NFDI 4
BIODIVERSITY

www.nfdi4biodiversity.org
@NFDI4Biodiv
#NFDI4Biodiv

What happens to all the data?

Insights into the new National Research Data Infrastructure in Germany

19.01.2022

Research data

Research data is data created in the course of scientific activity,
e. g. through observations, experiments, simulations, surveys, interviews,
the study of sources, records, digitisation, or evaluations.

In actual research, one differentiates, although not always clearly,
between **primary research data** and **secondary research data**, which
documents and **contextualises** the process of creating primary data.

> Rat für Informationsinfrastrukturen, 2016

Origin of data

publicly funded for example ...

... at over 400 universities and technical colleges

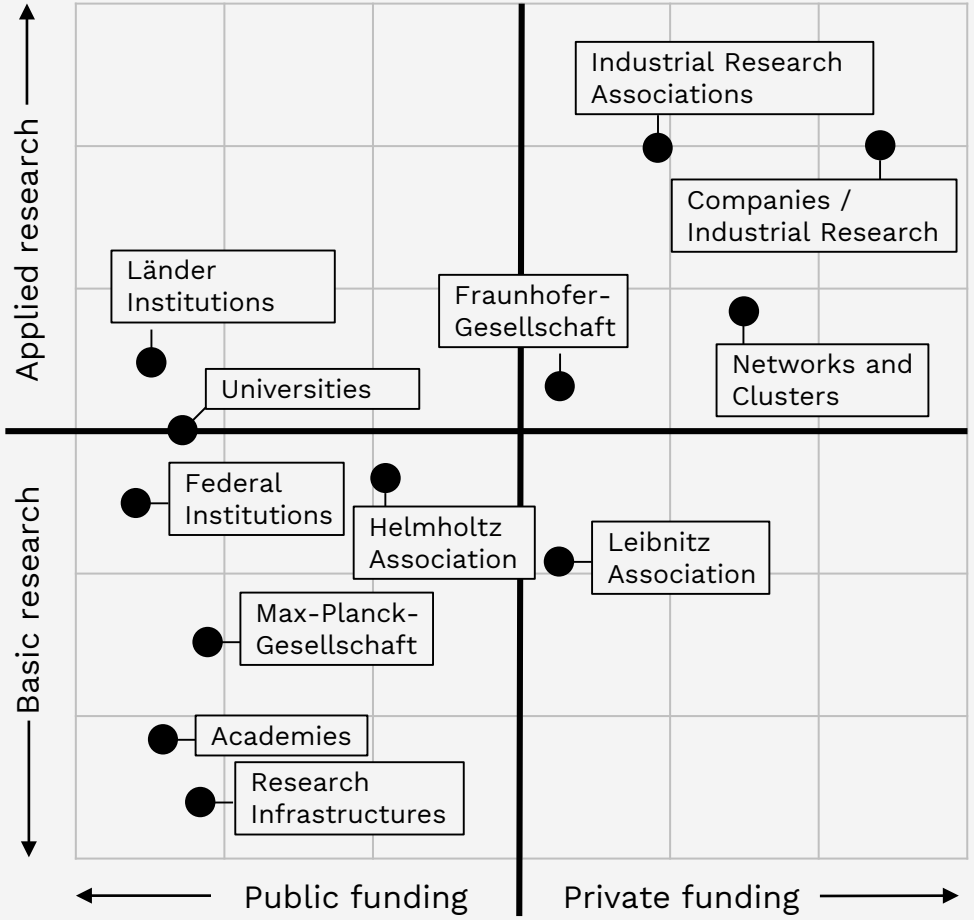
... at the non-university research institutions and the scientific academies

... in large, jointly funded research infrastructures

Area of activity of the Joint Science Conference of the federal government and federal states (Article 91b of the Basic Law)

German R&D under the microscope

A self-ranked assessment of how public and private research organizations in Germany are funded and their research priorities.



https://www.nature.com/articles/d41586-019-00910-7
Source: Federal Ministry of Education and Research

Acknowledging the need for coordinated action

The rising tide of data – nearly as many digital bits as there are stars in the universe.

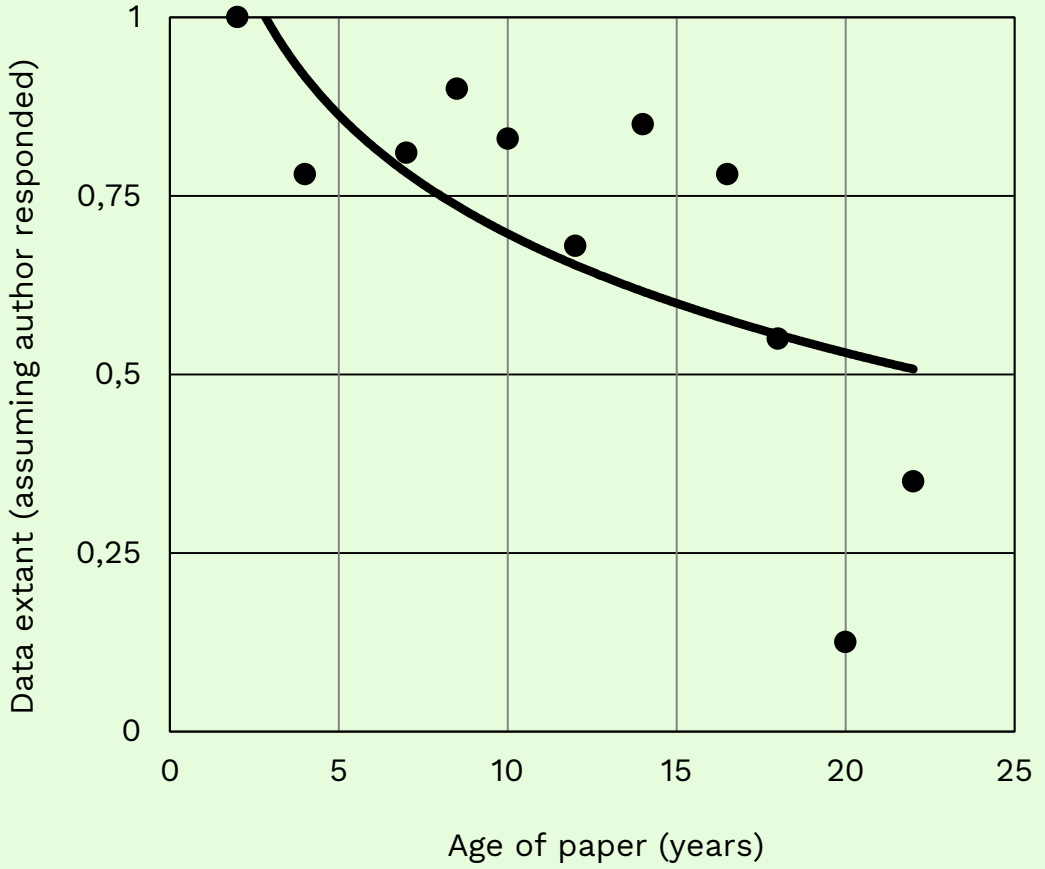
Source: IDC's Digital Universe Study 2014
https://www.emc.com/leadership/digital-universe/2014iview/executive-summary.html

Availability of Research Data with Time

- Data being lost are estimated to increase by **17% in every** year after publication.
- Find a working e-mail address for the first, last, or corresponding author fell by **7% per year**.
- Overall, we only received 19.5% of the requested data sets, and only 11% for articles published before 2000.

Missing data

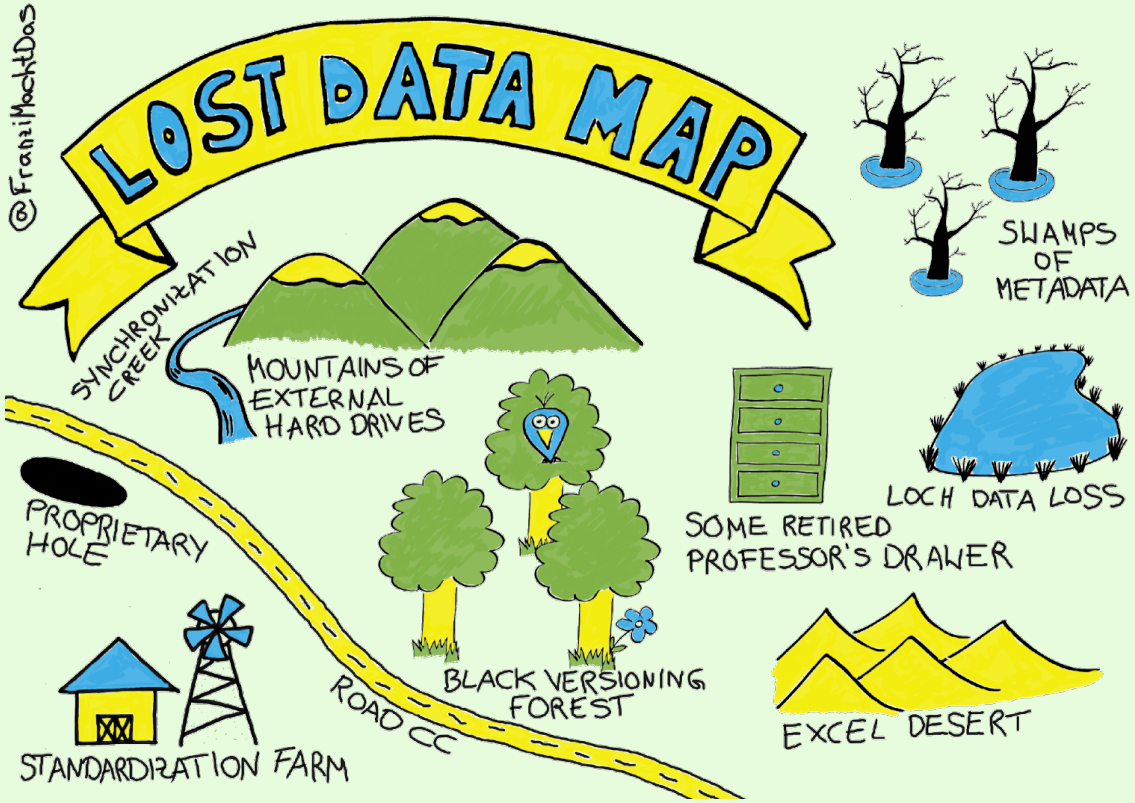
As research articles age, the odds of their raw data being extant drop dramatically.



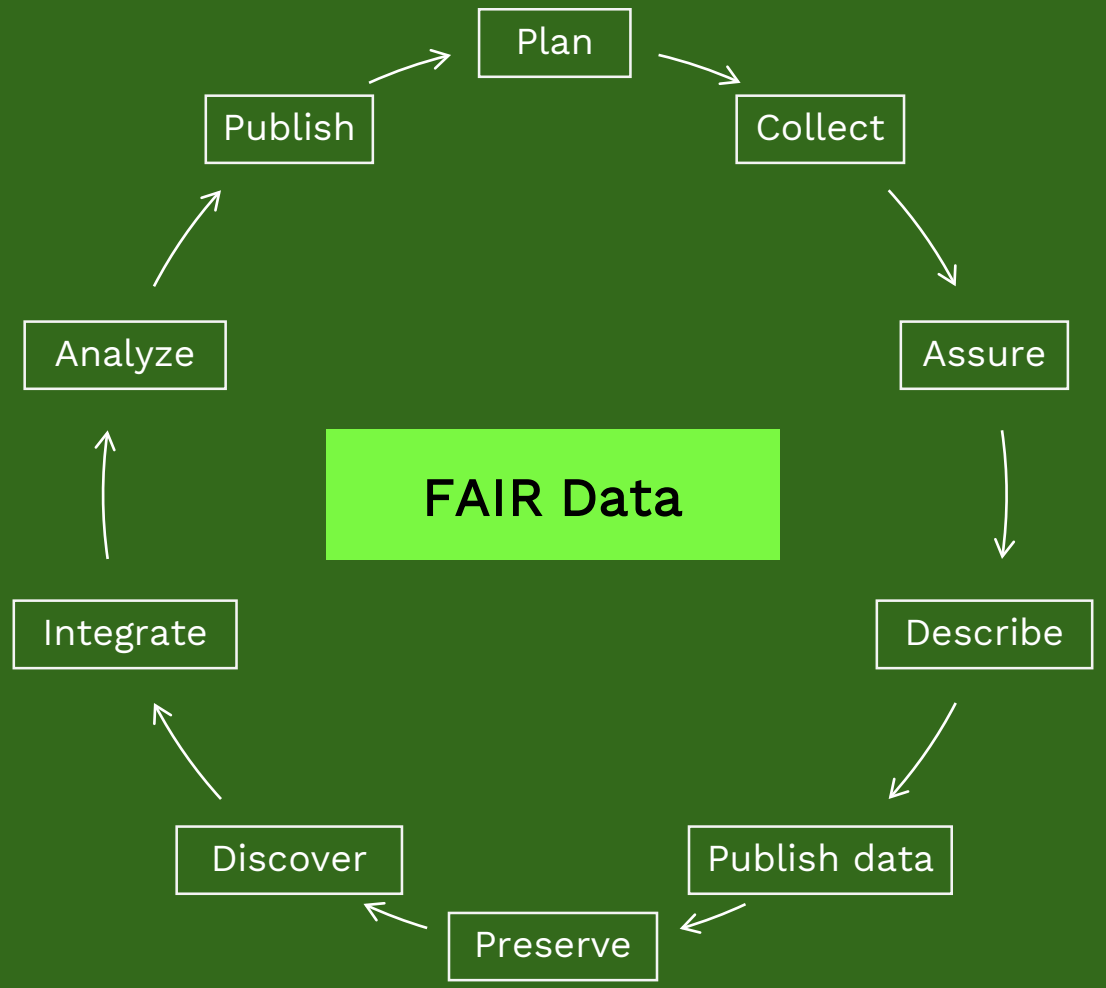
Vines, Timothy H. et al. Current Biology, 2014, Volume 24, Issue 1, 94-97

Our Goal: Making Research Data FAIR

F	FINDABLE
A	ACCESSIBLE
I	INTEROPERABLE
R	REUSABLE



Data Life Cycle



Successful community-driven federated initiatives in DE already exist

Three examples	
Since 2003: The German Astrophysical Virtual Observatory (GAVO) connects German research projects to the global VO	Funding ended 2017 now part of NFDI
Since 2004: The German Data Forum RatSWD develops a network of research data centres at public authorities and scientific institutions	Funding ended 2020 now part of NFDI
Since 2014: German Federation for Biological Data forms a network to host data from research projects on biodiversity	Funding ended 2021 now part of NFDI

NFDI Facts and Figures

- Joint funding by Federal government and the *Länder*, initial period 2019-2028 90 Mio EUR per year
- Founding of an association "NFDI e.V", to organise the data landscape
- Funding of up to 30 domain-related "consortia" (science-driven process)
- Scientific institutions join the NFDI association and support its goals

Source: DFG-Förderatlas 2021, Fig. 2.7

Structure of the NFDI - vision

Nationale Forschungsdaten Infrastruktur

From January 2023 ff.		
Up to **30** Consortia covering all research domains in germany	Several hundred organisations teaming up as suppliers and in decision processes	**> 1,000** Professional staff

Structure of the NFDI

January 2022

19 Consortia

220 member organisations

www.nfdi.org

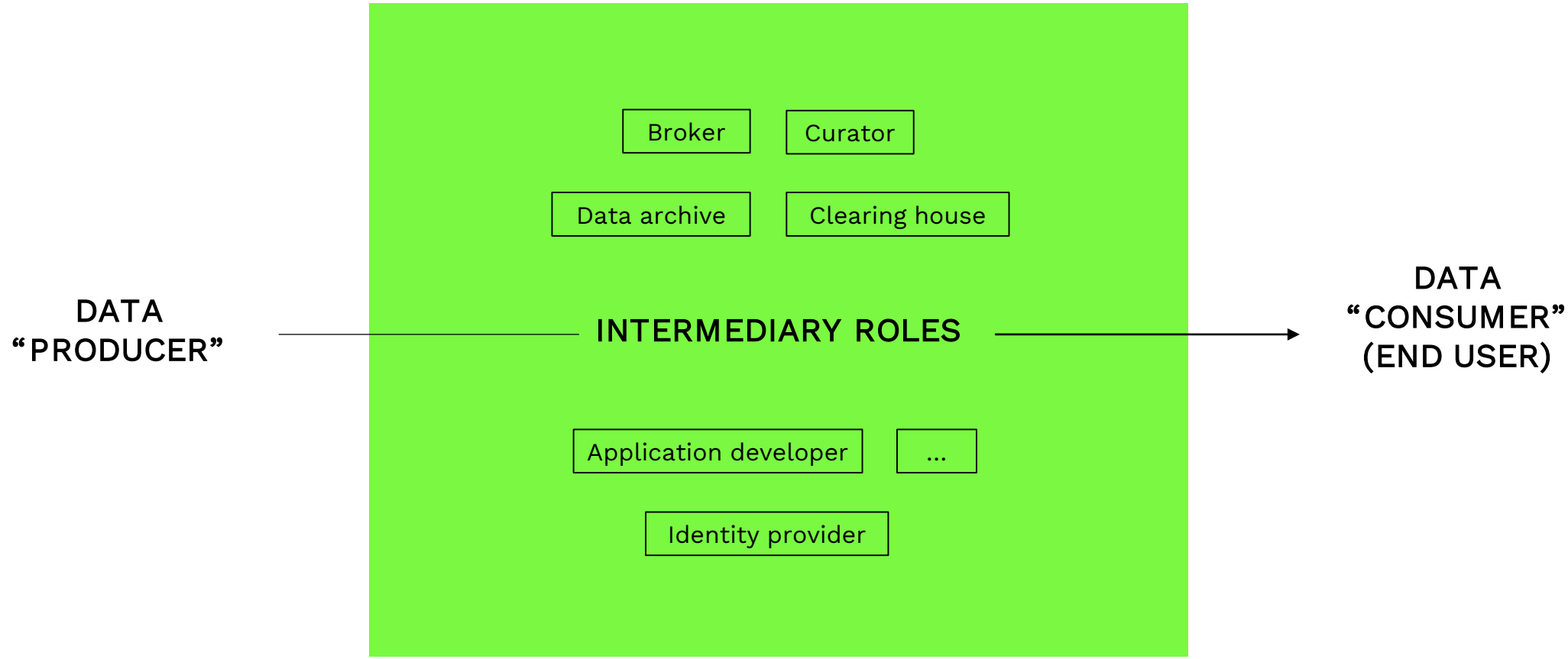
NFDI e.V. Association - Directorate -				
Data Plant	GHGA	NFDI4Biodiversity	NFDI4Cat	NFDI4Chem
NFDI4Culture	NFDI4Health	NFDI4Ing	KonsortSWD	BERD@NFDI
DAPHNE4NFDI	FAIRmat	MaRDI	NFDI4DataScience	NFDI4Earth
NFDI4Microbiota	NFDI4MatWerk	PUNCH4NFDI	Text+	...

What is NFDI for?

NFDI funding helps to organise
the roles between data producers and
data consumers in science

Organising the intermediary roles in NFDI

Rolemodel adapted from https://internationaldataspaces.org/use/reference-architecture/



NFDI4Biodiversity Community Services

Consortium for Biodiversity, Ecology and Environmental Data

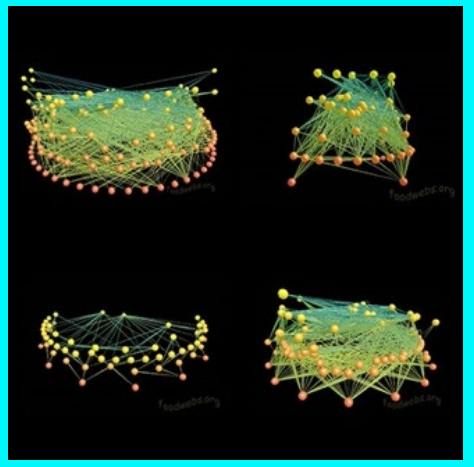
Facts of Biodiversity

Very specific data types – often combined with geospatial data. Time series, reference lists and taxonomies

species

genes

functions

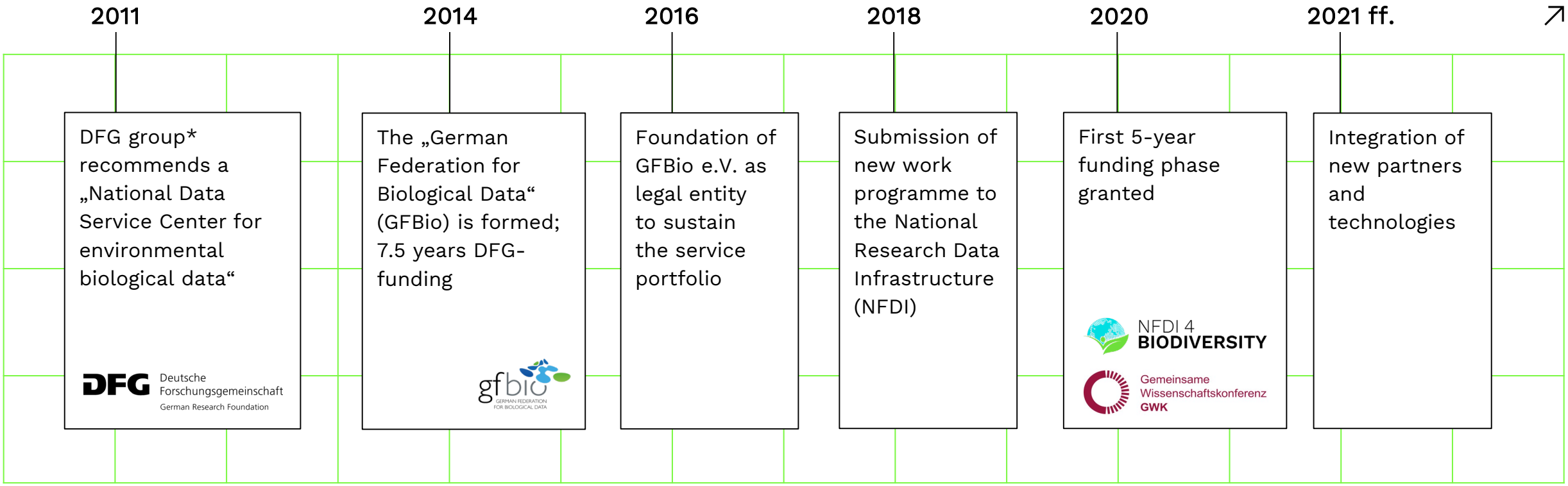


ecosystems

interactions

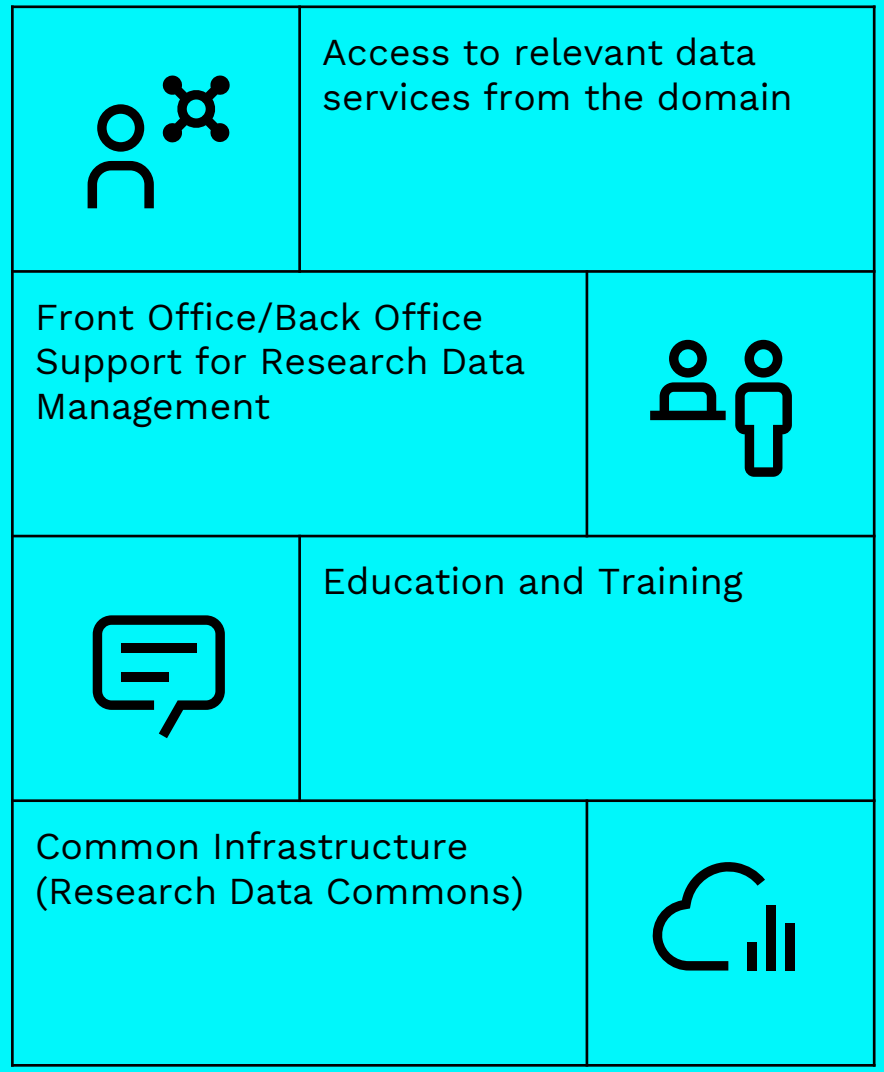
Joint services for environmental & biological data in Germany: A history

*AG Biodiversitätsdaten der Senatskommission für Biodiversitätsforschung



Services in NFDI4Biodiversity

With our experienced partners, we offer access to tried-and-tested tools for handling biodiversity and environmental data.



Data types and archives

Typical types of data our partners can handle

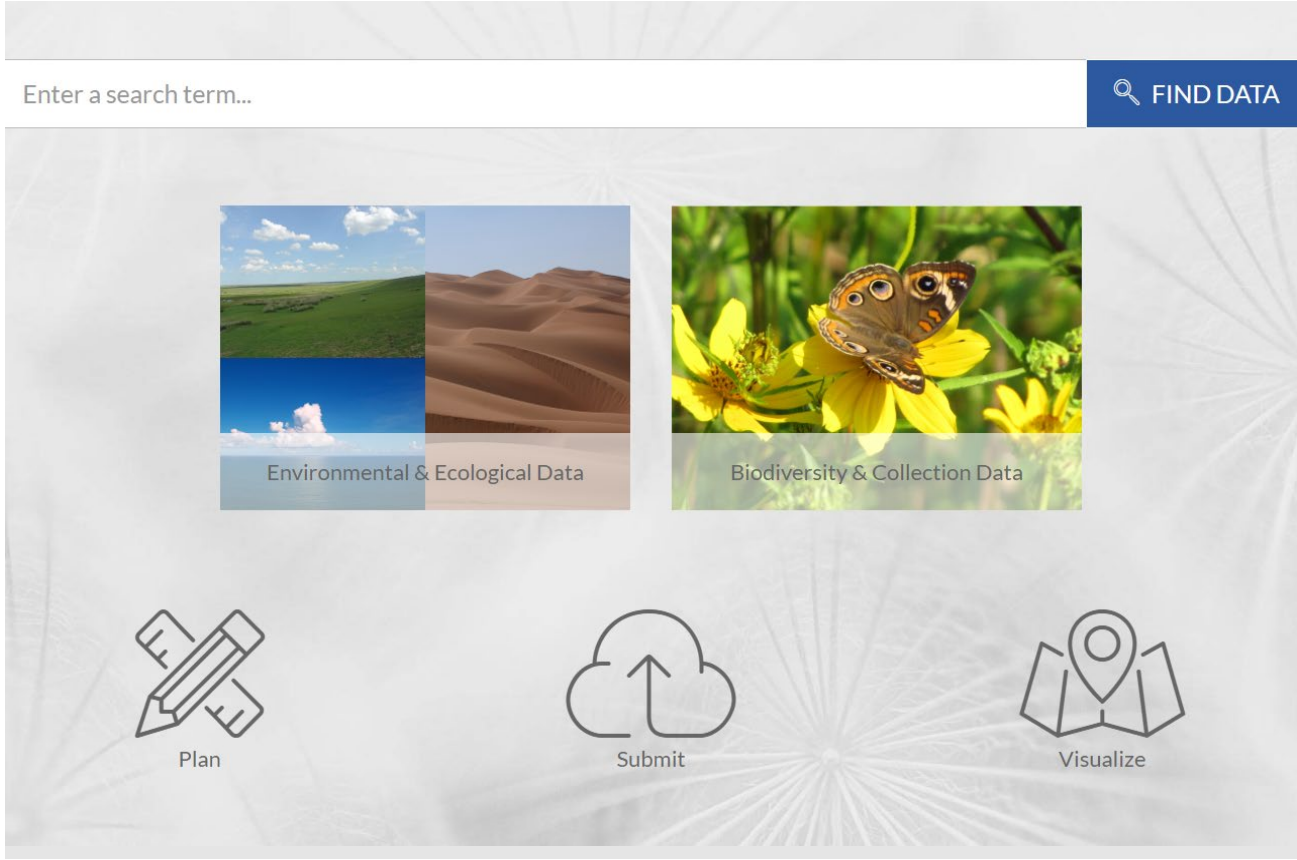
- Environmental data (e.g. temperature, rainfall, salinity)
- Trait data (e.g. seed number and mass)
- Molecular data (e.g. sequences)
- Experimental and laboratory measurements
- Multimedia (photographs, audio, video), f.e. of observed specimens
- Orthophotos produced using a drone
- Digital surface models produced using a drone and assigned software
- Model code
- Statistics

Data Centers specialized on Nucleotide, Plant and Environmental Data

Data Centers at Natural Science Collections

Research data services

A portfolio brought to NFDI4Biodiversity through the German Federation for Biological Data (GFBio e.V.)



Resources on Youtube

Data Management for Researchers [GFBio]

5 Videos • 21 Aufrufe • Zuletzt am 10.11.2021 aktualisiert

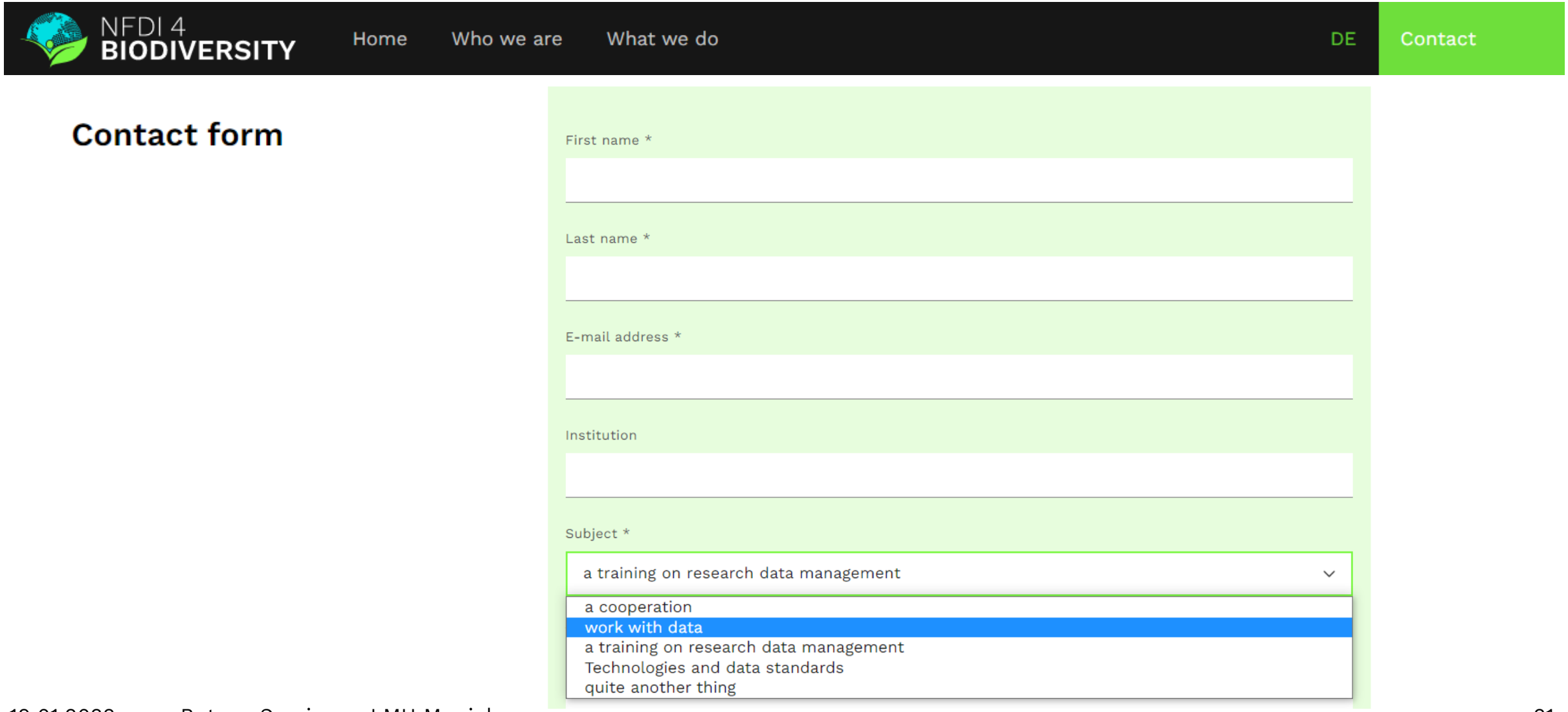
Data Visualization, Analysis and Transformation

5 Videos • 38 Aufrufe • Zuletzt am 07.12.2021 aktualisiert

Data Science 4 Ecologists in R

3 Videos • 96 Aufrufe • Zuletzt am 07.12.2021 aktualisiert

For practical support, please drop us a line!



NFDI 4
BIODIVERSITY

www.nfdi4biodiversity.org
@NFDI4Biodiv
#NFDI4Biodiv

Thank you – time for questions!

Make or buy? Vorüberlegungen zur Gesamtstruktur

- Daten als Sonderform wissenschaftlichen Wissens – mehrere große Wissenschaftsverlage erweitern ihr Spektrum
- aber: das Vertrauensverhältnis ist aktuell beschädigt
- Daten und wissenschaftliche Methoden sind eng verflochten. Vieles spricht für wissenschaftsinterne Lösungen