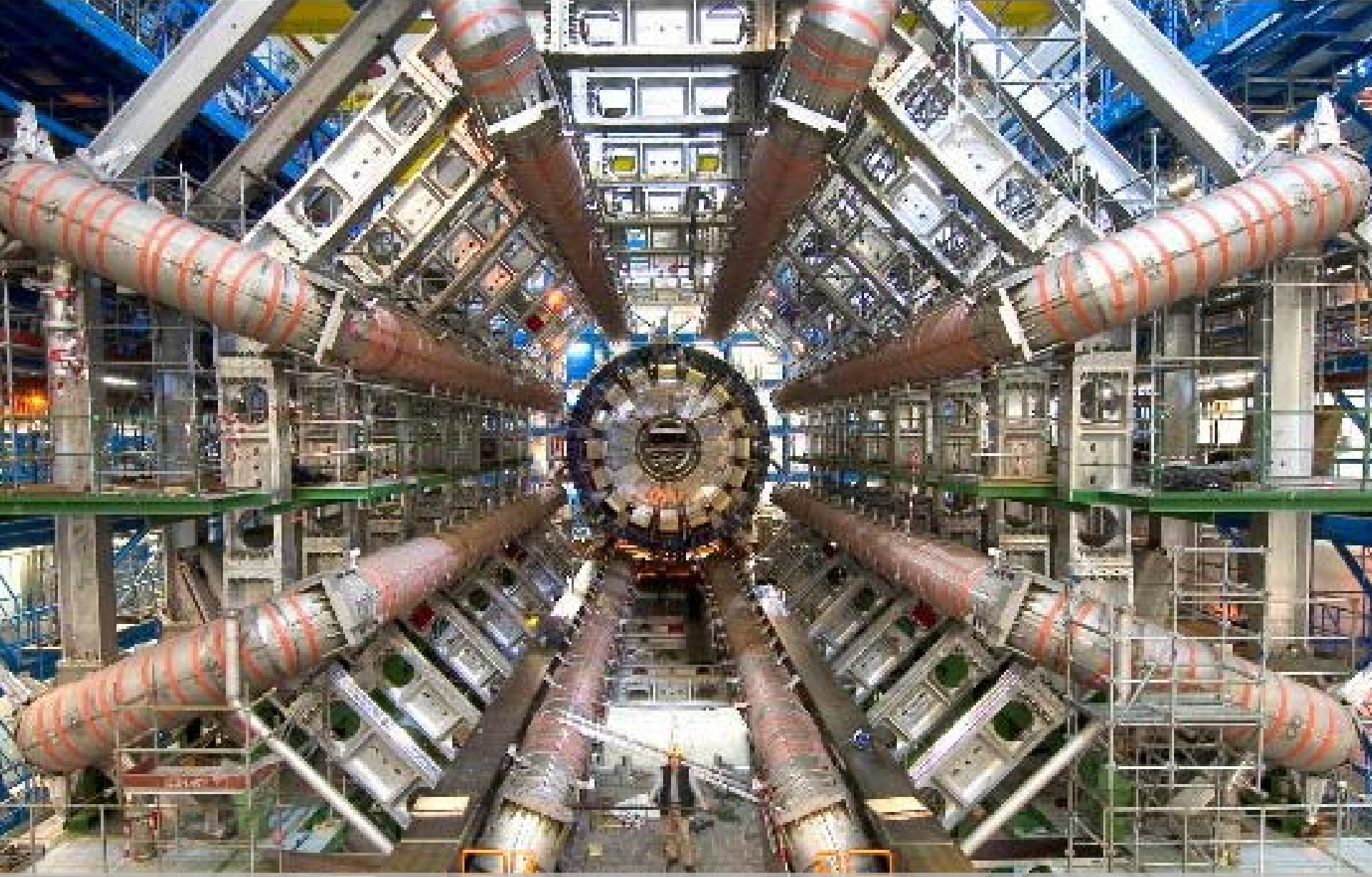
Verschiedenartig

Einzigartig, nicht wiederholbar

Teuer, wertvoll

… ein paar Beispiele aus verschiedenen Disziplinen

Large Hadron Collider @ CERN



LHC – ATLAS experiment

Grid View Developed by BARC, India

ABOUT GV

Monitoring and Visualization Tool for LCG

Gridftp Transfers | FTS Transfers | Service Availability | Service Availability (ACE)

New Gridview Interface

RELEASE INFO

LCG

What do you want ?

- (•) VO-wise graphs
- ( ) Site-wise graphs

- (•) Average Throughput
- ( ) Aggregate Data X-fer

VO
All VOs
aegis
Alice

- ( ) Use Site Abbreviations
- (•) Use Full Site Names

Source Site(s):
All Sites
######IN2P3-CC
AEGIS01-IPB-SCL
Tier(*)

Destination Site(s):
All Sites
######IN2P3-CC
AEGIS01-IPB-SCL
Tier(*)

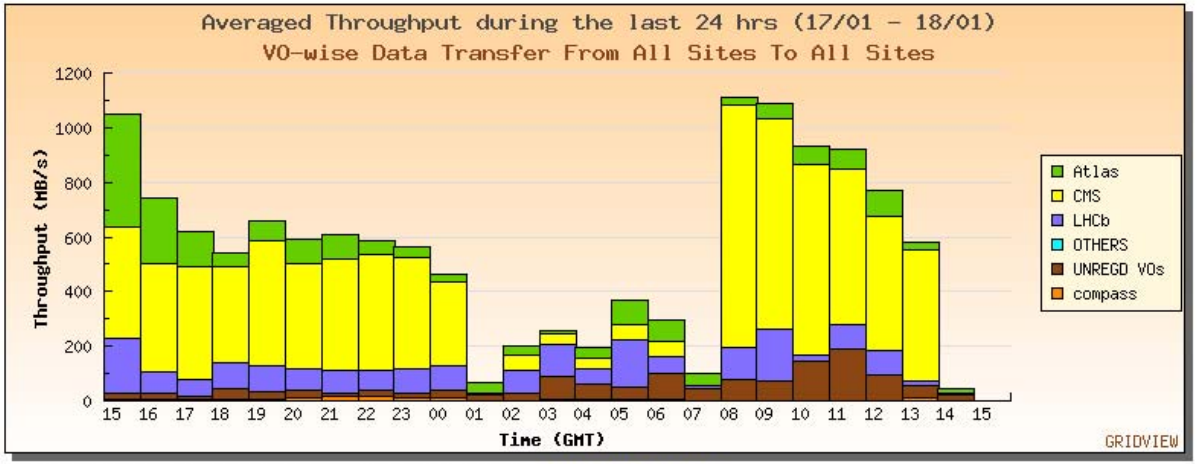
- (•) Current Summary
- ( ) Hourly Report
- ( ) Daily Report
- ( ) Weekly Report
- ( ) Monthly Report

From 00 Hours On 18 Jan 2012

To 15 Hours On 18 Jan 2012

Current Status

(VO-wise Data Transfer From All Sites To All Sites)

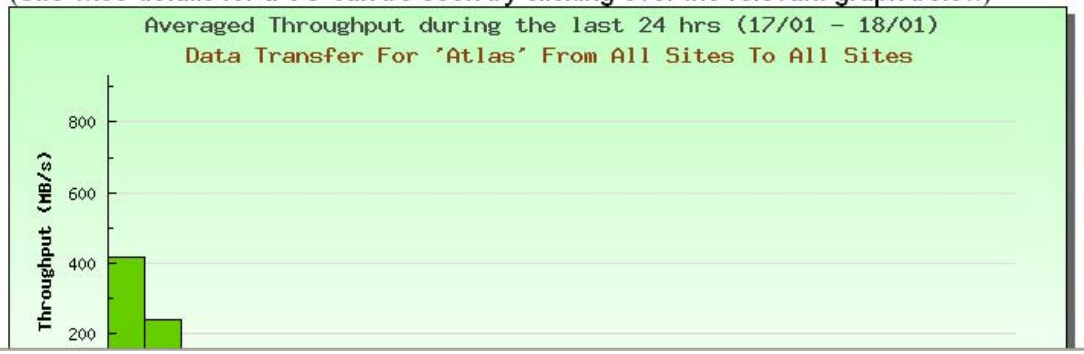


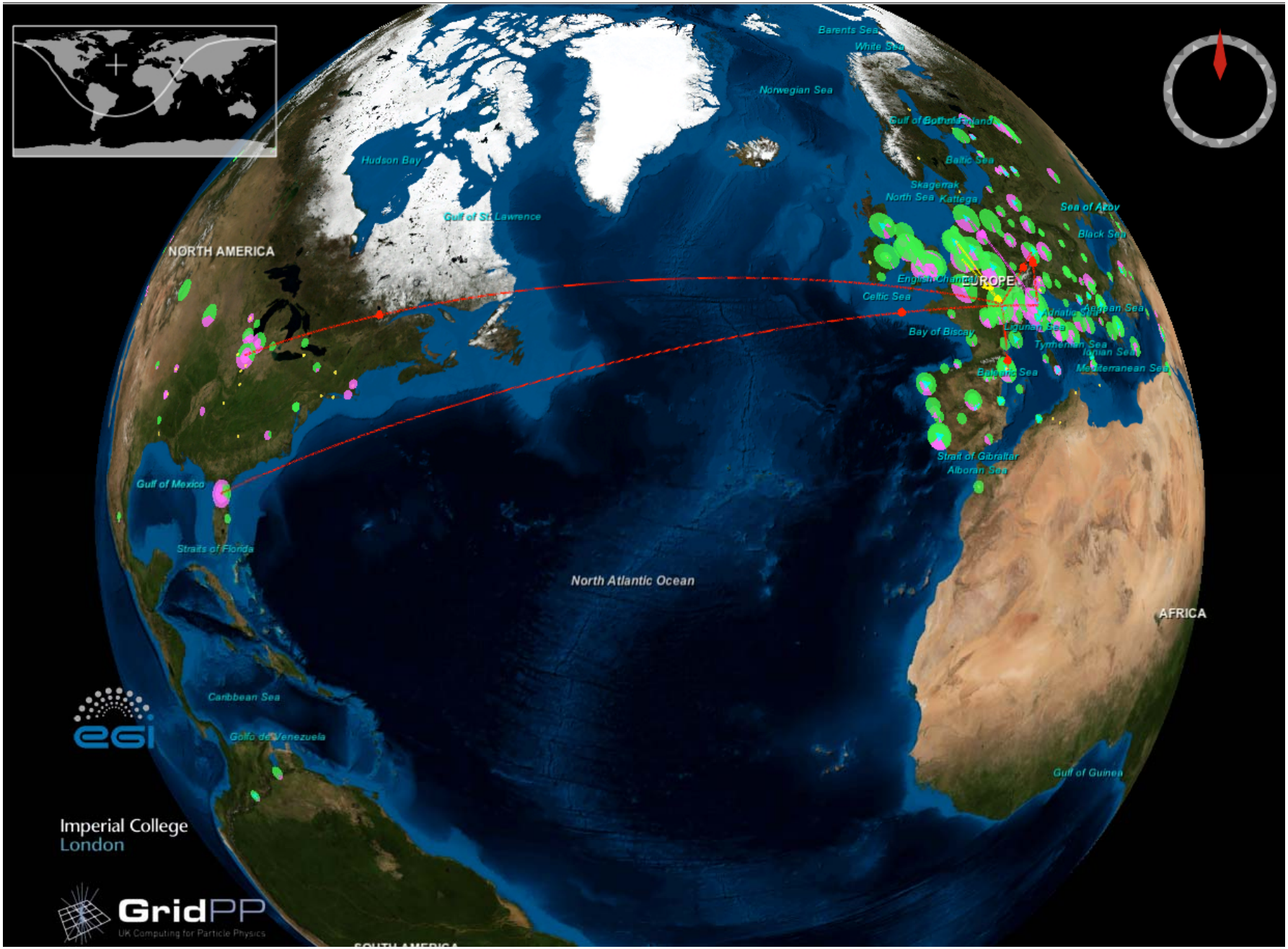
(**OTHERS**: VOs giving throughput less than 1% of max, click here for names)

Graphs for Individual VOs:-

Atlas | CMS | LHCb | OTHERS | UNREGD VOs | compass |

(Site-wise details for a VO can be seen by clicking over the relevant graph below)




Barents Sea
White Sea
Norwegian Sea
Gulf of Bothnia
Baltic Sea
Hudson Bay
Skagerrak
North Sea
Kattegat
Sea of Azov
Black Sea
Gulf of St. Lawrence
NORTH AMERICA
EUROPE
English Channel
Celtic Sea
Adriatic Sea
Aegean Sea
Bay of Biscay
Tyrrhenian Sea
Ionian Sea
Balearic Sea
Mediterranean Sea
Strait of Gibraltar
Alboran Sea
Gulf of Mexico
Straits of Florida
North Atlantic Ocean
AFRICA
Caribbean Sea
Golfo de Venezuela
Gulf of Guinea
Imperial College
London
GridPP
UK Computing for Particle Physics

Deep Sea Observation

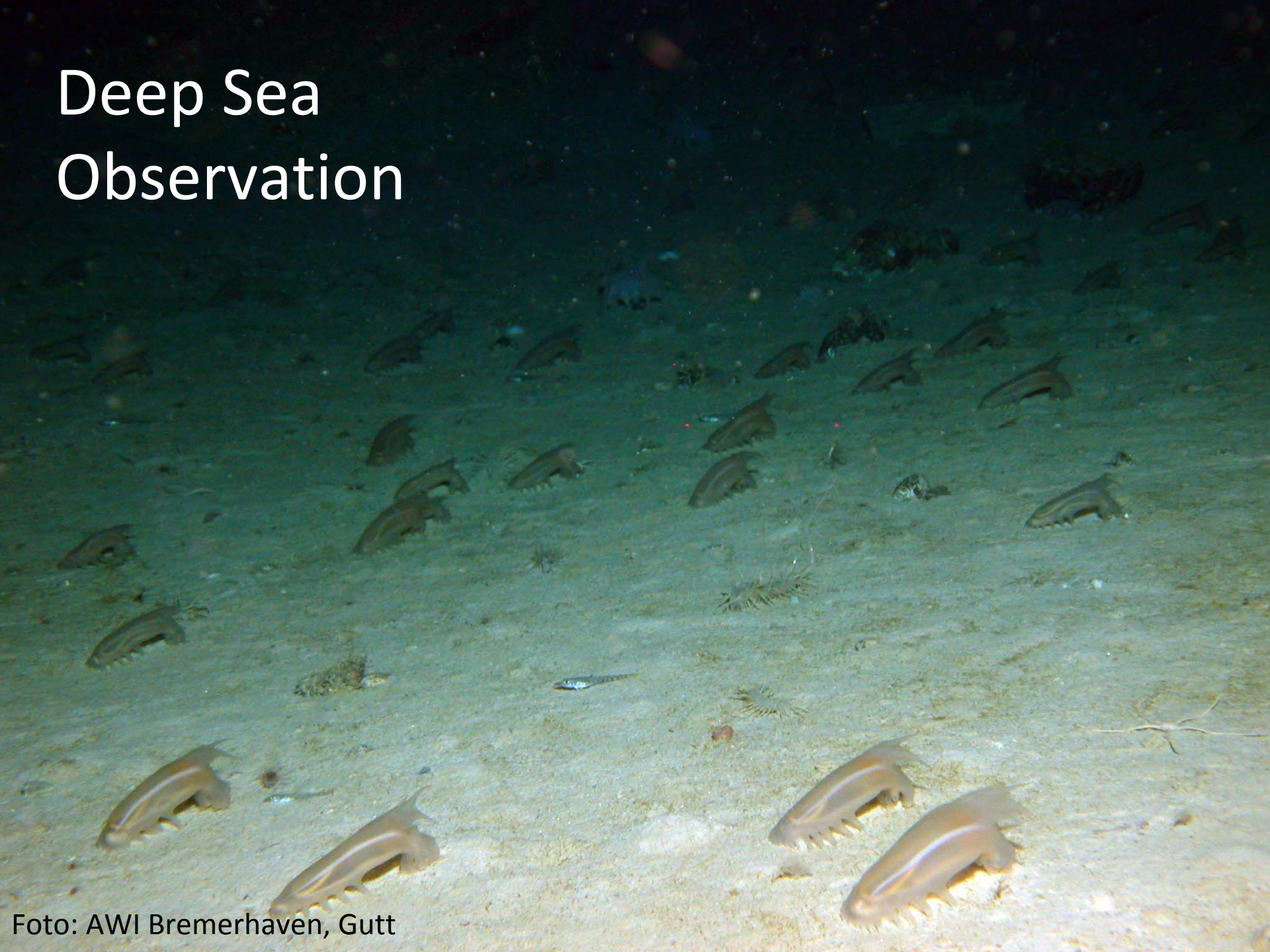
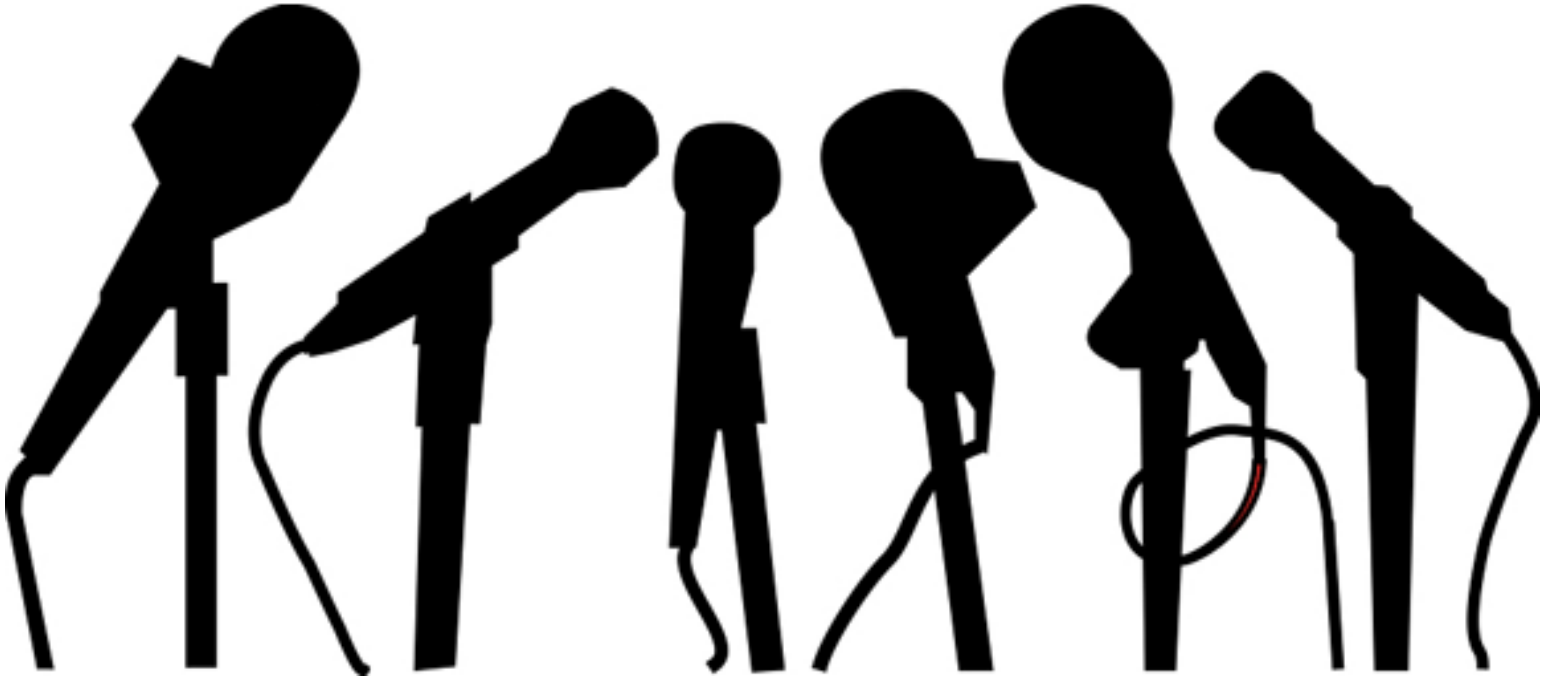


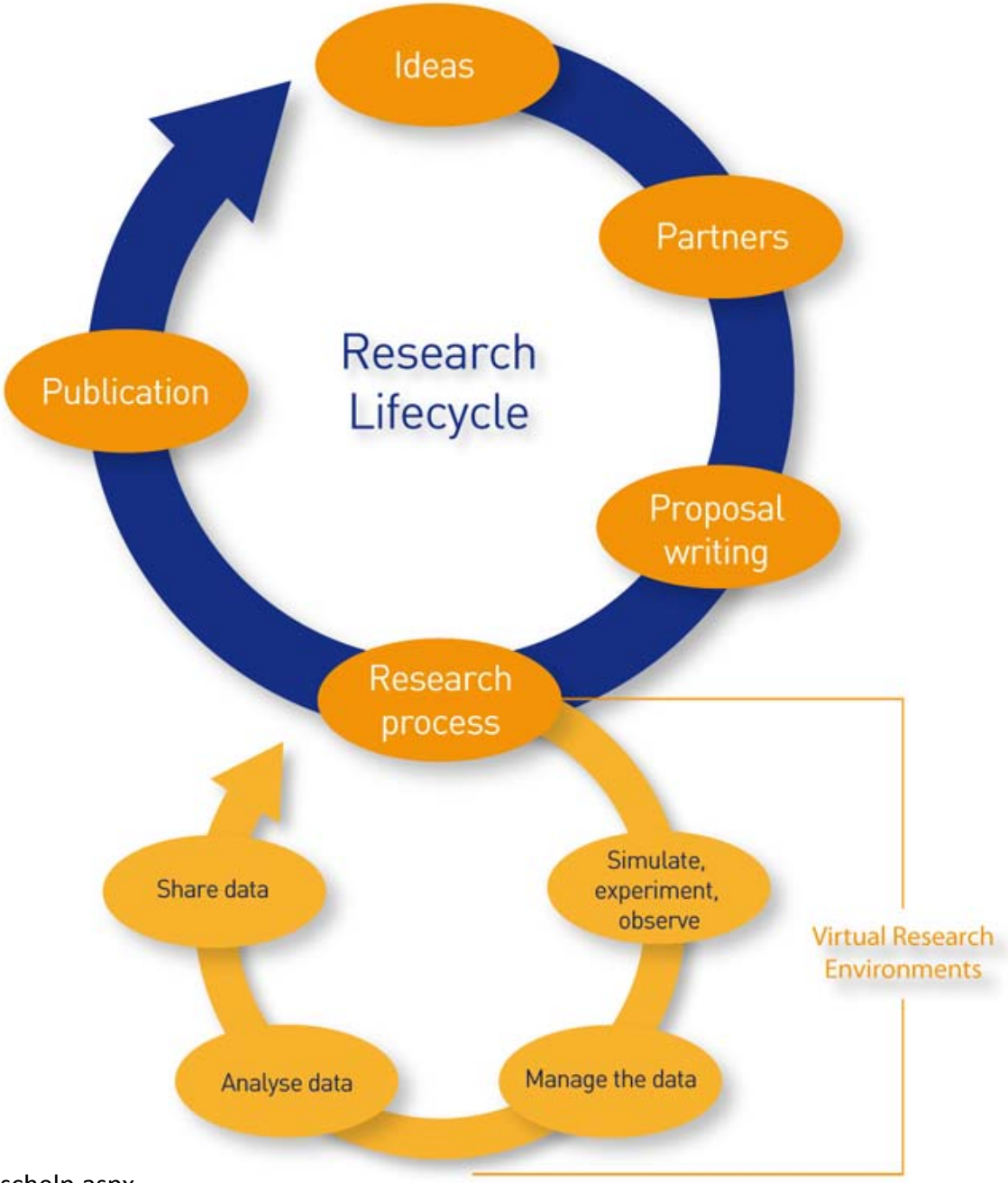
Foto: AWI Bremerhaven, Gutt

Forschungsdaten: Interviews



DATEN ALS TEIL DES WISSENSCHAFTLICHEN PROZESSES

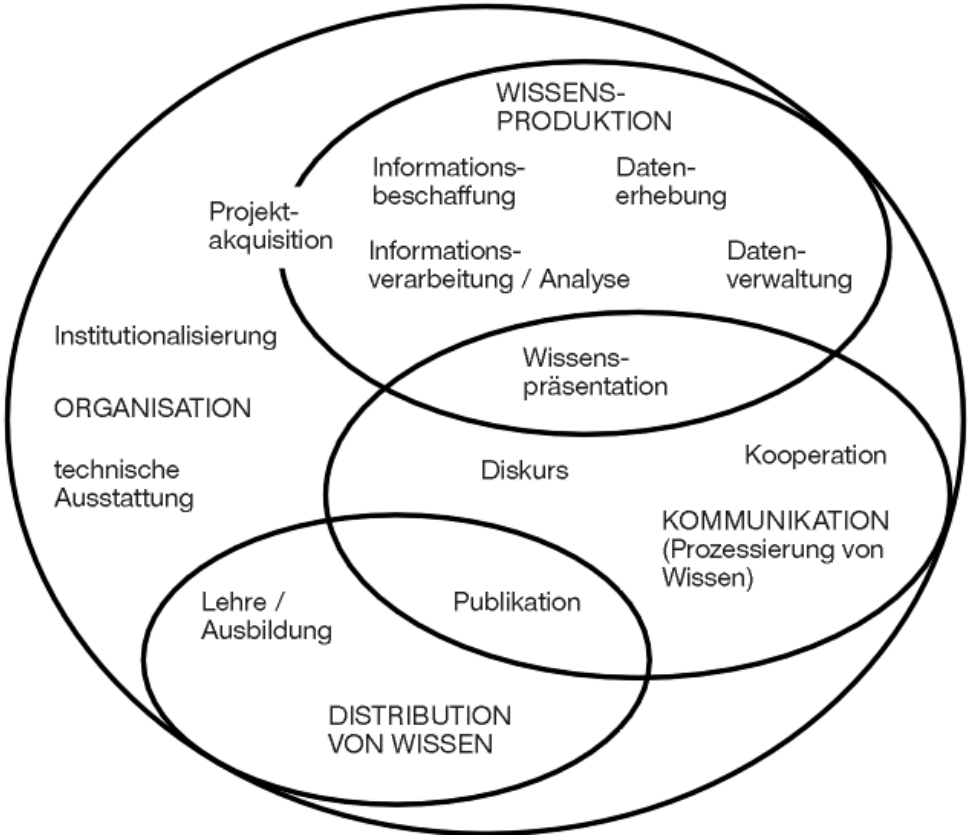
Forschungsdaten-management wissenschaftsnah



http://www.jisc.ac.uk/whatwedo/campaigns/res3/jischelp.aspx

Produktion, Kommunikation und Distribution von Wissen

Schaubild 1 Die wissenschaftlichen Aktivitätsformen und Rahmenbedingungen



Nentwich, 1999: http://www.mpi-fg-koeln.mpg.de/pu/workpap/wp99-6/bild1.html

“vague but exciting” @ CERN in ‘90s

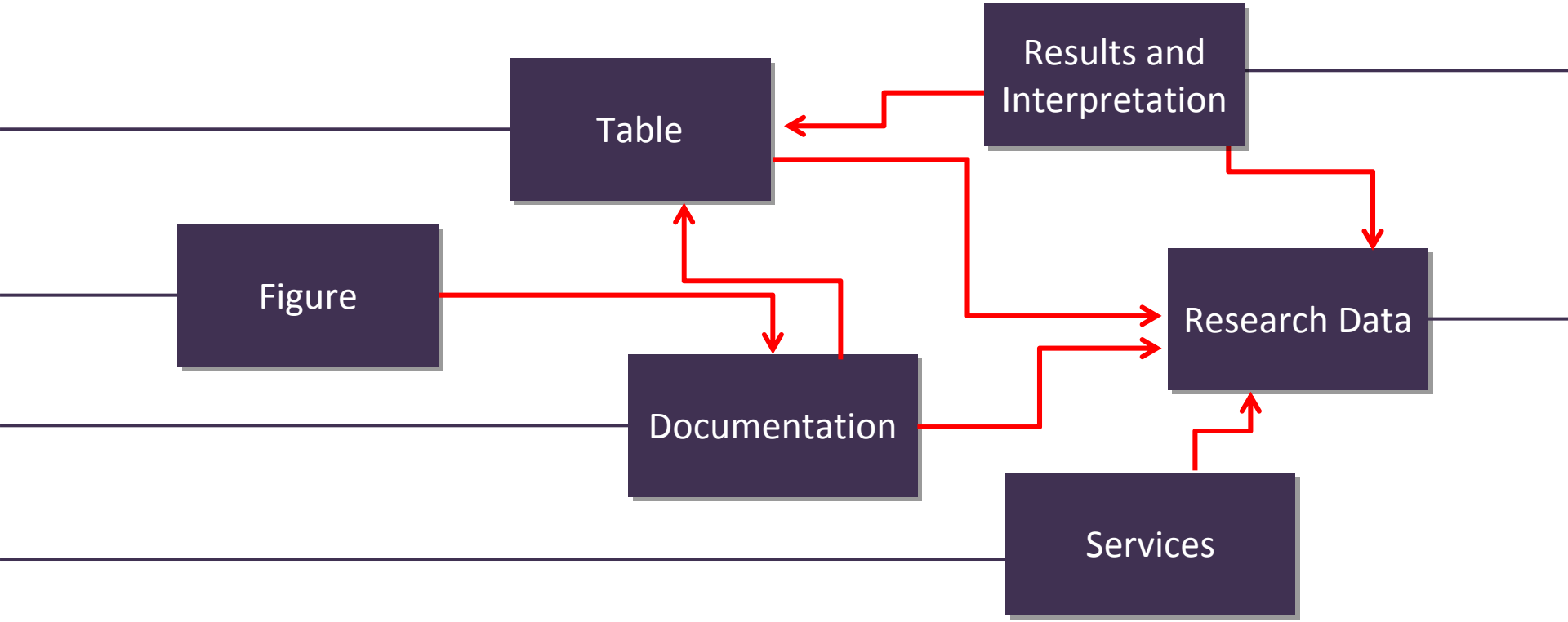
Und dann?

Forschungsdaten sind kein neues Objekt…

…aber können in digitalen Forschungsinfrastrukturen neu gehandhabt werden

- Digitale Langzeitarchivierung
- Permanenter Zugang zu Forschungsdaten
- Data sharing

Wo gehts hin?
Objektzentriert

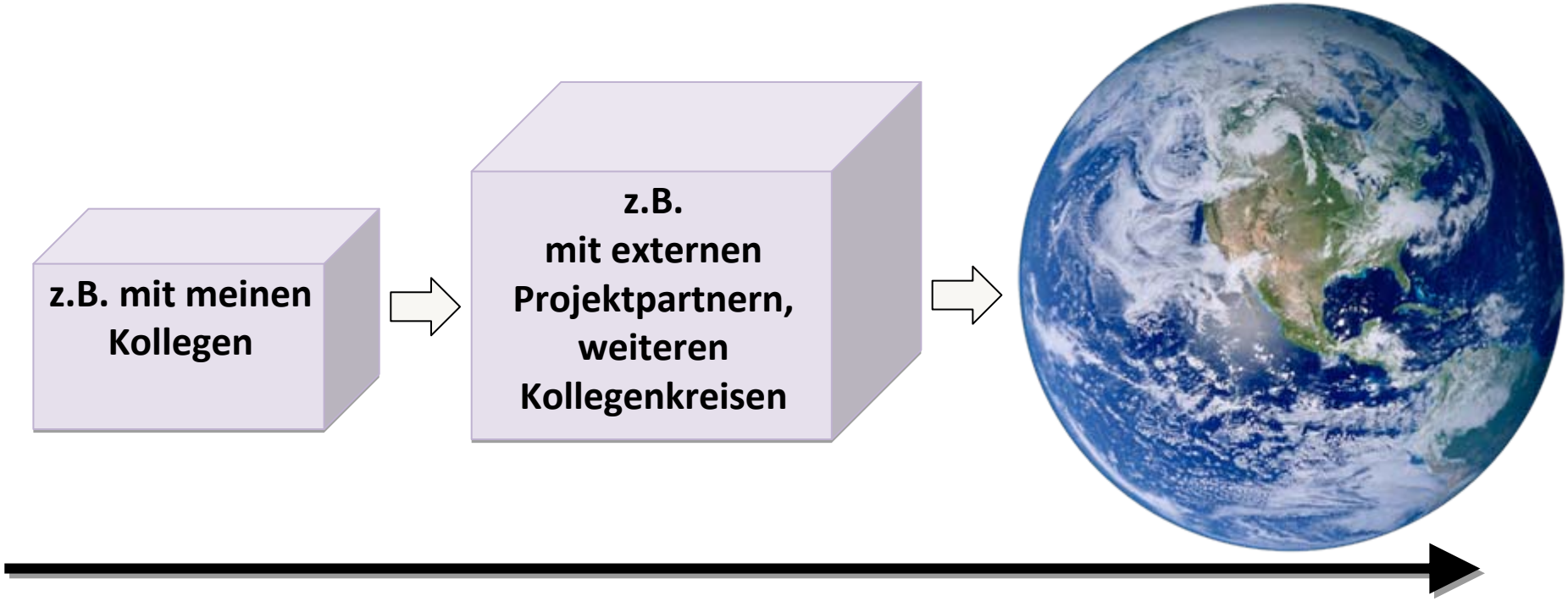


RAHMENBEDINGUNGEN

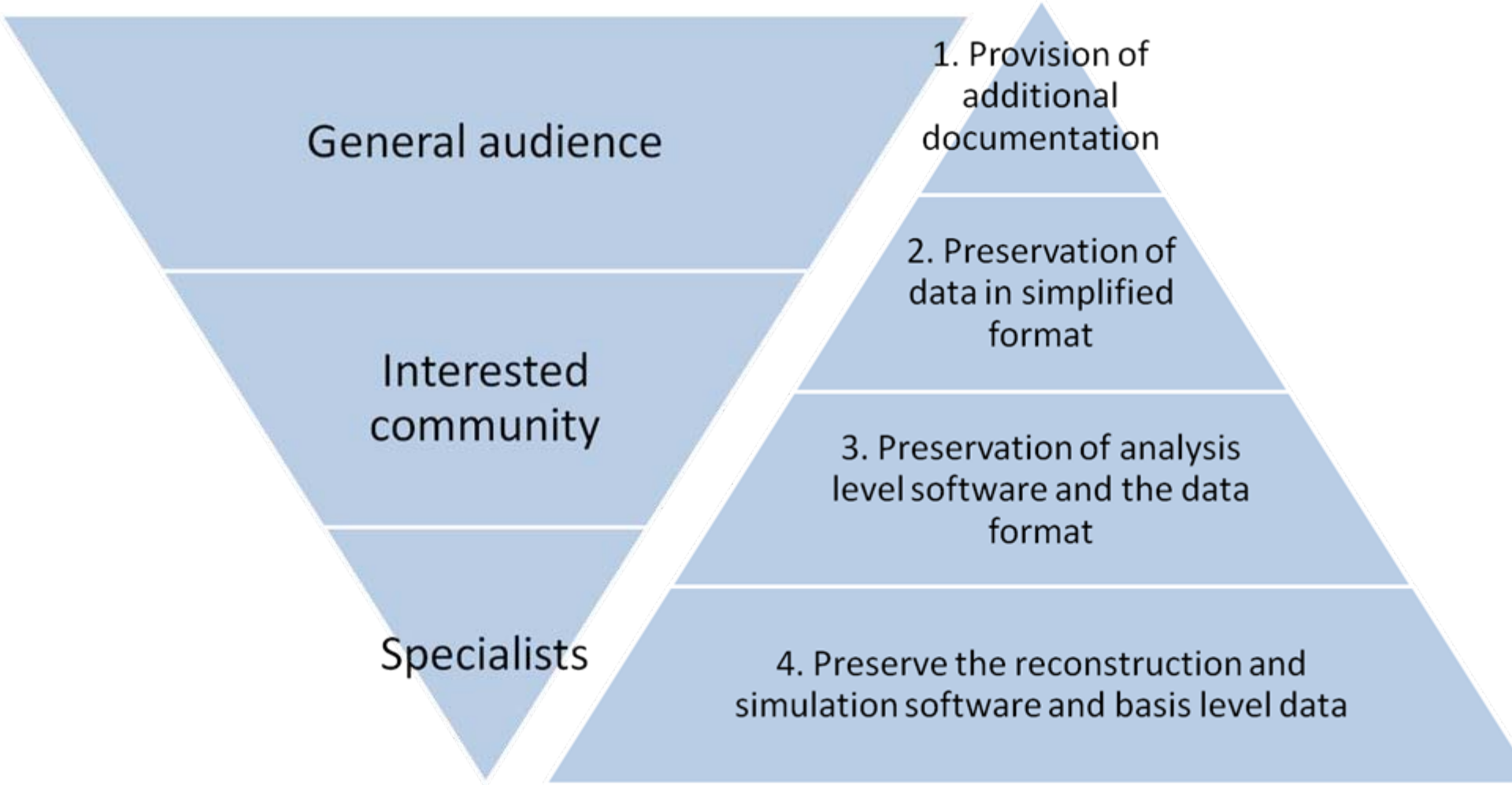
WIE GEHT DAS?

Wo?

[Raum der Datenbereitstellung]



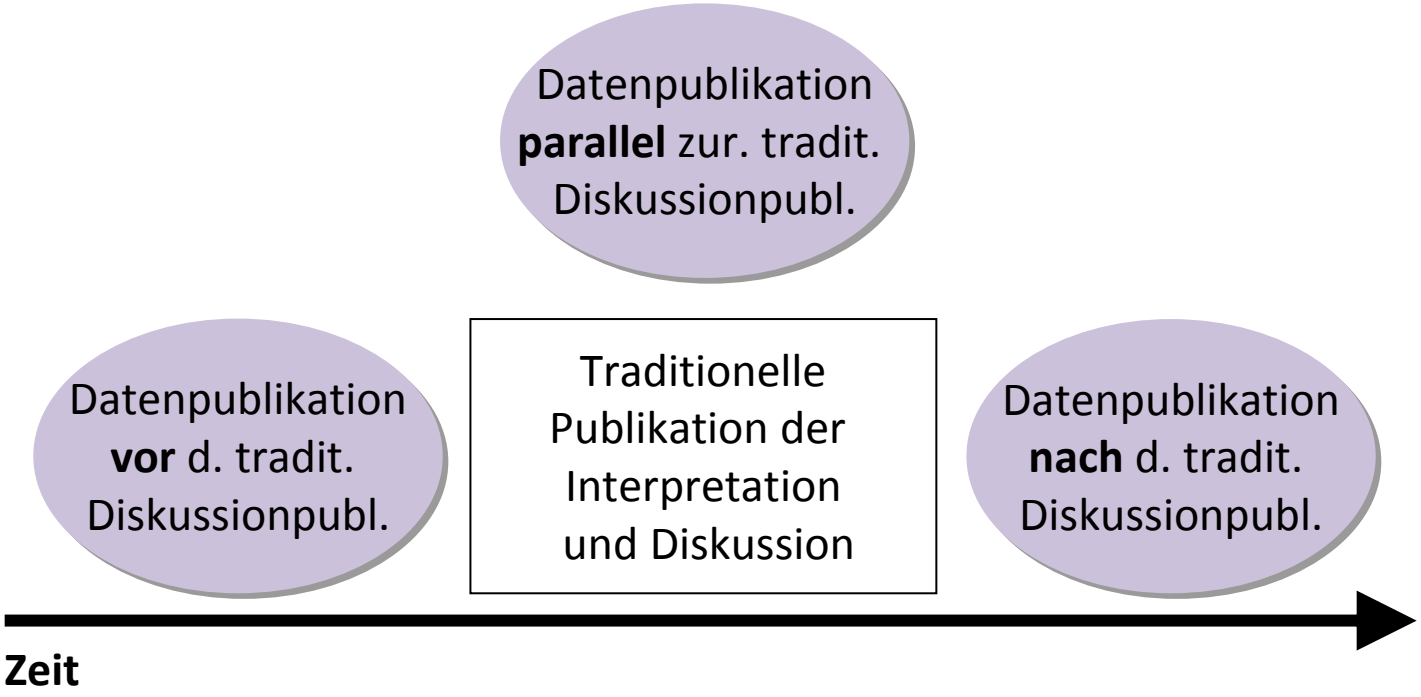
Wer und was?

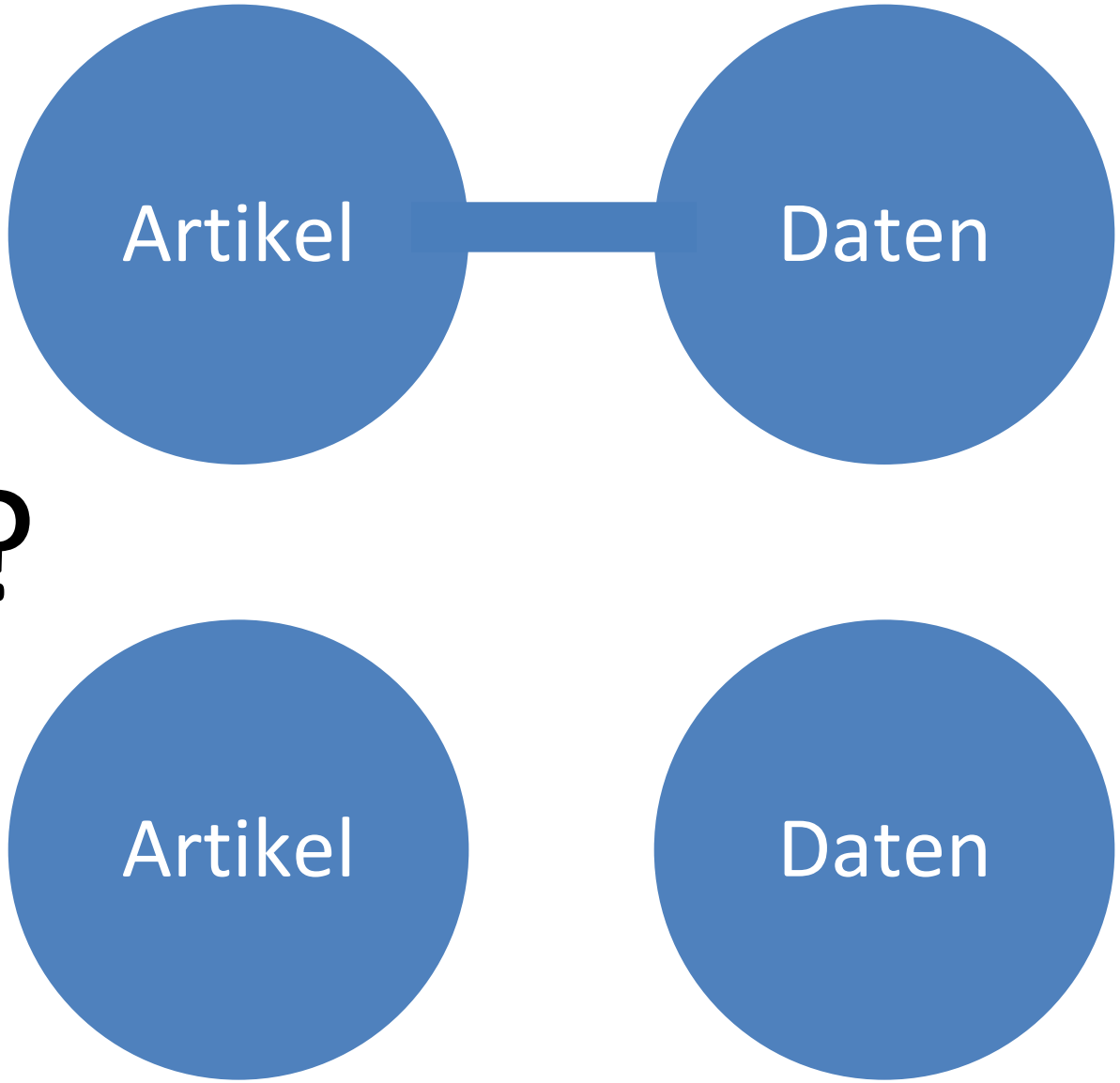


Modified after data preservation models by DPHEP, 2009

Wann? [Zeitpunkt der Bereitstellung]

- Direkt nach Datenproduktion/vor der Interpretation (Pre-)
- Mit der Analyse/Interpretation (with publication)
- Nach der Publikation der Interpretation (Post-)




Artikel
Daten
Wie?
Artikel
Daten

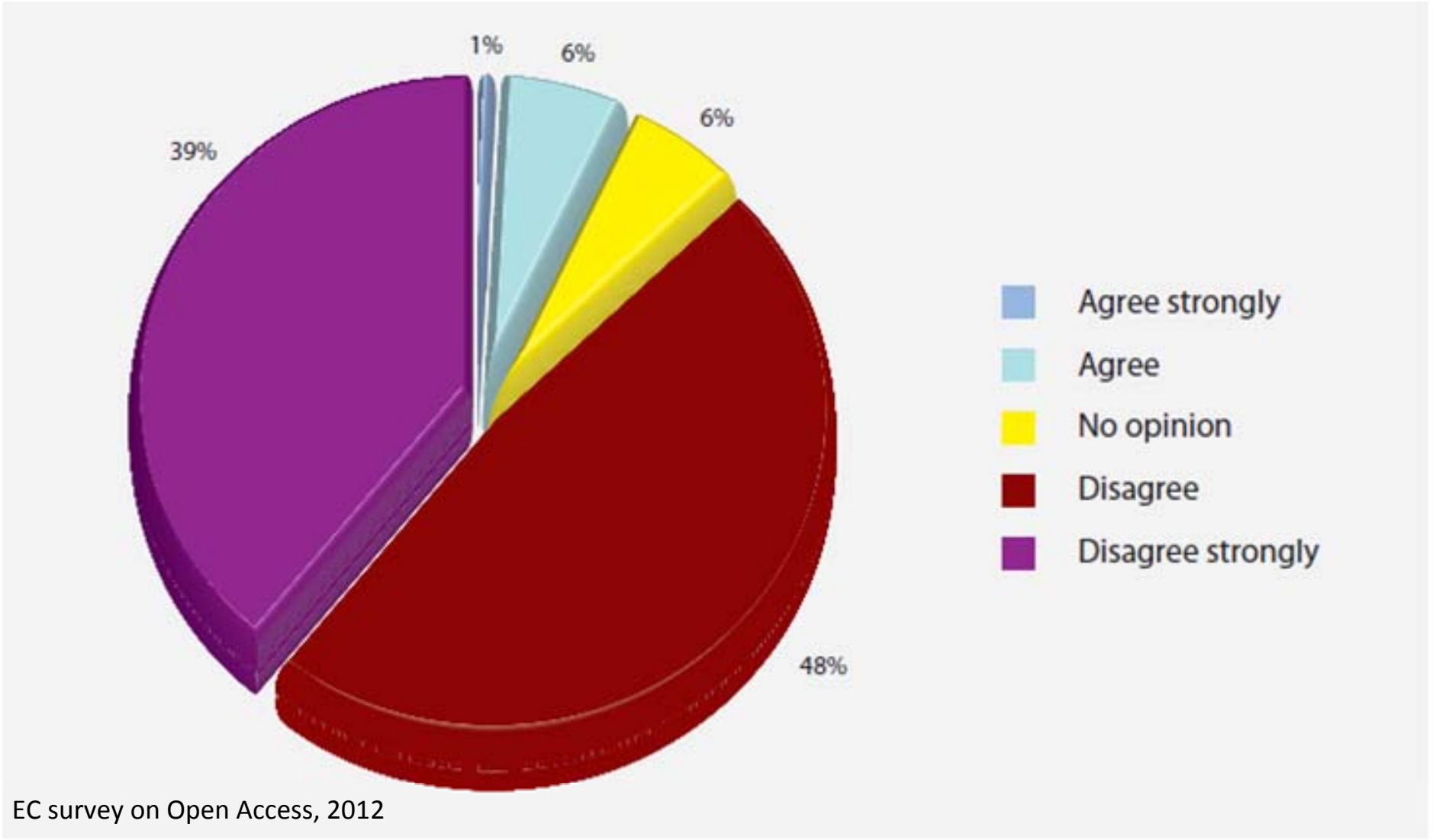
… wie…

- Langzeitarchivierung
- Standardisierung
- Interoperabilität
- Dokumentation
- Persistent Identifiers, DOIs
- Added-value services

…

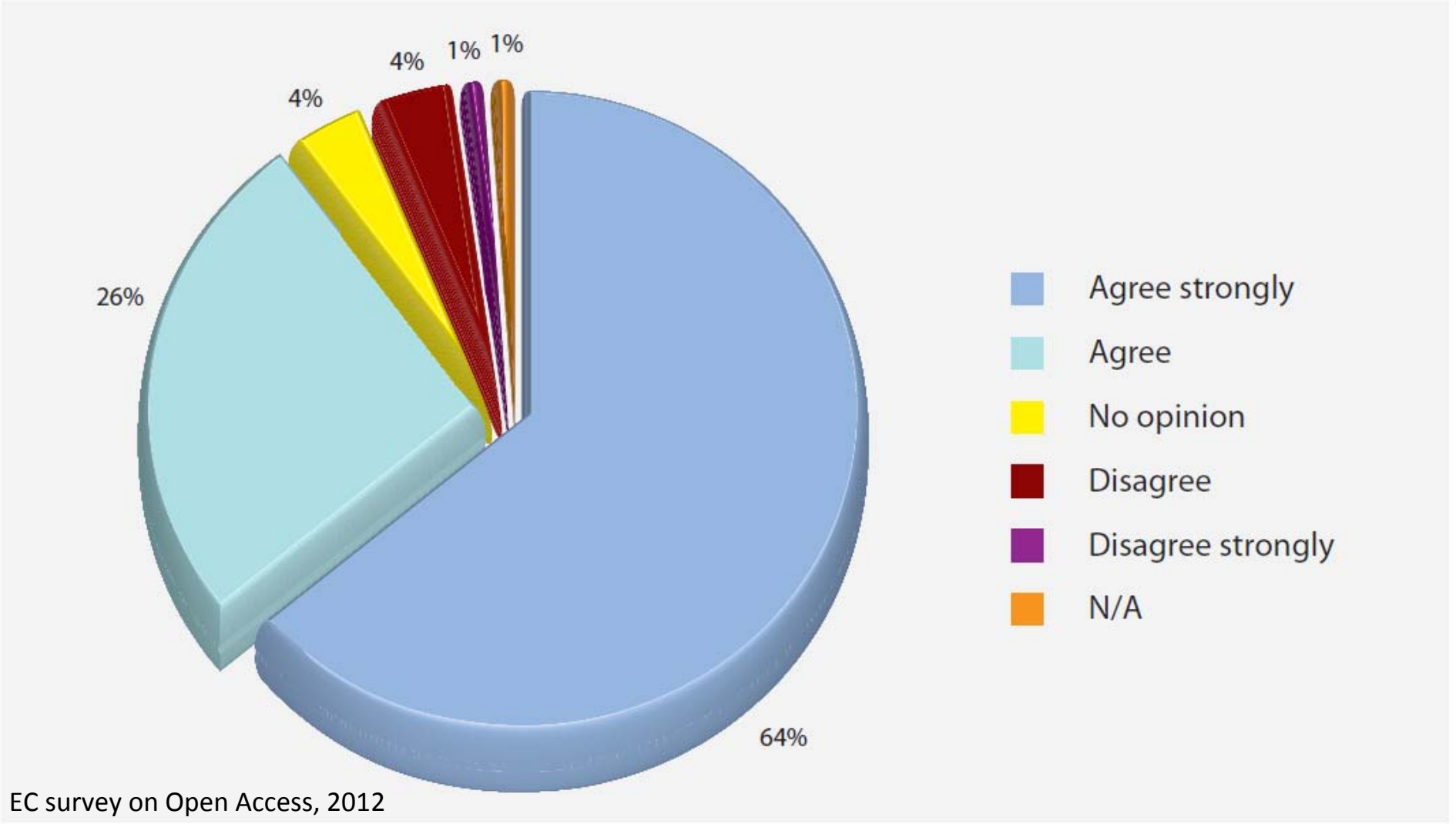
ABER: WAS PASSIERT HEUTE?

Do you agree with the following statement: "Generally speaking, there is NO access problem to research data in Europe"



EC survey on Open Access, 2012

Do you think that research data that is publicly available and that results from public funding should, …, be available for reuse and free of charge on the Internet?



EC survey on Open Access, 2012

Zugang zu Daten in den Disziplinen

	Others can access my data easily	
	Agree strongly	**Agree somewhat**
social sciences	11(5.4%)	36(17.8%)
computer science/engineering	12(10.3%)	29(24.8%)
physical sciences	17(11.3%)	41(27.3%)
environmental sciences & ecology	56(12.0%)	124(26.5%)
atmospheric science	12(23.5%)	13(25.5%)
biology	28(15.6%)	50(27.9%)
medicine	2(6.5%)	2(6.5%)
other	12(13.0%)	21(22.8%)

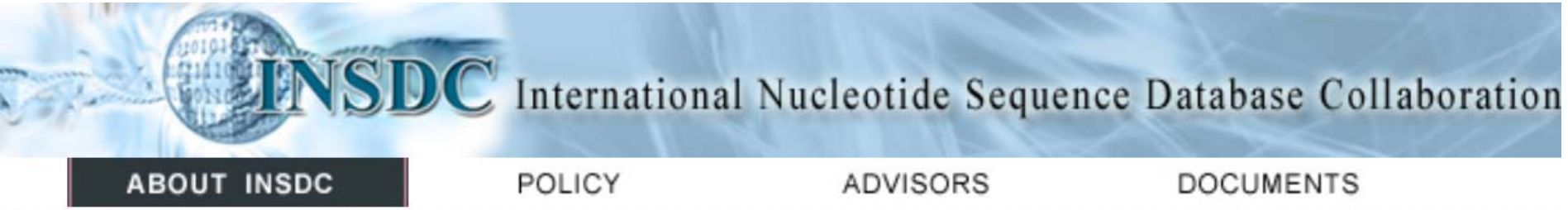
Tenopir et al., 2011

Opportunities for Data Exchange

Science & Technology
Facilities Council

UMSETZUNGSBEISPIELE
“DATA SHARING”

Molekulare Biologie



NCBI

DDBJ

ENA
European Nucleotide Archive

International Nucleotide Sequence Database Collaboration

- The International Nucleotide Sequence Databases (INSD) have been developed and maintained collaboratively between DDBJ, ENA, and GenBank for over 18 years.
- The INSDC advisory board, the International Advisory Committee , is made up of members of each of the databases' advisory bodies. At their most recent meeting, members of this committee unanimously endorsed and reaffirmed the existing data-sharing policy of the three databases that make up the INSDC, which is stated below.
- Individuals submitting data to the international sequence databases should be aware of INSDC policy.

How to submit data

- For full details of how to submit data to the databases, please select a collaborating partner.
- DDBJ, ENA, GenBank
- The INSDC Feature Table Definition Document is available here.

http://www.insdc.org/

OECD – iLibrary

http://www.oecd-ilibrary.org/economics/data/oecd-factbook-statistics-2010_fb10-data-en

Publishing Network for Geoscientific and Environmental Data

PANGAEA®
Publishing Network for Geoscientific & Environmental Data

You are not logged in

Always quote citation when u

Data Description

Citation: Spieß, V; Grobe, H (1996): Paleomagnetic measurements on sediment core PS1387-3, doi:10.1594/PANGAEA.51316,
In Supplement to: **Grobe, Hannes; Mackensen, Andreas; Hubberten, Hans-Wolfgang; Spieß, Volkhard; Fütterer, Dieter K (1990):** Stable record and late quaternary sedimentation rates at the Antarctic continental margin, *In: Bleil, U & Thiede, J (eds.), Geological History of the Pole Oceans - Arctic versus Antarctic, NATO ASI Series, Kluwer Academic Publishers, Dordrecht, Boston, London*, 539-571, hdl:10013/epic.1166

Project(s): **Paleoenvironmental Reconstructions from Marine Sediments @ AWI** (AWI_Paleo)

Coverage: *West:* -5.8667 * *East:* -5.8667 * *South:* -68.7333 * *North:* -68.7333

Date/Time Start: 1985-12-28T15:48:00 * *Date/Time End:* 1985-12-28T15:48:00

Minimum DEPTH, sediment: 0.1 m * *Maximum DEPTH, sediment:* 10.0 m

Event(s): **PS1387-3** (PS08/365) * *Latitude:* -68.7333 * *Longitude:* -5.8667 * *Elevation:* -2416.0 m * *Date/Time:* 1985-12-28T15:48:00 * *Recovery:* 10.0 m * *Penetration:* 11.2 m * *Location:* Atka Bay * *Campaign:* ANT-IV/3 (PS08) * *Basis:* Polarstern * *Device:* Gravity corer (Kiel type) * *Comment:* upper 30 cm lost in weight, parallel station PS1

Parameter(s):

#	Name	Short Name	Unit	Principal Investigator	Method	Comment
1	DEPTH, sediment	Depth	m			Geocode
2	Susceptibility	Susceptibility	sensor units	Spieß, Volkhard	Susceptibility unit AWI, MS2C, 145 mm	

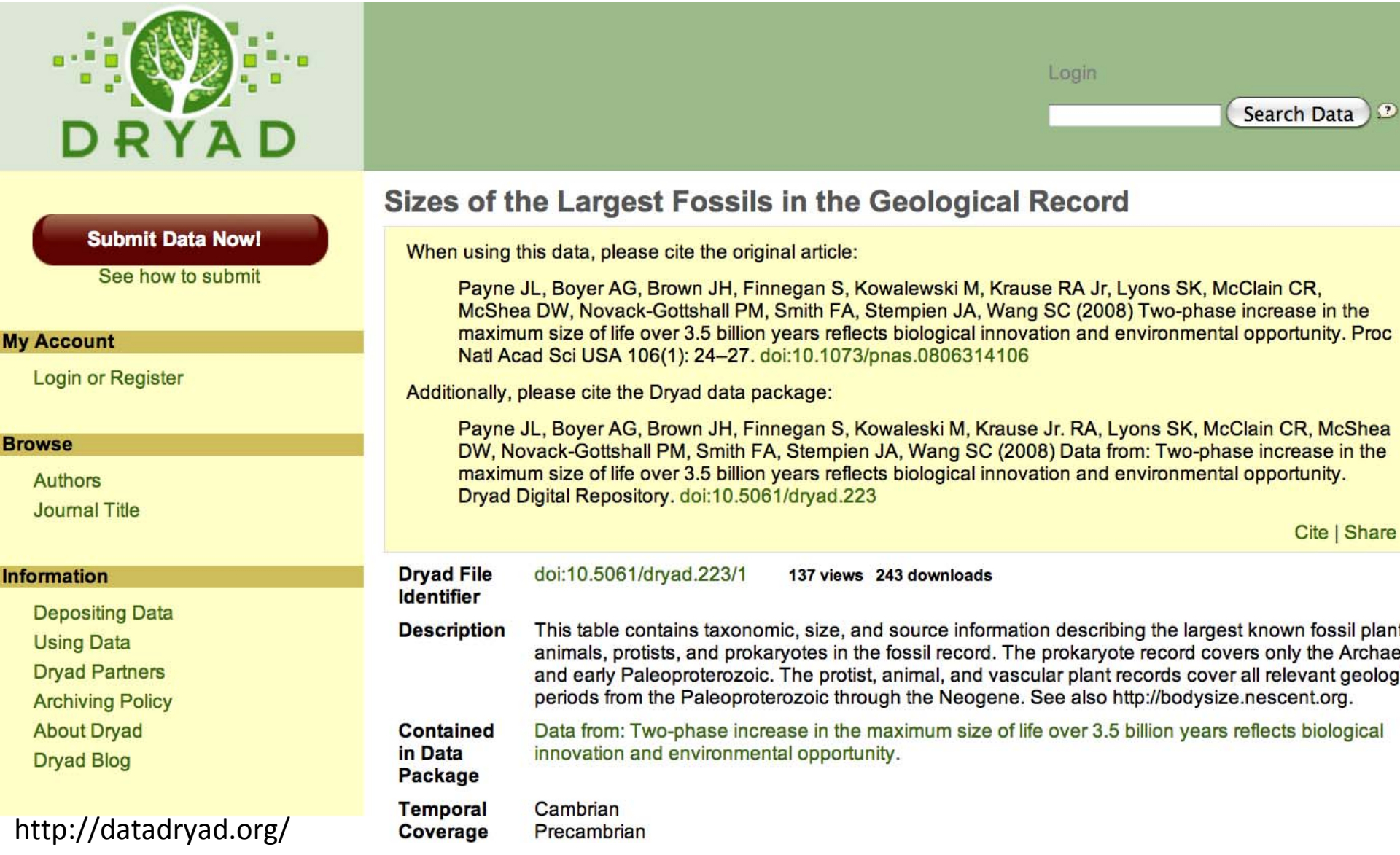
Size: 100 data points

Data

Download dataset as tab-delimited text *(use the following character encoding:* ISO-8859-1: ISO Western (PANGAEA default) *)*

1 Depth [m]	2 Susceptibility [sensor units]
0.06000	0.83
0.15000	1.04
0.25000	1.00
0.35000	1.00
0.45000	1.28

Dryad



http://datadryad.org/

Search for squarks and gluinos using final states with jets and missing transverse momentum with the ATLAS detector in $\sqrt{s} = 7$ TeV proton-proton collisions.

ATLAS Collaboration (Georges Aad (Freiburg U.) *et al.*) Show all 3024 authors.

Sep 2011 - 9 pages

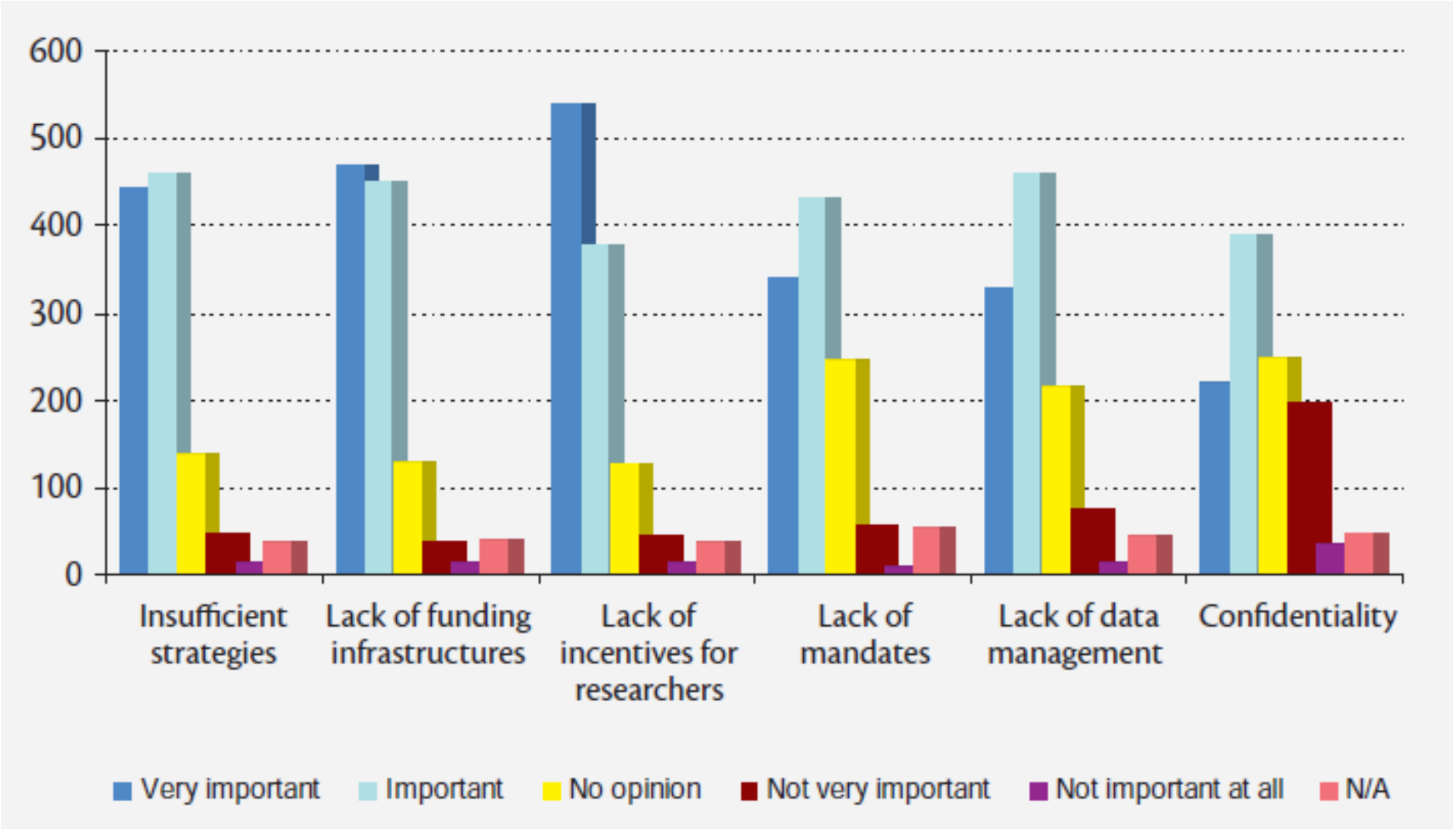
Phys.Lett. B710 (2012) 67-85
DOI: 10.1016/j.physletb.2012.02.051
CERN-PH-EP-2011-145
e-Print: arXiv:1109.6572 [hep-ex] | PDF
Experiment: CERN-LHC-ATLAS

Abstract: A search for squarks and gluinos in events containing jets, missing transverse momentum and no electrons or muons is presented. The data were recorded in 2011 by the ATLAS experiment in sqrt(s) = 7 TeV proton-proton collisions at the Large Hadron Collider. No excess above the Standard Model background expectation is observed in 1.04 fb^-1 of data. Gluino and squark masses below 700 GeV and 875 GeV respectively are excluded at the 95% confidence level in simplified models containing only squarks of the first two generations, a gluino octet and a massless neutralino. The exclusion limit increases to 1075 GeV for squarks and gluinos of equal mass. In MSUGRA/CMSSM models with tan(beta)=10, A_0=0 and mu> 0, squarks and gluinos of equal mass are excluded for masses below 950 GeV. These limits extend the region of supersymmetric parameter space excluded by previous measurements.

Note: 9 pages plus author list (20 pages total), 2 figures, 3 tables, matches published version in Physics Letters B

WAS FEHLT

“rate the importance of barriers to an enhanced access to research data”



EC survey on Open Access, 2012

In my discipline we started data preservation and sharing in the 19th century

In my discipline data sharing is not possible

I consider myself a data sharer rather than a classical researcher

The awareness of the problem is already building up (…) but the actions are lagging behind

Wo stehen die Bibliotheken?

ROLLENVERTEILUNG

Neue Objekte, neue Rollen, neue Aufgaben

- Akteure
  - Bibliothek
  - IT, Datenzentrum
  - Verlag
  - Wissenschaft

- Wissenschaft:
  - Bereitstellung der Daten
  - Dokumentation der Daten
  - Datenzitierung und Nachnutzung

- Herausforderung: Neue Workflows, Bereitstellung und Contentakquise

Rolle der Wissenschaft

Datenproduzenten

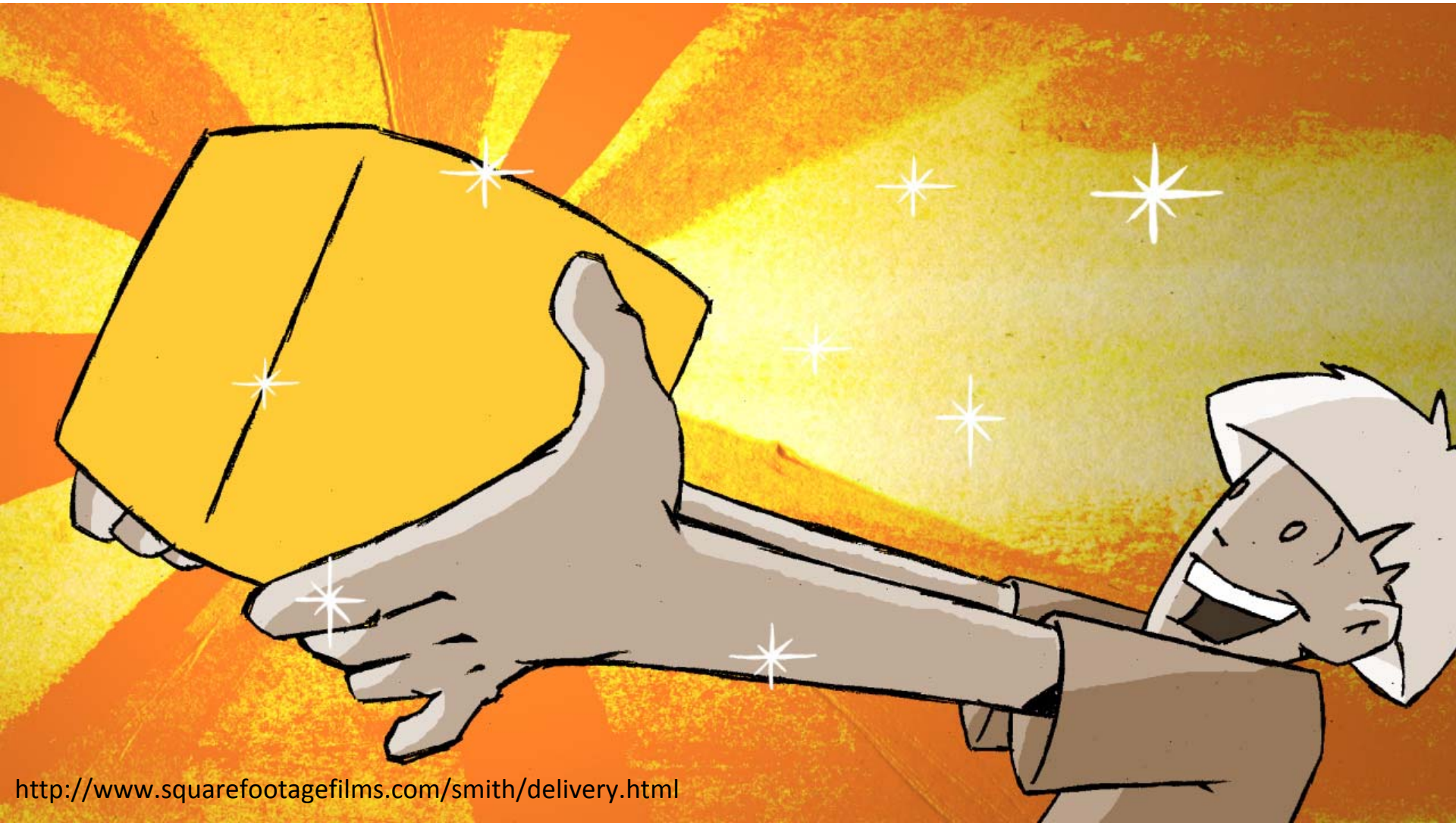
- Dokumentation
- Dateneinreichung
- Qualitätssicherung

Datennutzer

- Datenzitierung

Source: Nature 2006

Wer liefert die Inhalte?



http://www.squarefootagefilms.com/smith/delivery.html

Informationsmanagement

Service

- Zusammenarbeit und Service
- Management des “research material remix”
  - Neue Arbeitsabläufe
  - Integration von z.B. Datenzitierung
  - Metadaten
- Schnittstellenfunktionen
  - IT, Verlage, Repositorien, Wissenschaft

foto: flickr

Bibliotheken – DataCite

DataCite

Helping you to find,
access, and reuse data

One million DOI names minted

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

The 1,000,000th DOI name was minted by ETH Zurich on the 16th December 2010:

doi:10.5169/SEALS-130540

Read more

DataCite welcomes new members

ZBW KiSTi SND ICPSR osti.gov

On 2 December 2010 the British Library hosted the annual DataCite General Assembly, which brought together Member representatives from around the world to review progress and strategy. Following the Assembly, we are

Opportunities for Data Exchange

Science & Technology
Facilities Council

HELMHOLTZ
ASSOCIATION

APA
ALLIANCE FOR
PERMANENT ACCESS

Zusammenfassung

- Forschungsdaten: ein komplexes Objekt im gesamten Wissenschaftsprozess
- Generelle Rahmenbedingungen und flexible Modelle
- Praxis: interessante Umsetzungen, “Data Sharing” unterschiedlich etabliert
- Bibliotheken: neue Aufgaben für neue Objekte – Kooperation mit Wissenschaftlern und anderen Akteuren

Referenzen

DPHEP, 2009: http://arxiv.org/ftp/arxiv/papers/0912/0912.0255.pdf

Dryad, http://datadryad.org/

European Commission Recommendation: http://ec.europa.eu/research/science-society/index.cfm?fuseaction=public.topic&id=1301

INSDC, http://www.insdc.org/

JISC, Research Lifecyle: http://www.jisc.ac.uk/whatwedo/campaigns/res3/jischelp.aspx [accessed May 27th, 2011]

Opportunities Data Exchange (ODE), http://www.ode-project.eu/

OECD, Factbook:

http://www.oecd-ilibrary.org/economics/data/oecd-factbook-statistics-2010_fb10-data-en [accessed May 27th, 2011]