

Project NFDI4Earth

Erwartungen der Erdsystemwissenschaften an die Langzeitarchivierung
und Wiedernutzbarmachung archivierter Daten

Wolfgang zu Castell
(Director Department Geoinformation & CIO)

Introducing German Research Center for Geosciences

Helmholtz Association

- 43k employees in 18 centers
- addressing „big“ questions
- interdisciplinary research

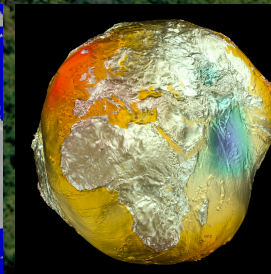
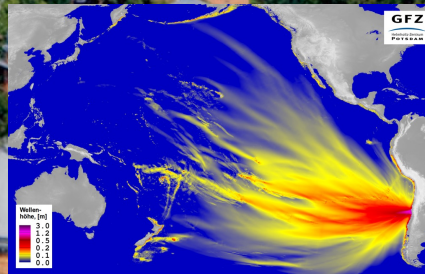


Changing Earth – Sustaining our Future

(7 centers)

GFZ Potsdam

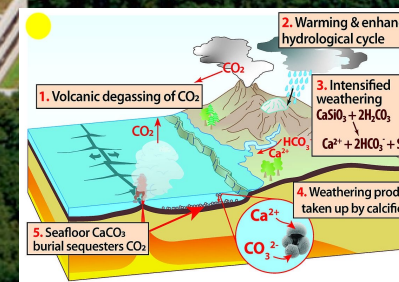
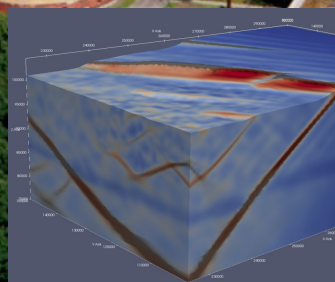
- 1.3k employees in 5 depts
- focussing on solid Earth
- geodesy, processes, risks



- observatories, satellites, monitoring networks
- geoscience labs
- modelling

Data Science

- maths, physics, comp. sci.
- modelling & data analysis
- numerics, ML/AI, complexity



- Library & Information Services
- eScience Center
- ICT Services
- Data Science Center

Outline

- NFDI4Earth – one of the 27 consortia forming the National Research Data Infrastructure¹
- Research in Geosciences – Challenges towards letting data flow freely
- Research Data Management and Digital Preservation – Friends or Foes?

¹ slides by the NFDI4Earth Consortium



NFDI4Earth

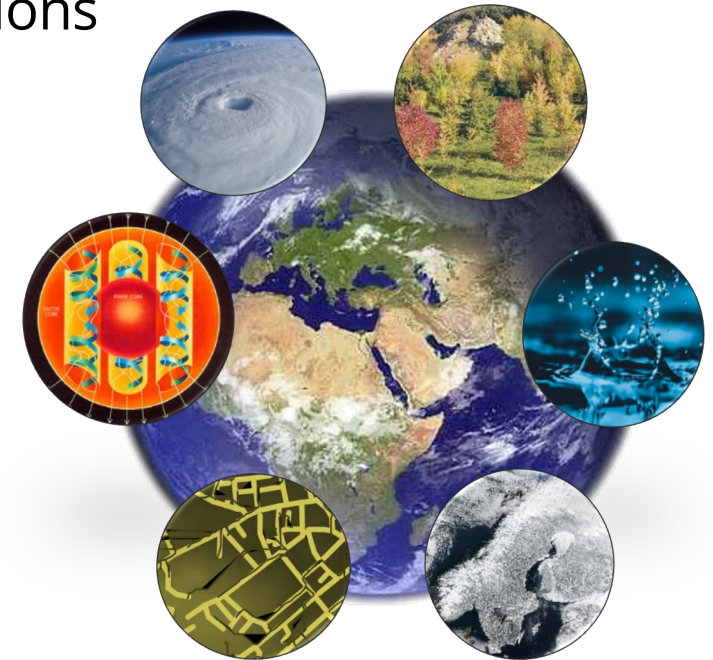
National Research Data Infrastructure for Earth System Sciences

NFDI4Earth Consortium 2022

<https://www.nfdi4earth.de>

Earth System Sciences (ESS)

- Geosphere, Atmosphere, Biosphere, Hydrosphere, Cryosphere, Anthroposphere and all their interactions
- From Local Processes to Global Challenges
e.g. earth systems dynamics, natural hazards, climate change, environmental pollution, water scarcity, land use change, scarcity of raw materials
- Observing, measuring, modelling, analyzing, predicting the Earth System
- In international and interdisciplinary settings using spatio-temporal data as the common reference

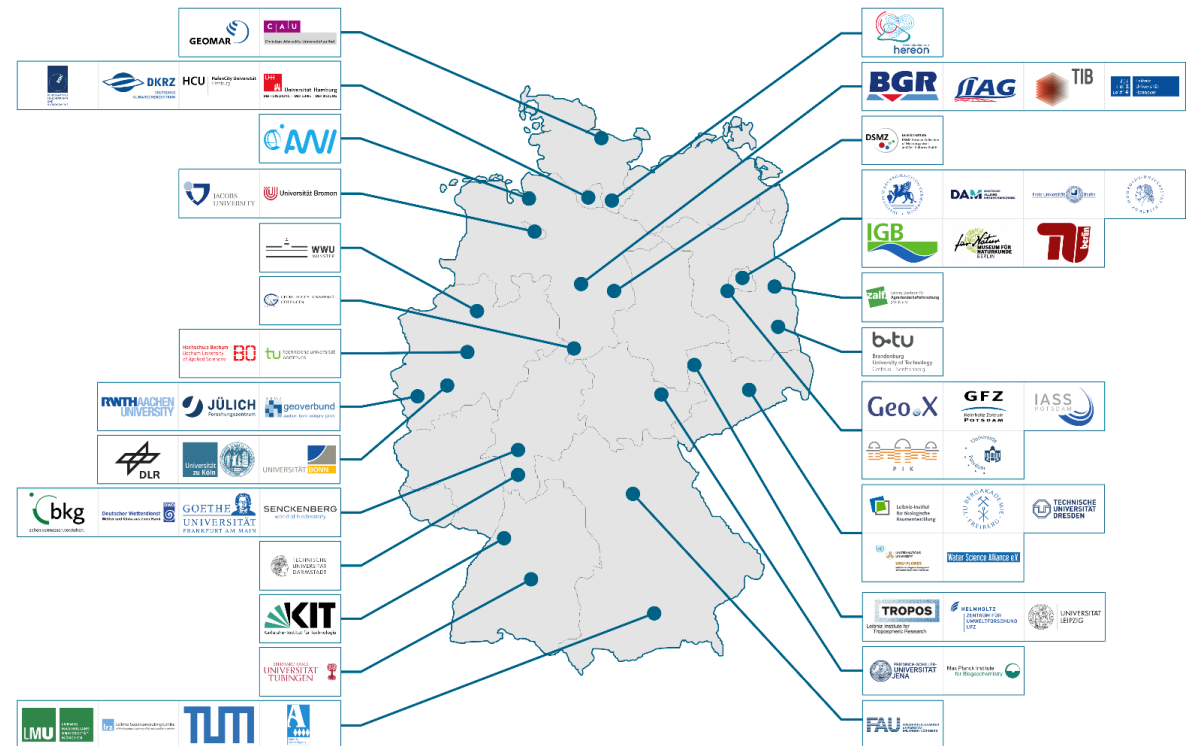


Who we are



- 59 institutions covering the breadth of Earth System Sciences (ESS)
 - Universities
 - Research Organizations
 - Infrastructure Providers
 - Governmental Institutions
 - Scientific Associations & Networks

- Established 2018 as **Open Consortium** and the ESS branch in the NFDI **DFG-Funded since 10/2021**



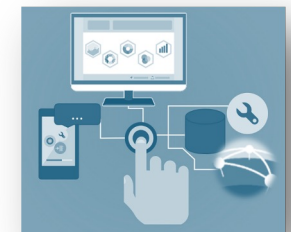
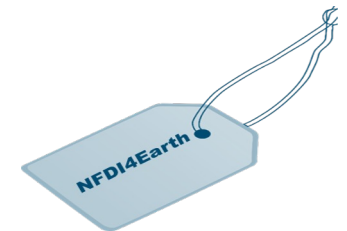
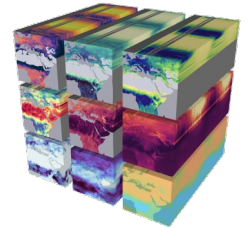
Our Issues

- Large, diverse number of data services (>> 100) and activities related to ESS...only a few sustainable
- Data heterogeneity
quality, curation levels, licenses, semantics, sizes, repeatability, scale, spatio-temporal resolutions,...
- Different data cultures and velocities to FAIR
from data streams to long tail data; RDM experts and novices
- Incomplete support along the data lifecycle
- Lack of support (platforms, tools) for joint and collaborative interpretation of heterogeneous and decentralized data

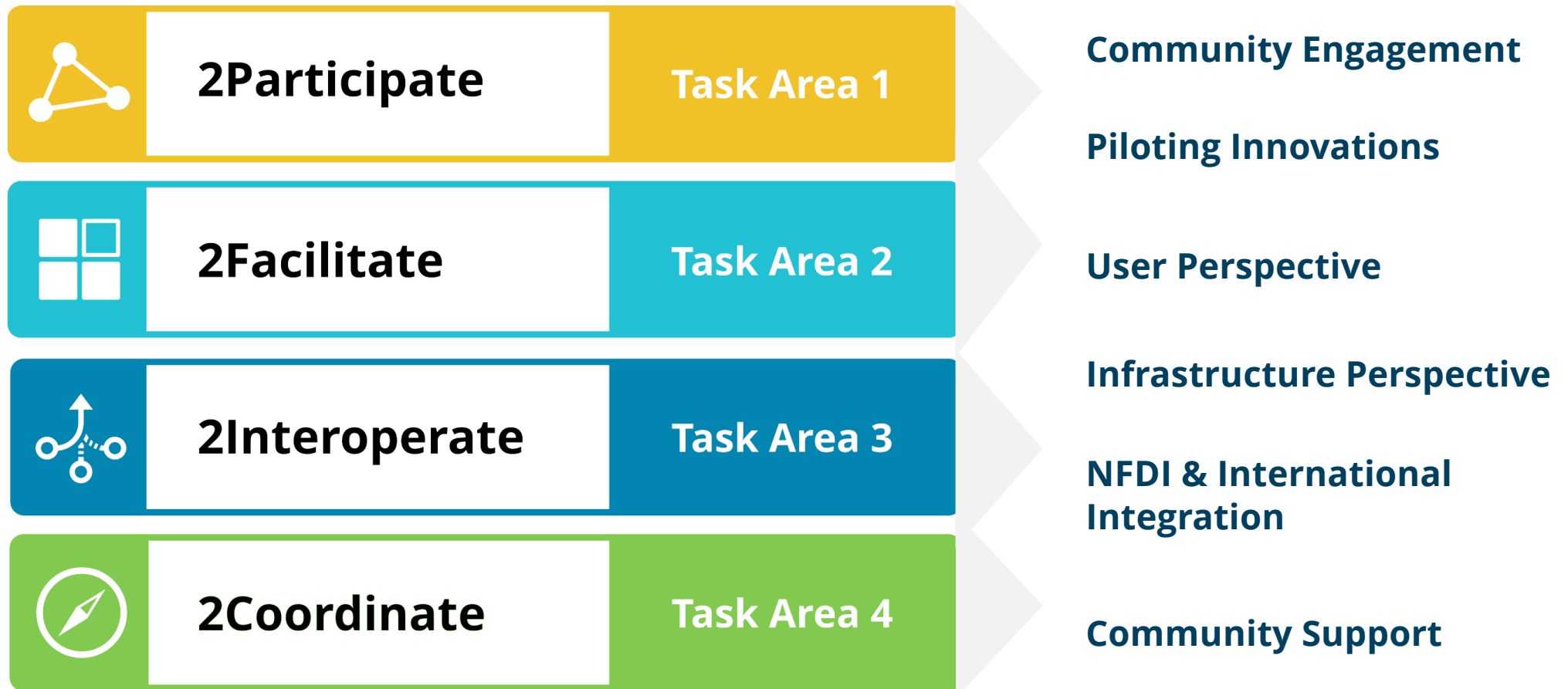


Key Goals

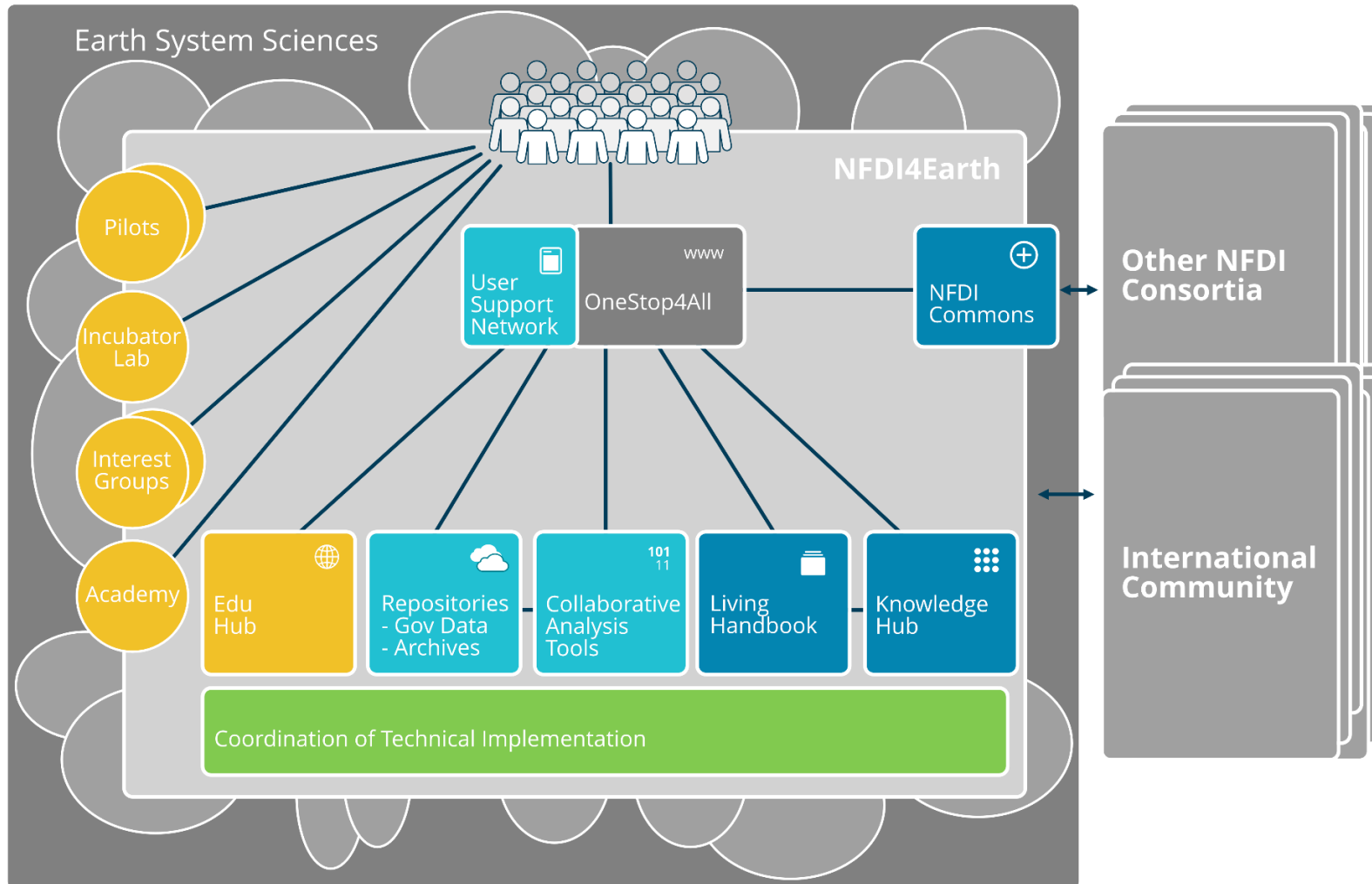
- *One Community* approach to sustainable, Open and FAIR Research Data Management in Earth System Sciences (ESS)
- Community driven agile development of innovative platforms for data integration and collaborative data analysis
- Qualification for people, data, tools and services as a basis for FAIR RDM and viability
- *OneStop4All* and *User Support Network* for ESS RDM as integral part of NFDI and International Infrastructures
- Key-driver of the build-up and operation of the NFDI



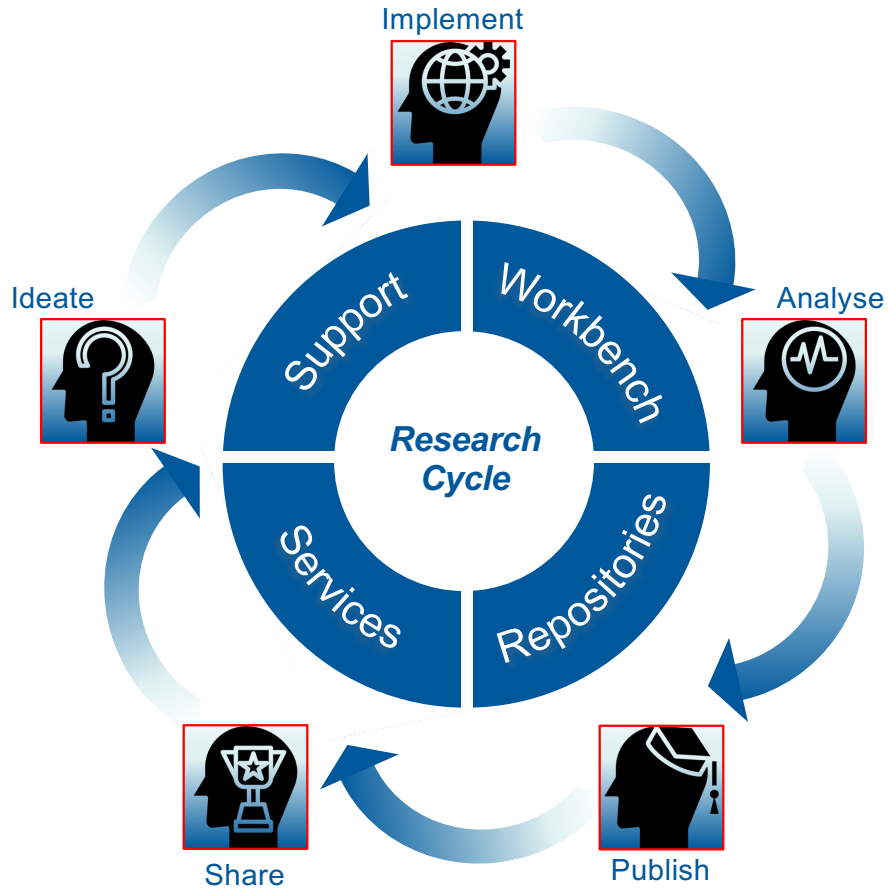
NFDI4Earth – Strategy 2021-26



Implementing the NFDI4Earth Strategy



Starting point is the research cycle ...

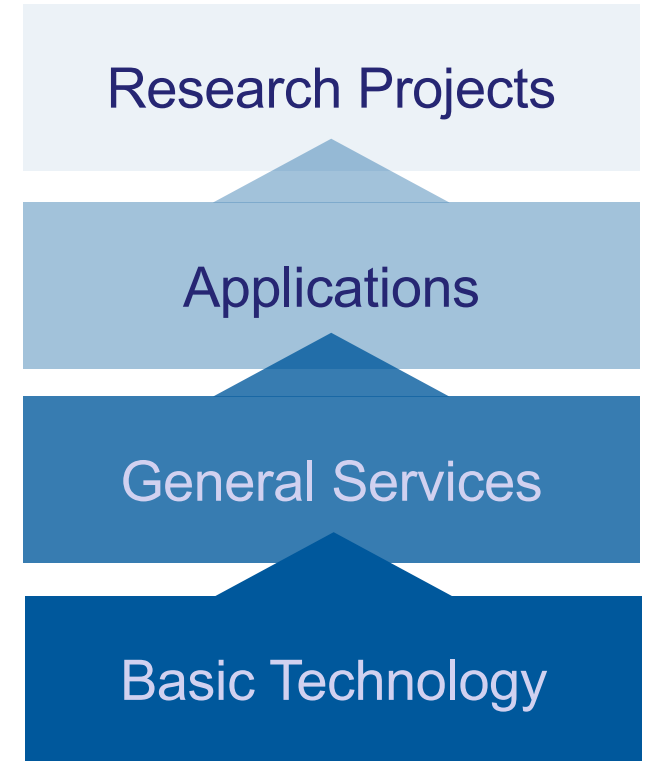


Almost all artefacts of scientific work today are digital.

Our goal is to build a digital ecosystem which allows to move through the steps of the cycle in arbitrary ways.

Therefore, we must integrate services on all levels of the technological stack.

Technological Stack



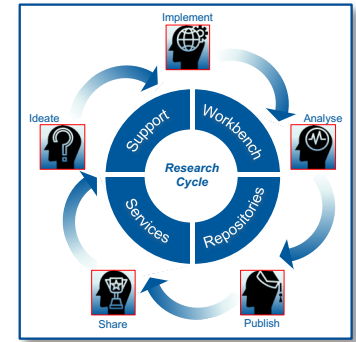
Our vision: a Digital Research Ecosystem ...

sample generation sample analysis

data analysis

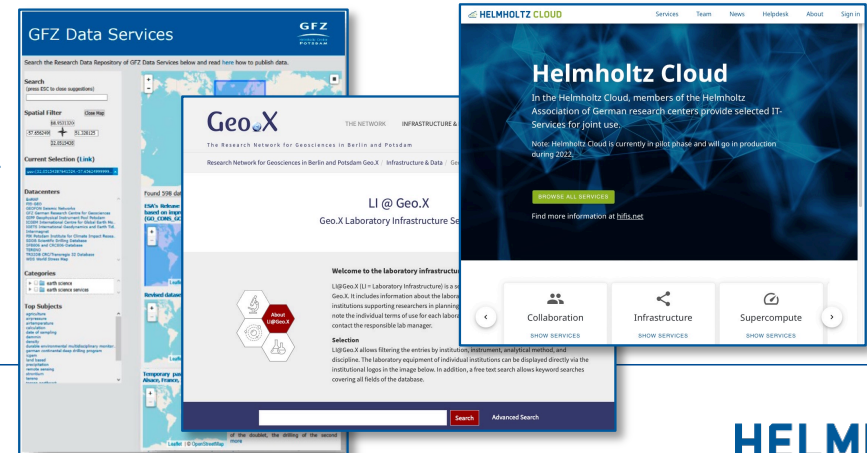
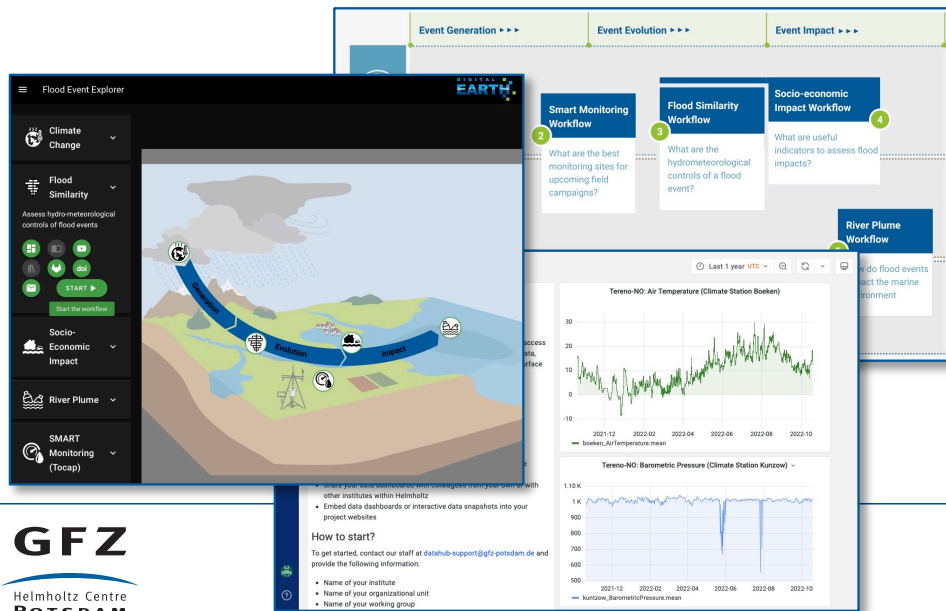
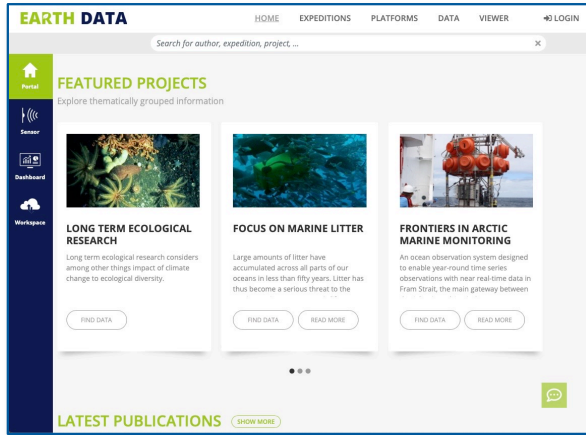
presentation publication

Sensor Management Instrument Management Sample Management Identity Management Data Repository Paper Repository

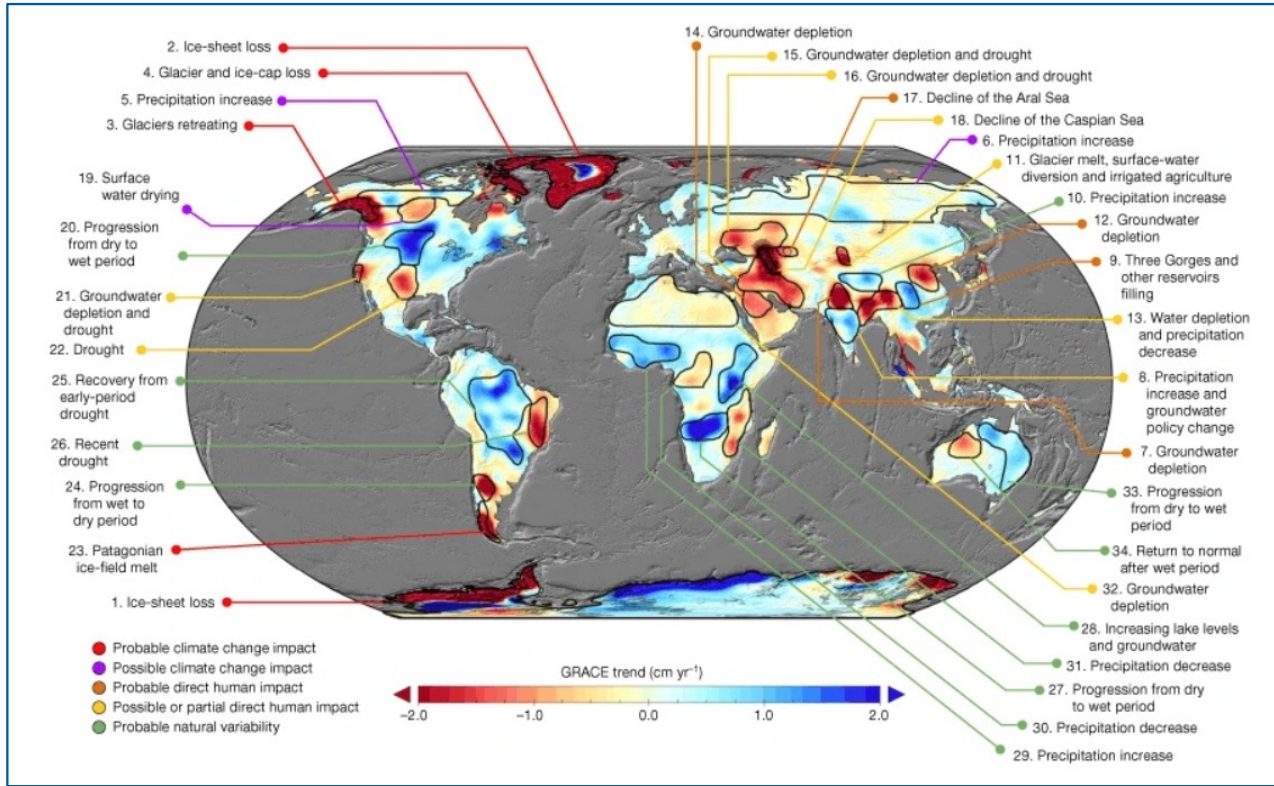


The **cloud** herby abstracts users from the explicit instantiation of technology.

Digitalization in the Research Field Earth & Environment



An example – analysis of terrestrial water storage



Rodell et al., Nat 2018

ANALYSIS

Corrected: Author Correction

<https://doi.org/10.1038/s41586-018-0123-1>

Emerging trends in global freshwater availability

M. Rodell^{1*}, J. S. Famiglietti^{2,5}, D. N. Wiese², J. T. Reager², H. K. Beaudoin^{3,4}, F. W. Landerer² & M.-H. Lo⁴

Freshwater availability is changing worldwide. Here we quantify 34 trends in terrestrial water storage observed by the Gravity Recovery and Climate Experiment (GRACE) satellites during 2002–2016 and categorize their drivers as natural interannual variability, unsustainable groundwater consumption, climate change or combinations thereof. Several of these trends had been lacking thorough investigation and attribution, including massive changes in northwestern China and the Okavango Delta. Others are consistent with climate model predictions. This observation-based assessment of how the world's water landscape is responding to human impacts and climate variations provides a blueprint for evaluating and predicting emerging threats to water and food security.

Analysis is build on using data of the GRACE satellite mission.

The example ctd' – GRACE/GRACE-FO satellite mission



GravIS
Gravity Information Service
GFZ GERMAN RESEARCH CENTRE FOR GEOSCIENCES

Dataset
GravIS GravIS RL06 Continental Water Storage Anomalies

File
Download data (FTP)
Previous versions (FTP)
License: CC BY 4.0

Dataset Description
Documented by
Dobslaw, H., Boergens, E. (2020) GFZ/COST-G GravIS Level-3 Products: Terrestrial Water Storage Anomalies (Technical Note)
GravIS Terrestrial Water Storage Anomaly Website

Related Work
Derived from
Dahle, C., & Murböck, M. (2019). *Post-processed GRACE/GRACE-FO crosspolarizer GSM Coefficients GFZ RL06 (Level-2B Product)* [Data set]. GFZ Data Services. https://doi.org/10.5880/GravIS_01_L3_TWS

Variant Form of
Boergens, E., Dobslaw, H., & Dill, R. (2020). *COST-G GravIS RL01 Continental Water Storage Anomalies (Data set)*. GFZ Data Services. https://doi.org/10.5880/COST-G-GravIS_01_L3_TWS

References
GravIS Website

Abstract
GRACE/GRACE-FO Level-3 product based on GFZ RL06 Level-2B products (Dahle & Murböck, 2019) representing Terrestrial Water Storage (TWS) anomalies provided as 1° latitude-longitude grids as defined over all continental regions except Greenland and Antarctica. The TWS anomaly grids are provided in netCDF format divided into yearly batches. The files each contain four different variables:
1) 'tws': gravity-based TWS
2) 'tws_err': gravity-based TWS uncertainties
3) 'leakage': spatial leakage contained in TWS
4) 'model_atmosphere': background model atmospheric mass
These Level-3 products are visualized at GFZ's web portal GravIS (<http://gravis.gfz-potsdam.de>). Link to data products: http://isdcftp.gfz-potsdam.de/gravis/COST-G/GravIS/Level-3/TWS/0000_release

Version History
9 December 2021:
Release of Version 0004. This is an update of Version 0003 of the same data set (see changelog).
9 September 2020:
Release of Version 0003. This is an update of Version 0002 of the same data set (see changelog).
9 June 2020:
Release of Version 0002. This is an update of Version 0001 of the same data set (see changelog).
All changes and updates are documented in the changelog available via the data download section. Previously released versions of this data set are available at http://isdcftp.gfz-potsdam.de/gravis/COST-G/Level-3/TWS/0000_release

Authors
Boergens, Eva GFZ German Research Centre for Geosciences, Potsdam, Germany
Dobslaw, Henryk GFZ German Research Centre for Geosciences, Potsdam, Germany
Dill, Robert GFZ German Research Centre for Geosciences, Potsdam, Germany

Contact
Boergens, Eva; GFZ German Research Centre for Geosciences, Potsdam, Germany;

Contributors
Dahle, Christoph; Murböck, Michael

Keywords
Gravity Recovery And Climate Experiment (GRACE), GRACE Follow-on (GRACE-FO), Level-3, Mass, Mass Transport, Total Water Storage, Time Variable Gravity, Mass Balance, Satellite Geodesy

GMD Science Keywords
EARTH SCIENCES > SOLID EARTH > GRAVITY/GEOMETRICAL FIELD
Earth Observation Satellites > NASA Earth System Science Pathfinder > GRACE

More Metadata
dataset: view online / download xml
isdc0135: view online / download xml

From data ...



... to data products ...

GFZ
Helmholtz Centre
Potsdam

GravIS
Gravity Information Service
GFZ GERMAN RESEARCH CENTRE FOR GEOSCIENCES

TERRESTRIAL WATER STORAGE | OCEAN BOTTOM PRESSURE | ANTARCTIC ICE-MASS CHANGE | GREENLAND ICE-MASS CHANGE | CORRECTIONS AND AUXILIARY PRODUCTS | RELATED LINKS

Climatically similar regions | Water Storage

Terrestrial Water Storage Anomalies
Terrestrial Water Storage (TWS) variability as observed by GRACE/GRACE-FO is an integrated signal from a number of different processes. The following individual components are provided as gridded Level-3 TWS product (available at ISDC for GFZ and COST-G) and are displayed in the map on the left:
GRACE: Water Storage
Water mass anomalies expressed in terms of equivalent water height from all water storage compartments including snow, surface water, soil moisture, and deep groundwater; not corrected for spatial leakage.
GRACE: Water Storage Uncertainty
Time-variable component of the uncertainty estimate for the GRACE-based water storage variability, given as standard deviation per grid point.
GRACE: Spatial Leakage
Terrestrial water storage variations likely caused by spatial leakage; intended as optional correction for the water storage component.
Model: Atmospheric Mass
Atmospheric mass variability as represented in the non-tidal de-aliasing model AOD1B expressed in terms of equivalent water height.

GRACE/GRACE-FO GFZ RL06 August 2022

8
6
4
2
0
-2
-4
-6
-8
-10
-12

2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022

Contact: Eva Boergens (boergens at) gfz-potsdam.de

GFZ | © Helmholtz Centre Potsdam
GFZ German Research Centre for Geosciences

HELMHOLTZ RESEARCH FOR GRAND CHALLENGES | AWI | TECHNISCHE UNIVERSITÄT DRESDEN | GFZ GERMAN RESEARCH CENTRE FOR GEOSCIENCES



... to services!

First challenge – finding data

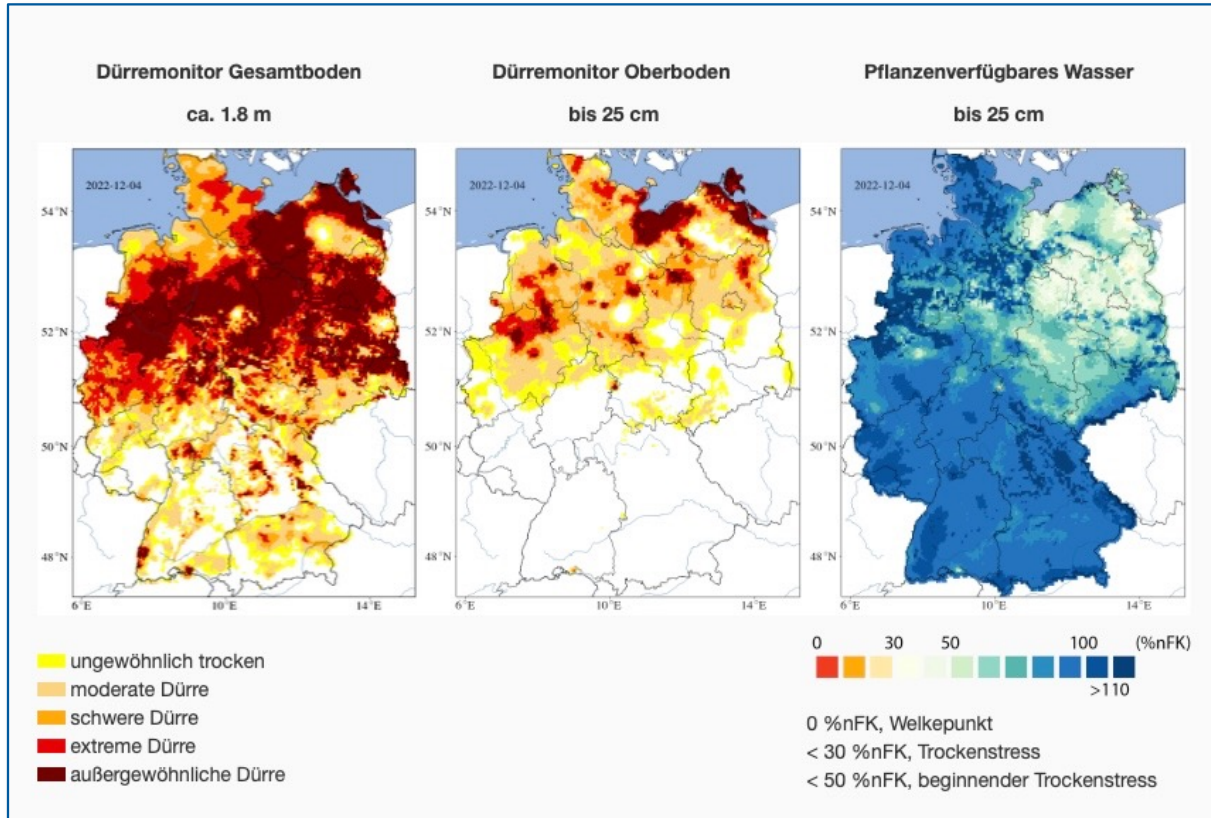
There is really a lot out there ...

... organized by scientific domains, organizations, (inter-)national agents, public authorities etc.

Integrative services and metadata search engines do help!

Second challenge – showing vs **accessing data**

UFZ Dürremonitor Deutschland



<https://www.ufz.de/index.php?de=37937>

Useful services are also a popular gateway to data.

Can the data actually be used ... ?

... in the sense of FAIR?

Download der Karten, Download der Daten & Nutzungsbedingungen

Nutzungsbedingungen
Die Rechte der Grafiken und Karten des UFZ-Dürremonitors liegen, sofern nicht anders angegeben, beim UFZ. Die Grafiken können im Rahmen von Wissenschaft und Forschung sowie für redaktionelle Zwecke unter Angabe des folgenden Vermerks uneingeschränkt genutzt werden: UFZ-Dürremonitor/ Helmholtz-Zentrum für Umweltforschung. Bitte beachten Sie, dass der Quellenvermerk direkt an den Karten stehen muss.

Download der Karten	Download der Daten & Nutzungsbedingungen
Karten Gesamtboden Deutschland (PNG, PDF) Nordrhein-Westfalen (PNG, PDF) Hessen (PNG, PDF) Baden-Württemberg (PNG, PDF) Bayern (PNG, PDF) Mecklenburg-Vorpommern (PNG, PDF) Sachsen (PNG, PDF) Sachsen-Anhalt (PNG, PDF) Thüringen (PNG, PDF) Berlin + Brandenburg (PNG, PDF) Schleswig-Holstein + Hamburg (PNG, PDF) Niedersachsen + Bremen (PNG, PDF) Rheinland-Pfalz + Saarland (PNG, PDF) Norddeutscher Raum (PNG, PDF) Mitteldeutscher Raum (PNG, PDF)	Karten Oberboden Deutschland (PNG, PDF) Nordrhein-Westfalen (PNG, PDF) Hessen (PNG, PDF) Baden-Württemberg (PNG, PDF) Bayern (PNG, PDF) Mecklenburg-Vorpommern (PNG, PDF) Sachsen (PNG, PDF) Sachsen-Anhalt (PNG, PDF) Thüringen (PNG, PDF) Berlin + Brandenburg (PNG, PDF) Schleswig-Holstein + Hamburg (PNG, PDF) Niedersachsen + Bremen (PNG, PDF) Rheinland-Pfalz + Saarland (PNG, PDF) Norddeutscher Raum (PNG, PDF) Mitteldeutscher Raum (PNG, PDF)
Karten pflanzenverfügbares Wasser Deutschland (PNG, PDF) Nordrhein-Westfalen (PNG, PDF) Hessen (PNG, PDF) Baden-Württemberg (PNG, PDF) Bayern (PNG, PDF) Mecklenburg-Vorpommern (PNG, PDF) Sachsen (PNG, PDF) Sachsen-Anhalt (PNG, PDF) Thüringen (PNG, PDF) Berlin + Brandenburg (PNG, PDF) Schleswig-Holstein + Hamburg (PNG, PDF) Niedersachsen + Bremen (PNG, PDF) Rheinland-Pfalz + Saarland (PNG, PDF) Norddeutscher Raum (PNG, PDF) Mitteldeutscher Raum (PNG, PDF)	Legenden ▲ Oberboden und Gesamtboden ▲ pflanzenverfügbares Wasser

Download der Daten

Aktuell:
Die aktuellen Daten des SMI Bodenfeuchteindex und pflanzenverfügbaren Wassers der letzten 14 Tage können Sie im netCDF-Format herunterladen:
 → SMI Oberboden bis 25 cm Tiefe
 → SMI Gesamtboden
 → pflanzenverfügbares Wasser bis 25 cm Tiefe (%nFK)

Historisch:
Die monatlichen SMI-Daten 1951-2021 können Sie im NetCDF-Format herunterladen. Die statistische Basis umfasst den 1951-2019. Die Daten basieren auf Zink et al. 2015 (ERL) und enthalten den Bodenfeuchteindex (SMI, soil moisture index) skaliert zwischen 0-1. Die Klassifikationsgrenzen für Dürren finden sich im Abschnitt unten 'Was bedeutet Dürre?'. Die Datensätze enthalten sowohl Koordinaten in den Formaten Gauß-Krüger Zone 4 (EPSG: 31468) (variablen easting/northing) als auch Lat/Lon Informationen (EPSG:4326).

Oberboden bis 25 cm Tiefe:
▲ SMI_SMI_LO2_Oberboden_monatlich_1951-01_2021-12_inv.nc (87.1 MB)

Gesamtboden:
▲ SMI_SMI_Lat_Gesamtboden_monatlich_1951-01_2021-12_inv.nc (87.5 MB)

Dürremagnituden und -intensitäten in der Vegetationsperiode (April - Oktober) 1952-2022
Die Dürremagnitudo ist ein dimensionsloses Maß, um die Stärke von Dürren unterschiedlicher Jahre verglichen zu können. In die Berechnung fließt die Länge der Dürreperiode und die absolute Trockenheit im zeitlichen Verlauf ein. Die Zahl steigt mit zunehmender Dürreintensität und zunehmender negativer Abweichung von 20 Prozent des Bodenfeuchteindex SMI. Im Vergleich zu den Dürremagnituden wird bei der Dürreintensität zusätzlich eine Normierung über die Zeit (also die Tage der Vegetationsperiode) vorgenommen. Damit kann die Dürreintensität einen Maximumwert von 0,2 erreicht werden.
Eine Beschreibung des Ansatzes findet sich in: Samaligo et al (2013) und Böing et al (2022).

Oberboden bis 25 cm Tiefe:
▲ Dürremagnitudo_Oberboden_1952-2022_Apr-Okt.nc
▲ Dürreintensität_Oberboden_1952-2022_Apr-Okt.nc

Digital Earth Flood Event Explorer – bridging scales!?

Significant advances in Earth system understanding will only be achieved through better integration of data and knowledge from the different Earth science disciplines and Earth compartments. **Digital Earth**

ELWIS
Wasserstände und Schiffahrtsverwaltung des Bundes

Wasserstände und Schiffahrtsverwaltung des Bundes
Binnenschifffahrt Seeschifffahrt Sportschifffahrt Untersuchung/Eichung Schifffahrtsrecht Service

Ort in Deutschland

Gültigkeitszeitraum auswählen: 06.12.2022 - 05.01.2023

100 km

1 4387833

schifffahrtsrelevanten Pegeln

E-Mail-Service ELWIS-Abo zur Verfügung:

Do.	Fr.	Sa.	So.	Mo.	Heute
01.12.22	02.12.22	03.12.22	04.12.22	05.12.22	06.12.22
141 (+1)	134 (-7)	127 (-7)	117 (-10)	118 (-1)	125 (+7)
124 (-4)	120 (-4)	116 (-4)	104 (-12)	104 (+/-0)	107 (+3)
122 (+13)	120 (-2)	117 (-3)	113 (-4)	103 (-10)	99 (-4)
137 (+4)	145 (+8)	145 (+/-0)	139 (-6)	133 (-6)	126 (-7)
118 (-2)	125 (+7)	127 (+2)	126 (-1)	122 (-4)	115 (-7)
101 (-2)	106 (+5)	110 (+4)	108 (-2)	105 (-3)	101 (-4)
87 (-2)	90 (+3)	95 (+5)	95 (+/-0)	92 (-3)	90 (-2)
91 (-2)	91 (+/-0)	96 (+5)	97 (+1)	96 (-1)	96 (+/-0)

<https://www.elwis.de>

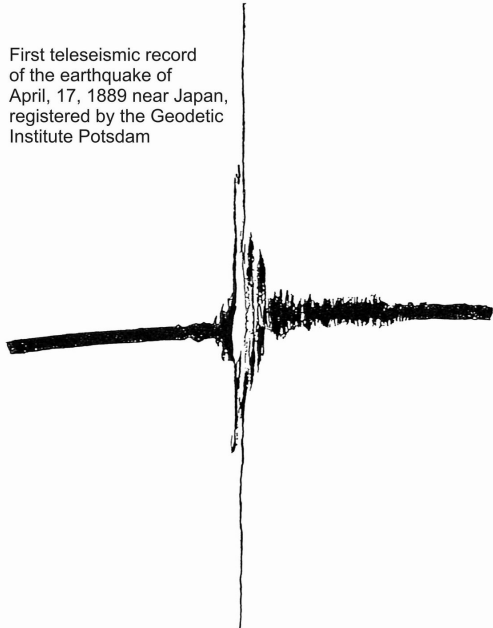
<https://via.bund.de/wsv/elwis>

Third challenge – **interoperable data**

You never know who will do what with your data ...

Aufzeichnung des Erdbebens von 1889

First teleseismic record of the earthquake of April, 17, 1889 near Japan, registered by the Geodetic Institute Potsdam



GFZ
Helmholtz-Zentrum
POTSDAM

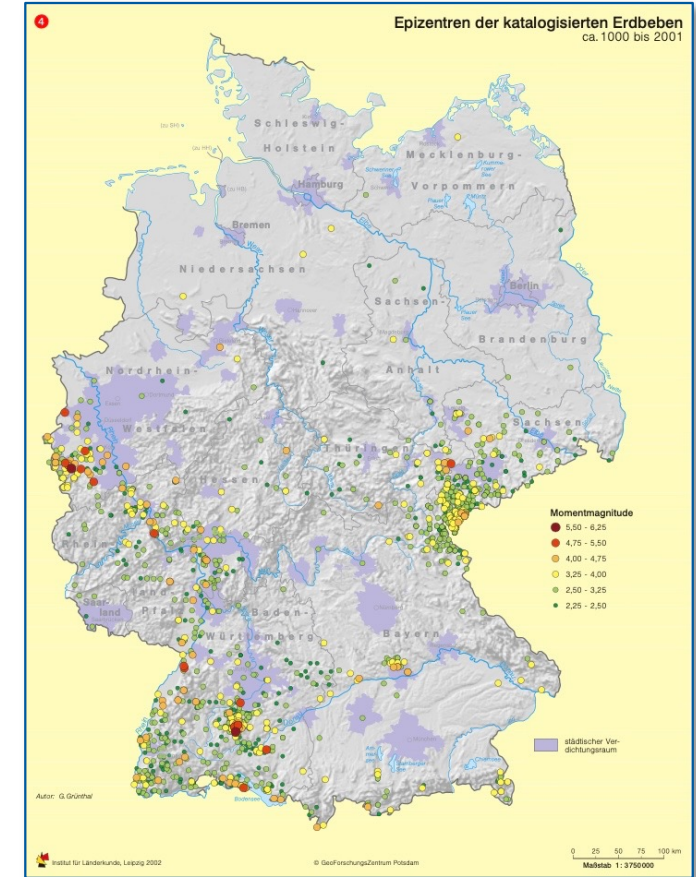
Fourth challenge –
reusable data

Working in geological timescales implies that mostly, proxies have to be used.

Historical recording of earthquake in the Bielefeld area 1612



Vogt & Grünthal, Geowissenschaften 1994



G. Grünthal, in: Nationalatlas Bundesrepublik Deutschland Band 2, 44-45

Long-term preservation – scientists use what they can get ...

That's why they tend to keep everything!

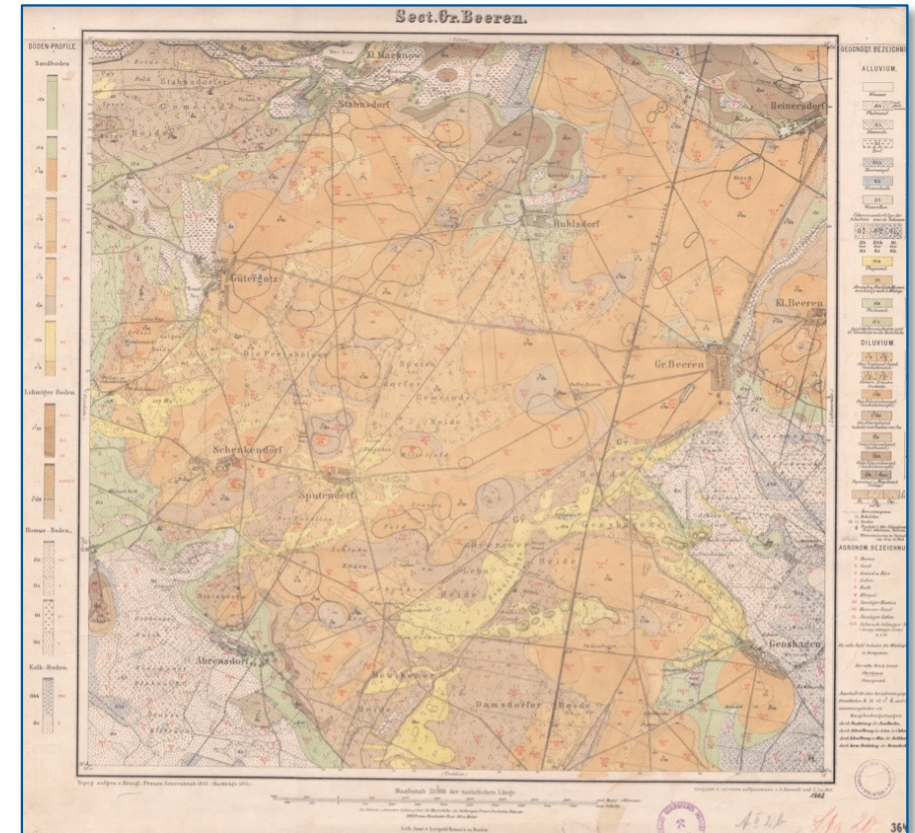
Historical recording of earthquake in the Bielefeld area 1612



Vogt & Grünthal, Geowissenschaften 1994

The clue of the data perspective is to use what is available rather than to wait for what can be created.

The treasure of time series!



Geologische Karte, G. Berendt 1875, in: Digitales Brandenburg

Preservation in science ... more than 10 years?

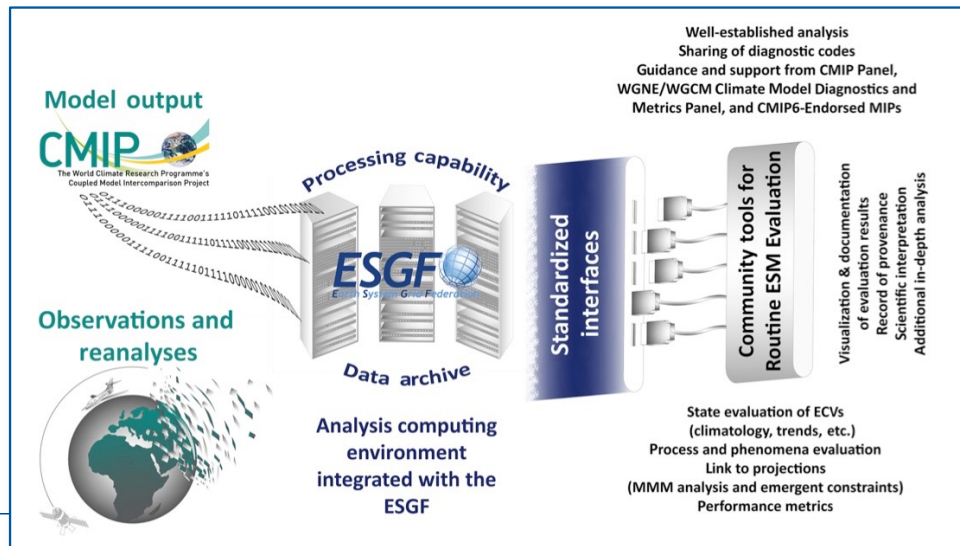
Geosci. Model Dev., 11, 3659–3680, 2018
 https://doi.org/10.5194/gmd-11-3659-2018
 © Author(s) 2018. This work is distributed under
 the Creative Commons Attribution 4.0 License.



Requirements for a global data infrastructure in support of CMIP6

Venkatramani Balaji^{1,2}, Karl E. Taylor³, Martin Juckes⁴, Bryan N. Lawrence^{5,4}, Paul J. Durack³, Michael Lautenschlager⁶, Chris Blanton^{7,2}, Luca Cinquini⁸, Sébastien Denvil⁹, Mark Elkington¹⁰, Francesca Guglielmo⁹, Eric Guilyardi^{9,4}, David Hassell⁴, Slava Kharin¹¹, Stefan Kindermann⁶, Sergey Nikonov^{1,2}, Aparna Radhakrishnan^{7,2}, Martina Stockhause⁶, Tobias Weigel⁶, and Dean Williams³

Coupled Model Intercomparison Project (CIMP)



Eyring et al., Earth Syst Dynam 2016

nature
human behaviour

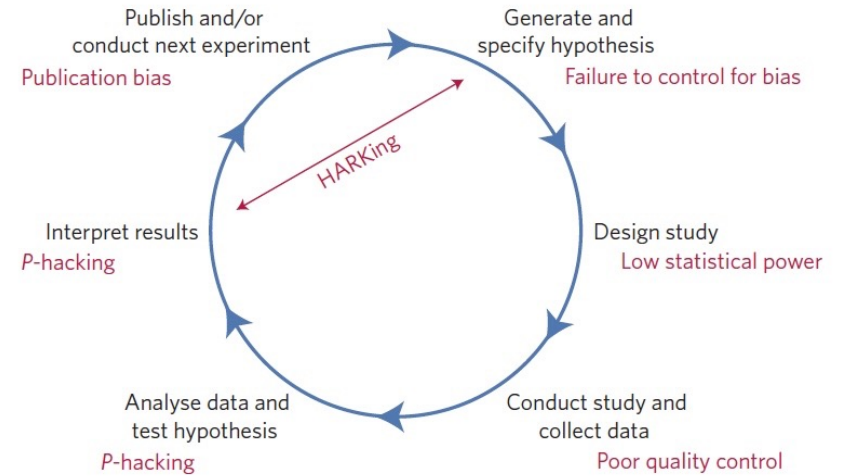
PERSPECTIVE

PUBLISHED: 10 JANUARY 2017 | VOLUME: 1 | ARTICLE NUMBER: 0021

OPEN

A manifesto for reproducible science

Marcus R. Munafò^{1,2*}, Brian A. Nosek^{3,4}, Dorothy V. M. Bishop⁵, Katherine S. Button⁶, Christopher D. Chambers⁷, Nathalie Percie du Sert⁸, Uri Simonsohn⁹, Eric-Jan Wagenmakers¹⁰, Jennifer J. Ware¹¹ and John P. A. Ioannidis^{12,13,14}



Munafò et al., Nat Hum Beh 2017

„We need to talk about reproducibility“

Crick, Hall and Ishtiaq cite studies, finding that **50% of published studies** (including top-tier journals) **cannot be repeated** by an industrial lab.



conclusion

- just publishing linked data is not enough
- „set the code free“
- welcome Web 2.0 technologies (use and share workflows, provide web services, use cloud ...)

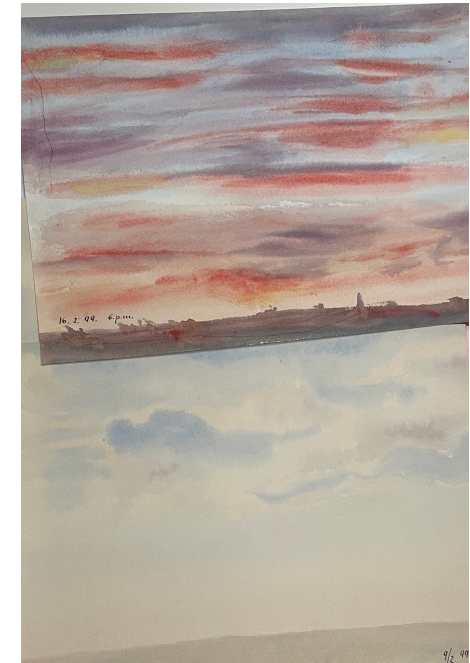
To what **level of detail** do we need to keep data and what are appropriate **time scales**?

Scientific data ... high diversity & low standardization

- need to deal with the „long-tail“ of research data
- lack of comprehensive criteria for long-term preservation
- joint issues: formats, accessibility, data migration

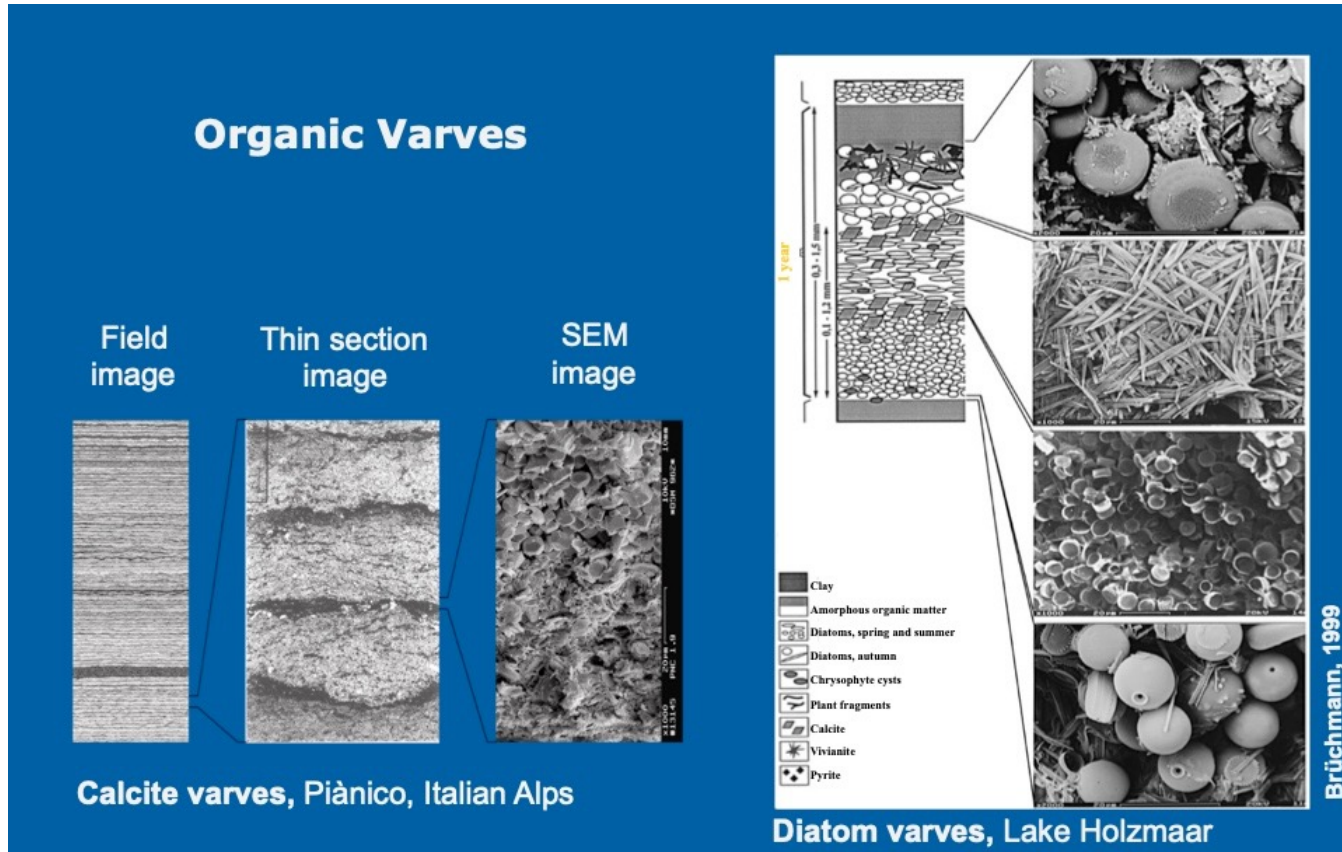
There are different challenges/requirements for

- closed projects
- real-time data products
- simulation data



Clouds above Potsdam
(by R. Süring)

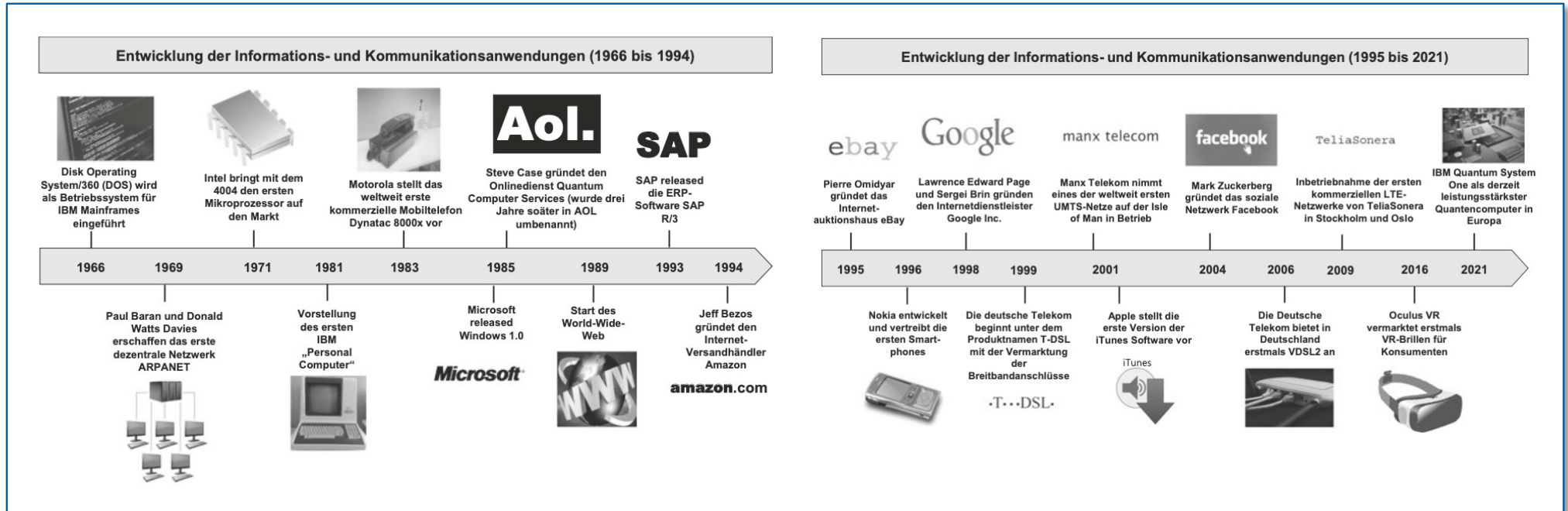
Digitalization – the challenge of formats and technology



paleoclimate reconstruction
from lake sediments

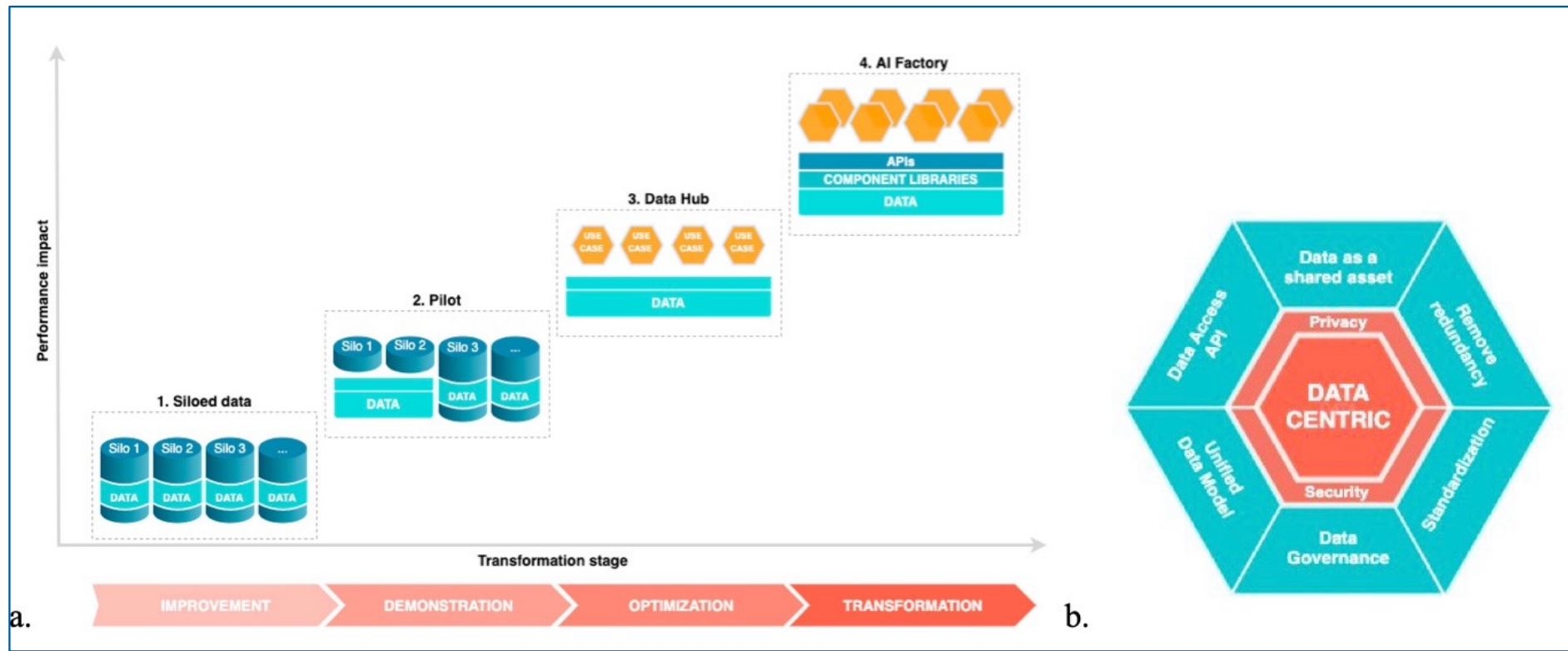
A. Brauer, IRTG-STRATEGY Lecture 2016

Digitalization – the challenge of formats and technology



B.W. Wirtz, E-Government, Springer 2022

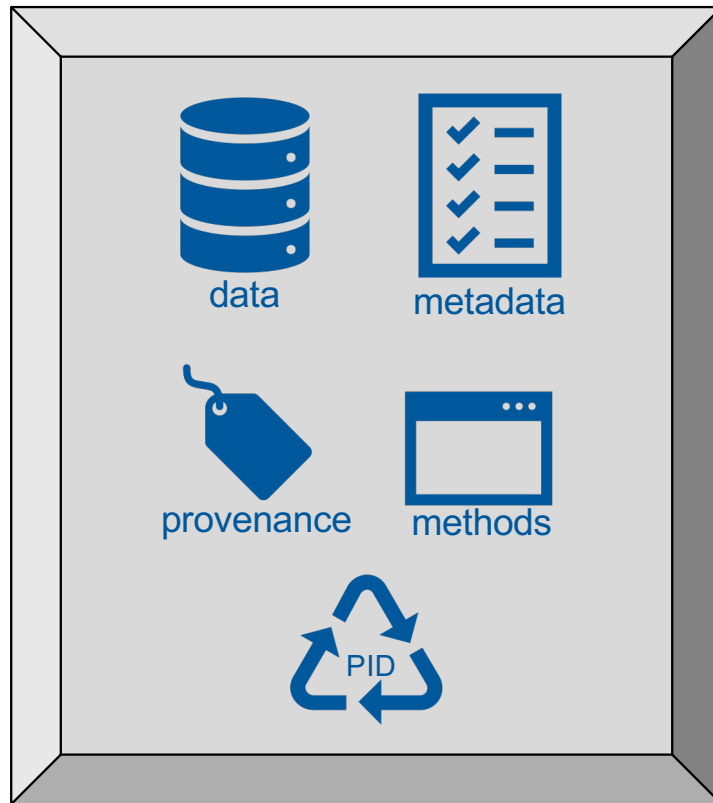
Towards a data centric architecture ...



Alvarez-Coello et al., Procedia Comp Sci 2021

- overcoming silos is the first step, only
- organizing silos as we did data means just scaling the challenge
- keeping all necessary information together will become a bottleneck

From thinking in applications to thinking in data ...



extended data containers

These are the key principles of the data centric manifesto:

1. Data is a key asset of any person, organization, and society.
2. Data is self-describing and does not rely on an application for interpretation and meaning.
3. Data is expressed in open, non-proprietary formats.
4. Access to and security of the data is a responsibility of the enterprise data layer or the personal data vault, and not managed by applications.
5. Applications are allowed to visit the data, perform their magic and express the results of their process back into the data layer.

The Data-Centric Manifesto

Thank you for your attention!