DFG Consortia Proposals National Research Data Infrastructure (NFDI) 2020

NFDI Consortium Earth System Sciences

Revision 1 from 2021-10-01

Applicant Institution: Technische Universität Dresden

Co-Applicant Institutions: Alfred Wegener Institute – Helmholtz Centre for Polar and Marine Research (AWI), German Climate Computing Center (DKRZ), Helmholtz Centre Potsdam – GFZ German Research Centre for Geosciences, Bochum University of Applied Sciences, Karlsruhe Institute of Technology (KIT), Gottfried Wilhelm Leibniz Universität Hannover, Goethe-Universität Frankfurt, Leipzig University, Max-Planck-Institute for Biogeochemistry, Senckenberg Society for Nature Research

Participants: ABC/J, BGR, BKG, BSA, BSH, BTU, DAI, DAM, DLR, DSMZ, DWD, FUB, GEOMAR, HCUNIHH, HUB, HZH, IASS, IGB, IOER, JÜLICH, JUNIHB, LIAG, LMU, LRZ, MFN, PIK, RWTH, TIB, TROPOS, TUB, TUBAF, TUDA, TUDO, TUM, UFZ, UNIBN, UNIERN, UNIGOE, UNIHB, UNIHH, UNIJ, UNIK, UNIKI, UNIMS, UNIP, UNITUE, UNU-FLORES, WSA, ZALF

Web: https://www.nfdi4earth.de/

Table of Contents

Table of Contents II

Acronyms IV

Acronyms of the Members of the NFDI4Earth Consortium VII

1 General Information 1

1.1 Name of the consortium in English and German 1

1.2 Summary of the proposal in English and German 1

1.3 Applicant institution 3

1.4 Spokesperson 3

1.5 Co-applicant institutions 3

1.6 Co-Spokespersons 4

1.7 Participants 4

1.8 Names and numbers of the DFG review boards (DFG Fachkollegien) that reflect the subject orientation of the proposed consortium 9

2 Scope and Objectives 10

2.1 Research domains or research methods addressed by the consortium, specific aims 10

2.2 Objectives and measuring success 14

3 Consortium 16

3.1 Composition of the consortium and its embedding in the community of interest 16

3.2 The consortium within the NFDI 28

3.3 International networking 31

3.4 Organizational structure and viability 32

3.5 Operating model 39

4 Research Data Management Strategy 39

4.1 State-of-the-art and needs analysis 40

4.2 Metadata standards 44

4.3 Implementation of the FAIR principles and data quality assurance 46

4.4 Services provided by the consortium 48

5 Work Programme 51

5.1 Task Area 1: 2Participate 53

5.1.1 Measure 1.1 (M1.1): Earth System Science Pilots 55

5.1.2 Measure 1.2 (M1.2): Incubator Lab 64

5.1.3 Measure 1.3 (M1.3): Education and Training Materials and Services 66

5.1.4 Measure 1.4 (M1.4): NFDI4Earth Academy 69

5.2 Task Area 2: 2Facilitate 72

5.2.1 Measure 2.1 (M2.1): OneStop4All 73

5.2.2 Measure 2.2 (M2.2): User Support Network 75

5.2.3 Measure 2.3 (M2.3): Governmental Data 78

5.2.4 Measure 2.4 (M2.4): Data in Long-Term Storage 80

5.2.5 Measure 2.5 (M2.5): Advancing Tools 82

5.3 Task Area 3: 2Interoperate 85

5.3.1 Measure 3.1 (M3.1): Synthesis of a Sustainable NFDI4Earth Architecture 86

5.3.2 Measure 3.2 (M3.2): Common Standards for FAIR ESS Data 89

5.3.3 Measure 3.3 (M3.3): NFDI Commons 92

5.3.4 Measure 3.4 (M3.4): International Networking & Embedding 95

5.4 Task Area 4: 2Coordinate 97

5.4.1 Measure 4.1 (M 4.1): Coordination, Collaborative and Sustainable Governance of NFDI4Earth 98

5.4.2 Measure 4.2 (M 4.2): Towards a Cultural Change in ESS Research Data Management 100

5.4.3 Measure 4.3 (M 4.3): Central Support Services for the federated NFDI4Earth Community 104

5.5 Bibliography and list of references 108

5.6 List of Information Infrastructures and Data Repositories, Software Repositories (developed and operated by members of the consortium) 112

Acronyms

AGU	American Geophysical Union, see https://www.agu.org/
API	Application Programming Interface
ARD	Analysis Ready Data
AvH	Alexander von Humboldt Foundation
AWS	Amazon Web Services
CAWR	Center for Advanced Water Research
CODATA	Committee on Data for Science and Technology, see http://www.codata.org/
COPDESS	Coalition for Publishing Data in the Earth and Space Sciences, see https://copdess.org/
CRC	Collaborative Research Centres
CRM	Customer Relationship Management
CTS	Core Trust Seal, see https://www.coretrustseal.org
DataCite	Global non-profit organisation providing persistent identifiers (DOIs) for research data and other research outputs, see https://datacite.org/
DGG	Deutsche Geophysikalische Gesellschaft e.V.
DIAG	Copernicus Data and Information Access Service
DINI	German Initiative for Network Information (Deutsche Initiative für Netzwerkinformation), see https://dini.de/
DMP	Data Management Plan
DOI	Digital Object Identifier, see https://doi.org
D-SDA	German Satellite Data Archive; via https://eoweb.dlr.de/egp/
ECMWF	European Centre for Medium-Range Weather Forecasts
EGU	European Geosciences Union, see https://www.egu.eu/
EIDA	European Integrated Data Archive, see https://www.orfeus-eu.org/data/eida/
ENVO	Environment Ontology, see http://environmentontology.org/
ENVRI	Environmental Research Infrastructures, see https://envri.eu/
ENVRI-FAIR	FAIR services for ENVRI, see https://envri.eu/home-envri-fair/
EOSC	European Open Science Cloud, see https://www.eosc-portal.eu
ePIC	Persistent Identifiers for eResearch, see https://www.pidconsortium.net/
EPOS	European Plate Observing System, see https://www.epos-ip.org/
ERIC	European Research Infrastructure Consortium
ESDL	Earth System Data Lab, see https://www.earthsystemdatalab.net/
ESRI	Environmental Systems Research Institute, see https://www.esri.com
ES	Earth System
ESA	European Space Agency
ESS	Earth System Sciences
ESGF	Earth System Grid Federation, see https://esgf.llnl.gov/

ESSRP	Earth and Solar system Research Partnership, see http://www.earthsystem.de
FAIR-DO	FAIR Digital Object
FDSN	International Federation of Digital Seismograph Networks, see http://www.fdsn.org/
FID GEO	Specialised Information Service for Geosciences (Fachinformationsdienst Geowissenschaften); see https://www.fidgeo.de/
FORCE11	The Future of Research Communication and e-Scholarship, see https://www.force11.org/
FTE	Full Time Equivalent
GDI-DE	GDI-DE is the Spatial Data Infrastructure Germany, see https://www.geoportal.de/EN/GDI-DE
GEE	Google Earth Engine, see https://developers.google.com/earth-engine
Geo.8	European Alliance for Earth Sciences, see https://www.gfz-potsdam.de/en/centre/international/geo8/
GeoROC	Geochemistry of Rocks or the Oceans and Continents; database http://georoc.mpch-mainz.gwdg.de/georoc/
GeoZG	Geodatenzugangsgesetz
GML	Geography Markup Language
HMC	Helmholtz Metadata Collaboration
HPC	High Performance Computing
IAG	International Association of Geodesy, see https://www.iag-aig.org/
ICDP	International Continental Drilling Program, see https://www.icdp-online.org/
ICOS	Integrated Carbon Observation System, see https://www.icos-cp.eu/
IEDA	Interdisciplinary Earth Data Alliance, see https://www.iedadata.org
IG	Interest Group
IGSN	International Geo Sample Number, http://igsn.org
INSPIRE	EU Directive for the INfrastructure for SPatial InfoRmation in Europe, see https://inspire.ec.europa.eu/
IPBES	Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, see https://ipbes.net/
IPCC	Intergovernmental Panel on Climate Change, see https://www.ipcc.ch/
ISO	International Organization for Standardization, see https://www.iso.org
IUGG	International Union of Geodesy and Geophysics, see http://www.iugg.org/
M	Measure
MESI	Modular Earth Science Infrastructure
MetBase	Meteorite information database and bibliography of meteorites and planetary science, see https://metbase.org
ML	Machine Learning
MOOC	Massive Open Online Course
MS	Milestone

NDACC	Network for the Detection of Atmospheric Composition Change
nestor	network of expertise in long-term storage of digital resources in Germany (Kompetenznetzwerk digitale Langzeitarchivierung), see https://www.langzeitarchivierung.de/
netCDF	Network Common Data Form; https://www.unidata.ucar.edu/software/netcdf/
NFDI	National Research Data Infrastructure (Nationale Forschungsdaten Infrastruktur), see http://www.nfdi.de
NFDI4Earth	NFDI Consortium Earth System Sciences / NFDI Konsortium Erdsystemforschung, see https://www.nfdi4earth.de/
O2A	Observations to Archive: http://sensor.awi.de
OGC	Open Geospatial Consortium, see https://www.opengeospatial.org/
OpenAIRE	Open Access Infrastructure for Research in Europe, see https://www.openaire.eu/
ORCID	Open Researcher and Contributor Identifier, see https://orcid.org/ and https://www.orcid-de.org/
ORFEUS	Observatories & Research Facilities for European Seismology, see https://www.orfeus-eu.org/
OSF	OneStop4All search functionality
PANGAEA	Open Access library aimed at archiving, publishing and distributing data from earth system research, see https://www.pangaea.de/
PID	Persistent Identifier
RDA	Research Data Alliance, see https://www.rd-alliance.org/
RDC	Research Data Commons Berlin Declaration on NFDI Cross-Cutting topics, https://doi.org/10.5281/zenodo.3457212
RDM	Research Data Management
RDF	Resource Description Framework
re3data	Registry of Research Data Repositories, see https://www.re3data.org/
RfII	Council for Information Infrastructures (Rat für Informationsinfrastrukturen), see http://www.rfii.de
RISE	Research Infrastructure Self-Evaluation Framework
ScaDS.AI	National competence center for Big Data and Artificial Intelligence, see https://www.scads.de/en/
SDI	Spatial Data Infrastructure
SDL	Spatiotemporal data literacy
SDS	Spatiotemporal data science
SMWK	Saxon State Ministry for Higher Education, Research and the Arts
SOP	Standard operating procedures
TA	Task Area, NFDI4Earth has 4 task areas detailed in chapter 4
TA1	Task Area 1: NFDI4Earth2Participate
TA2	Task Area 2: NFDI4Earth2Facilitate
TA3	Task Area 3: NFDI4Earth2Interoperate

TA4	Task Area 4: NFDI4Earth2Coordinate
TDWG	Biodiversity Information Standards, see https://www.tdwg.org
TERENO	Terrestrial Environmental Observatoria, see https://www.tereno.net
TUDD-ZIH	Centre for Information Services and High Performance Computing at TUDD
TV	Thematic Viewers
UBA	German Environment Agency (Umweltbundesamt)
USN	User Support Network (TA2)
VRE	Virtual Research Environment
W3C	World Wide Web Consortium, see https://www.w3.org/
WDCC	World Data Center for Climate, see https://cera-www.dkrz.de
WDS	World Data System, see https://www.icsu-wds.org/
WMO	World Meteorological Organization, see https://public.wmo.int/en
WCS	Web Coverage Service, see https://www.opengeospatial.org/standards/wcs
WFS	Web Feature Service, see https://www.opengeospatial.org/standards/wfs
WMS	Web Map Service, see https://www.opengeospatial.org/standards/wms
WPS	Web Processing Service, see https://www.opengeospatial.org/standards/wps
XML	Extensible Markup Language
Zarr	Python package providing an implementation of chunked, compressed, N-dimensional arrays, see https://zarr.readthedocs.io

Acronyms of the Members of the NFDI4Earth Consortium

Co-Applicants and Participants

ABC/J	Geoverbund ABC/J
AWI	Alfred Wegener Institute (AWI) – Helmholtz Centre for Polar and Marine Research (Co-Applicant)
BGR	Federal Institute for Geosciences and Natural
BKG	Federal Agency for Cartography and Geodesy (Bundesamt für Kartographie und Geodäsie)
BSA	Generaldirektion der Staatlichen Archive Bayerns (Bavarian State Archives)
BSH	Federal Maritime and Hydrographic Agency (Bundesamt für Seeschifffahrt und Hydrographie**)**
BTU	Brandenburg University of Technology Cottbus – Senftenberg
DAI	Deutsches Archäologisches Institut
DAM	Deutsche Allianz für Meeresforschung e.V.
DKRZ	German Climate Computing Center (Co-Applicant)
DLR	German Aerospace Center (DLR)
DSMZ	Leibniz Institute DSMZ – German Collection of Microorganisms and Cell Cultures

DWD	Deutscher Wetterdienst
FUB	Freie Universität Berlin
GEOMAR	GEOMAR Helmholtz Centre for Ocean Research Kiel
Geo.X	The Research Network for Geosciences in Berlin and Potsdam (Co-Applicant with GFZ)
GFZ	Helmholtz Centre Potsdam – GFZ German Research Centre for Geosciences (Co-Applicant)
HCUNIHH	HafenCity Universität Hamburg
HSBO	Bochum University of Applied Sciences (Co-Applicant)
HUB	Humboldt-Universität Berlin
HZH	Helmholtz-Zentrum Hereon
IASS	Institute for Advanced Sustainability Studies e.V.
IGB	Leibniz-Institute of Freshwater Ecology and Inland Fisheries
IOER	Leibniz Institute of Ecological Urban and Regional Development
JÜLICH	Forschungszentrum Jülich GmbH
JUNIHB	Jakobs Universität Hansestadt Bremen
KIT	Karlsruhe Institute of Technology (KIT) (Co-Applicant)
LIAG	Leibniz Institute for Applied Geophysics
LMU	Ludwig-Maximillians-Universität München
LRZ	Leibniz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities
LUH	Gottfried Wilhelm Leibniz University Hannover (Co-Applicant)
MFN	Museum für Naturkunde – Leibniz Institute for Evolution and Biodiversity Science
MPIBGC	Max-Planck-Institute for Biogeochemistry (Co-Applicant)
PIK	Potsdam Institute for Climate Impact Research
RWTH	RWTH Aachen University
SGN	Senckenberg Society for Nature Research (Co-Applicant)
TIB	Leibniz Information Centre for Science and Technology University Library
TROPOS	Leibniz Institute for Troposheric Research
TUB	Technische Universität Berlin
TUBAF	Technische Universität Bergakademie Freiberg
TUDA	Technische Universität Darmstadt
TUDD	Technische Universität Dresden (Applicant)
TUDO	Technische Universität Dortmund
TUM	Technische Universität München
UFZ	Helmholtz Centre for Environmental Research – UFZ
UNIBN	Rheinische Friedrich-Wilhelms-Universität Bonn
UNIERN	Friedrich-Alexander Universität Erlangen-Nürnberg
UNIF	Goethe-Universität Frankfurt am Main (Co-Applicant)
UNIGOE	Georg-August-Universität Göttingen
UNIHB	Universität Bremen

UNIHH	Universität Hamburg
UNIJ	Friedrich-Schiller-Universität Jena
UNIK	Universität zu Köln
UNIKI	Christian-Albrechts-Universität zu Kiel
UNIL	Universität Leipzig (Co-Applicant)
UNIMS	Universität Münster
UNIP	Universität Potsdam
UNITUE	Eberhards-Karls-Universität Tübingen
UNU-FLORES	UNU-FLORES – United Nations University Institute for Integrated Management of Material Fluxes and of Resources
WSA	Water Science Alliance
ZALF	Leibniz Centre for Agricultural Landscape

1 General Information

1.1 Name of the consortium in English and German

NFDI Consortium Earth System Sciences / NFDI Konsortium Erdsystemforschung

1.2 Summary of the proposal in English and German

Summary

NFDI4Earth addresses digital needs of Earth System (ES) Sciences (ESS). ES scientists cooperate in international and interdisciplinary networks with the overarching aim to understand the functioning and interactions within the Earth system and address the multiple challenges of global change. NFDI4Earth is a community-driven process providing researchers with FAIR, coherent, and open access to all relevant ES data, to innovative research data management (RDM) and data science methods. The NFDI4Earth 2021-26 work plan comprises four task areas (TA):

TA1 2Participate will engage as broadly as possible with the ESS community and secures that NFDI4Earth is driven by community requirements: Pilots, small agile projects proposed by the community leverage existing technologies and manifest the researchers' RDM needs. The Incubator Lab identifies promising new tools and scouts for trends in ES Data Science. EduHubs produce open, ready to use educational resources on implementing FAIR principles in the ESS. The Academy will connect young researchers and their data-driven research to NFDI4Earth.

TA2 2Facilitate realizes the OneStop4All as the web-based entry point to FAIR, open and innovative RDM in ESS. It supports on how to find, access, share, publish and work with ES data. Specific user requests beyond the scope of the OneStop4All will be routed to a distributed User Support Network. TA2 will also unlock the wealth of data that exists in governmental data repositories and will collaborate with all services on supporting long-term archiving. By advancing tools, TA2 showcases user-driven best-practise examples of novel ESS data integration and analysis methods (e.g. Earth data cubes).

TA3 2Interoperate aims at interoperability and coherence of the heterogeneous, segmented range of ESS RDM services. The ecosystems of ESS (meta-)data and software repositories, data science services and collaboration platforms get integrated iteratively into a common NFDI4Earth architecture. Based on commonly agreed-upon standards TA3 provide consistent methods for a self-evaluation of RDM offerings. TA3 works on NFDI cross-cutting topics and makes outcomes accessible as a Living Handbook. It ensures co-operation in international RDM initiatives and standardisation bodies.

TA4 2Coordinate facilitates the overall management of the NFDI4Earth consortium. TA4 acts as central support service and coordination of the technical implementations. It also offers virtual

research environments. The NFDI4Earth Coordination Office will support the NFDI4Earth community in day-to-day operations and acts as the NFDI4Earth point of contact. It develops a commonly agreed model for a sustainable operation of NFDI4Earth. A commonly accepted NFDI4Earth FAIRness and Openness Commitment is key to fostering a cultural change towards FAIR and Open RDM in the ESS community.

Zusammenfassung

NFDI4Earth befasst sich mit den digitalen Bedürfnissen in den Erdsystem- (ES) Wissenschaften (ESW). ES-WissenschaftlerInnen kooperieren in internationalen und interdisziplinären Netzwerken mit dem übergeordneten Ziel, Funktionsweisen und Wechselwirkungen im Erdsystem zu verstehen und die vielfältigen Herausforderungen des globalen Wandels anzugehen. NFDI4Earth ist ein Community-gesteuerter Prozess, der Forschenden einen FAIRen, kohärenten und offenen Zugang zu allen relevanten ES-Daten, innovativen Forschungsdatenmanagement (FDM) und Data Science Methoden ermöglicht. Der NFDI4Earth-Arbeitsplan 2021-26 umfasst vier Task Areas (TA):

TA1 2Participate bindet die ESW-Community allumfassend ein und stellt sicher, dass deren Anforderungen die NFDI4Earth Lösungen bestimmen: Piloten, kleine agile aus der Community vorgeschlagene Projekte etablieren vorhandene Technologien und manifestieren FDM-Bedarfe. Das Incubator Lab identifiziert vielversprechende neue Tools und Trends für innovative ES Data Science. Die EduHubs stellen offene, sofort einsatzbare Bildungsressourcen für die Umsetzung von FAIR in den ESW bereit. Die Akademie vernetzt NachwuchswissenschaftlerInnen und integriert ihre datengetriebene Forschung in NFDI4Earth ein.

TA2 2Facilitate realisiert den OneStop4All als webbasierten Einstiegspunkt für FAIRes, offenes und innovatives RDM in den ESW. Dieser unterstützt bei Suche, Zugriff, Nutzung und Publikation von ES-Daten. Anfragen, die über den Rahmen des OneStop4All hinausgehen, bearbeitet ein User Support Network. Die Einbindung behördlicher Dienste ermöglicht auch den umfassenden Zugriff auf öffentliche Daten und gemeinsame Strategien der Datenarchivierung. Die anwendungsgetriebene Entwicklung innovativer Tools ermöglicht neue Ansätze für Datenintegration und -analyse (z. B. Earth Data Cubes).

TA3 2Interoperate zielt auf Interoperabilität und Kohärenz des heterogenen, segmentierten Angebots von RDM-Diensten für die ESW. Die vielfältigen (Meta-)Daten- und Software-Repositorien, Data Science Dienste und Kollaborationsplattformen werden iterativ in der NFDI4Earth-Architektur zusammengeführt. Auf Basis gemeinsam vereinbarter Standards stellt TA3 konsistente Methoden zur Selbstevaluierung von RDM-Angeboten bereit. TA3 bearbeitet NFDI-Querschnittsthemen und macht Ergebnisse als Living Handbook zugänglich. Es unterstützt Zusammenarbeit mit internationalen FDM-Initiativen und Standardisierungsgremien wird.

TA4 2Coordinate umfasst das Gesamtmanagement des NFDI4Earth-Konsortiums. TA4 agiert als zentrale Unterstützung und Koordination der technischen Implementierungen und stellt

virtuelle Forschungsumgebungen bereit. Das NFDI4Earth-Koordinierungsbüro unterstützt die NFDI4Earth-Community im Tagesgeschäft und fungiert als Kontaktstelle. Es entwickelt ein gemeinsames Modell für den nachhaltigen Betrieb der NFDI4Earth. Ein breit akzeptiertes NFDI4Earth FAIRness and Openness Bekenntnis ist Kernstück zur Förderung eines kulturellen Wandels zu FAIR und Open RDM.

1.3 Applicant institution

Applicant institution	Location
TUDD: Technische Universität Dresden	01062 Dresden

1.4 Spokesperson

Spokesperson	Institution, location
Prof. Dr. Lars Bernard	Technische Universität Dresden, Faculty of Environmental Sciences 01062 Dresden

1.5 Co-applicant institutions

Co-applicant institutions	Location
AWI: Alfred Wegener Institute – Helmholtz Centre for Polar and Marine Research	Am Handelshafen 12 27570 Bremerhaven
DKRZ: German Climate Computing Center (Deutsches Klimarechenzentrum GmbH, DKRZ)	Bundesstraße 45a 20146 Hamburg
GFZ: Helmholtz Centre Potsdam – GFZ German Research Centre for Geosciences	Telegrafenberg 14473 Potsdam
HSBO: Bochum University of Applied Sciences	Lennershofstr. 140 44801 Bochum
KIT: Karlsruhe Institute of Technology	Kaiserstraße 12 76131 Karlsruhe
LUH: Gottfried Wilhelm Leibniz University Hannover (Gottfried Wilhelm Leibniz Universität Hannover)	Welfengarten 1 30167 Hannover
UNIF: Goethe-Universität Frankfurt am Main	60629 Frankfurt am Main
UNIL: Universität Leipzig	Ritterstraße 26 04109 Leipzig
MPIBGC: Max-Planck-Institute for Biogeochemistry (Max-Planck-Institut für Biogeochemie)	07745 Jena
SGN: Senckenberg Society for Nature Research	Senckenberganlage 25 60325 Frankfurt am Main

1.6 Co-Spokespersons

Co-Spokesperson	Institution, location	Task area
Dr. Hildegard Gödde	Geo.X Research Network for Geosciences c/o Helmholtz Centre Potsdam – GFZ German Research Centre for Geosciences Telegrafenberg, 14473 Potsdam	TA 1
Prof. Dr. Miguel Mahecha	Universität Leipzig, Faculty of Physics and Earth Sciences, Talstr. 35, 04130 Leipzig	TA 1
Prof. Dr. Monika Sester	Gottfried Wilhelm Leibniz University Hannover Welfengarten 1, 30167 Hannover	TA 1
Prof. Dr. Carsten Keßler	Bochum University of Applied Sciences (Fachhochschule Bochum) Lennershofstr. 140,44801 Bochum	TA 1
Prof. Dr. Peter Braesicke	Karlsruhe Institute of Technology (KIT) Kaiserstraße 12, 76131 Karlsruhe	TA 2
Prof. Dr. Markus Reichstein	Max-Planck-Institute for Biogeochemistry, 07745 Jena	TA 2
Hannes Thiemann	German Climate Computing Center (Deutsches Klimarechenzentrum GmbH, DKRZ) Bundesstraße 45a, 20146 Hamburg	TA 2
Prof. Dr. Frederik Tilmann	Helmholtz Centre Potsdam – GFZ German Research Centre for Geosciences Telegrafenberg, 14473 Potsdam	TA 3
Prof. Dr. Stephan Frickenhaus	Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research (AWI) Am Handelshafen 12, 27570 Bremerhaven	TA 3
Prof. Dr. Horst Marschall	Goethe-Universität Frankfurt am Main, Faculty of Geosciences and Geography, Altenhöferallee 1, 60438 Frankfurt am Main	TA 3
Dr. Claus Weiland	Senckenberg Society for Nature Research Senckenberganlage 25 60325 Frankfurt am Main	TA 3
Prof. Dr. Wolfgang E. Nagel	TU Dresden, Center of Information Services and High Performance Computing (TUDD-ZIH), 01069 Dresden	TA 4

1.7 Participants

Acronym	**Participating institutions**	**Location**
ABC/J c/o JÜLICH	Geoverbund ABC/J c/o Forschungszentrum Jülich GmbH	52425 Jülich
BGR	Federal Institute for Geosciences and Natural Resources	Stilleweg 2 30655 Hannover
BKG	Bundesamt für Kartographie und Geodäsie – Geodateninfrastruktur Deutschland (GDI-DE)	Richard-Strauss-Allee 11 60598 Frankfurt a.M.

Acronym	Participating institutions	Location
BSA	Generaldirektion der Staatlichen Archive Bayerns	Schönfeldstr. 5 80539 München
BSH	Bundesamt für Seeschifffahrt und Hydrographie	Bernhard-Nocht-Straße 78 20359 Hamburg
BTU	Brandenburg University of Technology Cottbus – Senftenberg, Information, Communication and Media Centre	Platz der Deutschen Einheit 1 03046 Cottbus
DAI	Deutsches Archäologisches Institut	Podbielskiallee 69-71 14195 Berlin
DAM	Deutsche Allianz Meeresforschung e.V.	Wissenschaftsforum | Markgrafenstraße 37 10117 Berlin
DLR	German Aerospace Center	Linder Höhe 51147 Köln
DSMZ	Leibniz Institute DSMZ – German Collection of Microorganisms and Cell Cultures GmbH	Inhoffenstraße 7b 38124 Braunschweig
DWD	Deutscher Wetterdienst	Frankfurter Straße 135 63067 Offenbach
FUB	Freie Universität Berlin Department of Earth Sciences	Malteser Str. 74-100 12249 Berlin
GEOMAR	GEOMAR Helmholtz Centre for Ocean Research Kiel Information, Data and Computing Centre Research Unit for Marine Geodynamics	Wischhofstr. 1-3 24148 Kiel
HCUNIHH	HafenCity Universität Hamburg Geodesy and Adjustment Theory	Überseeallee 16 20457 Hamburg
HUB	Humboldt-Universität zu Berlin Department of Geography	Unter den Linden 6 10099 Berlin
HZH	Helmholtz-Zentrum Hereon	Max-Planck-Str. 1 21502 Geesthacht
IASS	Institute for Advanced Sustainability Studies e.V.	Berliner Str. 130 14467 Potsdam
IGB	Leibniz-Institute of Freshwater Ecology and Inland Fisheries	Müggelseedamm 310 12587 Berlin
IOER	Leibniz Institute of Ecological Urban and Regional Development	Weberplatz 1 01217 Dresden
JUNIHB	Jacobs University Bremen gGmbH Department of Physics and Earth Sciences	Campus Ring 1 28759 Bremen
LIAG	Leibniz Institute for Applied Geophysics	Stilleweg 2 30655 Hannover
LMU	Ludwig-Maximilians-Universität München Department of Earth and Environmental Sciences	Theresienstr. 41/iV 80333 München

Acronym	Participating institutions	Location
LRZ	Leibniz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities	Boltzmannstr. 1 85748 Garching b. München
MFN	Museum für Naturkunde – Leibniz Institute for Evolution and Biodiversity Science	Invalidenstraße 43 10115 Berlin
PIK	Potsdam Institute for Climate Impact Research	Telegrafenberg 14473 Potsdam
RWTH	RWTH Aachen University, Division of Earth Sciences and Geography	Wüllnerstr. 2 52064 Aachen
TIB	TIB – Leibniz Information Centre for Science and Technology University Library	Welfengarten 1 B 30167 Hannover
TROPOS	Leibniz Institute for Troposheric Research (TROPOS)	Permoserstraße 15 04318 Leipzig
TUB	Technische Universität Berlin Institute of Ecology	Straße des 17. Juni 135 10623 Berlin
TUBAF	Technische Universität Bergakademie Freiberg Institute of Mine Surveying and Geodesy	Fuchsmühlenweg 9b 09599 Freiberg
TUDA	Technische Universität Darmstadt Institute of Applied Geosciences Institute of Fluid Systems	Karolinenplatz 5 64289 Darmstadt
TUDO	Technische Universität Dortmund Department of Spatial Management and Modelling	August-Schmidt-Str. 10 44227 Dortmund
TUM	Technische Universität München Big Geospatial Data Management	Willy-Messerschmitt-Str.1 82024 Taufkirchen/Ottobrunn
UFZ	Helmholtz Centre for Environmental Research – UFZ	Permoserstr. 15 04318 Leipzig
UNIBN	Rheinische Friedrich-Wilhelms-Universität Bonn, Center for Remote Sensing of Landsurface	Regina-Pacis-Weg 3 53113 Bonn
UNIERN	Friedrich-Alexander Universität Erlangen-Nürnberg Chair of Applied and Environmental Geology	Schlossgarten 5 91054 Erlangen
UNIGOE	Georg-August-Universität Göttingen Göttingen State and University Library Geowissenschaftliches Zentrum, Division of Geochemistry	Wilhelmsplatz 1 37073 Göttingen
UNIHB	University of Bremen – MARUM Center for Marine Environmental Sciences	Leobener Str. 8 28359 Bremen
UNIHH	Universität Hamburg Center for Earth System Research and Sustainability (CEN)	Bundesstr. 53 20146 Hamburg
UNIJ	Friedrich-Schiller-Universität Jena Department for Earth Observation	Löbdergraben 32 07743 Jena

Acronym	Participating institutions	Location
UNIK	Universität zu Köln Department of Geosciences	Albertus-Magnus-Platz 50923 Köln
UNIKI	Christian-Albrechts-Universität zu Kiel Institute of Applied and Experimental Physics	Olshausenstr. 40 24098 Kiel
UNIMS	University of Münster Institute for Geoinformatics	Heisenbergstr. 2 48149 Münster
UNIP	Universität Potsdam Earth and Environmental Sciences	Karl-Liebknecht-Str. 24-25 14476 Potsdam-Golm
UNITUE	Eberhards-Karls-Universität Tübingen Center for Data Processing Center for Applied Geoscience	Wächterstr. 76 72074 Tübingen
UNU-FLORES	UNU-FLORES – United Nations University Institute for Integrated Management of Material Fluxes and of Resources	Ammonstr. 74 01067 Dresden
WSA	Water Science Alliance e.V. c/o Technische Universität Dresden Institute of Urban Water Management	01062 Dresden
ZALF	Leibniz Centre of Agricultural Landscape Research (ZALF)	Eberswalder Str. 84 15374 Müncheberg

Contributions of participating institution

Acronym	Contribution
ABC/J	provides infrastructure, education and training activities, supports the emplacement of an initial Academy Site in TA1 (M1.4) and serves as a facilitator for an academic network
BGR	public authority related to ESS, provides infrastructure and governmental data and is involved in TA2 (M2.3)
BKG	public authority related to ESS, provides infrastructure and governmental data in TA2 (M2.3) and is involved in M2.4 and TA3 (M3.3)
BSA	public authority related to ESS, involved in TA2 (M2.3 and M2.4), provides infrastructure and serves as a facilitator
BSH	public authority related to ESS, provides infrastructure and governmental data, involved in TA3 (M3.3)
BTU	provides education and training activities and materials for TA1 and will be part of an Academy Site (M1.4)
DAI	serves as a facilitator (for exchange with the archaeology community)
DAM	coordinates the setup of a NFDI4Earth compatible marine data infrastructure and supports TA1 (M1.4) by emplacement of an initial Academy Site
DLR	runs two pilots in TA1 (M1.1) and is involved in TA2 (M2.5) and TA3 (M3.4), provides manifold data (bases), infrastructures as well as services and serves as a facilitator
DSMZ	provides data bases as well as infrastructures and serves as a facilitator
DWD	public authority related to ESS, provides manifold data (bases), infrastructures and services, runs a pilot TA1 (M1.1) and in TA2 (M2.3)

Acronym	Contribution
FUB	provides education and training activities and materials for TA1, contributes to an Academy Site in M1.4 and is involved in M2.2
GEOMAR	provides infrastructure, runs a pilot (M1.1) and is involved in TA2 (M2.2) as well as in TA3 (M3.2) and serves as a facilitator for the marine community
HCUNIHH	provides education and training activities and materials and serves as a national and international facilitator for geodesy
HUB	provides education and training activities and materials for TA1 and will be part of an Academy Site in M1.4
HZH	involved in TA2 (M2.1), provides various archives, infrastructures, repositories and services
IASS	provides education and training activities and materials for TA1 and will be part of an Academy Site in M1.4
IGB	runs a pilot in TA1 (M1.1), provides infrastructure and serves as a facilitator for abiotic and biotic data series from rivers and lakes, strongly engaged in education and training activities
IOER	involved in a pilot in TA1 (M1.1), provides infrastructure and esp. the IOER Monitor (as a long-term high resolution time series of land use)
JUNIHB	runs a pilot in TA1 (M1.1), provides education and training activities and materials (for TA1)
LIAG	runs a pilot in TA 1 (M1.1) and is involved in TA3 (M3.3), serves as a facilitator for the geological and geophysical community
LMU	contributes different community software, libraries and platforms related to ESS, serves as a facilitator
LRZ	involved in TA 2 (M2.2) and TA3 (M3.1), provides infrastructure and different services, serves as a facilitator
MFN	provides (international data) infrastructures
PIK	involved in TA1 in a pilot (M1.1) as well as in an Academy site (M1.4), serves as a link, i.e. to the global integrated assessment community et al.
RWTH	provides infrastructure as well as education and training activities in RDM
TIB	involved in TA3 (M3.2), provides infrastructure as well as services and serves as a facilitator
TROPOS	provides infrastructures and supports esp. the link to ACTRIS-D
TUB	provides education and training activities, involved in corresponding TA1 in an Academy Site (M1.4), serves as a facilitator
TUBAF	provides infrastructure as well as training and education activities, serves as a facilitator
TUDA	provides infrastructure and archives, represents the Special Interest Group ‘Geology & Geophysics’, serves as a facilitator to the engineering sciences
TUDO	provides infrastructure as well as education and training activities and materials
TUM	provides education and training activities and materials, involved in corres-ponding TA1 (M1.3)
UFZ	runs a pilot in TA1 (M1.1) and involved in TA3 (M3.2), serves as a facilitator and interface to national and international ESS related communities

Acronym	Contribution
UNIBN	provides infrastructure as well as training and education activities, involved in corresponding TA1 in an Academy Site (M1.4), serves as a facilitator
UNIERN	provides novel categories of interdisciplinary ESS data and infrastructure
UNIGOE	involved in a pilot in TA1 (M1.1) and in TA3 (M3.3), provides a database of rocks of the oceans and continents (GeoROC), serves as a facilitator
UNIHB	runs and involved in three pilots in TA1 (M1.1) and emplace an initial Academy Site in TA1 (M1.4), provides infrastructure and serves an interface to PANGAEA
UNIHH	involved in a pilot in TA1 (M1.1), offers training and education activities for M1.3, as well as involved in TA2 (M2.2) and in TA3 (M3.1), provides infrastructure
UNIJ	provides infrastructure and access to repositories (incl. higher-level products), offers innovative training and education activities
UNIK	involved in a pilot in TA1 (M1.1), hosts various repositories, provides infra-structures and offers training and education activities and materials
UNIKI	offers education and training activities, provides databases, infrastructure, portals and repositories, involved in TA2 (M2.4), serves as a facilitator
UNIMS	offers training and education activities, involved in corresponding TA1 (M1.3) as well as in TA 2 (M2.5)
UNIP	offers training and education activities and materials, involved in an Academy Site in TA1 (M1.4)
UNITUE	provides infrastructure, involved in TA3 (M3.3), serves as a facilitator
UNU-FLORES	offers training and education activities and materials, provides the Nexus Tools Platform and serves as a facilitator to UN institutions
WSA	serves as a facilitator for a scientific association representing all aspects of water research and its most relevant institutions
ZALF	provides different services as well as training and education modules related to ESS, serves as a facilitator and link to landscape research

1.8 Names and numbers of the DFG review boards (DFG Fachkollegien) that reflect the subject orientation of the proposed consortium

NFDI4Earth addresses the research areas: Atmospheric Science, Oceanography and Climate Research (313); Geology and Paleontology (314); Geophysics and Geodesy (315); Geochemistry, Mineralogy and Crystallography (316); Geography (317); Water Research (318), Astrophysics and Astronomy (311), Soil Sciences (207-01).

2 Scope and Objectives

2.1 Research domains or research methods addressed by the consortium, specific aims

The Earth System (ES) is composed of a set of highly intertwined subsystems, often divided into the solid Earth, the hydrosphere, the atmosphere, and the biosphere. Through constant exchanges of matter and energy, these spheres form a complex spatio-temporal dynamical system. Earth System Sciences (ESS) aims to decipher the driving processes and to detect emergent patterns of these dynamic systems. The overarching aim is to understand the planetary evolution and the functioning of all subsystems of the ES and their interactions. Data and its analysis are the key to understanding these complex processes.

ESS are increasingly in the focus of the wider public, given the pressing global challenges due to the strong imprint of human activity on many ES subsystems. Climate change, water scarcity, land-use change, environmental pollution, and natural hazards, for instance, require comprehensive interdisciplinary efforts at the national and international level. Here, data from many disciplines have to be integrated to achieve novel scientific insights of societal relevance. **These linkages across disciplines and scientific approaches can be substantially strengthened through collaborative research data management (RDM) and knowledge extraction, ultimately leading to a new quality of science.** These linkages also imply that NFDI4Earth naturally has multiple connection points to other NFDI consortia.

Even though the NFDI4Earth community is highly diverse, **a common characteristic of ESS data is their spatio-temporal contexts** described with comprehensive space-time reference systems across different scales. On the one hand, the ESS community produces and deals with rapidly growing large, dense streams of data (from, e.g., satellites, sensor networks or simulations), while on the other hand we face data sets that are rare, spatially or temporally coarse grained, distributed or thematically incomplete (from e.g. historical observations, surveys or laboratory data). **NFDI4Earth will therefore support FAIR and Open RDM for the full range from big data to long-tail data, and paths will be elaborated to jointly access and co-interpret any ESS-relevant data.**

The multitude of observations and model data in very high spatial and temporal resolutions across all ESS disciplines lead to rapidly increasing data volumes. Thus, describing and assessing ES processes, their dependencies, their interactions, and their changes urgently require RDM workflows that are as efficient as possible, foster Openness and fully follow the FAIR (findable, accessible, interoperable, reusable) guiding principles (Wilkinson et al., 2016). Open and FAIR ESS data will enable novel and more powerful data analytics frameworks – leading to novel insights into the Earth system and its evolution through time. Today, there is a broad spectrum of services supporting RDM in ESS. However, (a) these services are scattered and heterogeneous

and (b) only a few of them have a long-term perspective, many are project-based without stable prospects. In addition, appreciation and literacy of RDM, FAIR principles and related issues still differ among researchers, institutions and research fields. **Aligning existing and emerging RDM cultures, strategies, and services along FAIR and Openness principles and working towards a unifying and long-term perspective for services can only be achieved in a broad community effort as represented by NFDI4Earth.**

NFDI4Earth targets the consolidation and harmonization of research data-related services in ESS and will offer a user-driven, innovation-friendly as well as efficient and sustainable service environment to researchers, complemented by education and training. Collaboration with national and international stakeholders outside NFDI4Earth on cross-cutting topics will contribute to achieving this goal on a broad level. **NFDI4Earth will provide in five years' time an environment for simple, efficient, and – whenever possible – open access to ES data, RDM support, data integration and data analysis services guided by FAIR principles, hereby supporting the entire research data life cycle.** It will thus enable cutting-edge research in ESS and beyond.

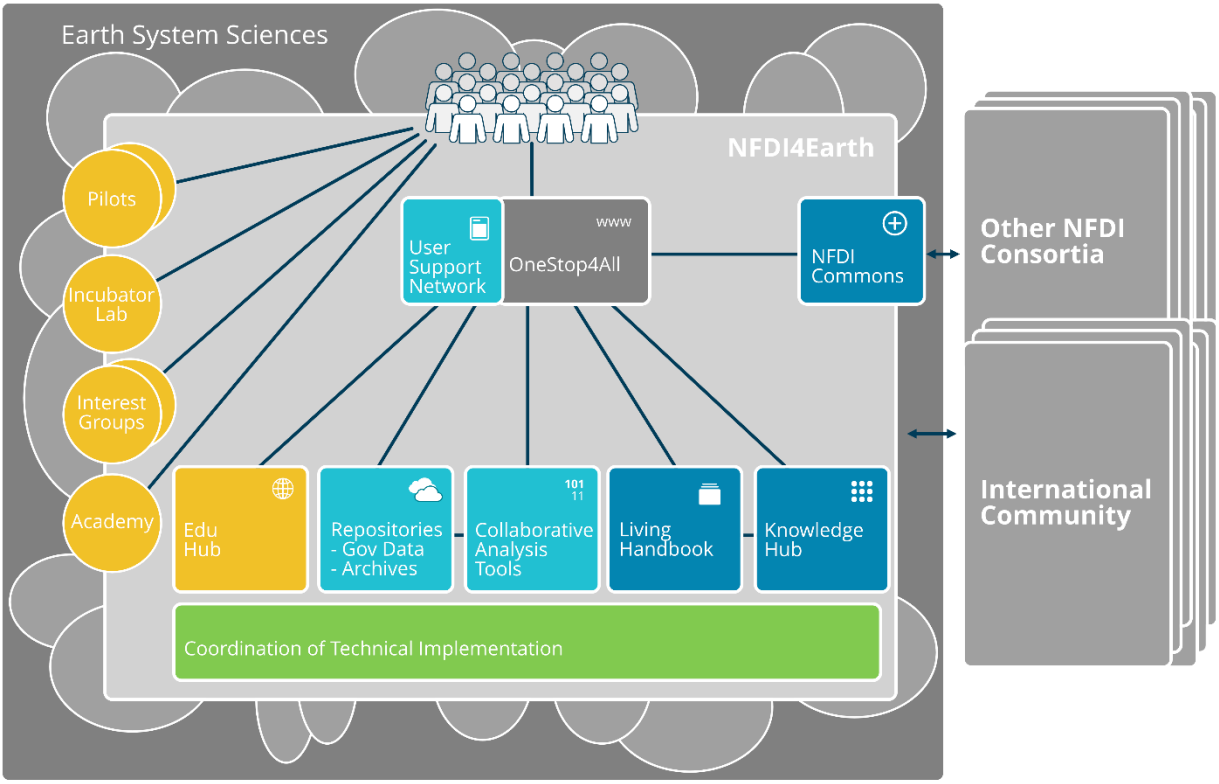

Figure 1: The NFDI4Earth community support for FAIR ESS RDM from a user perspective.

NFDI4Earth responds to the growing complexity of digital requirements within ESS. As a community effort, NFDI4Earth brings together – for the first time – scientists and infrastructures from universities, universities of applied science, research institutions and public authorities with

their wealth of resources and experience. As shown in Figure 1 and Figure 2, **NFDI4Earth will provide community support for FAIR ESS RDM tailored to individual needs.**

NFDI4Earth moderates between RDM supply and demand, and stimulates innovation and cultural change in the collaborative and connected use of all kinds of data in ESS. **NFDI4Earth supports open science by promoting the use of open data services and open software** repositories, and by fostering the implementation of institutional data policies and data management plans that enable collaborative research and seamless data publication.

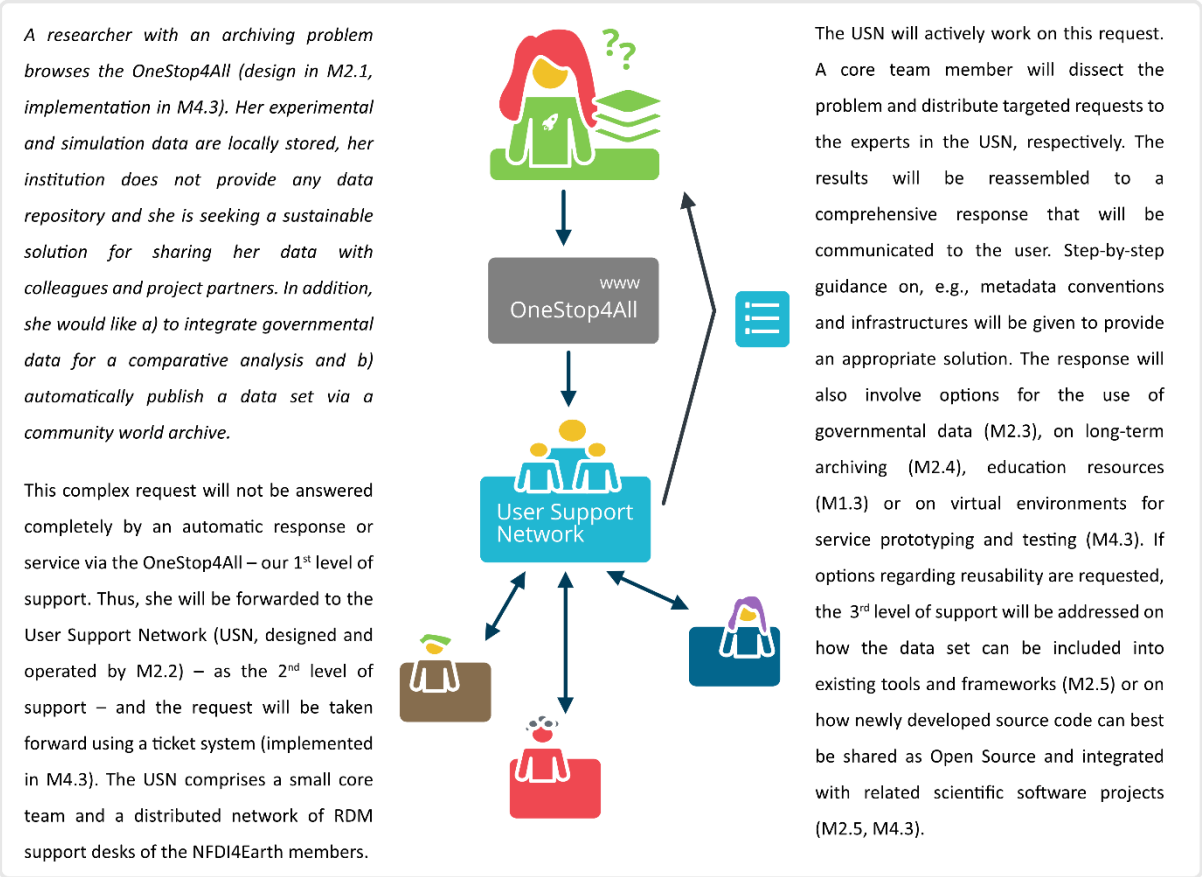

Figure 2: Example workflow from a user perspective – how a request will be routed through the NFDI4Earth measures.

NFDI4Earth will first consolidate its structure by implementing the governance described in chapter 3.4 and reinforcing community buy-in by starting the pilot projects. We will accompany this by establishing a common vision, the **NFDI4Earth FAIRness and Openness Commitment** (see chapters 4.3 and 5.4.2). The **NFDI4Earth OneStop4All** will be established and maintained as an efficient means to realize FAIR RDM and to make services findable. In conjunction, the **NFDI4Earth User Support Network (USN)** will be set up as a distributed ESS RDM help desks to provide researchers with individual support on specific RDM and Data Analysis issues. These two services are core assets to enable researchers to share their data and software, to find and publish data and software across ESS domains, assure and assess data quality, and easily utilize heterogeneous and dispersed data in their specific scientific

workflows (see Figure 2 for an individual example). Further, we will advance **innovative analysis environments (e.g. Earth data cubes, machine learning) as elements of best practice models** that can be aspired to (see chapter 5.2.5). The **NFDI4Earth EduHub** will link universities and provide **education and training** on FAIR and Open RDM, addressing researchers and other professionals at all levels (see chapter 5.1.3).

In addressing RDM infrastructure providers, we will focus on linking their services into a **Common NFDI4Earth Architecture, embedded in the NFDI and International Infrastructures**. Further, we will establish **Common Standards for FAIR ESS Data** as a foundation for evaluated and sustainable of FAIR ESS RDM services (see chapters 5.3.1 and 5.3.2). We will integrate information derived from these services and standards (e.g., their applicability for describing FAIR data, possible transformations, related vocabularies, etc.) into the **NFDI4Earth Knowledge Hub** (see chapters 5.3 and 5.4). By doing so, we will make the information (re-)usable not only for OneStop4All and the User Support Network but also for other NFDI services.

All these **developments will be supported by agile *ESS Pilots* as use cases from the ESS community**. These pilots will target specific interoperability issues for and between particular ESS subdomains. Pilots will also pioneer innovative approaches and provide blueprints for other domains. Similarly, an **Incubator Lab will foster lean and fast innovation projects to explore and prototype new methods in ES data science** and to link NFDI4Earth to cutting edge technologies (see chapters 3.4 and 5.1). Due to their integrative character and the opportunity to solve domain specific RDM and Earth Data Science problems, these elements will boost the participation of new members and spur growth of the NFDI4Earth community. In addition, the NFDI4Earth Academy guides junior scientists to implement and test new NFDI4Earth results and methods as part of their research projects, thus generating impacts beyond the original application field.

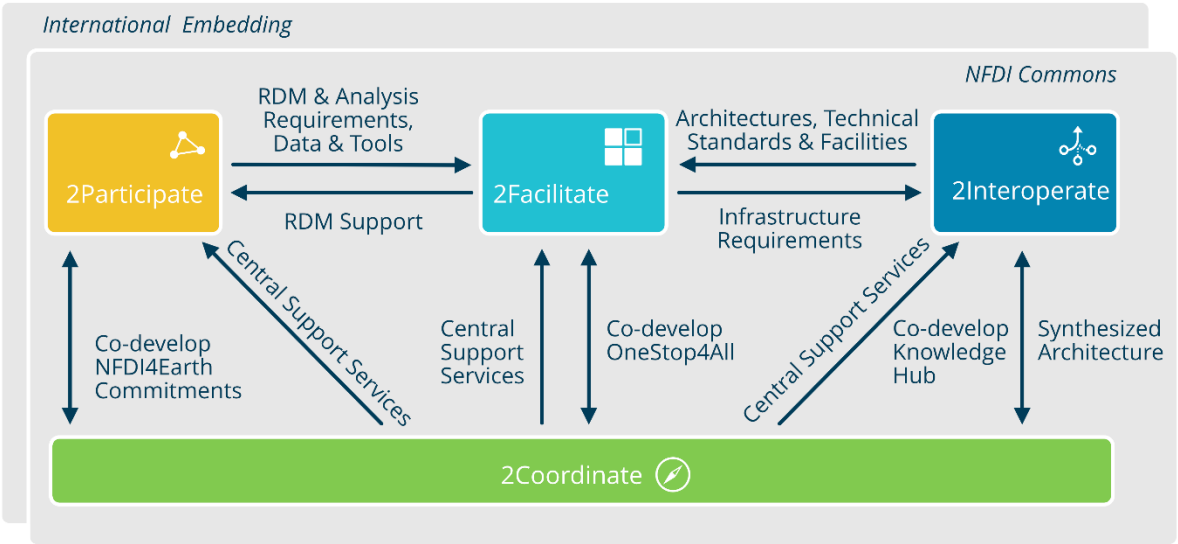

Figure 3: Major interactions between the NFDI4Earth task areas – their embedding via NFDI Commons and international networks is indicated by grey boxes.

Figure 3 illustrates the major interactions of the NFDI4Earth task areas (TAs) that we will implement. NFDI Commons and international embedding provide the framing for the TA interactions. Since a well-defined joined architecture of NFDI4Earth does not exist yet, a core community effort will be to synthesize such a common framework on the basis of existing contributions from the participants, and at the same time coordinate the necessary adaptation steps towards an interoperable functioning (2Interoperate). Coordination and central support services form the foundation of the NFDI4Earth community effort (2Coordinate). On this support layer, all other TAs will build their services and offerings. It will support different implementation needs and will e.g. secure a reliable offer of virtual environments, ticket services and education platforms. This includes the OneStop4All and the USN (both in 2Facilitate) that facilitate the provision of RDM support. Establishing and maintaining all pathways (and more) is a key element of building the NFDI4Earth in its first 5 years.

We expect – towards the end of the first 5 years – a series of new research questions in the field of Earth system data science and will strive for novel and forward-thinking projects, formulated as DFG Priority Programmes or DFG Graduate Schools to accompany the second NFDI4Earth phase.

Future work will further extend NFDI4Earth services to support research collaboration along the full data life cycle and will further elaborate on collaborative data analytics for interdisciplinary research related to ESS. The NFDI4Earth community envisions NFDI4Earth services which are competitive with, yet compatible to, other international initiatives, keeping Germany at the forefront of ESS and societal relevant innovation based on ES knowledge.

2.2 Objectives and measuring success

The following table lists NFDI4Earth's key objectives and a number of measures of success for each of these objectives:

NFDI4Earth Objective	Measure of Success
Active and continuously committed ESS stakeholders form the NFDI4Earth community and develop FAIR and Open RDM capacities in the German ESS in a community driven process.	▪ Start of the NFDI4Earth project ▪ Regular, well-attended plenaries ▪ Good response on call for Pilots, Incubators and Interest Groups ▪ Active Interest Groups (developing white papers, communications, related project ideas, etc.) ▪ Young researchers are part of NFDI4Earth Governance ▪ Academic societies sign the NFDI4Earth Commitment on FAIR and Open RDM
Enable students, scientists, lecturers, infrastructure providers and other professionals in ESS to feed and utilize an	▪ Frequent use of the OneStop4All / USN ▪ Attractive new ESS Pilots and Incubators proposed and realised

NFDI4Earth Objective	Measure of Success
open and sustainable research data life cycle and facilitate a cultural change towards a FAIR and collaborative RDM and open and reproducible ESS.	▪ ESS RDM education provided ▪ NFDI4Earth Education material provided ▪ NFDI4Earth EduHubs up and running ▪ NFDI4Earth Academy up and running and attracting junior researchers ▪ Academic societies implementing and promoting FAIR and Open activities
Establish a set of common principles, rules, standards and guidance for ESS researchers and as a qualification framework for NFDI4Earth research data, services, and infrastructures.	▪ Constitution of NFDI4Earth Commitment on FAIR and Open RDM ▪ Agreed NFDI4Earth Common Standards published ▪ Broad uptake of the NFDI4Earth commitment and agreements ▪ DMPs in ESS projects become a matter of course ▪ Uptake of NFDI4Earth Common Standards
Consolidate research data services in ESS and develop long-term perspective and strategy.	▪ Agreed Common Standards ▪ Synthesis of a Common Architecture ▪ Establishment of NFDI4Earth-based DMPs in ESS research projects ▪ Call for additional Academy and EduHub sites
Scrutinize the potential of existing and novel solutions (machine learning, interactive computations, model-data integration, and visual analytics) in a sandbox approach, continuously improving data-driven ESS research processes, to advance NFDI4Earth towards a collaborative environment for data analytics	▪ Successful ESS Pilots and Incubators ▪ Emergence and implementation of new tools ▪ Scientific papers on novel data science methods in ESS ▪ Joint publications/project proposals of Academy fellows
Contribute strongly and coherently to the overall development of the NFDI and ESS-related RDM in Germany, crosslink and interact with other NFDI consortia and develop a sustainable long-term operation model for NDFI4Earth embedded in the NFDI.	▪ Participation in the NFDI Senat ▪ Contribution to NFDI cross-cutting topics ▪ Contribution to and usage of NFDI commons ▪ Established ESS/4Earth-branch in the NFDI Verein ▪ Close cooperation with other NFDI consortia ▪ Joint workshops and publications ▪ Established network to all kind of national ESS stakeholders
Showcase examples of best practices on innovative data science approaches and tools in support of interdisciplinary research	▪ Success stories from cross-NFDI Pilots and Incubators ▪ Provide Living Handbook on the workings of NFDI4Earth ▪ Implementation of FAIR Earth data cubes linking different ES spheres

NFDI4Earth Objective	Measure of Success
	▪ Scientific papers on novel data science approaches
Establish NFDI4Earth as an integral part of international research data infrastructures and get at the forefront within international ESS RDM developments.	▪ International visibility of NFDI4Earth ▪ Establishment of one voice for German ESS RDM interests in international initiatives ▪ Strong contributions to RDA and international standardisation ▪ International Pilots

These measures of success will be further detailed in TA4 (see chapter 5.4.1). Additional success criteria that will be monitored include quantitative parameters (data usage in publications or published educational materials, data set publications that refer to NFDI4Earth, web analytics, page performance and engagement social metrics from social media channels, number of participants in NFDI4Earth activities, number of interest groups, number of signatories of the NFDI4Earth commitment, number of sites using NFDI4Earth educational material, number of collaborative projects requesting support, number of interoperability self-evaluations, etc.) and qualitative parameters (personal feedback, commentary functions, guest questionnaires, etc.). We will regularly evaluate all measures and parameters of success and will communicate the results in public reports (see chapter 5.4.1). We will also monitor and report on the cultural change towards FAIR and Open RDM in ESS (see chapter 5.4.2) and communicate these findings to the general NFDI community and other scientific communities.

3 Consortium

3.1 Composition of the consortium and its embedding in the community of interest

The NFDI4Earth consortium comprises a large representative cross section of the German ESS community and has adopted an open and inclusive membership strategy (Figure 4). The co-applicants cover all domains of Earth System Sciences or act as communication and education hubs in this field. The co-applicant institutions have diverse backgrounds and include universities, research centres, museums, the Helmholtz Association, the Leibniz Association, research data and computing centres, and cover the breadth of Earth System Sciences. Co-applicant and participant institutions also host secretariats of a variety of ESS-related German and international societies and research networks.

Equally, the participants in the NFDI4Earth consortium represent diverse backgrounds as well. They include, amongst others, several university groups with strong stakes in ESS, researchers that explore and develop novel ESS data science methods, and user support experts for (long-

term) preservation of digital ESS research results, as well as public authorities, libraries and archives (as users and providers of research data). The participants strongly support the core team of applicant and co-applicants with domain-specific expertise, additional repositories, and as multipliers in the ESS community. Co-applicants and participants often act in the dual role as providers and users of Earth system data. However, the data provided by NFDI4Earth partners is often used in a much wider international context. Since many of the participating repositories have been operational for several years, NFDI4Earth will not so much create new standards, data repositories and archives, and instead primarily concentrate its efforts on interoperability, gap analyses, consolidation and harmonization of existing standards and data infrastructures with the clear goal to bring all of them to the highest FAIRness levels. Many co-applicants and participants are heavily involved in teaching at the university and graduate level and will thus act as multipliers. Novel concepts of RDM will increasingly be included in existing curricula, and NFDI4Earth supports that with a dedicated measure (see chapter 5.1.3).

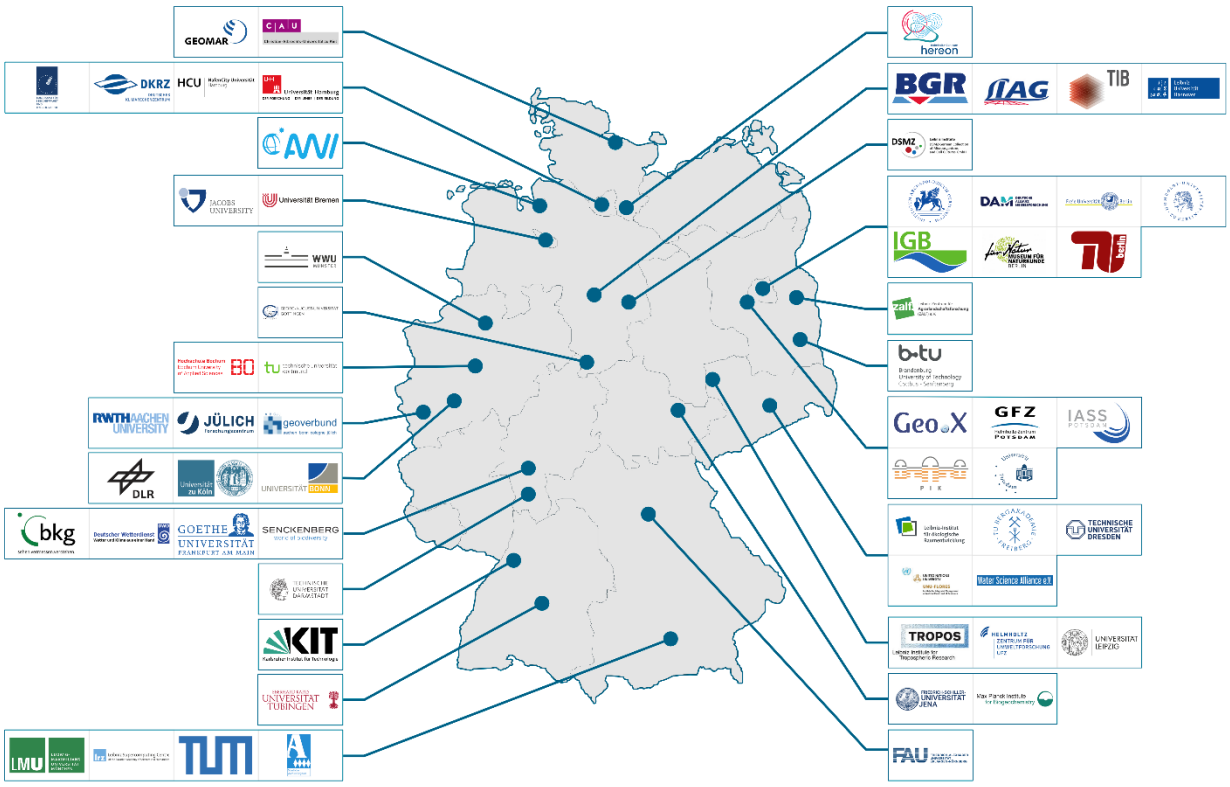

Figure 4: Map of partners in the NFDI4Earth Consortium

A community effort to screen the landscape of existing infrastructures, services and collaboration tools supporting ESS resulted in a list of more than 100 platforms and tools. Many of the screened platforms are hosted, co-operated or co-developed by one or multiple members of the NFDI4Earth consortium. Prominent examples are PANGAEA (AWI), GFZ Data Services (GFZ), Earth System Grid Federation (DKRZ), and also includes interdisciplinary services as re3data (GFZ, HUB, KIT). NFDI4Earth can further capitalize on a series of consolidation efforts, e.g. the DataHub run by the Helmholtz Research Field Earth and Environment (see chapters 3.4 and 4.2). This DataHub will

become an element of the NFDI4Earth community, supporting researchers outside the Helmholtz realm. NFDI4Earth includes several German centres for scientific computing and has close links and conditional access to HPC and storage resources (e.g. DKRZ, KIT, TUDD-ZIH, LRZ). These services will be streamlined further to offer easy accessible collaborative data science environments for NFDI4Earth.

Building the Community

The NFDI4Earth consortium was established in a comprehensive bottom-up process (NFDI4Earth). Starting in 2018, large parts of the German ESS community came together in physical and on-line meetings to shape the structure and working programme of NFDI4Earth. More than 50 partners from universities, research institutions, infrastructure providers, public authorities and different research organisations in the realm of ESS responded to a series of open calls and finally formed the consortium for a joint proposal in 2019.

Working groups were formed to map requirements and proposals by data infrastructure providers and to suggest a strategy towards building the NFDI4Earth. Following an open call for candidates, the entirety of coordination group, working group leads and consultation team unanimously confirmed the nominated NFDI4Earth Spokesperson Lars Bernard and TU Dresden as the applicant institution in May 2019. Further – from an open call for expressions of interest – an agreement was reached on the co-applicant institutions and co-Spokespersons representing the various ESS domains and related RDM stakeholders. The consortium was and is fully supportive of NFDI4Earth as a common undertaking to unite RDM activities in the ESS realm. In its early phase, the consortium elaborated and agreed on a core set of needs and aspects:

1. **Cultural change in ESS for the digital age**, requiring efforts towards responsible FAIR and Open RDM as a foundation for research data sharing and reproducible research, proper recognition of research data originators, broad use of persistent identifiers and adequate perception, support and tools for achieving interoperability.
2. **Levels of RDM knowledge are very diverse** in the community, ranging from digitisation frontrunners to novices. There is a need for **education on RDM in ESS** and in ESS-related data science at all graduate and professional levels.
3. **Many researchers in ESS require guidance** on where to find ESS-related research data and on how and where to publish and to sustainably archive research data.
4. **Consolidation of the existing plethora of repositories to achieve sustainability**: This includes long-term archiving and active guidance on how to find an easy path to the environments and repositories best suited for certain types of research data.
5. **Harmonizing standards and agreement on common core standards** is needed (e.g. in the context of metadata, vocabularies, encodings, application programming and service interfaces) as a foundation for interoperability.

6. **A common national umbrella – as intended by NFDI4Earth – does not yet exist**. A number of sub-disciplines are well integrated in international data-infrastructure networks (e.g. on providing coherent sensor, observational, laboratory and model data). In addition, a number of related initiatives and measures are currently underway by the Helmholtz Association, the Leibniz Association, and the Max Planck Society as well as on national and on international levels, which should be networked and join NFDI4Earth.
7. **A gateway and hub for ESS researchers to innovate approaches, methods and tools for interdisciplinary data science in ESS is required:** There is tremendous potential for establishing NFDI4Earth as a means towards more interoperability in ESS to more efficiently find, share, access, integrate and collaboratively analyse data across the boundaries of ESS spheres.

In response to these needs, we collaboratively developed the work plan for the NFDI4Earth proposal submitted in October 2019. The expertise of the consortium and the Spokespersons, the NFDI4Earth objectives, as well as the proposed RDM approach generally convinced the international reviewing board in the open and constructive DFG review process in late 2019. However, the reviewers also voiced concerns about (a) the maturity of the work plan and proposed milestones and risk assessments; (b) demonstrative benefits for the ESS community; and (c) outreach to the entire ESS community and integration of their data. In conclusion, the NFDI Expert Group recommended a resubmission of a revised proposal for NFDI4Earth.

To adequately reflect the reviewers and expert group's recommendations, we held a series of (largely virtual) plenary meetings and specific workshops, resulting in a carefully co-designed and reshaped NFDI4Earth structure, including a detailed adjustment of the work plan. We presented the reshaped concept for NFDI4Earth to the DFG Senate Commission on Earth System Research and received positive feedback.

Continuing as *one* ESS NFDI consortium

The reviewers highly appreciated the one-community approach of NFDI4Earth as the way to tackle the interdisciplinary research challenges. We were recently able to integrate two NFDI-related consortia (NFDI4SolidEarth/NFDI4Lithosphere, NFDI4NaturalRessources). This further strengthened NFDI4Earth and ensures the one-community approach towards common RDM for ESS. The ESS community clearly signalled that it is agreed and well understood that NFDI4Earth is designed as an open consortium and will grow to include interested partners developing and establishing NFDI4Earth.

We slightly reorganized the group of co-applicants to reinforce user perspectives: We strengthened the participation of universities to now also include the University of Leipzig and the University of Frankfurt. The GEOMAR Helmholtz Centre for Ocean Research and Forschungszentrum Jülich withdrew from their roles as co-applicants. Meanwhile, all participating

Helmholtz institutions consolidated and extended their commitment on substantial contributions from the Helmholtz Research Field Earth and Environment DataHub (see chapter 3.4).

(Re-)Shaping the NFDI4Earth work plan

We introduced the concept of the ESS Pilots, in particular to engage with our ESS community even more closely and more openly and to drive NFDI4Earth in an agile manner, triggered by specific needs and contributions from the broader community. We formulated tangible milestones and deliverables for the five-year implementation plan. We now additionally provide a general risk assessment and contingency plan. By reshaping, we are enriching the development plan in linking in a more obvious way to innovative concepts such as the Earth cubes and the iterative implementation of a federated Knowledge Hub as the backbone for our NFDI4Earth work and services (User Support Network, OneStop4All, EduHubs…).

ESS Pilots as drivers of agile development and Incubators as innovation scouts

ESS Pilots are individual one-year projects of participants funded through NFDI4Earth and will commence at the same time as NFDI4Earth. Pilots are use cases from the entire breadth of ESS domains and address all aspects of data harmonization, implementation, interoperability, setting up services, building tools, etc. We use pilots as instruments for developing best practices and deriving requirements. The pilots will be linked with the individual measures and serve as user-contributed building blocks to design the specific infrastructure in an agile process (Figure 5). The pilots will further initiate a user community from the start, thereby driving the bottom-up approach. Additionally the pilots facilitate cooperation with citizen science projects and commercial data and solution providers related to ESS (see list of pilots in chapter 5.1.1).

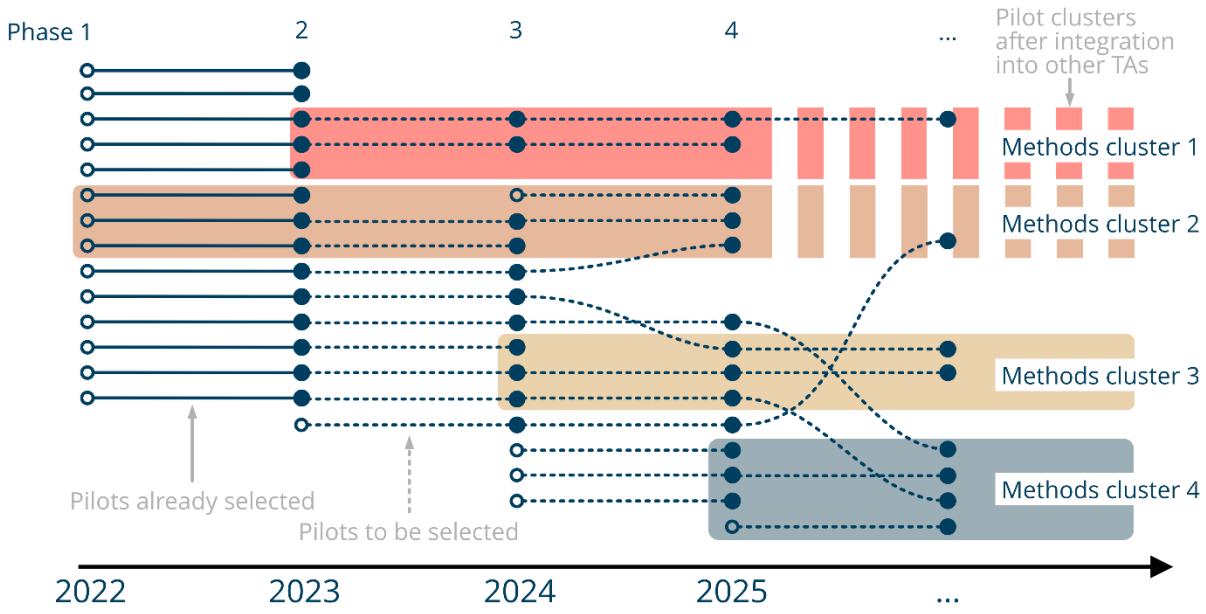

Figure 5: The NFD4Earth consortium calls for pilot studies in four phases. The first set of pilots has been already selected in 2020 to start shortly after project start. Depending on scope and performance, pilots will end after one phase or continue until maturity. Initially independent pilots shall converge in methodological clusters and mature solutions into NFDI4Earth services.

The response from the ESS community was tremendous – despite a short-term notice of a call for applications – to an open call for pilots in June 2020 with an impressive number of 38 contributions representing the plethora of ESS RDM aspects and related innovations. 14 pilots were selected in a transparent review process. We plan for three more open calls in the first five years allowing both the application for continuing a pilot and the (re-)submission of new pilots (see chapter 5.1.1 and www.nfdi4earth.de for details). We kept the instrument of NFD4Earth Incubators for hands-on trend-scouting and prototyping of ESS data science innovations and tools in short-termed projects. Again, we plan for four calls for Incubator projects in the next five years (see chapter 5.1.2 for details).

NFDI4Earth Interest Groups as cooperation and communication platforms

NFDI4Earth Interest Groups (IGs) offer a temporary or permanent forum for a specific information community or stakeholder group to e.g. identify, discuss and communicate requirements for NFDI4Earth, jointly advance NFDI4Earth-related concepts, technologies and processes, implement NFDI4Earth principles, or to disseminate specific NFDI4Earth offers.

NFDI4Earth IGs can contribute to NFDI4Earth in various ways: IGs are entitled to submit proposals to the NFDI4Earth Steering Group to address requests concerning NFDI4Earth development (e.g., for support of specific domain specific requirements, development of specific standards or their adoption, etc.). IGs can also elaborate on a communication or white paper, which, once approved by the NFDI4Earth Plenary, becomes a common NFDI4Earth position.

A group of NFDI4Earth stakeholders can initiate an Interest Group any time following a set of guidelines published on the NFDI4Earth website (see www.nfdi4earth.de for more details). Initial proposals for NFDI4Earth IGs cover topics such as Geophysics, Data Management Plans, Reproducible Software, Industry and Company Data, FAIR Data, and Interoperability.

NFDI4Earth as a long-term competence network attracting next-generation researchers

The NFDI4Earth work programme involves a sizeable number of participants and stakeholders such as academic societies, who are expected to continuously act as multipliers in their respective institutions and communities. The NFDI4Earth Academy and the NFDI4Earth EduHubs are means to establish a sustainable network of ESS-related education and data science competence centres, jointly advancing capacity building and developing innovative approaches. In addition, we will closely network with academic societies and use surveys, workshops and conferences to actively engage the NFDI4Earth consortium with research in Earth System Sciences and to allow several feedback loops. The NFDI4Earth education measures, the Academy and the activities of academic societies will specifically address next generation ESS researchers.

Profiles, Strengths and Roles in the NFDI4Earth Consortium

The following brief descriptions summarise the characteristics and strengths of the applicant and the co-applicant institutions and their planned role in the consortium:

Technische Universität Dresden (TUDD) is for the second time one of Germany's Universities of Excellence. TUDD is the largest university in the federal state of Saxony and one of the largest of the German Universities of Technologies. The TUDD Faculty of Environmental Sciences unites the three disciplines of forest, geo and hydro sciences, which is unique in Germany. Activities focus on monitoring and modelling the Earth System on the global, regional, and local scales, and use of this knowledge in the context of the sustainable development of the human environment. Research and teaching cover a wide range of topics and are firmly embedded in both regional and international networks. Here the NFDI4Earth Spokesperson holds the chair of Geoinformatics, which has extensive experience in research related to the design and establishment of (scientific) geodata infrastructures and interoperability for geodata analytics. Together with the Helmholtz Centre for Environmental Research (UFZ), the faculty established the Center for Advanced Water Research (CAWR), which bundles their water competences. Within its implementation of the excellence strategy measures, TUDD will further strengthen research on Smart Water Systems in an Extreme World. TUDD hosts the coordination office of the Water Science Alliance e.V. fostering the networking and collaboration of research institution and authorities to strengthen German Water Research. NFDI4Earth also builds on the competencies and services of the Center for Information Services and High Performance Computing (TUDD-ZIH), which operates IT services and IT infrastructure for TUDD and is the High Performance Computing centre of the state of Saxony. TUDD-ZIH is a founding member of the Gauß Alliance, an association of national science centres to promote High Performance Computing (HPC) at a national level as an independent strategic research activity. Since 2014, TUDD-ZIH has been leading the BMBF-funded national competence centre for Big Data and Artificial Intelligence *ScaDS.AI Dresden/Leipzig*, which conducts R&D for the handling and analysis of large or complex data sets and develops methods and tools. With the new hardware and software infrastructure HPC-DA (High Performance Computing - Data Analytics), ZIH offers researchers a unique tool for processing and analysing all kinds of data. TUDD will further enhance the Service Center Research Data, a joint undertaking of TUDD-ZIH and the Saxon State and University Library, founded in 2017 to support scientists in research data management. The Service Center Research Data acts as the coordination office of SaxFDM, the joint initiative of Saxon universities and research institutions on Research Data Management. In sum, TUDD's competences and planned measures perfectly complement TUDD's major engagement as an applicant for NFDI4Earth. TUDD acts as the NFDI4Earth Applicant. TUDD leads TA4 and implements the NFDI4Earth Coordination Office. Furthermore, TUDD will contribute to TA1, TA2, and TA3.

The **Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research (AWI)** is situated in Bremerhaven with research stations in Potsdam, Helgoland and Sylt. It is Co-founder of the Helmholtz-Institute for Functional Marine Biodiversity HIFMB in Oldenburg. Research is

structured in three research divisions: climate sciences, geo sciences and bio sciences. AWI operates the research ice-breaker POLARSTERN as its largest research infrastructure for research in the marine realm and the cryosphere, with access for the marine and polar science community. To support acquisition of research data from multiple sensors and research platforms, AWI has extensive experience in the development of professional data management and data publication environments – both as hardware systems and software services. Data publication is done in close collaboration with University Bremen by the well-established PANGAEA systems and workflows, including curation for quality control, parameter descriptions and further metadata. The software and workflow developments are complemented by introducing new data platforms, enabling cloud services within the Helmholtz Data Federation. The Observation to Archive and Analysis (O2A) system in particular is central software that covers the full data life cycle, from source to analysis and further to re-use. O2A is AWIs contribution to the DataHUB in Research Field Earth and Environment, opening up to science communities from polar terrestrial and marine research, also in NFDI. In the DataHUB AWI exchanges know-how on concepts of dataflows, distributed data processing and data portal technologies, aiming at presenting innovative data products and analyses. AWI is one of the Co-Applicants responsible for TA3. Beyond coordination of the architecture definition and improvement process in TA3, interfacing to DataHuB concepts, services and developments will be secured.

The **German Climate Computing Center (Deutsches Klimarechenzentrum GmbH, DKRZ)** is a central service centre for German climate and earth system research. Its high performance computers, data storage and services form the central research infrastructure for simulation-based climate science in Germany. Particularly in the field of data management, the DKRZ offers its scientific users services that span over the entire life cycle of research data from planning to generation, harmonization, publication, distribution, (inter)disciplinary re-use and long-term archiving. Noteworthy is the CoreTrust Seal (CTS) certified World Data Centre for Climate (WDCC), which in combination with the DKRZ-operated complete set of Earth System Grid Federation (ESGF) services plays a highly respected role worldwide. DKRZ participates in numerous national and international projects and collaborations that are of importance in climate research and beyond. The Excellence Cluster CLICCS (Climate, Climatic Change, and Society) in particular, but also projects such as the EOSC-hub that are involved in the build-up of the European Open Science Cloud (EOSC) are prominent examples and are highly relevant in the context of NFDI4Earth. DKRZ is one of the Co-Applicants responsible for TA2.

The **Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences (GFZ)** is Germany's national research centre for the solid Earth Sciences. GFZ has a comprehensive spectrum of expertise in geodesy, geophysics, geology, mineralogy, geochemistry, physics, geomorphology, geobiosciences, mathematics, and engineering. A core task of GFZ is the provision of research infrastructures, information infrastructures, and networks. The GFZ

research infrastructure comprises satellite systems, global real-time instrument networks and regional observatories, lab facilities, instrument pools, and data systems. The Modular Earth Science Infrastructure (MESI) helps to accomplish these large-scale tasks in a coordinated manner. A large part of GFZ infrastructure is available to external researchers for scientific usage. GFZ is one node of EIDA (seismic European Integrated Data Archive) engaged with the World Stress Map in WDC. GFZ offers a domain-specific service for citable data publication (GFZ Data Services), supports research data publication on a national level with the project Specialised Information Service Geosciences (FID GEO), and is experienced in systems for persistent identifiers, e.g. the central registry for the International Geo Sample Number (IGSN). GFZ contributes solid earth to the DataHub in the Helmholtz Research Field Earth and Environment and transfers solutions to NFDI. GFZ participates in a large number of national and international projects and collaborations related to data management (i.e. operation of data services for the International Association of Geodesy IAG or as Operational Support Group of the International Continental Scientific Drilling Program ICDP) and offers a strong link to associations such as IUGG and its members' data services. Furthermore, GFZ, as host of the Helmholtz Open Science Office, can serve as an intermediary to national and international Open Science initiatives.

GFZ hosts the general office of **Geo.X**, a research network of five universities and six non-university research institutions in the Berlin and Potsdam area. Covering the entire spectrum of geoscientific disciplines, Geo.X partner institutions join forces for close cooperation in the fields of research, teaching, shared infrastructure and data management, internationalization, and knowledge transfer. Within the joint structure of the Geo.X Young Academy, the partners enable early-career scientists to develop scientific profiles in the interdisciplinary research areas Geo.Data Science and Geo.Society. The network has communication paths and hosts various events tailored to the needs of diverse target groups. GFZ is one of Co-Applicants responsible for TA3 and here specifically to coordinate and foster the international embedding of NFDI4Earth. Further, GFZ/Geo.X is one of the Co-Applicants responsible for TA1 with a focus on establishing the NFDI4Earth Academy. In TA4, GFZ leads the measure M4.2 working on cultural change towards FAIR and Open RDM.

The **Bochum University of Applied Sciences (HSBO)** and its departments of Geodesy, Electrical Engineering and Computer Science as well as its currently being founded Interdisciplinary Institute for Applied Artificial Intelligence offer a broad range of competencies and resources for implementing the NFDI4Earth work programme. HSBO is well connected to the national landscape of universities of applied sciences and can thus act as a multiplier and interface to NFDI4Earth-relevant actors within this group (e.g. via the Graduates Institute NRW). The staff has a strong R&D background in developing and implementing spatial data infrastructures on a national and sub-national level, specifying interoperable and standards-based spatial data processing environments as well as data analytics and machine learning. HSBO

offers study courses – i.a. focusing on spatial data management and analytics – which provide a basis for the development of NFDI4Earth-specific education and training offers and which will be iteratively adapted to NFDI4Earth principles. HSBO is one Co-Applicant responsible for TA1 with a focus on the aspects of capacity building and education and contributes to M4.3 to build the technical EduHub environment.

The **Karlsruhe Institute of Technology (KIT)** is the Research University in the Helmholtz Association. It combines a state university with over 25,000 students with a federal research lab creating and imparting knowledge for the society and the environment. Recently, KIT was selected as one of the German Universities of Excellence. A range of institutes at KIT will support the NFDI4Earth endeavour. For example, the Steinbuch Centre for Computing (SCC) is an experienced provider of infrastructure and software services and the library is a supplier of catalogue and PID services (in conjunction with the SCC). The Institute of Meteorology and Climate Research (IMK) and its four departments provide ESS experience in atmospheric research and adjacent fields (additional ESS experience is available at KIT via the Climate and Environment Center – a structural element of KIT governance). IMK is part of the Helmholtz Research Field Earth and Environment and manages the university degree course Meteorology. IMK participates in the Research Field Earth and Environment DataHub initiative that complements DAM development and the institute is a premium user of SCC resources. There is an RDM service team which supports the institutes with respect to their specific implementation requirements. KIT is one of the Co-Applicants responsible for TA2.

Leibniz University Hannover (LUH) and its research centre FZ:Geo offer broad expertise in the Earth System Sciences domain at the intersection of Geoscience, Computer Science and Engineering. A major topic of joint exploration lies in digitisation, with a focus on advanced AI technologies, as well as and handling uncertainty. Being involved in large coordinated research projects involving massive data acquisition and processing, also data management is of elementary importance. As a University, the link to education and training of next generation Earth scientists is an immediate task. There is also very good experience in developing MOOCs for e-learning and blended learning purposes. LUH is one of the Co-Applicants responsible for TA1. LUH will coordinate the Incubator Lab.

Goethe-Universität Frankfurt am Main (UNIF) is the largest university in Hessen and currently hosts one excellence cluster. The department of Geosciences (Institut für Geowissenschaften – IfG) has 15 professors, approx. 35 scientific staff, 35 doctoral students and 400 students. The IfG covers a diverse field of research from biodiversity to climate science, geochemistry, mineralogy, geophysics, structural geology, paleaontology, petrology to geochronology. It recently established a new group ‘Digital Mineralogy’ with a focus on developing tools to access, visualize and analyse data from geo- and cosmo-chemical repositories directly in a web browser. This is based upon many years of experience with building the interactive web interface for the MetBase

meteorite repository. This Digital Mineralogy group cooperates on a national level with the University of Göttingen to apply these tools to the geochemical GeoROC repository, and communicates on an international level with e.g., IEDA and Astromaterials to allow for broad interoperability. The group has further extensive experience on implementing, maintaining and long-term archiving a curated data repository for long-tail data with MetBase. It is invested for several years with digital learning formats and recently released an entirely new and innovative blended learning approach that is directly linked with interactive tools to visualize and analyse data from repositories. Ultimately, the entire approach allows for data-driven research as well as learning. UNIF is one of the Co-Applicants responsible for TA3. UNIF will specifically contribute on NFDI Commons and long-tail data.

The **Universität Leipzig (UNIL)** has a strong tradition in the Earth System Sciences, and is particularly known for excellent research in climate feedback (cf. DFG Collaborative Research Centre Arctic Amplification, http://www.ac3-tr.de/), biodiversity and ecosystem research (cf. the German Centre for Integrative Biodiversity Research, https://www.idiv.de/), among others. In 2020, UNIL has substantially reinforced the area of Earth system research by founding the Remote Sensing Centre for Earth System Research (https://rsc4earth.de/) in collaboration with the Helmholtz Centre for Environmental Research, UFZ, and the appointment of multiple new professorships in the area (three appointed, new calls to be expected soon). UNIL's strategy is to bridge various sub-branches of Earth system research by exploiting the joint potential of novel satellite remote sensing data and big data technologies. The latter is possible, given that UNIL has been committed to foster research and education on artificial intelligence. For instance, UNIL is also one of the founding partners of ScaDS.AI and has recently established a new Master degree in Data Science - a degree on Earth System Data Science and Remote Sensing has been approved and will be established in 2021/22. To foster FAIR data developments, UNIL has established a RDM working group comprising the Department of Research and Transfer, the Universities Computer Center and the University Library. The RDM group is actively participating as spokesperson in SaxFDM. Furthermore, through the Arqus European University Alliance, in which UNIL is one of seven participating European Universities, the consortium will have access to a broad network of Open Data and Open Science experts in the framework on the Openness Taskforce. The focus of an additional newly granted H2020 project for the Arqus Alliance is on the fields of Artificial Intelligence/Digital Transformation and Green Deal/Climate Change as transversal priority areas for enhanced research collaboration. UNIL is one of the Co-Applicants responsible for TA1, and participates in TA2. UNIL will specifically contribute on NFDI community engagement and data cube technologies.

The **Max Planck Institute for Biogeochemistry in Jena (MPIBGC)** is a research institute of the Max Planck Society. Within the NFDI4Earth, the MPIBGC acts as the representative for the Earth and Solar System Research Partnership (ESSRP). The ESSRP additionally encompasses the

Max-Planck-Institutes for Chemistry in Mainz, for Meteorology in Hamburg and for Solar System Research in Göttingen. The mission of the ESSRP is to pool research excellence across disciplines to understand how the Earth functions as a complex system and to improve the predictability of the effects of human activity. The MPIBGC combines observational and process-based studies with global scale modelling and contributes long-term expertise in model-data integration, machine learning of global environmental processes, as well as developing novel approaches to combine data-driven approaches with physical modelling. The MPIBGC has successfully coordinated a range of data-driven EU projects (e.g. CARBOEUROPE-IP, CARBO-Extreme, BACI) and has also developed concepts for interactive computations on large gridded data in the Earth System Sciences (earthsystemdatalab.net) supported via various projects of the European Space Agency. MPIBGC is one of the Co-Applicants responsible for TA2. MPIBGC will especially focus on the provision of Earth cubes.

The **Senckenberg Nature Research Society (SGN**) is Germany's largest natural history research facility, including seven research institutes and three museums, which keep vast natural history collections in the fields of Geology, Paleontology, Botany and Zoology. Within its research field 'Biodiversity and earth system dynamics', SGN describe the dynamics of the total 'System Earth' focusing on interactions between the Earth's surface and the biosphere, the way the Earth has developed and it's early environmental conditions. SGN maintains and contributes to a range of IT portals, in particular to AquilaGeo (together with TUBAF) for geoscientific collection objects. Digesting data and communicating these to various scientific communities and the wider public is thus a focus of the institution. Within the ESFRI roadmap project DiSSCo, SGN coordinates design and implementation of the central infrastructure (encompassing multiple millions of digital geoscientific and biological specimens in European collections). SGN is one of the Co-Applicants responsible for TA3. SGN will coordinate the development of the semantic interoperability framework.

Roles of the Participants

The strengths of the 48 NFDI4Earth participants span a broad range of expertise, roles, duties, responsibilities, and competencies related to ESS research data management (see chapter 1.7). In open calls, NFDI4Earth participants were asked for their particular interests and expertise in contributing to one or more of the NFDI4Earth Task Areas (TA1 to TA4) and their specific measures. The responses show that the NFDI4Earth consortium covers the entire spectrum of task areas and provides a very good team to start the NFDI4Earth work programme (Figure 6).

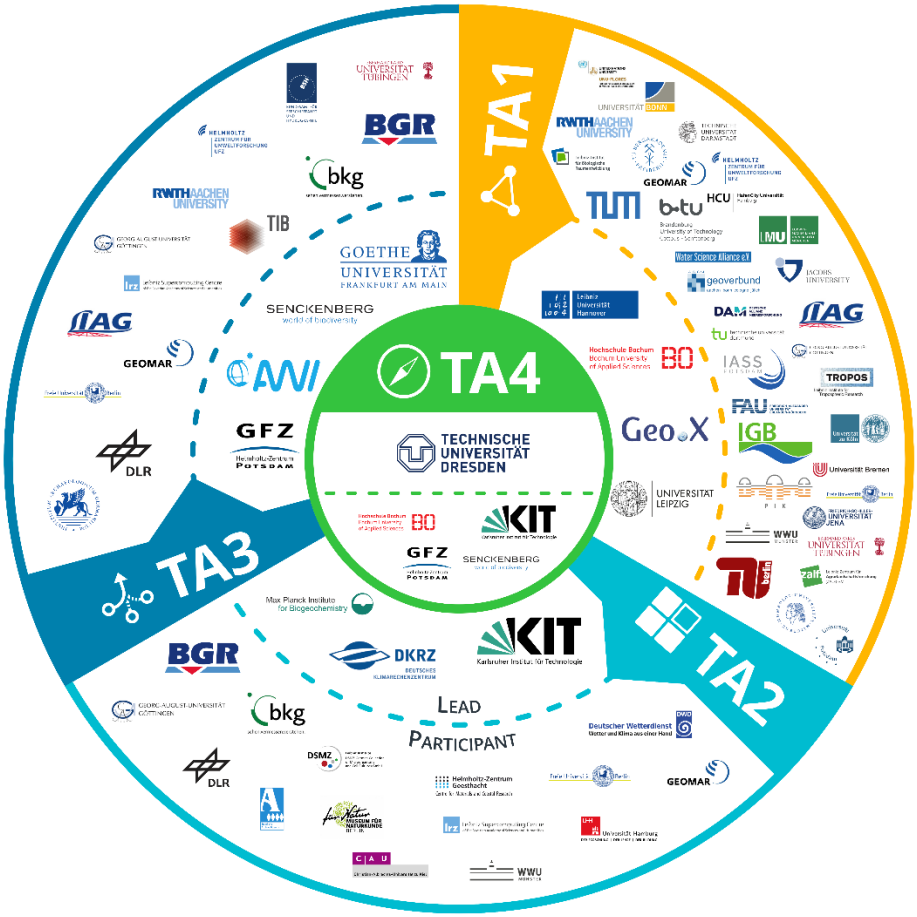

Figure 6: Overview of the partners' contributions in the NFDI4Earth task areas

3.2 The consortium within the NFDI

Interfaces to other (proposed) NFDI Consortia

The NFDI4Earth co-applicants have already participated in a number of cross-NFDI communications and workshops. In summer 2019, NFDI4Earth started jointly with other NFDI Consortia (DataPLANT. GHGA, KonsortSWD, NFDI4Agri, NFDI4BioDiversity, NFDI4Cat, NFDI4Chem, NFDI4Crime, NFDI4Cuture, NFDI4Health, NFDI4Ing) to develop NFDI Cross-Cutting Topics (see below) and to prepare collaborative development of the NFDI. We address common NFDI aspects and NFDI-wide cooperation in all NFDI4Earth Task Areas. M3.3 (in TA3) in particular is dedicated to ensure our contributing and bi-directional interfacing with other NFDI Consortia. We have already set up collaborations to initiate this process (see below and M3.3). M4.1 (in TA4) will work towards a long-term NFDI4Earth operation model embedded in the NFDI.

The NFDI4Earth consortium is fully committed to the idea of the Research Data Commons (RDC) as the NFDI-wide agreed and developed set of fundamental and common NFDI standards, technological approaches and infrastructure components and will streamline its design and developments in line with this concept. Here, we contributed to the recent position papers on NFDI cross-cutting topics as well the RfII recommendations papers on Data Quality, Digital Competencies (RfII, 2019; RfII, 2020).

Furthermore, NFDI4Earth will in particular act as the NFDI expert pool for dealing with Earth related spatio-temporal data – even beyond the creation of the NFDI4Earth Knowledge Hub as a community service. NFDI4Earth agreed jointly with KonsortSWD, NFDIAgri, NFDI4BioDiversity, NFDI4Objects, NFDI4Microbiota, NFDI4MobilTech and PUNCH4NFDI on cross-consortium activities and developments to demonstrate the NFDI benefit for interdisciplinary research. NFDI4Earth plans to run common summer schools on innovative aspects in ESS RDM and in ESS data science in close cooperation with NFDI4BioDiversity (see chapters 5.1.4 and 5.3.3).

In addition to the joint work and interfaces addressed as cross cutting topics (see below), NFDI4Earth consortium members also participate in other NFDI consortia, thus acting as direct interfaces. In sum our members participate in BERD@NFDI, DAPHNE4NFDI, DeBioData, DataPLANT, FAIRmat, GHGA, KonsortSWD, MarDI, NFDI4Agri, NFDI4BioDiversity, NFDI4Cat, NFDI4Chem, NFDI4Culture, NFDI4Datascience, NFDI4Health, NFDI4ING, NFDI4Memory NFDI4MobilTech, NFDI-Neuro, NFDI4Objects, NFDI4Phys, NFDI4RSE, NFDIxCS, PAHN-PaN, and Text+.

Common elaboration on NFDI Cross-Cutting Topics

The NFDI4Earth co-spokespersons co-drafted and signed the *2019 Berlin Declaration on NFDI Cross Cutting Topics* (Glöckner et al., 2019) and the *2020 Leipzig-Berlin Declaration on NFDI Cross-Cutting Topics* (Bierwirth et al.) to define a core set of NFDI topics. We identified the following aspects that need to be tackled in cross-NFDI activities:

- NFDI governance model and sustainability, operational, cost-covering and legal models for the coordination bodies/offices of the NFDI and the NFDI consortia
- NFDI-wide harmonization and coordination (1) on legal aspects (licensing, intellectual property rights and data protection) and ethical aspects of sharing research data and research software; (2) on terminologies (vocabularies and ontologies, reference systems, code lists); common data and metadata standards and encodings (e.g. NFDI core metadata) and unique identifier systems; (3) on data and service quality criteria, evaluation and qualification criteria, qualification and/or certification processes for NFDI service offerings and infrastructures, etc.
- International and European embedding of the NFDI consortia, safeguarding NFDI needs and requirements in European/international developments and initiatives (e.g. RDA) on research data infrastructures (e.g. EOSC) and in international standardization bodies (W3C, ISO, etc.)
- Establishment of a set of common NFDI shared services that exist in a scientific cloud solution that is supported by and supports German institutions and researchers (e.g. computing and storage services, collaborative working environments, authentication and access mechanisms, registries, long-term archiving etc.)
- Establishing coherent NFDI user-support, linking to existing and emerging research data help desks and support units in the various NFDIs, research institutions and universities

- NFDI capacity building and education with activities towards capacity building on RDM at all levels, establishing professional RDM education and careers
- Stimulating a cultural change of ESS data users and providers towards FAIR and Open research data, establishing scientific reputation for research data activities and engagement in scientific software developments, initiating common conferences, graduate schools and research projects related to innovative aspects in RDM and data analytics. This is a fundamental building block of the digitalisation strategy that our society requires.

How NFDI4Earth contributes to the NFDI Cross-Cutting Topics

The NFDI4Earth work plan contains several measures to establish common NFDI services and we will actively and strongly:

- Contribute to the NFDI governance model and sustainability models. Several of the NFDI4Earth co-applicants and participants operate research infrastructures of various sizes and with different operation models. Building on the broadness of existing practices and experience and fully acknowledging the need for linking and consolidating the existing plethora of services and tools, NFDI4Earth members have a vital interest in developing sustainable NFDI operation models (TA3 and TA4).
- Offer federated user support for other NFDI Consortia by implementing the NFDI4Earth User Support Network for researchers (TA2, M2.2) – which is proposed as a blueprint for a NFDI-wide User Support Network.
- Serve as the NFDI competence pool for accessing, using and analysing Earth-related spatio-temporal data – for all different disciplines and application areas that require access to such data and related competencies – and to warrant interoperability to related standards (metadata, PIDs, encodings, vocabularies, APIs; TA3).
- Provide strong links to engage and align with governmental data services, data infrastructures (e.g. GDI-DE) and archiving services (M2.4).
- Provide networks, engagements and experiences for NFDI international embedding (M3.4).
- Provide links to a variety of citizen science projects and commercial data and solution providers in ESS as partners in the ESS pilots (see M1.1).
- Participate in NFDI harmonization efforts such as in inter-consortium working groups. Several NFDI4Earth members can build on widespread and valuable experiences in contributing to international standards and agreements, specifically for metadata standards and spatial data encodings (TA3, all measures).
- Contribute to concepts concerning open scientific software, capitalizing on a well-established culture and skills-set on community software developments in ESS (TA2, M2.5 and TA4, M4.3).
- Demonstrate a new quality in interdisciplinary research in common, cross-consortium applications. We have already elaborated on such applications jointly with NFDI4Agri,

NFDI4BioDiversity and KonsortSWD. Here, the pilots provide the means to develop common cross-NFDI pilots in the course of NFDI projects (TA1, M1.1).

- Provide input to education and capacity building modules related to Earth system data management and analysis (TA1, M1.3).
- Incentivise cultural change towards increased awareness of FAIR principles and Open Science by improving the embedding of data publications in the scientific publication process and the recognition of data and software publications for their scientific reputation. For this purpose, NFDI4Earth will strongly get involved with academic societies and ESS-relevant Specialised Information Services (Fachinformationsdienste, FID), i.e. FID GEO, FID Karten, FID Montan (TA4, M4.3).
- NFDI4Earth and NFDI4Biodiversity agreed to commonly contribute to the vision, co-design and stepwise implementation of a NFDI-wide usable, common storage and compute environment. Here, NFDI4Earth is building on the commitment of its members who operate powerful HPC and storage infrastructures (TA3).

Meanwhile, NFDI4Earth enables all ESS researchers to participate in the NFDI, opening up the many opportunities it offers.

3.3 International networking

NFDI4Earth is, due to its members, actively embedded in international initiatives and in various international networks. This applies to ESS-related initiatives on research data infrastructures such as COPERNICUS, ECMWF, ENVRI-FAIR, EPOS, ICOS, IPCC, IAG, WMO, and WDS, and engagement in organisations such as the Open Geospatial Consortium (OGC) to ensure the establishment of international standards. It also includes accentuated roles in cross-cutting initiatives as the Research Data Alliance (RDA) by, for instance, organising a global plenary or annual RDA-DE meetings. Many participants are actually engaging in EOSC.

Thus, NFDI4Earth not only covers all relevant ESS aspects on international scale, it can also internationally promote best practices and solutions coming from NFDI4Earth and benefit from those networks and engagements in terms of the ongoing establishment of accepted standards and the implementation of international directives. Especially the measures within TA3 ensure and promote the international visibility and awareness of NFDI4Earth and support the development of novel avenues of RDM by connecting and actively participating in international initiatives. Thus, by forging cooperation with different stakeholders ranging from the ESS and RDM communities, NFDI4Earth strengthens networking and promotes its portfolio in and outside the German academic research system. This includes positioning NFDI4Earth activities in international Open Science efforts. NFDI4Earth intends to contribute actively and continuously to the establishment and operation of a network of international contact points within the NFDI as a

whole (TA3, M3.4) to support a coherent international positioning of (ESS-related) German RDM interests.

There are several large European and international research infrastructures and advanced data services for ESS data, e.g. Research Infrastructure Consortia (ERICs; European Commission, 2019a) such as the European Plate Observing System (EPOS) and the Integrated Carbon Observation System (ICOS) on the European level, the Global Earth Observation System of Systems (GEOSS) on an international level, and also discipline-specific approaches such as the International Federation of Digital Seismograph Networks (FDSN) and the Network for the Detection of Atmospheric Composition Change (NDACC). NFDI4Earth will need to build and maintain close collaboration with many of them in order to keep abreast of novel avenues of metadata standards, concepts, information technologies and interfaces. NFDI4Earth will equally connect to international efforts such as the climate data store (ECMWF), the European Space Agency (ESA) Data Cubes, the EOSC, and the US based IEDA and EarthCube data services efforts to develop novel avenues towards making scalable data analytics available to a wide user community in the cloud.

We will also inspire and contribute to the development of international RDM standardization. Participation in international standardization bodies will allow us to capitalize from international development and to safeguard the overall NFDI needs in terms of RDM developments. This also includes engagement in standardization and regulation work closely related to ESS as e.g. in CTS, INSPIRE, OGC, and W3C in general RDM and e-science related Initiatives such as RDA, CODATA, FORCE11, OpenAIRE and strong integration with general research services such as DataCite, ePIC, re3data, ORCID and WDS.

International assessment reports on ES-related topics provided by intergovernmental institutions such as IPCC or IPBES refer to scientific data and include ESS data as references in their assessments (e.g. the WMO/UN Environment sponsored Scientific Assessment of Ozone Depletion every four years). The FAIR data provision of NFDI4Earth will support such institutions and other stakeholders in referencing and using such data and thereby increase ESS outreach.

We will also engage with software companies and solution providers and, for instance, elaborate on cooperation in the development and adaptation of international interoperability standards and on common open source software developments.

3.4 Organizational structure and viability

The ESS community is large and diverse, with a broad range of sub-disciplines, multiple RDM-related infrastructures as well as with varying awareness and different levels of competence relating to FAIR RDM. This high diversity requires coordinated management as well as a collaborative governance structure from an early stage. This structure must be able to initially identify, subsequently respond and continuously adjust to the needs of the ESS community through

an open and flexible process leading to a sustainable structure within the NFDI. Here, we capitalize from the intensive initialisation process of NFDI4Earth (see chapter 3.1), were we have already provided several community building and communication elements, such as, for instance, NFDI4Earth plenaries, workshops, dedicated conference talks and the initial NFDI4Earth website (www.nfdi4earth.de).

NFDI4Earth Structure, Governance and Decision Making

The NFDI4Earth consortium is not yet an independent legal entity. NFDI4Earth intends to become part of the *NFDI Verein* as the umbrella organisation for NFDI4Earth and the other NFDI consortia, as discussed at the DFG Governance Workshop 2019 and at the NFDI Conference 2020. NFDI4Earth consortium members will join the NFDI Verein once it is officially constituted and form the NFDI4Earth section.

Figure 7 illustrates the agreed general model of the NFDI4Earth organizational structure.

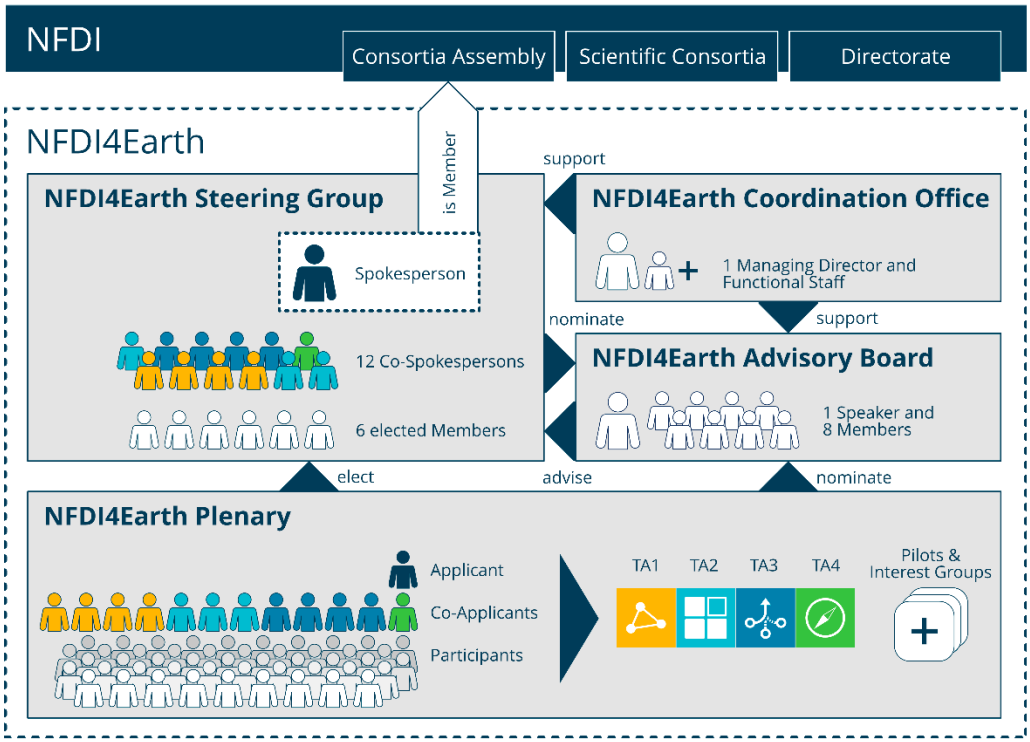

Figure 7: Overview of organizational Structure of NFDI4Earth and interfaces to NFDI

The **NFDI4Earth Spokesperson** represents the NFDI4Earth consortium externally, chairs and convenes the NFDI4Earth Plenary consortium assembly and the Steering Group, acts as the speaker for the NFDI Consortia Assembly and reports on NFDI4Earth to the **NFDI Organs**. The **NFDI4Earth Co-Spokespersons** are responsible for their task areas.

The **NFDI4Earth Steering Group** is formed by the NFDI4Earth (Co-)Spokespersons and six elected members from the NFDI4Earth Plenary. Supported by the NFDI4Earth Coordination Office, the Steering Group monitors and drives the overall progress and operation of NFDI4Earth. The Steering Group proposes and confirms (re-)allocations in the funding scheme, allocation of

the centrally-managed funds (see below on financial management) and supervises the activity of the task areas. If risks of failure or substantial deviations from the overall work plan are identified, the Steering Group develops adequate mitigation measures (e.g. reallocations) and decides on their implementation. The group will meet regularly, typically monthly (using videoconferences), it prepares the NFDI4Earth Plenary meetings and communicates closely with the NFDI4Earth Advisory Board. Decisions are taken by majority voting; proxy voting is possible. In case any decision violates DFG (funding) rules, the Spokesperson holds veto rights and will consult with the DFG afterwards.

The **NFDI4Earth Plenary** encompasses all representatives of participating institutions (including co-applicant institutions) of the NFDI4Earth consortium. The Plenary will meet at least once a year in a consortium assembly and discuss project status and progress as well as the embedding of NFDI4Earth into the NFDI. The Plenary proposes and admits new NFDI4Earth members (host institutions of participants) and decides on membership terminations. It nominates and elects from its membership the six participants (as representatives of their host institution) to join the Steering Group (election for annual terms, re-election is possible), representing the different Interest Groups and activities, respectively. In particular, **we will specifically invite young researchers to stand for election to the Steering Group**.

The Plenary will be regularly informed about recent developments and news in NFDI4Earth using different communications channels. Decisions in the Plenary are taken by majority voting, where each member institution (including applicant and co-applicant institutions) has one vote. Decisions can be taken by online voting. In the event of a tie, the Spokesperson casts the deciding vote.

Plenary members can propose **NFDI4Earth Interest Groups** to elaborate new issues and specific or cross-cutting topics in the development and operation of NFDI4Earth. The NFDI4Earth Steering Group receives the proposal for interest groups and the NFDI4Earth Plenary is asked to decide upon adoption of an NFDI4Earth Interest Group (IG) based on feasibility and relevance. The NFDI4Earth coordination office will support interest groups by providing communication facilities (virtual meetings, web space etc.) to facilitate their daily work und to ensure proper communication (see M4.1 in TA4). The activities of each IG will be reviewed on a yearly basis. The Plenary may decide to close an IG in case of inactivity or activities in contradiction with NFDI4Earth's overall goals and vision.

Every participant can respond to calls for **ESS Pilots**, **NFDI4Earth Incubators, Educational Pilots** and **Academy sites** and apply for funding. The calls will be approved by the Steering Group and subsequently by the Plenary (see TA1 for details on the preparation of calls). Thus, we jointly ensure that the calls meet the actual needs of the ESS community. Plenary members will be appointed to a review board and decide on granting the proposals. The review process will be open and follow the example of EGU journals, where non-reviewers may also comment. The

review process will be handled by M1.1, M1.2, M1.3, M1.4, respectively, based on a transparent set of criteria as described in TA1. The review results, recommendations and a ranked list of proposals will be presented to the Plenary. The Plenary will then be asked to approve the funding recommendation. If the recommendation is not approved by the Plenary, the Steering Group will work on conflict resolution.

The **NFDI4Earth Advisory Board** ensures external strategic advice to NFDI4Earth. It consists of the speaker and eight members, all external to NFDI4Earth and covering the breadth of ESS, representing different national and international perspectives and stakeholders (junior and senior researchers, universities, infrastructure providers, industry, state and public interests). The board reviews the status of NFDI4Earth and provides recommendations for the development and strategic operation to the Steering Group. Advisory Board candidates will be nominated by the NFDI4Earth Plenary members in response to an open call. The NFDI4Earth Steering Group will appoint four members of the Advisory Board. The NFDI4Earth Plenary will appoint the five additional members. The Advisory Board members meet annually and are invited guests at NFDI4Earth Plenary meetings.

The **NFDI4Earth Coordination Office** at TUDD (TA4, M4.1) supports and manages the handling of funds within the NFDI4Earth consortium and controls (financial) reporting. The office supports the Spokesperson, Co-spokespersons and the Steering Group in the coordination of NFDI4Earth. The office (1) coordinates and manages NFDI4Earth events and meetings and NFDI4Earth communications; (2) supports the NFDI4Earth Advisory Board, and (3) will provide information channels to inform the NFDI4Earth consortium and the wider public on the status and progress of NFDI4Earth. A Managing Director heads the NFDI4Earth Coordination Office.

We will complement this general model with detailed bylaws for the NFDI4Earth consortium agreement and the governance in TA4 (M4.1), covering the organizational structure and the related regulations and safeguarding their compliance with DFG regulations. We will review the organisational model at regular intervals and adapt to changing requirements, in particular with the view on the long-term establishment and operation of NFDI4Earth and embedding in the NFDI.

Risks and Contingency Approach on Resolving Conflicts

Related to the ambition and long-term success of NFDI4Earth, we see the following risks and related mitigation approaches:

1. **Community buy-in cannot be sustained.** NFDI consortia – by their very nature – rely on cultural change. If this cultural change is too slow, the NFDI idea will not succeed in creating a network of networks. Here, all NFDIs need to make sure that they appeal to their communities and create an environment for a vibrant exchange of ideas. NFDI4Earth in particular supports such an environment via interest groups, pilots, incubators, education and

training and the academy, as well as its continuous networking with ESS-relevant academic societies (TA1 and TA4).

2. **Staff cannot be found quickly enough.** To keep the community engaged, a fast ramp up of the NFDI4Earth project will be crucial. Here, in particular some highly qualified and motivated staff will have to be found to initiate the formation of structural NFDI4Earth elements (e.g. the User Support Network in TA2). Partly, such risk can be mitigated by (co-)applicant(s) and participants' own contributions. We will participate in joint national and international NFDI recruiting activities. We will also capitalize from the education and academy measures (TA2) and their close links to the potential of next generation of staff. Moreover, it is expected that the emerging NFDI network will lead to a common pool of shared competences and will foster mutual support and exchange between the different consortia. In the long run, we expect that students and young researchers will increasingly see the benefit of, and will engage in, advancing their education in Earth System data science. This will enhance the pool of qualified staff in the future.
3. **Because of its reliance on own contributions, structural change at a key-players institution could be challenging.** Here, we will make sure that competences and responsibilities are broadly shared, such that single points of failure cannot occur as a consequence of (presumably otherwise necessitated) structural change (TA4, M4.1 will develop the long-term operation and governance model accordingly).
4. **Share of flexible funds is rather high and risk that it does not achieve its goals.** The flexible funds scheme aims at attracting new partners with new topics that enrich NFDI4Earth. The NFDI4Earth build-up phase showed great interest in contributing actively to NFDI4Earth. It can be expected that response to the calls for participation (mainly by Pilots and Incubators) will be very positive, as it was for the first call (see above).
5. **Technical limitations might slow down progress.** We have made sure to involve a number of (co-)applicant(s) and participants with substantial technical resources that are also undergoing periodic renewals of their resources. Here, we see an opportunity (in the long run) to influence future decisions on technical resources. TA3 for instance will provide self-evaluation mechanisms to support these decisions.
6. **Alignment with national and international developments are critical.** NFDI4Earth brings the unique diversity of the German ESS community together in an unprecedented way. There are already many existing commitments in national and international communities and we will integrate them in NFDI4Earth and NFDI (all TAs).

In case of any undesirable developments relating to the NFDI4Earth work programme, or problems in achieving the specific objectives of the measures of the work programme, as task area leaders, the co-spokespersons shall inform the Spokesperson. With support from the NFDI4Earth coordination office, the Spokesperson develops and proposes mitigation strategies,

which – once endorsed by the NFDI4Earth Steering Group – are implemented and may lead to changes in the work programme. To this end, all NFDI4Earth co-spokespersons are willing to follow an agile management process in developing NFDI4Earth and are all in responsible positions at their institutions, enabling them to adapt to required changes in the task profiles.

The calls for ESS Pilots and NFDI4Earth Incubators in particular allow for a responsive process: in launching specific pilot and incubator calls, NFDI4Earth can react if critical gaps are identified or new challenges arise.

In case of a conflict situation, the Spokesperson will review the conflicting positions and call the NFDI4Earth Steering Group to agree on a strategy – complying with DFG rules – to resolve the conflict.

Financial Management of NFDI4Earth Project Funds

According to the DFG rules, the NFDI4Earth Coordination Office at TUDD (i.e. the Applicant) will manage finances centrally. However, co-applicants and leading participants (working in specific measures and selected pilots, see TA descriptions and funding tables) will receive forwarding contracts from the beginning.

NFDI4Earth is designed as an open consortium and is anticipated to grow during the funding phase. In close cooperation with the DFG, we will seek to implement augmentation of the consortium by applying for cost-neutral amendments to the funding contract.

The consortium agreed to dedicate a share of the funding (flexible funds) to secure a constant influx of new partners with their ideas and contributions. These funds shall also allow to involve more/different/new competences and to appropriately react to arising development and adaptation needs. ESS Pilots and Incubators are the main instruments for the allocation of flexible funds in the context of NFDI4Earth. Consequently, the financial plan already allocates 10% of the funding request for future Pilots and 5% for future Incubators. The money is accounted for in the centrally managed funds - see justification of costs for TA1. The mechanisms to invite project ideas and to select them as pilots or incubators are described above and in the corresponding measures (see chapters 5.1.1 and 5.1.2).

Benefiting from the Helmholtz Research Field Earth and Environment DataHub

The Helmholtz Research Field Earth and Environment DataHub integrates data repositories, infrastructures and services in the joint research programme *Changing Earth – Sustaining our Future* across Helmholtz centres (AWI, GEOMAR, GFZ, HZH, JÜLICH, KIT and UFZ). A comprehensive, common RDM in an open and cooperative network enables the harmonisation of data flows and metadata profiles. This will enhance data (re-)usability in and outside the research field and will foster better exchange with diverse national and international initiatives. Three compartment-specific associated hubs (ATMO, MARE, TERRA) form the core structure of the DataHub. They are connected by cross-cutting activities and well-defined working groups

pursuing joint developments. The creation and sustainable operation of the DataHub as a FAIR infrastructure furthers data science in the Helmholtz Research Field Earth and Environment for new research and insights relevant to the general public. At the same time, the DataHub allows a systematic positioning and networking of RDM themes across research fields in Helmholtz, including the Helmholtz Metadata Collaboration (HMC) platform. The DataHub will support and contribute to the development of NFDI4Earth and acknowledges NFDI4Earth as the platform to reach out to the ESS community and their needs, to collaboratively foster innovations in ESS data science and to commonly develop a model for long-term operations (M3.3). In addition, relevant international (e.g. EOSC and RDA) RDM initiatives and organisations will be jointly addressed (TA3, M3.4). The DataHub resources will form a valuable Helmholtz contribution to NFDI4Earth, including co-staffing in TA2 and TA3.

Contributions and long-term commitments

NFDI4Earth will have access to many existing infrastructures through co-applicants and participants. Their home institutions own and operate data-related infrastructures of varying sizes and will dedicate a share of the infrastructure to operate NFDI4Earth. A detailed overview of the infrastructures operated by the NFDI4Earth co-applicants is included in chapter 7.1. See chapter 4.1 for details on examples of these infrastructures. Some of the infrastructures already have a long-term perspective from an institutional commitment in a well-developed professional network. Libraries play a particularly important role as traditional brokers for long-term archiving and the preservation of knowledge.

NFDI4Earth also comprises members such as the Federal Agency for Cartography and Geodesy (BKG), the German National Meteorological Service (DWD), and the Federal Institute for Geosciences and Natural Resources (BGR) as important data providers for public geodata of high relevance in ESS. Here, we established a mutual agreement between NFDI4Earth and the German Geodata Infrastructure (GDI-DE) for close future cooperation (see https://www.gdi-de.org/en/cooperations/science).

The applicant and co-applicants will provide substantial support in terms of in-kind staff contributions, contribution of facilities and of additional expertise and support services. A detailed description of the (Co-) Applicant contributions is available in chapter 6.4.

As regards the long-term operation of NFDI4Earth, being fully embedded in a well-functioning NFDI operation model is clearly the primary target (see TA4, M4.1, M4.2). It is expected that embedding of this sort will specifically address the way researchers are guided by funding instruments and policies to make full use of the offerings within NFDI and NFDI4Earth and that common financing models are implemented to guarantee long-term operation of the NFDIs. A second important pillar is becoming linked and embedded into European and international research data infrastructures (TA3). Setting up the NFDI4Earth EduHubs and the NFDI4Earth Academy (TA1) as a network of competence on ESS data science forms an important cornerstone

for the long-term establishment of NFDI4Earth at universities, universities of applied science and research institutions. The implementation and establishment of NFDI4Earth OneStop4All (TA2) as *the* single-entry point for researchers in ESS for support regarding their various needs in research data use and management is an important asset for bonding the community in the long run. A criteria framework (see TA3, M3.2) will be established to assess data infrastructure offerings concerning their fit to NFDI4Earth requirements. This will allow providers and funders of services to evaluate trade-offs, including the range of offerings versus their costs and their potential for long-term operation. Continuous monitoring and applicability testing of new technologies and approaches followed by adaptation of matured solutions (see TA1, M1.2 and TA2, M2.5) into the consolidated NFDI4Earth services ensures an on-going innovation process.

We will provide a business plan for long-term operation of the NFDI4Earth Coordination Office to secure the durability of communication structures and community support (see TA4, M4.1). TUDD and the Saxon State Ministry for Higher Education, Research and the Arts (SMWK) assured (long-term) support for parts of the NFDI4Earth Coordination Office staff.

3.5 Operating model

The NFDI4Earth consortium agrees on amending FAIR general principles with the principles of sharing free open data, free open code and on implementing a no-fee policy for the services provided as part of NFD4Earth (see also chapter 4.3). A common *NFDI4Earth FAIRness and Openness Commitment* will be developed and detail these general principles (see chapter 4.3). NFDI4Earth Co-applicants will provide a range of resources in accordance with and fully acknowledging these principles. All current and future NFDI4Earth members will have to acknowledge this commitment.

The development towards a comprehensive governance model and on agreements on how to ensure the operation of NFDI4Earth in the long term is a core subject of TA4 (M4.1, M4.2). Meanwhile, this is understood as a collaborative measure within the NFDI (Bierwirth et al.; Glöckner et al., 2019). NFDI4Earth will contribute to the discussion regarding the possibly necessary redesign and adaptation of the funding regime for universities, research and infrastructure institutions running infrastructure services also as part of NFDI4Earth and the NFDI, such that these NFDI offerings can be guaranteed over the long-term (TA4, M4.1).

4 Research Data Management Strategy

In a community-driven process, NFDI4Earth will link existing ESS RDM services into a federated infrastructure as part of the overall NFDI and related international efforts. We will establish a Common NFDI4Earth Architecture based on state-of-the-art Policies and Standards as a foundation for sustainable FAIR and Open ESS RDM offerings. OneStop4All and the User Support Network will serve researchers as an entry point to NFDI4Earth – making FAIR and Open

RDM a matter of fact – and as an interface to assistance for innovative ESS RDM and powerful data analytics.

4.1 State-of-the-art and needs analysis

Research in Earth System Science (ESS) has increasingly extensive digital components throughout the research cycle. A number of ESS applications can be seen at the forefront of digital innovations. However, supposedly simple data discovery and data integration from heterogeneous sources is still tedious and the work and analysis processes are hampered by several discontinuities and a lack of information regarding data quality, provenance, adequate tools and analysis environments. Collaborative, also interdisciplinary, scientific data analytics in ESS – as also envisioned by the NFDI4Earth consortium – has been pioneered in individual projects. **However, their wider uptake and sustainable implementation suffers from several obstacles, including different maturity levels of digitisation in different sub-disciplines, heterogeneous nature of vocabularies, and availability of sustainable and commonly agreed research data infrastructure concepts and core components.**

Standards and Software for interoperable ESS RDM services

NFDI4Earth must deal with a wide variety of data types due to the range of subjects and subdisciplines in ESS. Data stems from observations, simulations and analysis in different combinations, forms and formats (numerical, textual, and graphical). Many different open and standardized as well as proprietary data formats as well as service interfaces are used. Relevant standardization is provided for instance from the World Wide Web Consortium (W3C), the Open Geospatial Consortium (OGC), ISO/TC 211 as well as from legal frameworks, such as the EU INSPIRE directive. Just to name a few examples: XML data are widely used for domain-specific schemata (e.g. GeoSciML, WaterML, BorholeML etc.), data formats such as netCDF/HDF5 and GeoTIFF are very commonly used to exchange meteorological, hydrological and remote sensing data. Topographic data is often exchanged using the GML, GeoJSON or ESRI's Shape format.

Thematic ESS data sets differ enormously in size and complexity and often require specific database operations to allow for spatio-temporal queries. Thus, different repository types and software-stacks exist. Interoperability is typically achieved by offering web-based Application Programming Interfaces (APIs) following OGC Standards (WCS, WFS, WMS, WPS, etc., see e.g. Bernard et al., 2014). A number of well-maintained and regularly updated Open Source packages (e.g. CKAN, Geonetwork, GeoServer, PostgreSQL, THREDDS, 52°North SOS and WPS) are available and broadly used to implement interoperable (meta-)data repositories in ESS.

Today’s landscape of Research Data Infrastructures and Repositories

The NFDI4Earth consortium has screened the landscape for existing infrastructures (see chapter 7.1), services and collaboration tools supporting ESS and – in a community process – mapped these against a first set of user requirements. The survey resulted in a list of more than 100

platforms and tools, which only partially overlap with the repositories listed in re3data (re3data.org, 2019) as *the* registry for research data repositories (developed with participation of co-applicants GFZ and KIT). Many of the screened platforms are the result of international projects with German participation, underscoring the need for aligning NFDI4Earth with international efforts. The survey showed that well-developed infrastructures exist across nearly all subdisciplines in ESS. Examples range from very specific platforms e.g. geochemical exploration tools for meteorites to more general data hubs for mineralogy or sediment analytics, to generic platforms for in-situ, space, and model data in the terrestrial, marine, atmospheric, and planetary domains. Many of these offerings are hosted, coordinated or co-developed by one or more members of the NFDI4Earth consortium. Thus, we will only describe a selection of examples of these platforms operated by NFDI4Earth co-applicants in some detail to highlight some of these characteristics. These firstly demonstrate the current state-of-the-art and existing relevant NFDI4Earth expertise, and secondly underpin the NFDI4Earth work plan to offer a consolidated route for finding and accessing infrastructures (of course, including many more) via OneStop4All (see TA2, M2.1):

- PANGAEA (PANGAEA, 2019) is amongst the most prominent data publication platforms related to ESS, recommended by high-level scientific journals. It is a key driver for FAIR research data management and sharing. PANGAEA has a high level of maturity, national and international visibility and usage. It hosts almost 400,000 quality controlled and managed data sets that are citable and linkable to other publications through digital object identifiers (DOI). PANGAEA has established a detailed curation process to guarantee data and metadata quality and re-usability.
- GFZ Data Services (GFZ Data Services, 2019) is a domain repository for geosciences research data and software covering all geoscientific disciplines as well as being an allocating agent for the International Geo Sample Number (IGSN). GFZ Data Services focus on the one hand on cooperation with projects, global networks and services, and on the other on the long-tail of research data, small individual data sets that are highly variable and rarely held in larger databases. Examples for cooperation are GEOFON and the GFZ instrument pool GIPP, the geodetic services ICGEM, IGETS and ISG of the International Association of Geodesy (IAG), TERENO, the Word Stress Map, EnMAP and ICDP. Furthermore, data publication services for INTERMAGNET and EPOS are currently available. GFZ Data Services is also a partner repository for the PIK and the Specialised Information Service (FID GEO). Software publications are closely connected with software repositories. PANGAEA and GFZ Data Services both use the Open Researcher and Contributor ID (ORCID) to identify individuals and thus to crosslink published data and software with researchers.
- DKRZ is a well-established node in the international Earth System Grid Federation (DKRZ, 2019). Its hosting of Earth System model data across sub-domains forms the basis for

international climate impact assessments and multi-model analyses. Due to its integration with the long-term archive WDCC, it forms a basis for the IPCC Data Distribution Center (IPCC-DDC) and for services within COPERNICUS. The example of the publication of CMIP6 data, which will form the basis for the upcoming IPCC report, clearly shows how, for example, RDA recommendations for the use of PID profiles are implemented.

- GeoROC and MetBase are more specific examples being the two of the largest geochemical and cosmochemical repositories hosted in Germany, together holding more than 600,000 data sets of chemical and isotopic data from rock samples that cover the entire Solar System's history. Both were established independently about 20 years ago (GeoROC by the Max-Planck Institute for Chemistry; MetBase privately) and are of high relevance for their respective communities – thousands of publications have made use of these repositories. All data sets are made internally interoperable by curators. The repositories are part of international networks through cooperation with e.g., EarthChem, IEDA Data or the Meteoritical Society. MetBase was recently equipped with innovative and interactive web-based tools that allow direct visualization and data analysis. These tools will now also be implemented into GeoROC. Both repositories will soon be migrated to the University of Göttingen, which ensures long-term maintenance and archiving, initiated by a recently awarded, substantial DFG-LIS to the university. Both repositories are excellent use cases for how to curate long-tail data and migrate it to a new technical framework.

These and many more platforms contribute to NFDI4Earth not only with their offerings but also with their substantial experience in supporting scientists in FAIR research data publications and archiving within ESS. Further initiatives have developed near real-time data services, large-scale sensor data management, online access, and metadata management systems.

Towards supporting the full research data life cycle

The majority of existing services are dedicated to direct data upload, data search and download. However, ESS data usage increasingly requires data integration across domains and across data centres. For instance, satellite remote sensing data needs to be combined with numerical model and in-situ data stemming from mobile or static monitoring platforms with the data being stored in different repositories. In a strongly user-driven, stepwise and agile process, NFDI4Earth aims to consolidate these offerings, for simplifying FAIR data publications, to ease data integration across ESS and to extend these offers with common workspaces for collaborative data-driven research.

NFDI4Earth can capitalize on a series of initial consolidation efforts that share this vision, e.g. the DataHub started by the Helmholtz Research Field Earth and Environment (see chapters 3.4.; 4.2). NFDI4Earth will connect to international efforts such as the climate data store (ECMWF), the European Space Agency (ESA) Data Cubes, the European Open Science Cloud (EOSC) to develop novel avenues to make scalable data analytics available to a wide user community in the cloud. In this context, NFDI4Earth represents a number of German centres for scientific

computing and has close links and access to HPC resources – jointly contributing to the NFDI Research Cloud (see chapter 3.5) – for example:

- DKRZ as an HPC centre for climate research provides HPC resources as well as data management services covering the entire data life cycle. Continuous participation and cooperation in projects enables the DKRZ to further develop the catalogue of services in the interest of the scientists. Current projects include, for example, the development of services for data analysis using artificial intelligence methods or PID services for the clear and permanent identification of data objects at a level of granularity below quoted data collections. In addition, there is active participation in cooperation and projects in the field of e-Infrastructures (e.g. EOSC, IS-ENES, EUDAT). NFDI4Earth will benefit from knowledge and experience gained in climate research, interdisciplinary research and e-Infrastructures.
- TUDD-ZIH hosts the HPC and Storage Complex (HRSK-II) and its extension HPC Data Analytics. Its architecture is specifically tailored to data-intensive computing and is an infrastructural base to ScaDS.AI Dresden/Leipzig (ScaDS, 2019). TUDD-ZIH coordinates ScaDS.AI as one of the Germany-wide Big Data centres for data analytics and artificial intelligence funded by the Federal Ministry of Education and Research. ScaDS.AI offers access to its Big Data/AI and HPC services and support for running and optimizing ESS software for HPC and in big data analytics. Furthermore, ScaDS.AI offers links to education and user support relating to Big Data analytics.
- KIT has the Steinbruch Centre for Computing (SCC) as its central institution for data-intensive computing that also supports the analysis of large-scale data. The SCC has high national and international visibility and operates comprehensive hardware configurations for simulations (e.g. ForHLR to be succeeded by HoreKa) and data storage (e.g. the Large Scale Data Facility, LSDF). Users from ESS institutes are frequent users of these resources, including scientists of the Institute of Meteorology and Climate Research (IMK), and aspects regarding state-of-the-art modelling and data handling (including compression) are explored together.

Building on existing offerings, NFDI4Earth members will jointly develop the next generation of services and tools enabling Earth system scientists to interactively explore a wide range of data and to enable automated data and metadata flows from sensors and models to data analysis, as well as to data and software publications. In particular, the integration of data publications with software publications is becoming increasingly important. In NFDI4Earth we will bring together several competencies of the co-applicants and participants. To name only a few: MPIBGC has developed a novel data cube approach (Mahecha et al., 2020) that allow users to interactively explore large gridded data sets in the cloud and map *arbitrary functions* on this data in the Julia programming language. TUDD has long-standing experience in designing and supporting the development of Spatial Data Infrastructures (INSPIRE, GDI-DE, regional SDIs), interoperable geoprocessing services and geoportals, and in testing and advancing their usability (Müller et al.,

2013; Bernard et al., 2014; Henzen, 2018). HSBO, KIT, LUH, MPIBGC, SGN, TUDD, and UNIL have developed several novel approaches for data integration of spatio-temporal data from heterogeneous sources and within various ESS application domains (Cayoglu et al., 2018; Buttlar et al., 2014; Wiemann and Bernard, 2016; Schultz et al.; Lemmens et al., 2007; Bodesheim et al., 2018). AWI has developed a data management framework from observation to archive and analysis (O2A), linking sensor management, data flow, cloud-services for data processing and data science to data publication (Koppe et al., 2015 - 2015). Due to the integration of the national German Competence Center for Big Data and AI (ScaDS.AI Dresden/Leipzig), NFDI4Earth profits from direct links into modern data analytics (e.g. machine learning) and into scalable data architectures and services.

Policies for FAIR Data in ESS

The European Community of Environmental Research Infrastructures (ENVRI; ENVRIplus Project Office, 2019b) has set up ENVRI-FAIR as the project *ENVironmental Research Infrastructures building Fair services Accessible for society, Innovation and Research* (ENVRIplus Project Office, 2019a) as a means to connect ENVRI to the European Open Science Cloud (EOSC) and to link to the EOSC-hub. ENVRI-FAIR targets a common European framework for policies and developments of FAIR data services of environmental research infrastructures. Focussing primarily on research (data) infrastructure providers, the ENVRI-FAIR objectives do line up with a number of the consolidation objectives within NFDI4Earth and here in particular in TA3. NFDI4Earth capitalizes from the fact that several NFDI4Earth members are participating in ENVRI-FAIR. Meanwhile, NFDI4Earth complements ENVRI-FAIR with its focus on including users, their requirements and needs.

COPDESS is the Coalition for Publishing Data in the Earth and Space Sciences and networks Earth and space science publishers and data facilities. In 2018, COPDESS started the project *Enabling FAIR data in the earth, space and environmental sciences* (COPDESS, 2019b) and formulated a commitment that adheres signatories to the FAIR principles. Several NFDI4Earth Members (e.g. DKRZ, GFZ) have signed the commitment (COPDESS, 2019a).

The measure on international networking (see TA3, M3.4) will ensure bi-directional communication and exchange between NFDI4Earth, ENVRI, EOSC and COPDESS.

4.2 Metadata standards

Initiatives to standardize digital metadata for geodata as a basis for implementing digital catalogue and discovery systems for geodata date back to the late 1980s and early 1990s, when strategies towards sharing digital geodata were first initiated, strongly promoted by the US act on the National Spatial Data Infrastructure (FGDC, 2019). **Thus, today most of the available data publishing portals and repositories in the ESS domain use a metadata schema based on international standards.** Most prominent is ISO 19115 – Geographic information – Metadata

(ISO, 2014). For administrative data, the European INSPIRE directive (European Commission, 2019b) defines a profile of ISO 19115 and, recently, the DCAT-AP (DCAT-AP, 2019) has been defined, which is mainly used for administrative Open Data in Germany. There are also metadata standards in use, which are specific to an ESS subdomain, such as the International Geo Sample Number Metadata (IGSN Metadata; IGSN, 2019). However, there is also use of quite generic metadata standards, e.g. the Dublin Core Metadata Element Set, Version 1.1 (ISO, 2017) or the DataCite Metadata Schema (DataCite e.V., 2019). Beside data set related metadata, some repositories also provide metadata to describe web services for geodata download, visualisation or processing. Schema mappings and content transformations between metadata sets that follow the different standards are usually possible. However, data and service discovery often suffers from incomplete or out-dated metadata.

A commonality of the vast majority of data sets in the ESS domain, and possibly a unique feature compared to the metadata of other disciplines, is the ever-present, required description of spatio-temporal context. Unlike other metadata standards, ISO 19115 enables the description of the spatial data representation and the linkage to space-time reference systems used for putting the data into spatio-temporal context.

As previously discussed GFZ Data Services, for instance, provides metadata following ISO19115, DataCite and NASA GCMD DIF Standards. Equally, WDCC and PANGAEA provide metadata following ISO19115. The data and geoinformation services available in the German GDI-DE are described following the ISO 19115 INSPIRE profile. Several of the existing applications for metadata acquisition and discovery make use of common vocabularies and gazetteers. Prominent and widely used examples are the General Multilingual Environmental Thesaurus (GEMET; EIONET, 2019), the Global Change Master Directory (GCMD; NASA, 2019) Keywords and the global gazetteer service GeoNames (GeoNames, 2019).

More detailed metadata descriptions depend on the type and origin of data. Such provenance data are fundamental to evaluate data quality and to improve the reproducibility, interpretability and fitness-for-use evaluation of ESS data. The ISO 19115 metadata schema can for instance be utilized for more structured provenance descriptions, which can then also support data exploration through interactive visualization of data provenance and use (Bernard et al., 2014). Observational data requires a description of the sensors and sensor platforms (including calibration information, accuracy, units of measure, etc.) and possible refinement processes on the raw data. For simulated data, the simulation process with its input data sets, parameter settings and possibly scenario descriptions etc. needs to be described using common vocabularies. However, such provenance metadata is usually not acquired by repositories or only available as textual descriptions. **Common vocabularies and comprehensive use of persistent identifiers for data, contributors, institutions, instruments and sensors are key to establishing provenance information.**

NFDI4Earth strategy on Metadata Standards

Even though metadata standards are well developed in a large number of ESS sub-disciplines, a consolidation effort is required across sub-disciplines to enable and accelerate scientific discovery as well as to improve usability and support for data integration. **TA2 (M2.2) and TA3 (M3.2, M3.3) will address harmonizing metadata vocabularies and management procedures (e.g. conventions for use of persistent identifiers, use of common vocabularies, structured description of provenance, etc.)**. Here, NFDI4Earth will benefit from the recently started Helmholtz Metadata Hub Earth and Environment (HMC Hub E&E). HMC Hub E&E is part of the Helmholtz Metadata Collaboration (HMC) Platform (coordinated by GEOMAR), to promote FAIRness, provenance and openness of ESS research data. In an initial step and available as a starting point for NFDI4Earth by the end of 2020, the HMC Hub E&E will develop a structured survey of human and technical expertise in the field of ESS metadata.

We will also **address metadata contents and management aspects that should be harmonized through NFDI-wide cross-disciplinary metadata standards** (see M3.2, M3.3). Examples are representation of the results of quality assurance processes, the use of persistent identifiers for data, for research organisations and for researchers (as ORCID), structured descriptions of possible restrictions through privacy and licensing aspects, and the link to common vocabularies. To provide full transparency of research results, NFDI4Earth will also further encourage and support cross-referencing data publications with scholarly publications, related research data, software, and/or physical samples whenever appropriate (see TA4; M4.2).

4.3 Implementation of the FAIR principles and data quality assurance

Establishing a common NFDI4Earth FAIRness and Openness Commitment

As stated above (see chapter 3.5), the NFDI4Earth consortium agrees on amending the FAIR principles with the principles of sharing free open data and free open source code and on implementing a no-fee policy for the services offered as part of NFD4Earth. For both the Incubator Lab and for the ESS Pilots (TA1) we will only fund applications that adhere strictly to either these NFDI4Earth principles, or approaches with the goal to mature an application in compliance with the most current NFDI4Earth FAIRness and Openness principles.

A common *NFDI4Earth FAIRness and Openness Commitment* will be jointly developed and detail these principles (TA4; M4.2). Openness needs, for instance, clarification in terms of the protection of privacy and safeguarding intellectual property rights (currently often enforced by embargo periods on newly submitted data). Here we will also align the commitment with guidelines from NFDI cross-cutting developments (TA3; M3.3). Furthermore, such a commitment must build on current legislation. Here, it can capitalize from the German law on geodata access (Geodatenzugangsgesetz; BMJV, 2019), which implements the EU INSPIRE directive and enforces federal free and open geodata. In a similar manner, the planned law on geological data

(BMWI, 2019) enforces open data. Accordingly, the federal agencies BKG, DWD and BGR provide their geodata as free open data. The recently revised EU Directive on Public Sector Information, renamed the EU Directive on *Open Data and Re-Use of Public Sector Information* (Directive (EU) 2019/1024, 2019) explicitly addresses research data and is another pillar for definition of the *Commitment.* Moreover, the NFDI4Earth commitment needs to address adherence to quality criteria for data and software publication.

The NFDI4Earth commitment needs to carefully balance the necessity of quality-assured data and software versus the procedures and tools at hand, to conduct quality checks and implement curation with affordable efforts. Thus, the commitment will be developed in close contact with the community (via ESS Pilots and online consultations) and the measures on common architecture and standards providing an up-to-date trade-off between the Openness and FAIRness demands against the (a) solutions, which are at hand for ESS researchers to fulfil these demands and (b) the researchers' acceptance, willingness and legal constraints to fulfil these demands.

Supporting Interoperability – Common Standards, Self-Evaluation

The agreed framework of common standards and rules is core to interoperability within NFDI4Earth and NFDI (M3.2). It will also serve us a starting point to **develop and establish an *NFDI4Earth Label* as indicator for the interoperability level of a digital resource in NFDI4Earth**. This **label will be awarded via a self-evaluation process** and will be supported by automated testing procedures.

We will also develop approaches on the **combination of persistent identifiers, metadata and data into self-contained machine-readable and exchangeable digital objects** (TA3, M3.2): Taking the RDA FAIR Digital Object approach as a starting point and in close international cooperation, these developments showcase ways to advance cross-community research data exchange and serve as a blueprint for the NFDI Commons.

Data Curation and Data Quality

Curation of data comprises different aspects in the data life cycle. Support in workflow management facilitates proper documentation in metadata. Curation services and experts provide advice to make data fit for storage, resulting in better usability for collaboration, reuse and citation within and beyond the original community. A main objective in the curation practice – e.g. of GFZ Data Services – is to **link data sets via their metadata to articles, related data, related objects, software or information for further reading**. Data is published with a clear, open license and meets community standards. **Provenance is provided through data descriptions based on the W3C PROV data model including recommendations by the joint WG *RDA / TDWG Metadata Standards for Attribution*** (Thessen et al., 2019). Even though there are best practices – as exemplified here for the GFZ Data Services – research data management in ESS generally still lacks common established approaches towards automatic metadata acquisition, smart data quality assurance and curation approaches, as well as a continuous support of

metadata complementation throughout the data life cycle. As good-quality data and metadata still require time-consuming manual acquisition and curation, existing data is only rarely shipped with quality information and metadata is often deficit in terms of completeness and being up to date.

Data Curation currently depends on the expertise of a few specialized experts –PANGAEA, WDCC or GeoROC serve as examples here. A fully automated curation process cannot be expected in the near future. Hence, with an ever-increasing amount of data being published, there is fast growing demand for appropriate data curation. Thus, proper curation would become a bottleneck for data publications. Meanwhile, there is an ever-increasing number of generic research data support services at universities to propagate and establish concepts of RDM. However, professionals running these support services are often neither specifically aware of existing ESS RDM offerings nor – in particular – of the specific curation needs, e.g. for geodata. Therefore, (1) **impetus for data curation** (as addressed in TA2; M2.4, M2.2), (2) **training concepts for data curation such as the NFDI4Earth data curation certificat**e course (as addressed in TA1; M1.3) and (3) related **certifications for different data centres** (addressed in TA3; M3.2) will disseminate expertise and coordinate federated curation efforts within NFDI4Earth.

The OneStop4All and the User Support Network will provide **structured access to curation expertise and support for researchers** (see TA2). Here, the User Support Network implements a network of help desks for ESS RDM comparable to front office–back office approach for RDM (Dillo and Doorn, 2014). Additionally, NFDI4Earth will **explore automated curation procedures (pilots, incubators) and consolidate promising approaches** for tools (see TA2, M2.5). Here, we can build, for instance, on two BMBF-funded research projects: (1) GEOKUR (coordinated by TUDD; GeoKur, 2019) to support curation and quality assurance of environmental data during the research process and (2) AtMoDat (coordinated by DKRZ; AtMoDat, 2019) addressing data quality, curation criteria and DOI-branding for atmospheric data.

4.4 Services provided by the consortium

NFDI4Earth is a catchment basin for a manifold of existing services (see e.g. chapter 7.1 for services of participants). Based on those, new services within the NFDI framework will be established. Their design and implementation will follow a step-wise, iterative process. These services (and mechanisms related to them) must be able to accommodate of various levels RDM knowledge and, in addition, different digital capacities and digital literacy levels. Thus, NFDI4Earth does not follow a one-size-fits-all-model: some services primarily address ESS researchers that require simple solutions for their RDM needs, others mainly address researchers who act as data champions (Higman et al., 2017) and are highly skilled in ESS-related RDM and/or data analytics. The services that will be critical to NFDI4Earth, and which are adopted from existing services or will be newly developed, can be summarized as:

1. Provision of the OneStop4All and the User Support Network to **provide first and second level RDM support** as a community service and to provide **integrated access to the distributed ESS RDM and data offerings** (TA2).
2. **Services collecting and providing knowledge and structured information** about FAIR and Open ESS RDM to the community in particular the Knowledge Hub and the Living Handbooks (TA2, TA3, TA4).
3. **Services for education and training** on several aspects of FAIR and Open ESS RDM and ES data science (TA1).
4. **Provision of new tools** that pave the way to next-generation solutions for RMD and research data analytics for ESS (TA1, TA2).
5. **Provision of collaborative work and development spaces** as an initial foundation service (TA4) with an intent to motivate other service providers to open up their resources as well.
6. Provision of **services that support repositories to perform self-assessments**. This ensures compliance to an agreed set of qualification criteria and supports providers in improving their services (TA3).
7. Provision **of services to establish and support common agreements and regulations** for ESS researchers and infrastructure providers towards FAIR and Open ESS data and software management (TA3, TA4).
8. Provision of **services to fully link** and embed the NFDI4Earth developments of technical standards and agreements **into the national – in particular the NFDI – and international developments** (TA3).
9. Provision of the Coordination Office as a **communication and networking service** to support NFDI4Earth as an open forum **to commonly design, develop and establish FAIR and Open RDM in ESS** (TA4, TA1).

Agile Adaptation to User Needs and Changing Requirements

The NFDI4Earth development process requires continuous user consultation. The active and holistic involvement of the scientific user community is key – but still far from current practice in ESS (see chapter 2). This is partly due to the diversity of disciplines, organized in different scientific communities.

The ESS Pilots, Incubators and the Interest Groups are pivotal means for user-driven development and community building (TA1, TA4). The EduHubs and the Academy act bi-directionally as multipliers and continuously collect user needs and feed them into NFDI4Earth. TA1 in particular focuses on collecting and managing user requirements from the community of ESS researchers. User needs and interests can, for instance, be revealed by analysing the responses to calls and the applications for ESS Pilots, Incubators and Interest Groups (TA1). Technological progress will be regularly assessed by TA1 in a trend scouting activity of the Incubator Lab. **TA2 directly interacts with users through OneStop4All**

and the User Support Network, and will monitor relevant data from interactions with users. TA3 links infrastructure providers among themselves and as a group to the community, also in the context of (inter)national alignment.

TA4 will regularly compile all monitoring results in reports and identify and address gaps/risks in the NFDI4Earth implementation. In addition, this reporting serves as the basis for potential adaptations of the work programme or as decision support for the launch of dedicated calls for pilots or incubators addressing specific aspects related to ESS RDM (see chapter 3.4).

Longevity of NFDI4Earth

Considering long-term availability and operation of the RDM services, **NFDI4Earth builds on the commitments and distributed existing infrastructures of the co-applicants and participants** (see chapter 3.4). We will develop a **Common Architecture** (TA3, M3.1) and a **sustainable NFDI4Earth Governance and Operation Model** (TA4, M4.1) as the blueprint for a long-term operation of the NFDI4Earth components and the related services embedded in the NFDI. This closely links to the planned development of a self-evaluation framework for ESS repositories (also considering long-term storage and archiving) against a set of qualification indicators for ESS FAIRness (TA3, M3.2). Besides the consideration of interoperability, standards such a criteria framework will also address adherence to currently available generic certification frameworks for repositories (e.g. Core Trust Seal, Data Seal of Approval, World Data System, ISO 16363, Nestor Seal), not least in terms of enhancing sustainability of the solutions. This effort, involving infrastructure providers, will be a key element of our sustainability strategy for the NFDI4Earth endeavour (TA4, M4.1).

Newly developed NFDI4Earth components will **build on existing Open Source Software** implementations, which are well established, documented and maintained, e.g. guaranteed by their long-term applicability. Publishing and (re-)using open source software is seen as a matter of course in ESS and also as part of the required cultural change. Further encouraging the ESS community to establish collaborative quality assurance and long-term maintenance of open software is a key pillar to sustainable, scientific software development and core to reproducibility in ESS data science (Downs et al., 2015; Schmidt and Wytzisk, 2019). **NFDI4Earth will provide guidance and best practices for Open Source usage and publication for ESS RDM and data analysis tools** (M2.5, M3.3 and M4.3). Accordingly, software extensions and enhancements, which result from the ESS Pilots and measures, will also be made available as Open Source.

5 Work Programme

NFDI4Earth2Participate M1.1: Earth System Science Pilots * M1.2: Incubator Lab M1.3: Education and Training Materials and Services * M1.4: NFDI4Earth Academy *	**Coordination** H. Gödde M. Mahecha M. Sester C. Keßler		Task Area 1
NFDI4Earth2Facilitate M2.1: OneStop4All M2.2: User Support * M2.3: Governmental Data * M2.4: Data in Long-Term Storage * M2.5: Advancing Tools	**Coordination** P. Braesicke M. Reichstein H. Thiemann		Task Area 2
NFDI4Earth2Interoperate M3.1: Synthesis of a Sustainable NFDI4Earth Architecture * M3.2: Common Standards for FAIR ESS Data * M3.3: NFDI Commons * M3.4: International Networking & Embedding *	**Coordination** F. Tilmann S. Frickenhaus H. Marschall C. Weiland		Task Area 3
NFDI4Earth2Coordinate M4.1: Coordination, Collaborative and Sustainable Governance of NFDI4Earth * M4.2: Towards a Cultural Change in ESS Research Data Management * M4.3: Central Support Services for the federated NFDI4Earth *	**Coordination** L. Bernard W. Nagel		Task Area 4

Figure 8: NFDI4Earth Task Areas, their leads and measures.

Figure 8 provides an overview of the NFDI4Earth structure and task areas. All measures marked with an asterisk will contribute to the various aspects of developing the NFDI and/or crosslinking with other NFDI consortia. The measure descriptions below provide further details on these aspects and crosslinks. Cost justifications are provided with the overview about the required funding for each task area.

Each measure comes with a list of milestones and deliverables. Figure 9 provides an overview of major milestones in the NFDI4Earth work programme.

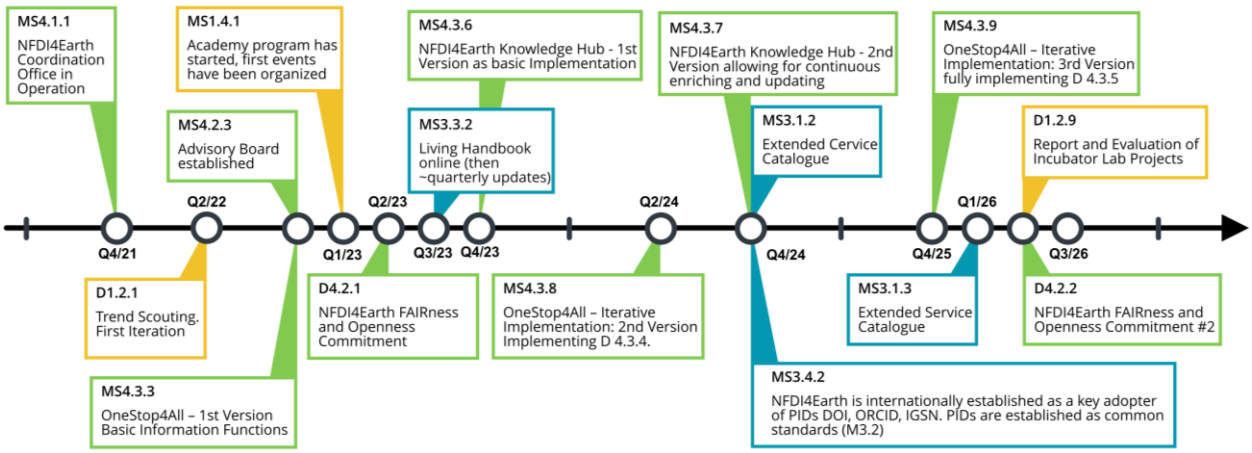

Figure 9: Major Milestones.

The task leads, jointly with the coordination office, will arrange for synchronisation between the measures and for monitoring progress in the measures. This will be done in a moderated dialogue process in the Steering Group (see chapter 3.4 and TA4). This process includes a continuous synchronisation with the community requirements (see chapter 4.4) and by the evaluation of agreed general measures of success (see chapter 2.2).

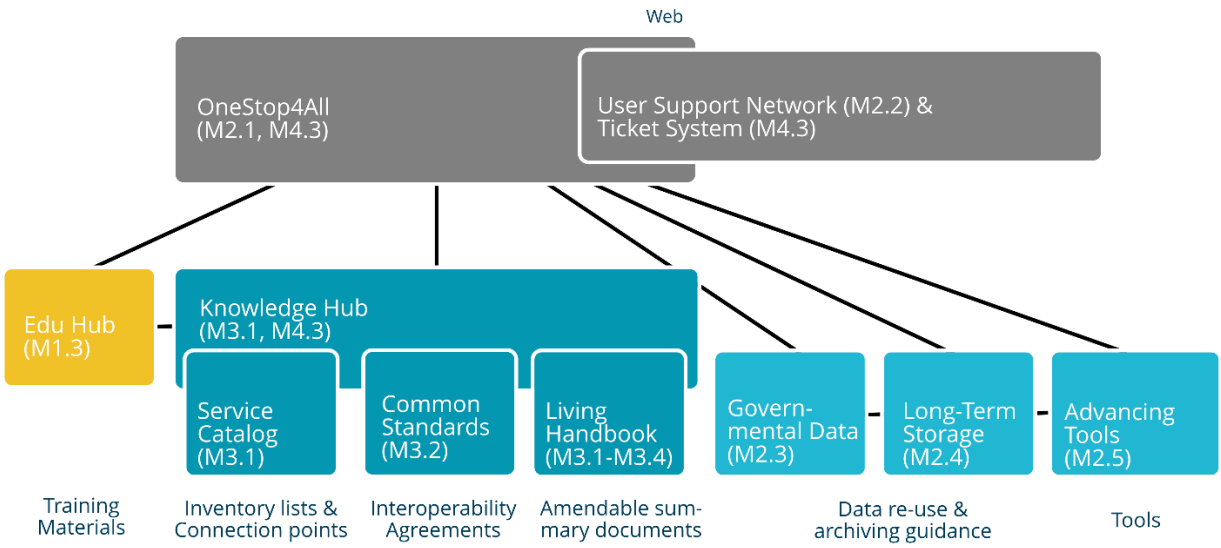

Figure 10: Overview on the linkages between the NFDI4Earth Components and the related measures.

Figure 10 outlines the measures' cooperation on core RDM components in the NFDI4Earth work programme. Both the OneStop4All and the User Support Network will be created in different stages within TA2 and TA4. In parallel, the Knowledge Hub will be designed (TA3, TA4), iteratively implemented (TA4) and filled (TA1, TA2, TA3), and serves as the fundamental information basis for subsequent versions of the OneStop4All and the User Support Network. The Living Handbook will be compiled based on the Knowledge Hub and designed as a human readable, easy-to-digest information product. Next, distributed services will be mapped via the single access point and subsequently (and together with the service providers) services will be made accessible through

the single access point. At the same time, other (distributed) partners will enlarge the team for the single access point. Their interaction will be facilitated by collaboration software and a ticketing system (provided by TA4) that can be fed by ESS researchers, looking for data and services.

5.1 Task Area 1: 2Participate

Lead: UNIL, LUH, Geo.X, HSBO

Overview

TA1 2Participate engages with the community as broadly as possible, from students to professionals and service providers to top scientists through four interrelated measures: Pilot studies, an Incubator Lab, educational materials and resources, and establishing an NFDI4Earth Academy. These measures guarantee that the development of NFDI4Earth is entirely driven by the community's needs and requirements and takes into consideration that the uptake of FAIR principles happens at different speeds. **TA1 aims to equally attract those practitioners and researchers who may, for the first time, work on making their data interoperable and accessible, as well as those who aim to couple the latest data analytics tools to a wide array of heterogeneous data sources and explore novel avenues of science. In this spirit, we attract all sub-branches in the Earth system sciences** (Figure 11). TA1 also aims to leverage novel approaches, methods, and technical solutions for ingesting research data into the scientific workflow. TA1 identifies and tests essential building blocks for a sustainable research data infrastructure that is prepared to integrate the full workflow from data collection all the way down to data exploration across domains so that reproducible analyses become possible and natural. Active scouting for novel developments is likewise an essential element via the Incubator Lab. It is therefore no coincidence that TA1 is also responsible for education: on the one hand, **the next generation of Earth system researchers and practitioners should be prepared to deal with the NFDI4Earth infrastructures**; on the other hand, this TA supports integration of the latest developments in their research activities. Hence, we need to enable the learning fundamental data management and exploitation concepts and enable future researchers to couple any relevant set of data (streams) to high-end analytic tools. With this big picture in mind, **TA1 invites the community to explore research paths that were not anticipated by the time data were created**. This requires that new participants are flexibly able to be integrated into the NFDI4Earth team via the open calls along the entire life of the project. This element is essential for both challenging, but also attracting Earth scientists and practitioners.

The measures use two different approaches to engage with the community: firstly, open calls will invite the community to integrate existing and test novel ideas relevant to NFD4Earth. Secondly, training opportunities and materials as well as scientific exchange platforms foster the development of an interface for Earth system research and data science. Measures 1.1 and 1.2 follow the first approach by inviting the community to propose ideas and test their potential for the

NFDI4Earth. Every researcher in Germany can bring in new ideas through open calls which, in case of success, may be integrated into the NFDI. While ESS Pilots (M1.1) leverage existing technologies and are used for requirements engineering in subsequent different task areas, the Incubator Lab (M1.2) invites proposals on novel cutting-edge tools and the latest innovations. The most promising new tools and methods shall fuel the cycle of discovery and explore their suitability for specific geoscientific applications. Incubator projects are short and focused on individual building blocks (duration 3-6 months), whereas pilots have a broader scope and are thus typically scheduled to be completed within one year. Active trend scouting in M1.2 supports this activity. Measures 1.3 and 1.4, instead, are collective actions designed to deliver NFDI4Earth materials for education (EduHub, M1.3) and offer and connect high-level expertise (Academy, M1.4) to the community. Curriculum and course materials targeting different stakeholder groups are developed, updated, synchronized, and distributed. By steering the development of spatio-temporal data literacy and science in the Earth system sciences, the EduHubs enable the training of the next generation of ESS scholars. The NFDI4Earth Academy enhances these efforts by establishing a think tank where young researchers interact on data science and RDM topics in the Earth system. Together, M1.3 and M1.4 enable scientists at any level to exploit sustainable data resources, conduct reproducible research, and anticipate next-generation interfaces of Earth system data sciences. Given that TA1 heavily relies on integrating additional participants via open calls, it is essential that their structure be fully transparent: this entails calls being published on a regular basis, communicated widely, and evaluation be participative and transparent.

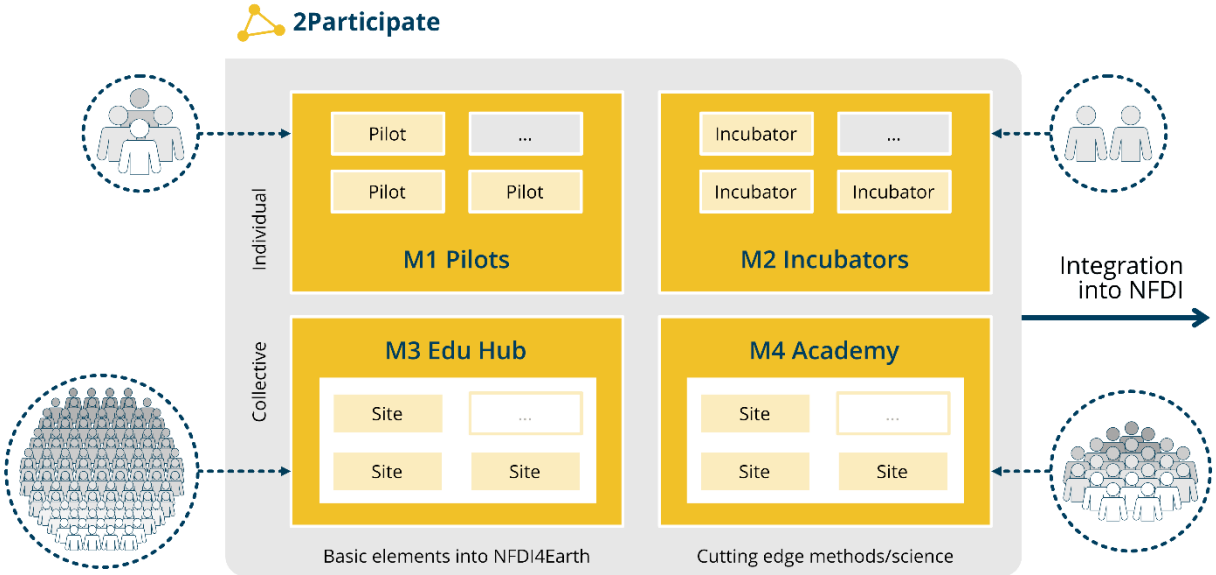

Figure 11: TA2 2Particpate – Overview.

The interaction among measures happens in multiple ways: Successful pilots can be extended via innovative ideas to incubators or, vice versa, incubators can become pilots in a second step. The aim of M1.1 and M1.2 is that their project-based approaches should become integral

elements of NFDI4Earth as a whole and - after some time of maturation - become part of the task areas TA2 and TA3. In the long-term, TA1 aims to pave the way towards making multiple approaches to research data management in the ESS fit for interdisciplinary exploitation. This TA follows the fundamental principle of *Data S*cience: One must think of the scientific benefit of RDM from the very beginning, such that data curation and its transformation into knowledge become a single integrated concept.

5.1.1 Measure 1.1 (M1.1): Earth System Science Pilots

Contributors: UNIL (lead), all leads of the selected ESS Pilots, additional via Open Calls

Links to all other measures

Objectives

Earth System Science Pilots are a wide array of small projects selected in an open review approach. Pilots fuel the cycle of discovery by leveraging specific methods and transforming them into essential ingredients in NFDI4Earth widely used in the ESS. Each pilot is continuously accompanied by us to guide the development of generic solutions and develop a technical roadmap. Pilots and their accompaniment in particular are used for requirements engineering in the different task areas.

Actions

Action 1: Open Pilot Calls for the community and Selection

Throughout the five-year funding phase of NFDI4Earth, **four calls for pilots will be launched and the respective community process will be coordinated**. Note that we already launched the first call prior to the start of the project (June 2020) to be able to include an initial set of pilots from the very beginning (see selected pilots below and all 38 submissions on www.nfdi4earth.de). We will prepare future pilot calls in close cooperation with the NFDI4Earth Steering Group and discuss drafts with the NFDI4Earth Plenary (see chapter 3.4). We will thus ensure that the calls match the actual needs of the community and are on the NFDI4Earth development path. **Pilot proposals are evaluated in an open review process**. Criteria for the evaluation of pilots will be in equal parts (scored on a 5 points scale for each criterion):

- Quality and Innovation (relevance beyond state-of-the-art)
- Feasibility (leveraging existing technologies)
- Potential for engaging a broad user community (also potential for linking scientific users, information infrastructures, or citizen science)
- Potential for integration into NFDI4Earth (relevance to NFDI4Earth as whole e.g. in terms of contributing to measures such as education, interoperability, archiving, and clearly addressing relevant aspects of the research data life cycle)

The selected set of **pilots must represent a broad range of disciplines within the Earth System Sciences**. Hence, pilots may be selected that score a bit lower than others on the criteria mentioned above in order to cover a wider range of disciplines and interrelations.

Action 2: Pilot steering

Pilots will be accompanied by the measure lead via weekly online meetings to ensure that all developments enable early integration within the NFDI4Earth TAs (e.g. OneStop4All, TA2). Additionally, such close communication aims to detect critical issues early on and allows us to test developments in the development phase. The objective here is to prevent solutions that are not relevant beyond the proposing community. M1.1 follows the idea that **pilots shall team up and form clusters of pilots that in the long-term become essential building blocks of NFDI4Earth in their own right**. For instance, the first round of pilots happened to contain several instances of high-dimensional gridded data processing approaches relevant to TA2. Here, M1.1 will steer a process that brings these pilots operationally together and works towards making them an integral element of TA2 as early as their state of maturity permits.

Action 3: Pilot output assessment and roadmap

Each pilot must provide two deliverables: a concrete implementation and a roadmap for future developments. M1.1. not only oversees the provision of the deliverables, but also synthesizes their outcomes in interaction with the NFDI4Earth community. Specifically, we will **elaborate an integrative roadmap for the practical needs and emerging vision that shall trigger relevant developments in NFDI4Earth** and will be included in subsequent calls (see Action 1).

Deliverables and Milestones

Mile-stone	Delive-rable	Type*	Description	Due end of
MS1.1.1	-	S	ESS Pilot approach discussed with community and pre proposal call successful (34 contributions; 14 selected)	Q3 2020
-	D1.1.1 D1.1.2 D1.1.3 D1.1.4	R	NFDI4Earth annual reports on progress of pilots - including their respective deliverables. Contains detailed documentation on pilot progress, independent testing of pilots by M2.1, weekly meetings, in-person meetings etc.	Q4 2022 Q4 2023 Q4 2024 Q4 2025
-	D1.1.5 D1.1.6 D1.1.7 D1.1.8	R	Living document synthesizing the roadmap from pilot experiences and community feedback	Q4 2022 Q4 2023 Q4 2024 Q4 2025
MS1.1.2 MS1.1.3 MS1.1.4 MS1.1.5	-	S	Pilots / clusters of pilots successfully integrated in NFDI4Earth TA2	Q4 2022 Q4 2023 Q4 2024 Q4 2025

* R-Report, S-Service

Successful Pilots – first round (call June, 2020)

In the following, we outline the pilots evaluated positively by the 2020 pre-project call. Selection criteria and review approach were presented, discussed, and adjusted in two teleconferences with a large attendance (>30 participants each). The call, criteria and all submissions are published on www.nfdi4earth.de).

ID	Pilot Description
P1	**Research Domain: Geophysics and Geodesy**, *Lead JUNIHB*
	Bathy4All: Workflows for Multibeam Processing and Visualization Vikram Unnithan[1], Peter Baumann[1], Carlos dos Santos Ferreira[2] 1 Jacobs University Bremen, D-28759 Bremen 2 University Bremen, MARUM - Center for Marine Environmental Sciences, D-28359 Bremen Easily accessible, high resolution bathymetry data is required for all marine sciences. Workflows for efficient processing, analysis and visualisation of multibeam data using open source tools and data cubes, which can be easily combined with satellite or climate data, are scarce. Bathy4all aims to design and implement workflows to process raw multibeam data present in data archives using the open source MB-System and rasdman to create bathymetry data cubes based on open standards and FAIR access. The combination of the best-of-breed pre-existing components (MBsystem, CUBE/CHRT, rasdaman) provide significant added value and progress beyond the state of the art. The deliverables include the integration of an advanced processing algorithm, a software demonstrator showcasing bathymetry data cubes, and a roadmap for inclusion of Water Column Data (WCD) and acoustic backscatter data. Stakeholders include not only marine sciences but also government agencies (e.g. BSH) and industry (e.g. maritime transport). The intelligent processing of bathymetry, building data cubes with standards-based APIs, inclusion of bathymetry in the EarthServer federation will promote the massive use of bathymetry by several earth sciences and remote sensing communities, in future including INSPIRE and thereby promoting NFDI4earth interests.
P2	**Research Domain: Atmospheric Science, Oceanography and Climate Research,** , *Lead DLR*
	Enhancing Earth system model evaluation with data cube enabled machine learning Veronika Eyring[1,2], Björn Brötz[1], Fernando Iglesias-Suarez[1], Axel Lauer[1], Miguel D. Mahechal[3], Markus Reichstein[4,5], Jakob Runge[6] [1] Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen [2] University of Bremen, Institute of Environmental Physics (IUP), Bremen [3] Remote Sensing Centre for Earth System Research, Universität Leipzig, Leipzig [4] Department of Biogeochemical Integration, Max Planck Institute for Biogeochemistry, Jena [5] Michael-Stifel-Center Jena for Data-driven and Simulation Science, Jena [6] Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Datenwissenschaften, Jena The application of machine learning methods to scientific questions in Earth system sciences is opening new ways to explore and better understand Earth system processes such as dynamic interdependencies between modes of climate variability or extreme events. One major technical challenge of machine learning applications in Earth system science is to efficiently handle the required large volume of input data. The objective of the pilot is to enable the use of cloud-ready data and machine learning methods (here we focus on causal discovery) for routine Earth system model evaluation. This will be done by enhancing the well-established Earth System Model Evaluation Tool (ESMValTool) with the novel Earth System Data Lab's (ESDL) data cube concept and infrastructure. This interface will also enable the efficient integration of machine learning algorithms that is currently unfeasible due to memory limitations. The result of the pilot will be the release of an enhanced ESMValTool version that includes the data cube concept and innovative machine learning diagnostics. The stakeholders of this pilot are (1) the Earth system science community including groups participating in the Coupled Model Intercomparison Project (CMIP) and ESMValTool development, (2) the climate informatics community, and (3) technology and infrastructure groups such as HPC centres and Earth science data providers. This pilot will

ID	Pilot Description
	contribute to improve efficient handling of Earth system data, which is of key importance for application of machine learning methods by the NFDI4Earth community.
P3	**Research Domain: Geochemistry, Mineralogy and Crystallography**, *Lead UNIF*
	Developing tools and FAIR principles for the GeoROC and MetBase databases Horst R. Marschall[1], Dominik C. Hezel[2], Gerhard Wörner[3], Matthias Willbold[3] [1] Goethe-Universität Frankfurt, Institut für Geowissenschaften, Altenhöferalle 1, 60438 Frankfurt [2] Universität zu Köln, Institut für Geologie & Mineralogie, Zülpicherstr. 49b, 50674 Köln [3] Universität Göttingen, Geowissenschaftliches Zentrum, Goldschmidtstr. 1, 37077 Göttingen GeoROC and MetBase are the two largest geochemical and cosmochemical databases hosted in Germany. Both databases are currently migrated to a new home, now with continued development and maintenance. A current DFG-LIS project aims to host and develop GeoROC at the University of Göttingen for re-building the technical infrastructure, data input, maintenance and interoperability. In this proposal, we aim to contribute to this endeavour by (i) adopting the recently built interactive visualisation and analysis tools for MetBase to GeoROC, and (ii) equipping the existing and future data with metadata according to international FAIR standards. We further want to (iii) build community awareness for the need for FAIR-guided curation of geochemical databases. Community workshops will initiate an NFDI4Earth geochemical special interest group to deeper embed the databases within the German geo-/cosmochemical community, and promote FAIR and user-friendly GeoROC and MetBase with improved accessibility, appropriate metadata, interoperability where data can be better visualised and analysed. GeoROC and MetBase have the potential of becoming a valued data source for geochemists, cosmochemists, petrologists, geologists and mineralogists for research and teaching as well as the ESS community and beyond.
P4	**Research Domain: Geochemistry and Paleontology**, *Lead GEOMAR*
	German marine seismic data access Christian Berndt[1], Andreas Lehmann[1], Christian Hübscher[2], Hanno Keil[3] [1] GEOMAR Helmholtz Zentrum für Ozeanforschung Kiel [2] Universität Hamburg [3] Universität Bremen Reflection seismic data are the paramount source of information for the subsurface structure as they provide the highest resolution of any geophysical technique. As such, they have been used for a large variety of academic and commercial purposes. For many decades, reflection seismic data were the largest data sets in Earth sciences, which created significant storage and archival problems and until today there is no international, European, or German data portal that is used throughout the community, hampering all new scientific projects that would like to use legacy data. Within this pilot project we will develop a unifying data infrastructure, kick off data archaeology for existing data, and prepare for future archival of reflection seismic data from future research cruises. The initiative will serve primarily Germany's marine geophysics community as represented by AGMAR of the Fachkollegium Physik der Erde of DFG. The pilot for reflection seismic data contributes to the efforts of NFDI4Earth to establish a distributed infrastructure for data curation by harmonized data workflows with connections to international data repositories. Interfaces and storage solutions for data volumes in the order of GBs to TBs are urgently needed in many more scientific fields.

ID	Pilot Description
P5	**Research Domain: Water Research**, *Lead IGB*
	Getting freshwater spatio-temporal data on track Sami Domisch[1], Giuseppe Amatulli[2], Vanessa Bremerich[1], Luc De Meester[1], Mark Gessner[1], Hans-Peter Grossart[1], Rita Adrian[1] [1] Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB), Müggelseedamm 301, 12587 Berlin, Germany [2] Yale University, Centre for Research Computing, New Haven, CT, 06511, USA Spatio-temporal freshwater-related earth system data are currently poorly organized and the full potential for research or management is rarely exploited, due to data disparity and a lack of interoperability given the different data standards and formats. It is especially the spatial structure of water bodies, i.e. the river network and lakes with time legacy effects that require a specialized workflow for earth system data integration into freshwater research. Engaging different freshwater-related research disciplines such as Hydrology, Chemistry, Geography, remote sensing, Climatology and Ecology effectively within a single framework poses an urgent prerequisite for a effective FAIR data management of environmental and biodiversity data, especially in light of climate and land use changes and feedback mechanisms between earth systems. Here, we propose a full start-to-finish pilot project that optimizes the integration of Earth system data into freshwater research and for Earth system science in general by accounting for the crucial, yet often neglected connectivity within freshwater water bodies themselves and their terrestrial catchments. The central tool will be the new GeoFRESH online platform that provides integration, processing, management and visualization of various standardized spatio-temporal, freshwater-related earth system data. The platform allows exploitation of the full potential of environmental, physical and biodiversity freshwater data along the hydrographical network, and will deliver a critical tool for stakeholders from research, sustainable water management and for monitoring freshwater ecosystem services alike. In addition, data interoperability will allow informing NFDI4Earth by enhancing the freshwater-specific data exchange among realms. The proposed pilot project represents a long-overdue component in freshwater-related FAIR data management, and once established, will prove itself as an essential key element for NFDI4Earth given its high potential for advancing seamless freshwater data integration across Earth system disciplines.
P6	**Research Domain: Geophysics and Geodesy**, *Lead LIAG*
	Interoperability and reusability of geoscientific lab data Matthias Halisch[1], Andreas Weller[2], Francois Holtz[3], Monika Sester[4] [1] Leibniz Institute for Applied Geophysics, Stilleweg 2, D-30655 Hannover [2] Institute of Geophysics, TU Clausthal, Arnold-Sommerfeld-Str. 1, D-38678 Clausthal-Zellerfeld [3] Institute of Mineralogy, Leibniz-Universität Hannover, Callinstr. 3-9, D-30167 Hannover [4] Institute of Cartographics and Geoinformatics, Leibniz-Universität Hannover, Appelstr. 9a, D30167 Hannover Geoscientific research has become a highly complex and interdisciplinary task that often needs several data sets to answer complex questions, and that produces huge amounts of manifold data, associated data types and related documentations accordingly. Although a broad range of possibilities exist to make these data findable and accessible (e.g., by assigning digital object identifier and by storing data in repositories), interoperability and reusability are mostly not guaranteed or even not possible, due to non-existing interdisciplinary standards and highly variable scales of research (i.e., from laboratory to fieldwork). Accordingly, the main objective of this pilot project is to develop an urgently needed interoperability and reusability framework, including inherent and meaningful standards, by exemplarily utilizing and assessing a broad

ID	Pilot Description
	range of available interdisciplinary geoscientific laboratory data and data types across different scales (i.e., petrophysical, geological, mineralogical and image-based data), with regard to hydraulic transport in porous rocks. Based on these different data sets and data types, this pilot will develop methods and assign standardized metadata, in order to ensure exemplarily completion of data integration for many independent research questions. A web-based platform will be set up to demonstrate, distribute and provide the developed standards and research data as intended for the NFDI in general, and for NFDI4Earth in particular. This project is primarily aimed at scientists, data curators, university teachers, and decision-makers; results are also relevant for infrastructure providers and system integrators for the development of new or the improvement of existing archives and repositories. Based on data sets created, compiled and exchanged within the FZ:GEO research network (Leibniz Forschungszentrum GEO at Leibniz Universität Hannover) and associated project partners, this pilot project will create the basis for interoperability and reusability of geophysical and geochemical data sets. For the future, such a web-based interface could be easily extended towards arbitrary fields of geoscience, including arbitrary scales of research and will become a powerful tool for providing information and data for new research concepts and fields.
P7	**Research Domain: Ecology, Biogeochemistry**, *Lead UFZ*
	Linking Environmental Data into European Scale RIs Jan Bumberger[1], Thomas Schnicke[1], Steffen Zacharias[1], Michael Mirtl[1] supported by the eLTER Information Management Team [1] Helmholtz Centre for Environmental Research - UFZ Leipzig, Germany In order to enable targeted measures to mitigate and adapt to global change, analysis and assessment of the effects of climate change and environmental pressures on ecosystem processes and biodiversity and access to long-term monitoring as well as experimental data are needed (Mirtl et al., 2018; Mollenhauer et al., 2018). While eLTER strives for harmonization of measurements at its sites, existing environmental and biological monitoring time series are not always easy to combine in large-scale assessments due to historic differences in measurement protocols and varying access mechanisms and policies (Hoffmann et al., 2014; Haase et al., 2018). The pilot aims to enable access to data from long-term monitoring by (a) identifying and aligning key data sources and variables across a range of different stakeholders and establishing common workflows for the documentation, harmonization and integration of biogeochemistry data supporting near real-time data provision; (b) establishing and adopting data documentation, data curation and data access procedures aligned with key stakeholders on European as well as global scale, ensuring data ownership and data licensing and leveraging newly observed as well as legacy data. Within the pilot, we will prototype the workflows to link environmental observation data from selected national data providers (e.g. TERENO, MOSES) using common services as provided by the EOSC and showcase issues on applying FAIR data management. The pilot will involve a broad range of community actors including the scientific community as well as data providers in order to ensure user-tailored implementation of data quality control and analysis. Providing FAIR environmental data from selected eLTER sites will an important asset within the national research data infrastructure, enabling the linkage of site-based observation data in the analysis by leveraging access to data from long-term sites. This contributes to the uptake and

ID	Pilot Description
	implementation of the European Open Science Cloud (EOSC) by research infrastructures as well as the scientific community.
P8	**Research Domain: Atmospheric Science, Oceanography and Climate Research**, *Lead TUDD*
	NFDI for seamless Earth system model-data integration Matthias Forkel[1], Kirsten Thonicke[2], Nuno Carvalhais[3] [1] Technische Universität Dresden, Junior professorship for Environmental Remote Sensing [2] Potsdam Institute for Climate Impact Research, Earth System Analysis [3] Max Planck Institute for Biogeochemistry, Department for Biogeochemical Integration Global Earth observation data is invaluable to evaluate, parameterize and further develop Earth system models and its land components, dynamic global vegetation models. However, such an integration of Earth observation data with Earth system models requires the expertise of multiple institutions, and in practical terms often involves converting file formats, copying large data sets between institutions and related computing systems, and manual and time-consuming application of multiple scripts and program code. The objective of the pilot is to identify requirements and to define a generic framework for a seamless integration of Earth observation data with Earth system models. Specifically, we will outline and develop a prototype for a seamless workflow to apply satellite observations for benchmarking and parameterisation of the LPJmL dynamic global vegetation model as part of the Potsdam Earth Model (POEM). The expected results will enable a comprehensive and continuous use of satellite observations for the development of LPJmL and POEM. Moreover, the developed framework and prototype will guide the development of seamless infrastructures to integrate Earth observations and models in the global biogeochemical, hydrological, ecological, and climate science communities and hence in NFDI4Earth as a whole.
P9	**Research Domain: Atmospheric Science and Climate Research**, *Lead DWD*
	Observations closer to Model Data (OcMOD) Frank Kaspar[1], Ivonne Anders[2], Christopher Kadow[2], Joaquim Pinto[3], Kevin Sieck[4] [1] Deutscher Wetterdienst, DWD, Offenbach [2] Deutsches Klimarechenzentrum, DKRZ, Hamburg [3] Karlsruher Institut für Technologie, KIT, Karlsruhe [4] Climate Service Center Germany, GERICS, Hamburg Working with climate model output nearly always includes a validation of this data compared to reference data from observations to ensure that the model data chosen is suitable for the individual research question or application. Model data and commonly used observational data have to be obtained from different sources: model data from e.g. the DKRZ and observational data from e.g. servers from public authorities. All data must be prepared to be in the same formats and standards before it can be used for further analysis. A broad range of users (e.g. climate modellers, impact modellers, climate scientists, providers for climate services and education) are dealing with same the effort and troubles with the same sets of data. The aim of the pilot is to bring observational data close to the model output, to easily access data from public authorities and increase the number of users of various disciplines, and to provide this data in standardized formats for easy usage. As stakeholders, we have chosen two big national climate change and climate impact projects, both need observations and climate model data for evaluation and climate change analysis. The pilot will integrate the DWD reanalysis data set

ID	Pilot Description
	COSMO-REA6 (Kaspar et al., 2020) into the infrastructure of ESGF (Earth System Grid Federation; https://esgf.llnl.gov/) as one example. After the end of the pilot, the developed workflow will work as a base for integrating further data sets from DWD or other public authorities.
P10	**Research Domain: Geography, Human-Environment-Interactions**
	On demand enhancement of Earth system data cubes with high-resolution socioeconomic data streams, *Lead UNIL* Guido Kraemer[1], Miguel D. Mahecha[1], Fabian Gans[2], Markus Reichstein[2] [1] Remote Sensing Centre for Earth System Research, Leipzig University, Talstr. 35, Leipzig, Sachsen 04103 [2] Max Planck Institute for Biogeochemistry, 07745 Jena Many subsystems of the Earth are constantly monitored in space and time with a large number of different data streams (e.g. gridded climate data, biophysical parameters of the land surface, or of aquatic bodies etc.). Interoperability among these data streams can be achieved via data cube approaches that allow efficient implementations of user-defined workflows. Today the human-environment nexus is affecting all aspects of Earth system functioning and should be considered in any environmental analysis. In fact ,there is also a rapidly growing wealth of socio-economic data sets available, enabling scientists to address human-environment interactions in a rapidly changing world. However, although these data sets are often freely available, they are not yet available as part of any given data cube and often come e.g. as annual shape files at some administrative units. It is therefore inconvenient to work jointly with gridded environmental data and socio-economic vector data. In this pilot, we explore novel opportunities to spatially disaggregate coarse socio-economic data with novel Machine Learning methods. The idea is that data integrals are maintained, while the spatial detail is realistically represented. Our aim is to implement algorithms that allow us to enhance any given Earth system data cube with socio-economic data streams, enabling the joint analyses of societal, biospheric and atmospheric data sets through a unified interface. Allowing the formation of joint data cubes containing biospheric, atmospheric, and social data is the aim. Integrated cubes of this kind will provide standardized interfaces for managing, accessing, and analysing these data streams jointly through Jupyter notebooks.
P11	**Research Domain: Landscape Ecology**
	PAMbase: a Repository of Soundscape Recordings to study Earth's Phonosphere, *Lead TUDD* Jan O. Engler[1], Anna F. Cord[1], Léa Courteille[1], Björn Andres[2] [1] Chair of Computational Landscape Ecology, Technische Universität Dresden, Helmholtzstr. 10, 01069 Dresden [2] Chair of Machine Learning and Computer Vision, Technische Universität Dresden, Nöthnitzer Str. 46, 01187 Dresden Recent breakthroughs in passive acoustic monitoring (PAM) now allow collecting acoustic environmental information at unprecedented spatio-temporal scales, quickly transforming it into a big and highly valuable data pool for Earth System Sciences. Applications include, for example, the cryosphere (monitoring the calving of glaciers), the hydrosphere (noise pollution in aquatic ecosystems), the biosphere (monitoring biodiversity) and the anthroposphere (human health). However, there is still no digital infrastructure that (i) can act as a central data repository for soundscape recordings, (ii) provides easy and rapid access, and (iii) allows standardized data analysis for a wider scientific community, so that the potential of acoustic data can be fully

ID	Pilot Description
	realized across disciplines. Building on existing open-source repositories, we aim to conceptualize such a repository for passive acoustic monitoring data (called PAMbase), which will allow curated data and metadata storage and quality control of acoustic data. We intend to address with this repository several user groups, including researchers, the public, planning authorities and decision-makers. The open-access beta release of PAMbase developed within the proposed pilot project will include a demonstration workflow for citizen science participation as well as a benchmark of user-friendly machine learning (in particular Deep Learning) tools for automated data analysis. In addition, the pilot will include a roadmap addressing user groups, opportunities for community engagement and a time plan for further technological and methodological development during the NFDI4Earth consolidation and advancing phase. PAMbase will complement the NFDI4Earth research areas and will enable macro-scale analyses of phonospheric information for multidisciplinary research and a wide range of applications.
P12	**Research Domain: Geology, Paleontology**
	Reusability of data with complex semantic structure, *Lead UNIHB* Michal Kucera[1], Robert Huber[1,2] [1] MARUM - Center for Marine Environmental Sciences, University of Bremen [2] PANGAEA - Data Publisher for Earth and Environmental Science Data on the occurrence and abundance of fossils provide invaluable insights into past climates and past ecosystem response to perturbations. Since such data are generated with substantial operator input (taxonomic identification) and use complex vocabularies (names of taxa, changing in time and inconsistent among operators), their reusability is severely limited, hindering global syntheses as a basis for global assessment of biotic response to climate change. Here we propose to facilitate reusability of new and old fossil occurrence and abundance data by developing a community-based workflow. Using data on Quaternary planktonic foraminifera as a model, we will combine community-driven development of semantic standards with technical implementation making use of NFDI core service platforms. We intend to involve selected researchers, learned societies, data scientists and data providers with the overarching aim to develop and demonstrate an approach to enhancing reusability of data with complex semantic structure that could be transferred on different types of long-tail data within the broad remit of NFDI4Earth. The approach considers upfront the full data life cycle, allowing integration of legacy data with workflows for new data submissions, and provides an intersection to NFDI4BioDiversity, thus fostering synergy within the broader NFDI community.
P13	**Research Domain: Atmospheric Science, Oceanography and Climate Research**
	Statistical learning to assess the factors underlying environmental changes, *Lead MPIBGC* Alexander J. Winkler[1,2], Fabian Gans[2], Miguel Mahecha[3], Ranga B. Myneni[4], Christian Kadow[5], Markus Reichstein[2] [1] Max-Planck-Institute for Meteorology, Bundesstrasse 53, 20146 Hamburg [2] Max-Planck-Institute for Biogeochemistry, 07745 Jena [3] Remote Sensing Centre for Earth System Research, Leipzig University, 04103 Leipzig [4] Department of Earth and Environment, Boston University, Boston MA 02215, USA [5] Deutsches Klimarechenzentrum GmbH, Bundesstraße 45a, 20146 Hamburg The Earth system is currently undergoing profound environmental changes including but not restricted to climate, biogeochemical flows, and biodiversity. A growing body of multivariate Earth

ID	Pilot Description
	observations can now diagnose these changes. Assessing the drivers underlying these changes is a challenging task, both technically and scientifically. The objective of this pilot project is to introduce a toolkit that facilitates statistical driver attribution by combining extensive resources of spatio-temporal data (Earth System Data Lab, ESDL) and the capabilities of new tools in the Big Data geosciences (Pangeo Project and statistical learning libraries). This toolkit is designed to be universally adaptable to problems of driver attribution in Earth system sciences. To showcase the functionality of the proposed tools, it will be applied to a key question in current climate research: What are the regional drivers of the observed changes in the Earth's ecosystems and their respective roles in driving the terrestrial sink of anthropogenic carbon? The expected result of this pilot project is an integrated workflow towards the identification and regional quantification of the factors underlying observed changes in the Earth system. This knowledge will be needed above all by the modelling community to assess the plausibility of current Earth system projections; the developed workflows can be likewise applied as diagnostics for model evaluation. The envisaged toolkit will interlink the ESDL to modern statistical learning libraries and will be integrated into the Pangeo ecosystem for open-source analysis software for Earth system sciences, guaranteeing interoperability of the NFDI with international developments in the field of analysing big n-dimensional data arrays in Earth system analysis.
P14	**Research Domain: Geography**
	World Settlement Footprint (WSF), *Lead DLR* Hendrik Zwenzner[1], Thomas Esch[1], Julian Zeidler[1] Gotthard Meinel[2], Jonas Eberle[1] [1] German Aerospace Center (DLR), Muenchener Str. 20, 82234 Wessling [2] Leibniz Institute of Ecological Urban and Regional Development (IOER), Weberplatz 1, 01217 Dresden Urbanization is the cause and consequence of most environmental and societal changes on Earth. In order to understand and manage its negative impacts, it is of utmost importance to gain detailed knowledge of the distribution and evolution of human settlements. With this pilot, a novel and unique data set - the global World Settlement Footprint (WSF) Suite - is proposed to the NFDI4Earth community together with the IOER Monitor, a German national land cover database. The objective of this pilot is to integrate these data sets into the NFDI4Earth infrastructure and make them accessible to a broad user community according to FAIR principles and common standards. Stakeholders are data and infrastructure providers, data and system integrators as well as users from different scientific domains, public authorities, and decision-makers. We expect that NFDI4Earth will advance and highly benefit from this pilot since the offered data sets are a valuable data source for many research applications. Besides, it can be seen as a precursor for numerous other satellite data applications, services and interoperability approaches to be integrated into NFDI4Earth.

5.1.2 Measure 1.2 (M1.2): Incubator Lab

Contributors: LUH (lead), additional partners via Open Calls

Links to all other measures*, in particular M2.5*

Objectives

The Incubator Lab fosters novel data science developments for ESS in dedicated focused projects. The objective of this task is to steer the **exploration of new, potentially relevant building blocks** to be included in NFDI4Earth and related NFDIs. Examples are tools for

automatic metadata extraction and annotation, semantic mapping and harmonization, machine learning, data fusion, visualization, and interaction. The Incubator Lab also serves as a forum where novel requirements can be formulated and trends presented in terms of a user consultation process. In this way, **scouting for new trends and opportunities** is achieved. The forum will materialize in annual meetings of *NFDI4Earth-Experiment*, where both achievements will be presented (e.g. from Lab projects but also from Pilots) and demands will be formulated (e.g. from the participants) which will trigger new ideas and potential projects. The results of the projects as well as the consultation process will be continuously monitored, evaluated and updated, resulting in a living document that describes current and future trends and records their implementation. The measure lead must oversee and monitor that compliance rules concerning the software and infrastructural developments are fulfilled while at the same time **innovative *blue sky* developments** should also be encouraged.

Thus, the action will facilitate an intensive exchange between experts in Earth system and data science (e.g. the national competence center for Big Data and AI Scads.AI Dresden/Leipzig, see chapter 4.1) to ensure **cutting-edge technology development**; it will create a **repository of advanced tools** to improve the use and usability of NFDI4Earth data, advancing the state-of-the-art to better meet (current and future) user needs.

Actions

Action 1: Open calls for the community, proposal selection and Trend Scouting

M1.2 shall organize open calls for the community (Milestone 1.2.2). Scientists within and outside the NFDI4Earth consortium can propose projects based on new technologies, and explore their potential for NFDI4Earth. The calls will be partly addressing **methodological gaps** as identified during the annual meetings and documented in a living trend scouting document (D1.2.1-5), and partly **open topics**. As with the pilot calls (see Action 1 in M1.1), the calls will be developed in close cooperation with the NFDI4Earth Steering Group and Plenary. Among the first activities for the Incubator Lab will be to refine **criteria that qualify for a useful project**, and a mechanism to ensure that the **selection process is open, transparent, and addresses multiple disciplines**, as well as data types. In addition, criteria for successful projects are developed (Milestone 1.2.1). All developments should show that there is **high potential for interdisciplinary and general usage**. For the software developments, a comprehensive set of compliance rules will be defined in close cooperation with TA2 (M2.5) and TA4 (M4.3). Some general characteristics of the developments must be specified, e.g. data types, functionality, software environment, related disciplines, and potential users.

Action 2: Monitoring and steering the Incubator Lab

The achievements of the projects will be monitored and documented. The selected projects from the open calls and their outcomes will be **published on the NFDI4Earth website and made available through public repositories**. Lab projects must report on their activities and on

compliance to the given rules on a regular basis. In this way, the measure lead can identify problems and opportunities early enough to take appropriate action; e.g. consult co-applicants and link closely related projects with each other. An excellent forum for exchange is the annual *NFDI4Earth-Experiment* event, which will be organized and show the latest developments and discuss them with a broader community. Experts from Computer Science and Data Science will also be invited to share their views and experience. Thus, an **annual user consultation** is facilitated and conducted during these events. These findings and developments are documented in the trend scouting living document (D1.2.1-5).

Deliverables and Milestones

Mile-stone	Delive-rable	Type*	Description	Due end of
MS1.2.1	-	R	Criteria and requirements for Incubator projects	Q2 2022
MS1.2.2	-	S	Calls for proposals, evaluation and selection	Q3 2022 Q3 2023 Q3 2024
-	D1.2.1 D1.2.2 D1.2.3 D1.2.4 D1.2.5	R	Trend Scouting: Living document describing research gaps (Final report as consolidated report of all intermediates)	Q2 2022 Q2 2023 Q2 2024 Q2 2025 Q2 2026
MS1.2.3	-	S	Organization of *NFDI4Earth Experiment-Event*	Q3 2022 Q4 2023 Q4 2024
	D1.2.6 D1.2.7 D1.2.8 D1.2.9	R	Report and Evaluation of Incubator Lab Projects (Final report as consolidated report of all Projects)	Q2 2023 Q2 2024 Q2 2025 Q2 2026

* R-Report, S-Service

5.1.3 Measure 1.3 (M1.3): Education and Training Materials and Services

Contributors: HSBO (lead), TUDD, UNIHH, UNIL, UNIMS, TUM, additional via Open Calls

Links to M1.1, M1.2, M1.4, M2.1-5, M3.1, M3.2, M4.2, M4.3

Objectives

The overall goal of Measure 1.3 (M1.3) is to **enable NFDI4Earth stakeholders – Bachelor and Master students, scientists, lecturers, data and infrastructure providers as well as other professionals – to conduct their work using an open and sustainable research data life cycle.** Such enablement is a prerequisite for widespread uptake and utilization of NFDI4Earth and facilitates cultural change towards FAIR and collaborative RDM in ESS. New career opportunities and qualification profiles such as *Research Data Manager/Steward, Research Data Curator* or even *Earth System Data Scientists* may arise. For this purpose, two primary objectives will be addressed: (1) develop competencies, necessary to conduct open and reproducible research in ESS; (2) enable ESS stakeholders to fully leverage the NFDI4Earth ecosystem in their respective (sub-) domains.

Based on the community's education and training needs, **modular *NFDI4Earth-ready* course materials and curricula** will be developed and published under an open licence (e.g. as Open

Educational Resources (OER) under a Creative Commons attributions license). The modules can be integrated into existing university curricula or serve as the basis for independent stand-alone courses. The materials cover the usage of NFDI4Earth services and the application of NFDI4Earth principles in research data management (RDM), research data curation (RDC), spatio-temporal data literacy (SDL) and spatio-temporal data science (SDS). All educational resources will be made available via a **single point of access**, which will be integrated into NFDI4Earth's OneStop4All. **Physical and virtual education and training events** will be organized on a regular basis in order to test and improve the course materials developed as well as to advance NFDI4Earth-related competencies within the ESS community. In particular, an **NFDI4Earth research data curation certificate course** will be offered. A concept for an open network of education and training hubs (**NFDI4Earth EduHubs**) will be developed and implemented. EduHubs are partner sites, which offer NFDI4Earth-related education and training services on a long-term basis. The EduHub-Network will at the same time ensure sustainable NFDI-related competency development beyond the life of the project.

Actions

Action 1: Development of curricula and course materials

High-quality training material is a key factor for the successful use of NFDI4Earth principles and technology throughout the ESS community. The material must be kept up-to-date and adjusted during the entire life of NFDI4Earth. As a starting point, existing **relevant training material will be identified and analysed** – this includes cooperation with other NFDI consortia on joined materials and education strategies. First cooperation agreements exist with NFDI4Biodiversity and NFDI4Health. The collection of training material will be a continuous effort and structured and organized according to its relevance to NFDI4Earth, the audience being addressed, the thematic field, cross-cutting topics, etc. The **identification of target group-specific education and training needs – and thus the definition of tailored learning objectives and curricula** – will be based on input from NFDI4Earth partners, as a rather complete representation of the ESS community. The collected training material and the structured user needs will be analysed and training gaps, overlaps, or incompatibilities identified. Based on this synopsis, **existing course materials will be adapted and new materials will be developed**. These materials consistently implement best practices of open and reproducible research. **Participants will be invited to contribute by providing support to make their educational offers NFDI4Earth-ready as well as by publishing regular calls for *Educational Pilots*,** i.e. calls for developing educational materials, required by the ESS community. All materials will be accessible via a single point of access. Course statistics, user feedback, and experience from trainers will be monitored based on online surveys. The results will be used to improve the quality of material and curricula.

Action 2: Single Point of Access

To share education and training related content (materials, online courses etc.), a **single point of access will be created as an integral part of NFDI4Earth's OneStop4All**. This access point will act as a structured hub to group, manage, search for and access material that might be also distributed on different platforms. The single access point will be designed and implemented according to FAIR principles and NFDI4Earth's openness guidelines. To save effort, technical components of existing systems (like CRM components, wiki systems, search engines, etc.) will be re-used. Beside the technical implementation, **procedures for content management and target group-specific quality assurance will be developed in order to ensure high quality, openness, sustainability, and accordance with NFDI4Earth's focus**. The single point of access will be linked with other platforms of related NFDI initiatives and might be integrated within other communities over the long term.

Action 3: Implementing NFDI4Earth education and training

To quickly develop and enhance NFDI4Earth-related RDM-, SDL- and SDS-competencies within the ESS community as well as to test and to improve the developed educational resources, **an NFDI4Earth education and training activity plan will be developed and maintained**. Based on this plan, **physical and virtual education and training events for different target groups** will be conducted on a regular basis (workshops, summer schools, hackathons, train-the-trainer etc.). This way also **higher education institutes will be supported in integrating NFDI4Earth educational resources and curricula into existing and future study programmes** to ensure sustainable long-term adoption of NFDI4Earth principles. Furthermore, an **open network of nationally distributed NFDI4Earth EduHubs will be established**. EduHubs are run by partners, who commit themselves to permanently offer education and training services for specific aspects of ESS-related research data management and spatio-temporal data science (e.g. by setting up joint study programmes, regular summer schools or vocational trainings). For the implementation of the network, a concept will be developed that describes basic services, quality criteria requirements as well as guidelines for implementing and maintaining an EduHub. Based on this concept, initial EduHubs with a clear topical focus and long-term educational offers will be established. UNIL, TUDD and HSBO will be the initial nuclei.

Deliverables and Milestones

Mile-stone	Delive-rable	Type*	Description	Due end of
-	D1.3.1	R	Mapping of existing educational resources and initial education and training needs within the ESS community	Q4 2022
-	D1.3.2 D1.3.3 D1.3.4	R	Target group specific curricula (initial report and revisions 1 and 2)	Q1 2023 Q2 2024 Q4 2025
-	D1.3.5 D1.3.6 D1.3.7	R	Best practices guidelines for NFDI4Earth ready course materials (incrementally updated and revised according to progressing NFDI4Earth implementation status)	Q1 2023 Q2 2024 Q4 2025

Mile-stone	Delive-rable	Type*	Description	Due end of
-	D1.3.8 D1.3.9 D1.3.10 D1.3.11	R	Evaluation of NFDI4Earth education training events based on NFDI4Earth educational resources	Q3 2023 Q3 2024 Q3 2025 Q2 2026
-	D1.3.12 D1.3.13	R	Technical specification for single point of access and process model (initial document and revision)	Q4 2022 Q3 2024
MS1.3.1 MS1.3.2	-	S	Single point of access for educational materials (initial service and revision)	Q2 2023 Q1 2025
MS1.3.3 MS1.3.4 MS1.3.5 MS1.3.6	-	S	Initial set of modular *NFDI4Earth-ready* course materials published under an open license	Q2 2023 Q1 2024 Q1 2025 Q1 2026
-	D1.3.15	R	Concept for establishing and maintaining a network of NFDI4Earth education and training sites (EduHubs)	Q3 2024
MS1.3.7	-	S	First NFDI4Earth EduHub implemented	Q4 2025

* R-Report, S-Service

5.1.4 Measure 1.4 (M1.4): NFDI4Earth Academy

Contributors: Geo.X (lead), DAM, ABC/J

Links to M1.1, M1.2, M1.3, M2.5, M3.3, M4.1

Objectives

This measure aims to establish the NFDIDI4Earth Academy as a **think tank to connect young researchers from all ESS sub-disciplines and their data-driven research**. The NFDI4Earth Academy allows its intrinsically motivated Academy Fellows to advance their data analysis skill set at the intersection of mathematics, statistics, computer sciences, and Earth science in a peer-mentored environment. Concurrently, NFDI4Earth findings such as standards, tools, algorithms, workflows particularly from pilots and incubators (M1.1, M1.2) and data cube technologies (M2.5) can be tested and implemented in the Fellows' ongoing **research projects** and, therefore, serve as an **initial proof of concept**. Academy Fellows are key players in the NFDI4Earth Academy's agile organizational frameworks. To contribute to the continued scientific and future-oriented development of NFDI4Earth, **academy fellows are encouraged to play an active role in the governance structures of NFDI4Earth** (see chapter 3.4). The open **NFDI4Earth Academy programme** aims to connect diverse scientific topics and sub-disciplines within the ESS community to foster interdisciplinary data science. The NFDI4Earth Academy concept is transferable to other scientific disciplines. It may serve as a **blueprint for an NFDI Academy** that unites all NFDI consortia's efforts to advance young researchers' NFDI-compliant data science skills (link to M3.3).

Actions

Action 1: Setting up the NFDI4Earth Academy

The networks Geo.X, DAM, and ABC/J will establish **three complementary Academy sites that serve as the NFDI4Earth Academy's backbone**. Together they encompass 14 universities, 22 research institutions from Max-Planck, Leibniz, and Helmholtz, authorities, museums, as well as Senckenberg and represent a broad and interdisciplinary research spectrum of ESS. The networks comprise a significant part of the NFDI4Earth consortium, which is sufficiently large to ensure the recruitment of a critical number of fellows from different ESS-subdiciplines. Due to their communication channels, experience in state-of-the-art promotion of young scientists, and their involvement in inter-/national geo-data management activities, the networks are suitable facilitators for setting up the NFDI4Earth Academy.

NFDI4Earth Academy coordinators of each site jointly **refine the Academy concept and realize the implementation** by (a) organizing the fellows' recruitment and their involvement in NFDI4Earth governance, (b) implementing the Academy programme, (c) applying quality assurance mechanisms, (d) managing further development of the Academy, and (f) ensuring compliance with NFDI4Earth objectives.

Action 2: Fellow recruitment

Open calls at each Academy site will address early career researchers at the Ph.D. and PostDoc level of their career, who want to accelerate their current projects by developing and testing novel approaches of advanced data analysis in a peer-mentored environment. A **scientific panel** consisting of scientists from the entire ESS community in Germany **will select** scientifically promising and project proposals compatible with NFDI4Earth in a diversity-sensitive, criteria-based recruitment process. This process will be conducted for three fellow cohorts by the Academy sites. Fellows of each cohort must bring along individual funding. To contribute to the continued scientific and future-oriented development of NFDI4Earth, **fellows from all sites will elect Academy representatives** to reflect the interests of young scientists in the management structures of NFDI4Earth.

Action 3: Establishing the Academy programme and involving the ESS community

The Academy programme relies on **agile, self-organized, and bottom-up processes** of the fellows. They will be guided by the coordinators towards collaboration that produces **joint outcomes of interest in the NFDI4Earth community**, e.g., publications, validation models, data sets, software, tools, etc. (see also Action 4). Programme elements are (virtual) talk series, workshops on self-chosen topics, autumn school, and Hackathons to advance professional skills in data science. We implement training services offered by M1.3, and organize the exchange with M2.5 developers to stimulate the application of innovative data cube technologies in the fellows' projects. Participation at events with scientists from the ESS Pilot and Incubator projects (notably the NFDI4Earth Experiment Event in M1.2) is mandatory for fellows.

Action 4: Monitoring quality and NFDI4Earth compliance

The coordinators develop a **quality assurance mechanism and document criteria for success**: (1) Number of fellows and ESS researchers involved, (2) NFDI4Earth Academy events, (3) joint fellow publications, project proposals, and (4) participating ESS researchers in the academy network. Qualitative feedback from fellows, PIs, and participating ESS researchers will be part of the coordinators' monitoring throughout the Academy. Based on this monitoring, the coordinators will **adjust and modify the Academy concept** to directly incorporate these improvements for following cohorts. The final result will be a White Paper proposing the improved NFDI4Earth Academy concept as a **blueprint for an NFDI Academy**.

Action 5: Expanding the Academy network

Setting up additional Academy sites within other NFDI consortia is possible on condition that consent to the basic principles of the Academy is given: : (1) Focus on promoting young researchers (Ph.D. candidates and PostDocs), (2) commitment to the NFDI4Earth mission (see TA4, M4.2), (3) openness for involvement of the ESS community in Germany, and (4) suitable management structures for supporting self-organization processes.

Deliverables and Milestones

Mile-stone	Delive-rable	Type*	Description	Due end of
-	D1.4.1	S	Open call for the first cohort has been published	Q2 2022
-	D1.4.2	S	Academy concept is refined; fellows have been recruited, and Academy representatives are elected	Q3 2022
MS1.4.1	-	S	Academy program has started, first events have been organized	Q1 2023
-	D1.4.3 D1.4.4 D1.4.5	R	Academy concept has been evaluated and adjusted	Q1 2023 Q1 2024 Q1 2025
-	D1.4.6	S	Open call for second cohort has been published	Q3 2023
MS1.4.2	-	S	Fellows of second cohort have been recruited; revised Academy program has started	Q1 2024
-	D1.4.7	S	White Paper proposing an NFDI Academy with NFDI4Earth Academy as blueprint	Q3 2024
-	D1.4.8	S	Open call for the third cohort has been published	Q4 2024
MS1.4.3	-	S	Fellows of third cohort have been recruited; revised Academy program has started	Q1 2025
MS1.4.4	-	S	Three fellow cohorts have passed the NFDI4Earth Academy	Q3 2026

* R-Report, S-Service

5.2 Task Area 2: 2Facilitate

Lead: DKRZ, KIT, MPIBGC

Overview

TA2 2Facilitate has two main aims: (1) to provide community support on all aspects of FAIR data and related infrastructures in NFDI4Earth and (2) to advance new technologies for FAIR data as best practice examples / templates.

Thus, task area TA2 will offer a first port-of-call for all participants and community members to start and/or continue their journey to FAIR data management in Earth System Sciences. The TA will provide a **OneStop4All (M2.1) to act as an initial visible contact point**. In a structured way, this web-based contact point will provide basic information on the general principles of FAIR data, e.g. how to find and access existing data sets, how to contact existing repositories, how to take first steps in making data FAIRer, and how to find other services provided by NFDI4Earth. **User requests going beyond the basic requirements will be routed to the User Support Network (M2.2)**. Here, experts will directly interact with the people asking the questions. The experts will also provide a bridge to the existing support infrastructures of the relevant partners, including offers such as data curation and data stewardship. To demonstrate best practices and foster the interaction between long-tail data and comprehensive data sources, the other TA measures will closely interact with M2.1 and M2.2. **Governmental Data (M2.3)** will provide the wealth of data that exists in governmental data bases and is strictly regulated by national and international laws and standards. Data in **Long-Term storage (M2.4)** provides an additional sustainability perspective that will also be of interest for new comprehensive and long-tail data sets. Thus, making long-term storage more accessible for use (data-in) and re-use (data-out) will be an aim of M2.4. Our **Advancing Tools (M2.5) measure provides a bracket to the best-practise exploration of data in all sorts of repositories**. In close collaboration with TA1, state-of-the-art solutions based on data cubes and other emerging technologies will be explored (e.g. creation of data cubes from governmental data and analysing it in conjunction with other remote sensing data). Some results will be presented in Thematic Viewers (TVs) that form part of the RF E&E DataHub to the general public. This best practices model should be seen as an example, and other data models (also as part of other emerging technologies) will be added at a later stage.

Certain measures form obvious points of contact to other TAs. M2.1 will act in close collaboration with TA4 to create a comprehensive NFDI4Earth website. It will be hosted at TUDD and will include general information on NFDI4Earth, an internal section for the organisation of NFDI4Earth, and **OneStop4All, where *clients* can start their ESS RDM journey by searching for information**. If the question cannot be answered by the search or a guided query (including parts of the Knowledge Hub), user support in M2.2 is activated. This distributed user support will be well embedded in the NFDI4Earth community and has linkages to the educational and Academy measure in TA1. KnThe owledge Hub is the central exchange platform created by

NFDI4Earth (the technical foundation resides in TA4). Here, knowledge is provided about what is available, how to best implement FAIR-enabling processes, and how data curation can be supported. In addition, analysis of requests will provide information on the most pressing issues clients have. Additional linkages (also via the Knowledge Hub) between M2.3, M2.4, M2.5 and corresponding Measures in TA3 will deal with sustainability, consolidation and implementation issues.

5.2.1 Measure 2.1 (M2.1): OneStop4All

Contributors: DKRZ (lead), KIT, HZH

Links M2.2, M2.3, M2.4, M2.5 and all other TAs

Objectives

This measure develops, provides and maintains the web pages and services of OneStop4All - together with TA4 (hardware, web content management system and general NFDI4Earth project web pages) and as part of the overall NFDI4Earth website. **OneStop4All acts as a key entry point for all support requests and inquiries to NFDI4Earth.** The OneStop4All (which is the 1st level support of NFDI4Earth) and its search functionality (OSF) is a key offering to the ESS community in Germany. The OSF will search local resources (e.g. the NFDI4Earth Knowledge Hub) and remote sites (pre-caching information from existing community offerings) using user provided keywords and will present results in an appealing human readable web page. The OSF itself will be continuously expanded by M2.1, whereas the technical foundations will be provided by M4.3 (TA4). Thus, FA of the FAIR principles are primarily addressed by the OSF and foster further alignment of the services in and close to NFDI4Earth by making them centrally searchable. **The purpose of OneStop4All is to directly answer questions of researchers inside and outside NFDI4Earth about research data management and to promote the implementation of FAIR principles in ESS in Germany**, however by doing so inconsistencies will be revealed. Typical questions that are anticipated are, for example: Where can certain ESS data be found? Which data standards should I adhere to? What options exist to make data permanently available and reusable? In particular, the requests of the pilots will be adressed.

Questions that cannot be answered directly via the OSF (triggered by a user query), are passed on to M2.2, which represents 2nd level support in NFDI4Earth. The information content accessible to the **OSF benefits directly from information available in the Knowledge Hub** (designed by M3.1 and M3.2; technically implemented and operated by MA4.3), which pools and distils all information collected by NFDI4Earth - covering a broad range of disciplines, methods and more.

Actions

Action 1: Setup, operation, and further development of OneStop4All

The OneStop4All and its search functionality (OSF) act as the central access point for all support requests to NFDI4Earth. First, we will identify the requirements and develop a first concept for the

structure of the OneStop4All and the OSF in particular, which is going to be checked and aligned several times during the project. Static local and pre-cached information will be generated and will be combined with an user friendly search function that will be linked (with increasing complexity over time) to the Knowledge Hub. In this way, the growing wealth of knowledge of NFDI4Earth will be made available to the inquirers in an agile, smart form.

Action 2: Transfer of complex requests to M2.2

Advanced questions which cannot be answered by the OSF will be handed over to the support team in M2.2. For this purpose, we link the OneStop4All to the ticketing system operated by TA4 (M4.3). M2 experts will retrieve the tickets. To make the handover to M2.2 efficient, we will jointly establish meaningful support categories together with M2.2, which are regularly reviewed.

Action 3: Contribution to the structure and interfaces of the Knowledge Hub

The Knowledge Hub is a key component of NFDI4Earth in general and a valuable resource for the OSF. It integrates the knowledge of NFDI4Earth, e.g. on available services and accepted data standards as well as on data management plans and advanced technologies as covered in M2.5. In order to unlock the potential for knowledge of the OneStop4All and to keep up with the rapidly growing information content over time, it is essential that the structure and interfaces of the Knowledge Hub are designed in a way that make them usable for the OSF and the colleagues in M2.2. Against the background of the support processes and with the experience gained from the questions handled by the OSF, we will contribute to structuring the Knowledge Hub operated by TA4.

Action 4: Systematic review of support

The central goal of M2.1 is to ensure high-quality support from the OSF. This applies not only to the range of topics addressed by OneStop4All, but also to the quality of the information presented and the promptness with which researchers reach a solution (in appealing human readable form, or as, e.g., reusable code). Furthermore, user satisfaction is a high priority. In M2.1 we will first compile the set of criteria by which we can assess the quality of OneStop4All. While respecting data privacy, we will collect data on these criteria from and with OneStop4All, evaluate it systematically at regular intervals and derive recommendations for the further development of support in all facets.

Deliverables and Milestones

Mile-stone	Delive-rable	Type*	Description	Due end of
MS2.1.1	-	S	OneStop4All fist version online (Action 1)	Q4/2022
MS2.1.2	-	S	Ticketing System integrated (Action 2)	Q4/2022
-	D2.1.1	R	Criteria for evaluating the support quality of the OneStop4All (Action 4)	Q1/2023
MS2.1.3 MS2.1.4 MS2.1.5	-	S	Definition and Review of Support Categories (Action 2)	Q3/2022 Q3/2024 Q3/2026
-	D2.1.2 D2.1.3 D2.1.4	R	Review of support requests (Action 4)	Q32023 Q3/2024 Q3/2026
MS2.1.6	-	S	Technical integration of Knowledge Hub into OneStop4All (Action 1 + 3)	Q3/2024

* R-Report, S-Service

5.2.2 Measure 2.2 (M2.2): User Support Network

Contributors: GEOMAR (lead), LRZ, GFZ, TUDD, UNIHH, FUB, DKRZ, KIT

Links M2.1, M2.3, M2.4, M2.5 and all other TAs

Objectives

This measure will develop, coordinate and operate a distributed, cross-institutional User Support Network (USN) for NFDI4Earth, based on the existing and well-embedded user support structures of the participating institutions. Dillo and Doorn (2014) – for instance – presented a comparable concept for RDM in the Netherlands. The USN serves as a single point of contact for requests that could not be handled via M2.1 and require individual consulting. Hence, M2.2 represents 2nd level support for NFDI4Earth**. At the USN, *real people* interact with and help ESS researchers to find answers to questions related to all aspects of NFDI4Earth services** (in collaboration with TA3), FAIR data (in collaboration with M2.3, M2.4, and M3.2) and software tools (in collaboration with M2.5). The USN approach covers the diversity of NFDI4Earth by involving many scientific backgrounds. Here, wide methodological knowledge and experience of heterogeneous data sets from observation, experiment and simulation can be found. The USN coalesces RDM support, i.e. institutional RDM help desks from many different (infrastructure) providers. Providing user support is seen as a community effort, with the ambition to provide a high-quality support network (including data stewardship) for ESS researchers and to create incentives for participating user groups to share their experience to achieve further quality improvement. Largely, the project resources will be used for organising the distributed request flow through the network and orchestrating the USN along the actions defined below. Technically, the USN will operate on a ticketing system, provided by TUDD (TA4) and triggered via OneStop4All and will use the Knowledge Hub, containing the Living Handbooks, as its *long-term memory* - adding and retrieving information.

Examples of tangible support actions to enable FAIR data user-interactions

- (findable) discover data that exists and store your own data so that it is findable for others (relevant, standardized metadata, well-chosen repository, etc.)
- (accessible) retrieve and use (discovered) data and provide your own data with qualified description in an accessible form (choice of repository and technology of provisioning)
- (interoperable) interconnect and pool data (enabling new discoveries on the boundaries of disciplines by reaching out to other disciplines), annotate and prepare your own data so that it becomes more accessible and usable in other disciplines (good metadata, efficient software for reprojection and merging of data)
- (reusable) find what data (and services/software) already exist (not reinventing the wheel saves both time and money) and how it can be understood (and used) by correct interpretation of the metadata (software manual) and how to ingest and publish your own data (and software / tools) in a form that can be easily discovered and reused by others.

As indicated by the examples above, support will encourage self-reflection regarding FAIR principles, i.e. is my own data FAIR and if not, what can I do, to achieve this? This part of the FAIR game is strongly associated with data curation, data quality (i.e. quality of description / metadata) and ingest procedures and options, since they provide the automated validation of appropriate data format and metadata description. Hence, to make comprehensive and long-tail data sets FAIRer and future-proof, and by doing so promote the cultural change towards FAIR and Open research data, a good understanding of data ingest procedures and options needs to be facilitated. Within TA2 this is supported by M2.3 (governmental perspective) and M2.4 (data in long-term storage – which includes curation). However, this also relates to many aspects of TA3 and how metadata will be developed (ontologies, vocabulary, mapping, etc.) to allow for a sustainable use of curated data.

By combining the distributed RDM knowledge of experts in the USN in conjunction with the Knowledge Hub, M2.2 will convey the notion (knowledge) of a best practice for dealing with data and how data can be made FAIRer and open (by acknowledging privacy and legal issues).

Actions

Action 1: Establish the User Support Network (USN)

Develop and implement a concept for the integration of existing RDM support structures into NFDI4Earth and operate the USN, substantiated with the implementation of a lightweight USN cooperation agreement between USN members, defining the rules and roles of each member. The USN cooperation agreement will supplement the NFDI4Earth consortium agreement (M4.1). The USN establishment process will compile partners and expertise and will define rules of participation as well as define and implement USN internal standard operating procedures (SOPs) for USN members. Annual workshops and interested RDM support groups will bring together

people and initiate and frame the continuous improvement process of the USN. The initial workshop will be held in Q2/2022, in order to make the USN ready to support the first NFDI4Earth pilots starting.

Action 2: Customise the Ticketing System (together with M2.1)

The ticketing system, technically provided by the TA4, will be customised by the USN. Action 2 will define terms of use and the rules and roles (responsibilities) among the partners for the basic internal support platform. Additionally, Action 2 defines (together with M2.1) and implements support categories for automated provision (distribution) of support requests and content tagging for automated publication of e.g. a FAQ list at OneStop4All (M2.1).

Action 3: Provide content for the NFDI4Earth Knowledge Hub

The USN will provide content for the Knowledge Hub, in particular for the NFDI4Earth Living Handbooks, edited by M.3.3, produced according to previously defined documentation standards for USN supporters. Additionally, user-relevant gaps in the Knowledge Hub documentation will be identified, collected and communicated to the corresponding measures in TA3.

Action 4: Implement a USN quality improvement cycle

The quality improvement cycle will ensure continuous facilitation of NFDI4Earth user support, based on collected USN statistics and request monitoring as well as on the regular user surveys implemented by this action. The aim is the development and establishment of an annual quality improvement cycle, including a supporter meeting (see action 1) and an annual report.

Deliverables and Milestones

Mile-stone	Delive-rable	Type*	Description	Due end of
-	D2.2.1.a	W	Workshop of RDM Support Groups in ESS (Action 1)	Q2 2022
-	D2.2.1b	R	Concept for User Support Network (USN) organisation (Action 1)	Q3 2022
MS2.2.1	-	S	USN core team established (Action1)	Q4 2022
-	D2.2.2	R	Concept for USN Ticketing system - customisation and implementation (Action 2)	Q3 2022
MS2.2.2	-	S	Ticketing System for USN support available (Action 2)	Q4 2022
-	D2.2.3	R	Concept for USN Documentation (Action 3)	Q4 2022
MS2.2.3	-	S	Framework for USN content production (Action 3)	Q1 2023
-	D2.2.4	R	Concept for an annual quality improvement cycle (Action 4)	Q4 2023
MS2.2.4	-	S	Framework for an annual quality improvement cycle (Action 4)	Q2 2024
-	D2.2.5 D2.2.6 D2.2.7	R	Annual USN Improvement Report (Action 4)	Q2 2025 Q2 2026
MS2.2.5	-	W, P	Continuous USN Improvement Process (Action 4)	Q2 2025 Q2 2026
-	D2.2.8	R	User Support Summary and Perspectives Report (Action 1-4)	Q2 2026

* R-Report, S-Service, P-Process, W-Workshop

5.2.3 Measure 2.3 (M2.3): Governmental Data

Contributors: DWD (lead), BKG, BGR, BSA

Links to M1.1, M2.1, M2.2, M3.2, M3.3

Objectives

In addition to **improving the findability and widening the use** of governmental data (note that in Germany, federal states and the federal government have distributed responsibilities – and governmental data is used as a generic term for both originators) in ESS, the aim of this measure is to establish a bidirectional communication channel between ESS researchers organized in NFDI4Earth and governmental authorities. This communication channel can lead, for example, to the authorities **receiving suggestions and recognizing needs from the NFDI4Earth community** in a more targeted way or - in the other direction - to the development of **processes in the ESS research community** for better standardized workflows mirroring best practices of the authorities. An important part of this networking strategy is the GDI-DE coordination office based at BKG that links federal states and the federal government. GDI-DE has endorsed NFDI4Earth as a prime partner (https://www.gdi-de.org/en/cooperations/science) to foster an even close relationship with all NFDI4Earth participants, thus rapidly increasing use and uptake of governmental data.

In Germany, systematic observation, analysis and prediction of selected Earth System components are the official duties of specialized government agencies. For example, in Germany long-term atmospheric observations are performed by DWD. Such data are often used for operational activities, which result in further derived products, e.g. numerical weather predictions or climatological data sets. Several such data sets are highly relevant for ESS, but were not easily accessible in the past. The general shift towards open data policies resulted in a change of data policies for governmental data, e.g. via general regulations such as GeoIDG, PSI directive, GeoZG and INSPIRE (see chapter 4); or via special regulations such as the recent Act on the Tasks of DWD in 2017 (which allows unrestricted use, e.g. integration in other databases or commercial use). In addition, the federal states (and their agencies and authorities) started opening up their data policies. However, the potential of the new situation is stilly only being exploited to a very limited extent. Integrating governmental data in NFDI4Earth requires some special considerations. Typically, the data sets are generated (either by observing systems, or other processes, such as prediction models) to be used for operational tasks of the agencies as well as in international activities that rely on standardized data exchange (e.g. for weather prediction, earth movement forecast). Additionally, data sets should be findable in other official portals, e.g. the German GovData portal that aims to provide a central catalogue for all governmental data. These activities are based on the standardized exchange of metadata between government agencies (and will be realized in conjunction with TA3). Several ongoing efforts aim at improved discoverability and access to data sets, but are often discussed

independently from standardization efforts in the ESS community. Thus, an alignment is urgently needed and can also be used to improve cooperation in other areas.

Actions

Action 1: Improving discoverability and use of governmental data

Governmental data provides a rich (under-utilised) resource for ESS research. It has its own legal framework and offers a unique opportunity to interact with a perspective that is more product orientated. Governmental data is also characterised by being very diverse, complex and of large volume. Thus, the full complexity of the metadata definition that is needed for the appropriately consistent integration of these multi-dimensional data sets into various kinds of external infrastructures has not yet been established or implemented.

Starting with data from the DWD and supported by the measure, for participants that cover different ESS domains the first step is to compile information on the data made available for research by governmental authorities at OneStop4All. This is facilitated by the fact that much of the data is already available on official public portals (such as the German Gov-Data Portal and the Geoportal.de), accessible via open APIs, e.g. enabling distributed discovery and federated data access. With the development of the Knowledge Hub in TA4 and TA3 and by feeding information into it, we will increasingly pool this information, making it more easily available.

In a further step, again starting with publicly available offerings from DWD, we will initiate an improved and consistent interaction of the products with the tools developed in NFDI4Earth. Based on initial experience (e.g. through pilot projects), general suggestions for the consistent integration of governmental data into NFDI4Earth will be developed. These developments will include strategies on metadata exchange between selected data catalogues. The focus will be on generalized concepts that can be generally applied for governmental ESS data. This will include technical, organizational and legal considerations (e.g. on licensing) as well as a general strategy for the coupling of governmental data catalogues to NFDI4Earth (in particular via OneStop4All). Here, DWD will also support cooperation with the international data centres hosted at DWD as well as international (European) organisations. This also applies to the BGR, which provides data to the national and European environment via the GDI-BGR and the GDOI-DE. To ensure success, we will work closely with TA3 (M3.2) in this step. In conjunction with M2.5, selected governmental data will be used to create data cubes that thematically (and by data volume) enhance existing data cube environments.

Action 2: Analyse and structure feedback from ESS researchers

We will use close interaction with OneStop4All (M2.1) and the support team in M2.2 together with the measure representatives in TA3 to generate impetus for improving the offerings of governmental data. We will ensure this by a systematic and regular evaluation of all interactions, including the uptake of the newly-created data cubes for research purposes.

Action 3: Develop recommendations to improve standardised workflows

The close observation of and stronger involvement with the activities in NFDI4Earth will also allow us to harvest the wealth of experience from governmental research for the consortium. For this purpose, we will compile white papers and handbooks in conjunction with TA3 and make them available to NFDI4Earth via the Knowledge Hub.

Deliverables and Milestones

Mile-stone	Delive-rable	Type*	Description	Due end of
-	D2.3.1 D2.3.2	R	Report on data made available for research by governmental authorities (Action 1, 2)	Q3 2024 Q1 2026
-	D2.3.3	S	Best practice data cube example from DWD (Action 1)	Q2 2023
-	D2.3.4 D2.3.5	R	General suggestions for the consistent integration of governmental data into NFDI4Earth and recommendations for improvement of offerings of governmental data (Action 1,2)	Q3 2024 Q1 2026
-	D2.3.6	R	Workflow recommendations white paper (Action 3)	Q1 2026

* R-Report, S-Service

5.2.4 Measure 2.4 (M2.4): Data in Long-Term Storage

Contributors: BKG (lead), BSA, UNIKI

Links to M1.1, M2.1, M2.2, M2.3, M3.1, M3.2, M3.3

Objectives

The aim of this measure is to bring together **different perspectives on long-term storage and archiving and to generate recommendations for improved preservation of ESS data** (irrespective of size and source) in long-term repositories. Perspectives to be considered are those of the scientists generating the data, those of the subsequent scientific users and those of the repositories and archives, also in the role of data stewards.

In order to achieve sustainable reusability of data, long-term storage (~8-12 years) and long-term archiving are key. In ESS in Germany, long-term preservation and archiving of data has already been implemented or is being aspired to in many places. For example, in World Data Centers such as PANGAEA or WDCC, inter-institutional data infrastructures (i.e. Helmholtz Data Federation, DLR's German Satellite Data Archive (D-SDA)), and in institutions/agencies (i.e. DWD, Landesbetrieb Geoinformation und Vermessung Hamburg, State Archives as the Bavarian State Archives). Nevertheless, huge amounts of data from various sources are still lost, because of insufficient measures to preserve the data of e.g. temporary research projects. Long-term preservation and archiving, including sustainable storage capacity, regular media and format migration, is a challenging task due to the large volumes and complex structures of ESS data sets. The main challenge is the preservation of the scientific reusability and interpretability of migrated research data. Several factors are important in this context (see e.g. Rönsdorf et al., 2016 for a set of principles). Metadata and data formats need to be carefully considered from the very beginning of the data creation processes, with long-term preservation in mind. Together with

TA3 (M3.2), **we contribute to the development of guidelines for long-term archiving within the framework of the Knowledge Hub**. These will also inform the entire task area TA1, including the EduHub and the Academy. Also, in interaction with M3.1, we will compile suggestions for repository providers and add these to the Knowledge Hub. Additionally we will give impetus to sustainable preservation and availability of research data towards the entire NFDI (in cooperation with M3.3). In ESS, the decision which data is relevant for long-term archiving (and which is not) is of utmost importance – not least because data volumes are significant – but is not yet coherently addressed. Here, we will develop criteria that are helpful for both researchers and repositories.

Actions

Action 1: Status of long-term preservation and long-term archiving in ESS in Germany

With this action, we will achieve a comprehensive overview of the current status and indications of strengths and weaknesses of the current setup. A questionnaire for researchers and providers will be developed and deployed at the beginning of the project. We will work in close collaboration with TA3 (M3.1 and M3.2). The survey will be adapted over time (if required) and will be repeated several times during the project.

Action 2: Consolidate guidelines for the preparation of data for archiving

Based on the surveys in action 1, the information from TA3, the experience of first and second level support (M2.1 and M2.2) and the experience gained in M2.5 preparing exemplary data cubes for long-term storage, we create white information sources for the Knowledge Hub, that will provide researchers with easily accessible information and best practice examples of long-term data preservation. Data formats and metadata will play a major role in this. Amongst others, researchers with long-tail data will be addressed. We will prove the practicability of these guidelines together with selected pilots (M1.1) and the expertise in M3.3.

Action 3: Consolidate guidelines for repository providers

This action helps to jointly improve long-term storage and archiving processes (making them faster and more efficient for users and providers), especially when new data is added. For this purpose, we will consult the survey from Action 1 and in cooperation with M3.3 we will link to similar activities in NFDI and beyond. Results will be made available in the Knowledge Hub.

Action 4: Consolidate guidelines for the appraisal of ESS data for long-term preservation and archiving

Selection criteria to qualify data for long-term archiving are manifold. They include, amongst others, requirements by funders, assessment processes and financial considerations. By means of a survey and the review of existing guidelines of funders, institutions, assessments, etc. we will generate guidelines, ingest these in the Knowledge Hub and, together with M3.3, check their applicability jointly with all NFDIs and, together with M3.4, their relevance in the international context. For this action, we will establish an interest group that will be dedicated to this topic.

Deliverables and Milestones

Mile-stone	Delive-rable	Type*	Description	Due end of
-	D2.4.1 D2.4.2	R	Survey covering status of long-term preservation and long-term archiving in ESS in Germany (Action 1)	Q3 2024 Q1 2026
MS2.4.1	-	S	Start of the interest group Criteria for Appraisal (Action 3)	Q4 2022
-	D2.4.3 D2.4.4	R	Guidelines for the preparation of data for archiving (Action 2)	Q3 2024 Q1 2026
-	D2.4.5	R	Guidelines for repository provider (Action 3)	Q1 2026
-	D2.4.6	R	Guidelines for the appraisal of ESS data for long-term preservation and archiving. (Action 4)	Q1 2026

* R-Report, S-Service

5.2.5 Measure 2.5 (M2.5): Advancing Tools

Contributors: MPIBGC (lead), UNIL, KIT, DLR, UNIMS

Links to M1.1, M1.2, M1.3, M1.4, M2.1, M2.2, M2.3, M2.4, M3.2

Objectives

This measure will integrate advanced software technologies with the NFDI4Earth community. We will investigate, assess and improve new technologies and their potential value for NFDI4Earth users and infrastructures. The main focus of this measure will be on “data cube technologies for gridded data”, but not limited to these. Here, we use data cubes as an umbrella term for software solutions that enable users to map functions and workflows on large data cubes.

Traditionally, working with gridded data is a large part of data-driven ESS. Examples of these are climate data sets such as climate projections or reanalysis, or a variety of satellite-based data products (level-3 and beyond). All these data sets share the property that they can be described as a multidimensional dense array, where the dimensions may have a physical interpretation such as space (X, Y, lon, lat, altitude), pressure level or time, or represent other categories such as ensemble members or sensor frequency bands. In the ESS community, there are several formats used for these types of data sets such as NetCDF, HDF, GeoTiff etc with different preferences for metadata conventions depending on the respective research field.

With the size of data sets becoming ever increasing, handling these large data sets has become increasingly more difficult. One problem is that data sizes are becoming so large that it is not possible for end users to simply download a whole data set to their computers, load them into memory and do their analysis, meaning that the way users interact with the data has changed significantly in the last decade. **The new often-cited approach here is *bringing the algorithm to the data***. In addition to traditional HPC solutions, where computer clusters have access to shared file systems, cloud computing services with their associated storage systems such as AWS, GEE, DIAS, promise users a simple approach to remote data sets and allows convenient distributed analysis. An additional obstacle here is that these services often rely on the data being available on object storage systems, which is not compatible with the storage formats mentioned

above which are widespread in the community. To alleviate this, format extensions (cloud-optimized GeoTiff, HDF5-cloud) as well as new cloud-compatible storage formats (zarr, TileDB) have been developed and are gaining increasingly more user attraction.

The next and most important group of advancements to watch and support in this measure is **how to efficiently apply complex analyses (including ML approaches) to the large and distributed data sets**. Ideally, a scientist should be enabled to prototype an analysis chain on a small subset of the data set and then, without large modifications to their method, apply it to a huge distributed data set using some computing infrastructure. Here, modern software frameworks such as the xarray+dask python library or the ESDL.jl Julia package play a key role in providing distributed complex analytics (beyond means and sums) across different machines. In the current state and in order to make this efficient, it is necessary that data storage anticipates common data access patterns and is optimized for them, so there should be constant interaction between data creators and users to maximize scientific outcome.

Actions

Action 1: Identification and assessment of new technologies enabling complex analysis of large ESS data

The main task for M2.5 will be to provide best practice guides towards usage of these new developments around new data formats and processing tools for large gridded data with an emphasis on their usability for NFDI4Earth users and their application of FAIR principles. With input from the incubator lab (M1.2), the pilots we will examine existing open source solutions and identify gaps and possible improvements that will be tackled either in the measure or in conjunction with future pilot proposals. We will compile the state-of-the-art solutions to cloud based data storage and analysis into a document and make this available to the internal knowledge base for further usage among NFDI4Earth partners. The document will be updated annually as new technologies emerge and are improved.

Action 2: Supporting relevant pilot studies and academy members in the application and extension of these technologies

M2.5 will work in close collaboration with Pilots from TA1. From the first pilot call, we identified Pilots P1, P2, P8, P9, P13 and P14 that plan to use (also) gridded data and some advanced data storage technologies (already including data cubes). We will create and maintain synergies between the pilots and help them in finding and developing fitting pairs of storage (with a perspective regarding long-term storage for selected data cubes, see M2.4 Action 2) and development tools for their respective use cases in conjunction with M1.1. For future pilot calls, the focus of M2.5 may be adjusted, depending on the technologies applied by the approved Pilots (e.g. other data models for long-tail data). Collaboration with the pilots will be summarized in a Pilot interaction report.

In addition, we will participate and engage in events organized by the NFDI4Earth Academy (M1.4) to remain in contact with the broader community of young Earth system scientists.

Action 3: Early adoption of NFDI common standards for data cube technologies

In conjunction with M3.2, we will work towards a cloud-based storage/analysis solution to meet the standards that will be defined in TA3 as NFDI common standards. As new storage formats such as zarr and TileDB emerge, they do not necessarily follow common conventions such as the CF conventions for NetCDF data, which usually take a while to manifest themselves within a community. The aim will be to choose one or more Pilot studies that use data cube technologies and seek to implement the common standards for them as soon as the standards become available. Following this, we will generate a best practices guide to implement common standards for new data formats and therefore guiding users in using these new formats in a FAIR way. In addition to pilot interaction, we will adopt data cubes so they can be made available to a broader audience through Thematic Viewers (TVs).

Action 4: Enhance existing open-source solutions where necessary

For all the actions mentioned above it might be necessary to implement new functionality in open source software projects. M2.5 will actively contribute to the development of community-driven open source projects wherever necessary to support the Pilot studies or to consolidate the implementation of metadata standards to meet the NFDI common standard. The open source software contributions and their added value for NFDI users will be summarized in an annual report on the software improvements that have been implemented by the team.

Deliverables and Milestones

Mile-stone	Delive-rable	Type*	Description	Due end of
-	D2.5.1-5	R	Overview of data cube technologies and review of other emerging technologies	Q4 2022 Q4 2023 Q4 2024 Q4 2025 Q3 2026
-	D2.5.6-9	S/R	Summary of software contributions to advancing tools in the Earth System Sciences (Action 4)	Q2 2023 Q2 2024 Q2 2025 Q2 2026
-	D2.5.10-13	R	Review of pilot support and interaction with the academy	Q2 2023 Q2 2023
-	-	S/R	Implementation of common standard labels into emerging data cubes and other services as they become available and annual revision	Q4 2024

* R-Report, S-Service

5.3 Task Area 3: 2Interoperate

Lead: AWI, GFZ, SGN, UNIF

Overview

At the core of TA3 is the further development of FAIRness with a special focus on the “I” (interoperability) of NFDI4Earth infrastructures and services. The NFDI4Earth architecture will include the characterisation of core infrastructure and service types and describe the standards required to make them interoperable. These descriptions will be based on the extended FAIR principles NFDI4Earth is committed to. An NFDI4Earth label for interoperability will be developed that encapsulates key requirements for infrastructures and services to be fully integrated. TA1 (Pilots, Education, Academy) and TA2 (User support, OneStop4All) will deliver relevant information on FAIR RDM from their interactions with users. The architecture, FAIR standards as well as the label will be designed cooperatively. To the extent possible, all kinds of stakeholders (other NFDIs, universities, research/governmental institutions) will be involved, thus **embedding NFDI4Earth in the NFDI and internationally**. The Knowledge Hub will provide a comprehensive, highly technical architecture description. A Living Handbook (derived from Knowledge Hub content) will provide a comprehensive summary of NFDI4Earth services and products (e.g. user workflows, implemented standards, interoperability evaluation criteria for an NFDI4Earth Label, etc.). Both Knowledge Hub and Living Handbook are provided in close interaction with TA4.

We will perform a continuous gap analyses of ESS user workflows, regarding missing interfaces, needs for services and performance improvements, convergence to common standards, demands for new services and suggestions to cross-link with external services. This requires communication with ESS users in their pilots. Gap analyses will trigger actions in TA4 to implement missing services on a prototype basis. A transition towards using data cubes and other database services will be supported, in close collaboration with TA2 (advancing tools). We will mandate implementation of common standards within participating infrastructures.

Measure 3.1 is responsible for the content of the Knowledge Hub and Living Handbook regarding infrastructure descriptions, interoperability issues, gap analyses list and a services catalogue.

Measure 3.2 will define the common standards for the NFDI4Earth architecture substantiated by the NFDI4Earth label, which will support interoperability of the services by provision of indicators, tools and metrics for self-assessment of the FAIRness of digital resources.

Measure 3.3 will provide a hub to exchange information on RDM architecture, interoperability, tools, etc. with stakeholders in Germany to engage on cross-cutting RDM topics. We will define guidelines to build interoperable architecture elements and collect, edit and provide inspiring best practice examples. This will serve as reference for future RDM designs in the ESS.

Measure 3.4 will ensure and promote the international visibility and awareness of NFDI4Earth. It will include Open Science principles and support the development of novel avenues of RDM by connecting and actively participating in international initiatives in and outside the ESS community.

5.3.1 Measure 3.1 (M3.1): Synthesis of a Sustainable NFDI4Earth Architecture

Contributors: AWI (lead) LRZ, UNIHH, TUDD

Links to M1.1, M1.2, M2.1, M2.2, M2.5, M3.2, M3.3, M3.4, M4.3

Objectives

The major objective of this measure is the collection and analysis of information on NFDI4Earth services and scientific ESS workflows to define an architecture for the further development of a sustainable and FAIR NFDI4Earth infrastructure.

The NFDI4Earth architecture will be defined as a set of components/services and their interactions (Action 1). The architecture will cover all aspects of NFDI4Earth such as applications, storage, data integration and access. It will be designed based on input from multiple stakeholder groups, both internal (e.g. pilots, participants, NFDI4Earth community) and external (e.g. NFDI Research Data Commons, international ESS community, compatibility with initiatives such as EOSC). The input, existing and future scientific workflows, services and other requirements will be analysed and abstracted. Openness, FAIRness and common standards (M3.2) will be basic design principles. In parallel, a comprehensive catalogue of all NFDI4Earth services will be set up (Action 2). It will include detailed information about the services, their interactions, documentation, standards used, usage etc. The services will also be mapped on the architecture and their requirements to allow individual analyses (see Action 3). Both the architecture and the service catalogue will be hosted (and the information interlinked) in the Knowledge Hub developed in cooperation with M4.3. Based on the architecture, gaps in the current service infrastructure (e.g. regarding interoperability or missing services) will be identified and roadmaps for further development of NFDI4Earth services will be proposed (Action 3). Information on the architecture and the service catalogue will be published as part of a Living Handbook for the NFDI4Earth community in OneStop4All (Action 4) and as supporting knowledge for user support (M2.2) and education (M1.3).

Actions

Action 1: Definition of a sustainable NFDI4Earth architecture of components

In this action, we will **design and define an architecture to interoperable combine and link the variety of services** of NFDI4Earth (e.g. repositories, analysis services, visualisation services, pre-processing services). We will assess typical ESS research workflows, services and more regarding their technical requirements - both of stakeholders internal and external to NFDI4Earth. Among others, this includes community-specific requirements from the pilot projects of M1.1 and the data cubes of M2.5, technical requirements from infrastructure providers (e.g. Cloud

provision), technical FAIRness criteria from M3.2, NFDI wide research data commons agreements via M3.3 (e.g. from the special interest groups), specifications of the international ESS community (in cooperation with M3.4), RDA recommendations, and the requirements to become part of EOSC. For the specifications, we will consider conformity with standards and recommendations, sustainability, agile evolution, openness, interoperability and accordance with FAIR general principles as well as Open Data as general guidance.

In interviews and workshops **with the stakeholders** (e.g. interviews with the pilots in M1.1, workshops within the ESS community at NFDI4Earth conferences), we will analyse the scientific workflows to identify and abstract a set of component/service types and representative examples as well as their interactions, interfaces and data flows.

An initial **NFDI4Earth architecture** draft will be created and continuously improved and extended over time. Selected workflows will be included for illustration. The architecture will be a blueprint for the further development of services to be integrated into NFDI4Earth. We will monitor the consolidation activities and provide necessary guidance. The architecture with all its component/services types and their characteristics, the standards used, APIs etc. **will be mapped in the Knowledge Hub**.

Action 2: The NFDI4Earth service catalogue

The service catalogue will be the **collection of all existing and new services provided by NFDI4Earth**. It will include detailed **descriptions** of the services, their features, documentation, **FAIRness**, their usage and **relation to user communities, used standards,** relation and **interoperability** with other services, code/libraries to access/use the service and more. The information will be continuously collected in the Knowledge Hub and interlinked to other information there, e.g.: the services will be mapped on the architecture and on the NFDI4Earth relevant standards (M3.2) they use and training materials on their usage (M1.3) and answered support questions (M2.2).

In cooperation with M4.3, we will define the **catalogue structure** and its interactions in the **Knowledge Hub** and start gathering service information from the beginning of the project. We will also define **procedures to update the service** information in the Knowledge Hub in case of changes or for the inclusion of new services and to provide tools for the (semi)automatic gathering and ingestion of information.

Action 3: Extensions and Improvements based on the NFDI4Earth Architecture

Gap analyses will be performed for the **workflows from the pilots** contributed by M1.1, with the existing architecture from Action 1 as a background. In these analyses, feedback from M3.2 will be incorporated. As a consequence, **new features and necessary performance improvements** will be suggested to pilots and service providers. These may concern components of the NFDI4Earth architecture, but also the way pilots are using it (e.g., transition of file-based services to databases/data cubes).

Special emphasis in gap analysis will be put on the interoperability of data processing services from the provider side. Data processing capacities are currently developed at many sites, e.g., by combining cloud-computing approaches with High-Performance Computing (HPC). However, the seamless integration of individual services is not fully achieved yet, e.g., with respect to AAI, data accessibility, or usage models for granting compute time for processing. **Missing links in terms of interoperability** will be identified and discussed in **close cooperation with service providers**. Also, community data products are often not properly enriched with metadata, and thus lack FAIRness. This is in particular an issue for Big Data from e.g. HPC, experiments or satellites, which cannot be moved from a particular site. Here, care will be taken for **integrating such data properly with NFDI4Earth**. Semi-automated assessments of interoperability from M3.2 for existing services are collected and ingested in the Knowledge Hub. When proposing new services and tools, we will devise ways for realising them in close collaboration with M4.3. Implementation can then be planned e.g. as an experimental service at TUDD or as extensions of existing services within the architecture. In addition, requirements, obstacles and possible solutions for **embedding existing services** from other NFDIs or national and international initiatives (e.g. CODE-DE, Copernicus) in the architecture are analysed in close collaboration with M3.3 and M3.4.

Action 4: External interfaces to the architecture description

The action gives access to synthesis of the architecture components descriptions in (1) a machine-readable and (2) human-readable/visual representation usable by others. The outcome of this action will be a service that **provides external interfaces to the architecture description**, based on the standards and results of actions 1 to 3 and embedded in the NFDI4Earth web environment (OneStop4All, TA2) by M4.3. The **targeted users of the architecture description** are data stewards and the help desks of the NFDI4Earth User Support Network (M2.2), OneStop4All (M2.1), as well as users from other NFDI consortia. The service will provide the information needed for the architecture chapter of the Living Handbook (M3.3). The information about the NFDI4Earth architecture and services is made available (1) for human users in a structured, easily understandable and easily accessible form via a user frontend. An **extended search function** for architecture components and services including the level of standardization will be provided. For machine-interaction, an **interface for automated retrieval** will be provided.

This information can be used, for example, for analysis and to connect the components, and promote interoperability. In order to keep the content continuously up to date, interfaces will be provided for the user community and help desk with which information can be added to the Knowledge Hub (crowdsourcing), for example, changes to the components or emerging services. Filters can be used to show the latest changes to systems and services. Furthermore, the service catalogue (Action 2) will be published on OneStop4All (web portal). Together with M2.1 and M4.3,

we will work on a concept and technical realisation for its integration as an information source for NFDI4Earth users (e.g. looking for services with special features).

Deliverables and Milestones

Mile-stone	Delive-rable	Type*	Description	Due end of
-	D3.1.1	R	Concept and first version of architecture and service catalogue based on discussion with the various stakeholders	Q4 2022
M3.1.1	D3.1.2	R	Improved version of the architecture and service catalogue integrated in the Knowledge Hub	Q1 2024
-	D3.1.3 D3.1.4	R	Improved version of the architecture based on new pilots and other community services	Q4 2024 Q1 2026
M3.1.2 M3.1.3	-	S	Extended service catalogue	Q4 2024 Q1 2026
-	D3.1.6	R	First version of interface for presentation (concept and first implementation)	Q4 2023
M3.1.5 M3.1.6	-	S	Further improvement of interfaces for information presentation	Q4 2024 Q1 2026
-	D3.1.7	R	First results from gap analysis to improve NFDI4Earth services and workflows	Q3 2023
-	D3.1.8 D3.1.9	R	Annual reports on gap analyses and service/workflow improvements	Q4 2024 Q4 2025

* R-Report, S-Service

5.3.2 Measure 3.2 (M3.2): Common Standards for FAIR ESS Data

Contributors: SGN (lead), TIB, GEOMAR, UFZ

Links to M1.1, M2.1, M2.3, M2.4, M3.1, M3.3, M3.4, M4.3

Objectives

The overall objective of M3.2 is to ensure FAIR ESS data in NFDI4Earth by developing and supporting integration and interoperability of the NFDI4Earth architecture and the adoption of a standards framework for ESS data and metadata services that addresses all FAIR sub-principles in research communities.

As a central management tool and main driver for the service adoption and transformation process, we will establish an NFDI4Earth Label providing indicators for the interoperability of integrated data infrastructures to assess the FAIRness of a digital resource and for harmonization of the service (Action 1). It is awarded as part of a self-evaluation process according to proven standards, e.g., RISE (Research Infrastructure Self-Evaluation Framework) or the FAIR Metrics group (Wilkinson et al., 2018), and is supported by automated procedures wherever possible. We will evaluate and assess different domain specific (meta)data and identifier schemes as well as provenance, quality, credit and attribution methods as essential prerequisites towards FAIR data services (Action 2). Leveraging on existing community standards, we will assess technical standards (e.g., data formats, exchange/access standards and persistent identifiers) and organizational frameworks (e.g., curation, sustainability, preservation, FAIR metrics) and propose

(if not already present) open and therefore future-proof substitutes for community wide and cross community interoperation. Furthermore, we will design and prototype an interoperability layer for NFDI4Earth infrastructure based on RDA's FAIR Digital Object approach FAIR-DO (Lannom and Wittenburg). Building on their inherent capability to implement FAIR data (Strawn, 2019), we will adopt FAIR-DO and the associated ecosystem of inter alia data types, identifier schemes and registries for Earth System Science (Action 3).

Actions

Action 1: Develop the NFDI4Earth Label as key driver and indicator for interoperability of services

Action 1 will **provide a framework for FAIR attribution of research data** during the entire research life cycle. Approaches to automating FAIR assessment have for instance been demonstrated by ENVRI-FAIR with tools (see https://envri-fair.github.io/knowledge-base-ui/) that support the **identification of gaps in FAIR principles implementation** in research infrastructures and their data repositories. Label development will follow a growing shell model in an iterative process together with M3.1. The labelling process will be guided with corresponding guideline documentation and workshops. The NFDI4Earth consortium will also - together with M3.3 and M3.4 - connect with other support networks, e.g., EOSC, GO FAIR. The functionality of labelling is tested and evaluated in the process of **populating the Knowledge Hub** (M3.1) together with the growing number of community contributions. Furthermore, the planned ESS Pilots (M1.1) will be essential input and the testbed for the development process. Descriptions and design studies of all NFDI4Earth entities such as infrastructures, services, data sets, organisations, communities, APIs, etc. contribute as semi-structured data to the NFDI4Earth Knowledge Hub (M3.1).

Action 2: Assess technical standards (e.g., data formats, exchange/access standards and persistent identifiers) leveraging on existing community efforts.

This includes actions on **compiling information on current data standards, preparing a data modelling framework, best practice guidance for semantic enhancement, designing PID schemes and investigating interoperability** between new NFDI4Earth e-services. Measure 3.2 notably covers data with more complexly structured connected entities, relations and properties such as ontologies, RDF triples and interlinked digital objects. We will represent such graph type data in a **graph database component**, which will be a supplementary resource for the NFDI4Earth Knowledge Hub **presenting content in both human and machine-actionable form**, and acting as an interoperability layer to the One NFDI substantiated as Research Data Commons.

The graph type representation of our data will enable us to employ state-of-the-art data operations such as **predictions based on semantic similarity** (Smaili et al., 2019), **ontology alignment** (Karam et al., 2020) **and statistical relational learning** (Schlichtkrull et al.) on ESS data.

This evaluation of technical standards will integrate the results of a corresponding mapping process currently being implemented across the research field Earth and Environment of the Helmholtz Association in the context of the Helmholtz Metadata Collaboration (HMC) initiative, coordinated by the participant GEOMAR. We will (1) **compile the ESS resource data information types** (e.g., time series, geological maps, multi-scale kinematics, etc.), (2) **assess ESS (meta-)standards regarding matching and coverage of the identified resource information types**, if necessary extension of existing data models, (3) measure compliance with legislation and community standards (INSPIRE, GeoZG, W3C) and FAIR principles and (4) **enable persistent identification of data sets, data exchange protocols and services,** instruments (Stocker et al., 2020), etc.

TA3 will to a large extent take up recommendations from research community organisations such as the Research Data Alliance (RDA), e.g. FAIR assessment criteria from the FAIR Data Maturity Model WG (Research Data Alliance FAIR Data Maturity Model Working Group, 2020) and forward relevant output to the Living Handbook (M 3.4).

Action 3: Design intraoperative machine-actionable data operation layer based on FAIR Digital Object Model

In close cooperation with Measure 3.1 and TA4, we will develop **serialization formats** (leveraging on established approaches such as gRPC/protobuf; see https://developers.google.com/protocol-buffers) to advance the **integration of NFDI4Earth into hyperinfrastructures** such as the One NFDI (respectively the research data commons, M3.3) or the European Open Science Cloud (EOSC, cp. M3.4). Complementary to the FAIR policy framework targeting semantic data compatibility (addressed in Action 1 and Action 2), we will provide a **technical framework for interoperability of services based on FAIR Digital Objects** FAIR-DO (Schwardmann, 2020; Kahn and Wilensky, 2006).

FAIR-DO combine metadata and data into a machine-actionable container, where content is made referenceable and accessible by globally unique persistent identifiers (Smedt et al., 2020). To enable reusability of the data, metadata enhancements and context descriptions based on open standards (D3.2.1) are provided (Stocker et al., 2018), thus **enabling machine agents to infer proper operations and valid permissions** for actions such as update or creation of encapsulated data. Relations and linking both across and within FAIR-DO will be expressed based on classes and properties adopted from established standards, in particular the OBO Foundry ontologies, e.g. the Environment Ontology (ENVO), the Basic Formal Ontology (BFO) and the Information Artifact Ontology (IAO), **forming in this way an interconnected relation graph** (Action 2).

NFDI4Earth will align these efforts (in close cooperation with M3.4) with international developments in EOSC, GEOSS, ENVRI-RI's and new recommendations developed in the Research Data Alliance.

Deliverables and Milestones

Mile-stone	Delive-rable	Type*	Description	Due end of
-	D3.2.1	R	First Draft of resource information types (protocols, vocabularies, metadata schemes) in use in NFDI4Earth	Q2 2023
MS3.2.1	-	R	Documentation and first workshops for NFDI4Earth label are available, best practices from the pilots may be shown here	Q4 2023
-	D3.2.2	R,S	Report on operational capability for ESS Graph Database supplementing the NFDI4Earth Knowledge Hub	Q4 2024
MS3.2.2	-	R	Iterative assessment of the NFDI4Earth core infrastructure using the NFDI4Earth label, second workshop focussing on common standards and data provenance.	Q2 2025
-	D3.2.3	R,S	Integration of information types in inter-process communication model (FAIR-DO), integration into Knowledge Hub, report for Living Handbook	Q4 2025
MS3.2.3	-	R	Iterative assessment using the NFDI4Earth label including interoperability of the core architecture based on D3.2.5	Q2 2026

* R-Report, S-Service

5.3.3 Measure 3.3 (M3.3): NFDI Commons

Contributors: UNIF (Lead), UNITUE, UNIGOE, LIAG, GFZ

Links to M1.1, M1.4, M2.3, M3.1, M3.2, M3.4, M4.2, M4.3

Objectives

The overall goal of the M3.3 is to develop a common vision for the NFDI, to integrate and link NFDI4Earth into this and to facilitate cross-disciplinary FAIR data sharing. Therefore, it promotes the communication between NFDI4Earth TAs and other NFDI consortia in all research fields, establishes, and maintains relations between NFDI4Earth and relevant national RDM stakeholders with focus on ESS data.

The NFDI4Earth infrastructure, its architecture, tools, websites, interfaces, etc., has **counterparts within, as well as outside the NFDI**, e.g., DFG-funded CRCs, governmental institutions, libraries, etc. The **NFDI4Earth will become the connecting ESS node** in this network of German ESS RDM infrastructures. Hence, it is a key necessity to fully **embed and connect the NFDI4Earth architecture with all national ESS RDM stakeholders**. Furthermore, many RDM topics benefit from or even require **common solutions**, e.g., legal, privacy or ethical issues, managing long-tail data, PIDs or describing data provenance (see chapters 3 and 4).

We will **identify, communicate and coordinate RDM aspects that are independent of the scientific target group but relevant across other NFDI consortia to address cross-cutting issues and contribute to the NFDI Research Data Commons**, in accordance with the initial declarations on cross-cutting topics (Bierwirth et al.; Glöckner et al., 2019). We will monitor concepts of other stakeholders and communicate new developments to the appropriate TA/Measures. We will also develop transfer strategies for NFDI4Earth components that qualify as general concepts for the NFDI. (Action 1).

Similarly, we will liaise with **governmental institutions, research institutes, libraries, IT centres, and related networks** (e.g., DWD, UBA, BSH, BKG, LG GDI-DE, RfII, RDA Germany, the Alliance of science organisations, DINI), with interests in FAIR (geo)spatial research data.

Note, that academic societies are dealt with in TA4 (M4.2), as they are important ESS-stakeholders and potential drivers of cultural change, but typically do not engage in RDM on their own. We will **develop these cross-cutting tasks** by reviewing concepts and user demands, coordinate their adoption in NFDI4Earth TAs, and ensure cross-discipline interoperability with a focus on spatial data. Conversely, **NFDI4Earth concepts and methods will be communicated to other NFDI consortia and other German ESS RDM stakeholders** (Action 2).

The work with other NFDI and RDM stakeholders on common approaches and cross-cutting issues **requires an accessible resource** documenting the architecture of the NFDI4Earth, its progress, products and services. We want the inner **workings of the NFDI4Earth** – its architecture, structure, etc. – to be **easily understandable**, accessible, transparent, and inviting for everyone. We will **compile a Living (i.e. continuously updated) Handbook providing low-threshold access to all relevant information on NFDI4Earth** as well as for overall guidance and orientation on working with spatio-temporal research data. (Action 3).

Actions

Action 1: Connecting to NFDI cross-cutting topics and to other NFDI Consortia

We will **organize topical workshops and recurring meetings** to establish an interoperable RDM infrastructure in Germany and work towards a common NFDI Research Cloud. **Seed interest groups** comprise: (1) **NFDI4Agri, NFDI4BioDiversity, NFDI4Microbiota, DataPLANT NFDI, NFDI4Objects, KonsortSWD** in **collaborating on harmonised standards for spatio-temporal (meta-)data, APIs, curation methods**, and in general towards **Research Data Commons aligned with international approaches** (Barker et al., 2019; Bierwirth et al.; Ferrari et al., 2018; Glöckner et al., 2019; European Commission, 2020) (2) With **NFDI4Objects**, we share archaeometric methodology and cooperate on **interoperability and reusability for geoscientific lab data.** Similar ties exist with **NFDI4Agri and PUNCH4NFDI**. (3) With **KonsortSWD**, we will work on e.g., **legal, ethical, and privacy aspects, with a focus on spatial data**. (4) A number of ESS Pilots (M1.1) (e.g., PAMbase, GeoRoc-MetBase Tools) use long-tail data. These examples, in combination with the archiving strategies developed in M2.4, will be used to derive common, **cross-NFDI management strategies for long-tail data** (Horstmann et al., 2017).

We will **engage in and/or initiate an NFDI working group to unify the structure and concept of key NFDI resources and access points.** We already identified the Knowledge Hub and the Living Handbook (within TA3), the User Support Network (TA2), and the Academy (M1.4) as candidates which are transferable into general concepts and products of the entire NFDI. M1.4 has already initiated inter-consortia collaborations between **NFDI4Biodiversity, NFDI4Health, NFDI4Earth, the Federal State of Bremen and the University Bremen Research Alliance (UBRA) to establish a concept that can serve as a template for a NFDI Academy**. We will

follow a similar approach for the Knowledge Hub, the Living Handbook and the User Support Network.

For NFDI4Earth, we envisage the collaborative use of data analysis services, making, for instance, extensive use of AI techniques. These services will require access to storage and computing resources of High Performance Computing centres. The current application practice for HPC resources (regular proposals for a predefined computing time, yearly reports etc.) is not flexible enough for distributed and open consortia services. Hence, talks between German HPC centres (e.g. via the Gauß Alliance), the NFDI directorate and other NFDI consortia will be initiated to define a suitable usage model for the sustainable operation of data analysis services originating from NFDI (in cooperation with M4.1).

Action 2: Connecting to other ESS-related RDM stakeholders

Actions will be implemented with **workshops and recurring meetings**. Seed interest groups comprise – among others: (1) Joint working groups to collaborate on protocols and code lists to enable cross-disciplinary sharing of **3D object data** (e. g. buildings, subsurface infrastructure, wells, geologic bodies etc.) building on OGC standards. Here we will collaborate with M2.3 as facilitator for governmental 3D geodata. (2) The Helmholtz initiative **DataHub** (see chapter 3.1) aims at common research data management that will enhance data (re-)usability beyond the Helmholtz research area and foster exchange with NFDI and international Earth Science data infrastructure initiatives. (3) The Berlin-based **CRC 1404 FONDA** will research workflow infrastructures for large scale scientific data analysis that allow reduction of workflow development time, an important issue for NFDI4Earth and other NDFI consortia. FONDA will support NFDI4Earth with its workflow technology, while NFDI4Earth will add specific workflows and data analysis to FONDA's use case repositories.

Action 3: Providing and editing the Living Handbook

The Living Handbook will describe the NFDI4Earth **architecture, its structure, innovations,** etc. It will be realised in close cooperation with M3.1, M3.2, and M4.2 (and published through M4.3) Entries will cross-link with the Knowledge Hub for in-depth technical information and formal descriptions. The Living Handbook will **include best (and failed) practices** from NFDI4Earth and other RDM stakeholders. Both the Knowledge Hub and the Living Handbook will be shared with other stakeholders and will serve as a basis to **identify common areas** for a future deeper connection, embedding and harmonisation/sharing of services, workflows, metadata, standards, vocabularies, etc. We will ensure these are of coherent length, balanced in terms of detail and quality, and have respective links to the Knowledge Hub. We will actively work with all TAs and measures to ensure homogenous contents and regular updates. We will further contact other NFDI consortia to exchange best practices that will also be included in the handbook. Chapters of this Living Handbook will be, for example: 'NFDI4Earth architecture', 'Geodata Science', 'FAIR',

‘Research Data Commons’, ‘Managing Long-Tail Data’, ‘Best (and failed) practices (incl. from other NFDIs)’.

Deliverables and Milestones

Mile-stone	Delive-rable	Type*	Description	Due end of
-	D3.3.1 D3.3.2 D3.3.3	S,R	Establish/participate in an NFDI-network to identify and continually work, update and expand on cross-cutting RDM topics through regular workshops. The results are issued in about annual whitepapers.	Q4 2023 Q4 2024 Q4 2025
-	D3.3.4 D3.3.5 D3.3.6	S,R	Establish relations with other national stakeholders in the NFDI4Earth context and work towards a common infrastructure for RDM in Germany (Research Data Commons) through regular workshops, and the results issued in about annually updated and expanded whitepapers.	Q4 2023 Q4 2024 Q4 2025
MS3.3.1	-	S,R	Define recommendations and guidelines on how to implement interoperable architecture elements for cross-cutting RDM topics through topic-related workshops. The results are updated, expanded and issued in about annual whitepapers.	Q4 2023 Q4 2024 Q4 2025
MS3.3.2	-	S	Provide and edit the Living Handbook with regular, ~quarterly updates)	Q3 2023
-	D3.3.7	R	In collaboration with other NFDIs develop general concepts and applications for the harmonization of metadata schemes and formats across domains in the NFDI4Earth context.	Q2 2025
-	D3.3.8	R	In collaboration with other NFDIs develop a whitebook for the provision of georeferenced object data.	Q4 2025

* R-Report, S-Service

5.3.4 Measure 3.4 (M3.4): International Networking & Embedding

Contributors: GFZ (Lead), TUDD, DKRZ

Links to M1.1. M1.4, M2.1, M2.2, M2.5, M3.1, M3.2, M3.3

Objectives

The overall goal is to ensure and promote international awareness of NFDI4Earth and to support the development of novel avenues of RDM by connecting and actively participating in international initiatives. M3.4 will in particular crosslink with M3.2 und M3.3 on the international embedding and development of regulations and standards.

On the one hand, this will lead to an improved level of information for users and ESS service providers in Germany about the opportunities available, and on the other , it will open up the possibility of developing structures such as the EOSC in a targeted manner.

NFDI4Earth seeks a high level of visibility for NFDI4Earth and its members in international ESS and RDM initiatives and will actively participate in shaping future RDM worldwide for researchers. We will **contribute to improved harmonisation of ESS-related RDM and standards across Europe and the world as a base for scientific work through integration of FAIR standards**.

The activities are organized along different target groups (ESS community with RDM focus, ESS community in general, RDM community beyond NFDI).

Actions

Action 1: International ESS Community with RDM focus

NFDI4Earth is, due to its members, actively embedded in international initiatives such as COPERNICUS, ENES, ENVRI-FAIR, EPOS, FluxNet, eLTER, ORFEUS/EIDA, ICOS. Thus,

NFDI4Earth does not only cover all relevant ESS aspects on an international scale, it also benefits from these networks and engagements in terms of the ongoing establishment of accepted standards and the implementation of international directives (such as INSPIRE). Based on the existing engagement of participants in NFDI4Earth, the measure will provide and use capacities and services for acquiring, sharing and processing ESS data and ensure the development of novel avenues of RDM by connecting and actively participating in international initiatives. **Participation in international standardization bodies will enable capitalisation from international development and safeguard the overall NFDI needs in terms of RDM development.** The ESS Pilot Linking Environmental Data into European Scale RIs (P7) will be closely integrated with this action item.

Action 2: International ESS Communities/Associations

NFDI4Earth participants are **active members of international ESS communities and associations** (e.g. AGU, EGU, IUGG, IPCC, Geo.8 and many more). These communities are all dealing with RDM on a broad scale, from policies to services. We will provide an **overview of all ESS-relevant networks and initiatives** and their link to NFDI4Earth and will actively use existing collaborations, linkages, memberships to **connect and report/spread** NFDI4Earth ideas. Conversely, we will **foster the integration of NFDI4Earth members** to RDM-related partners in a broad range.

Action 3: International RDM associations and communities beyond ESS scope

NFDI4Earth will become an **integral part of the international RDM community** and ensure the development of novel avenues of RDM by **connecting and actively participating** in international initiatives. Participation in **international standardization bodies** will enable capitalisation from international development and safeguard the overall NFDI needs in terms of RDM development. This includes engagement in bodies such as RDA, Codata, WDS, FORCE11, W3C, OGC, OpenAIRE and **strong integration in and with general data-related services** such as DataCite, ePIC, re3data or ORCID.

The possibilities that the **EOSC** will offer, such as **using services** or, from the provider's point of view, providing/offering services, have so far only been perceived sporadically by Earth system researchers and infrastructures or have been poorly coordinated. In cooperation with other NFDI consortia, NFDI4Earth will help to lessen/reduce this deficit. For example, platforms are being established and moderated for service providers, which offer an exchange of information about opportunities for participating in the EOSC. In particular, NFDI4Earth can make a **strong coordinated contribution to shape the EOSC**.

Deliverables and Milestones

Mile-stone	Delive-rable	Type*	Description	Due end of
-	D3.4.1	S	A comprehensive overview of all ESS relevant RDM networks and initiatives in which NFDI4Earth participants are active will be established and available (Knowledge Hub and OneStop)	Q4 2022
-	D3.4.2	S	Key stake holders and contact persons are included	Q3 2023
-	D3.4.3	S	Establish 'ambassadors' for the different networks as guides for interest and newcomers	Q4 2023
M3.4.1	-	S	Re3data: all NFDI4Earth Data repos (M3.2 Label) are included and up-to-date	
M3.4.2	-		NFDI4Earth is internationally established as a key adopter of PIDs DOI, ORCID, IGSN. PIDs are established as common standards (M3.2)	Q4 2024
-	D3.4.4 D3.4.5 D3.4.6 D3.4.7	S,R	Foster international participation through regular workshops and organising NFDI4Earth contributions to international conferences, workshops, etc. Provide an annual report about international Activities	Q3 2023
M3.4.3		S	Establish a newsletter on international RDM related news, starting Q4 2022	Q3 2024

* R-Report, S-Service

5.4 Task Area 4: 2Coordinate

Lead: TUDD

Overview

TA4 will **establish permanent and transparent governance and an operational structure for NFDI4Earth representing and integrating the various disciplines of ESS**. It will manage a sustainable and future-oriented organisational, operational and governance model for NFDI4Earth, being fully linked and embedded in the NFDI as a whole. The NFDI4Earth Coordination Office will administer, support, stimulate and motivate the NFDI4Earth organisational bodies and the NFDI4Earth community in the day-to-day running of NFDI4Earth. It will act as *the* NFDI4Earth point of contact for current and prospective members, the NFDI and other NFDI consortia, the DFG, the scientific community, policy and the general public. All coordination activities are organised in Measure M4.1.

NFDI4Earth targets a cultural change in ESS, which we will stimulate and monitor in Measure M4.2. Such cultural change will guide the NFDI4Earth community towards responsible research data management as well as open and FAIR sharing of the communities' research data and scientific software. **Designing, establishing and carefully progressing the *NFDI4Earth FAIRness and Openness Commitment* as a commonly agreed and accepted joint vision** is a key strategic element of M4.2. Being culture-shaping bodies, academic associations and community networks are natural partners to be firmly integrated in the work of M4.2. Besides monitoring, analysing and communicating the different facets of the cultural change will be constitutive for M4.2.

With Measure M4.3 we **manage and support the federated implementation of common NFDI4Earth services**. The Measure will **coordinate the implementation** of OneStop4All (TA2),

the NFDI4Earth Knowledge Hub and Living Handbook (TA2/3), the ticket system for the user support (TA2), and the virtual environment for the EduHub (TA1). M4.3 will assist and accompany common quality-assured open source developments in NFDI4Earth with a technical support infrastructure (e.g. for continuous integration). Additionally, M4.3 will provide a virtual research environment for the development of the ESS pilots and incubators. M4.3 will also provide the virtual environment to operate central NFDI4Earth services.

5.4.1 Measure 4.1 (M 4.1): Coordination, Collaborative and Sustainable Governance of NFDI4Earth

Contributors: TUDD (lead)

Links to all other measures

Objectives

With M4.1 we will initiate, develop and **secure the path from a project consortium to a long-lasting and sustainable NFDI4Earth community**. We will detail and set up governance mechanisms and define the interplay within and amongst the NFDI4Earth organisational bodies, the NFDI4Earth members and the NFDI – also considering that NFDI4Earth will function as an open and flexible initiative.

In M4.1 we will **establish the NFDI4Earth Coordination Office** (Action 1) at TUDD to administer the NFDI4Earth operation, to support the common governance and to control progress of the project, the reporting, the consolidation and quality assurance of NFDI4Earth outputs such as recommendations, reports and white papers. In developing and applying clear, DFG-compliant funding rules and flexible funding schemes, we will enable the framework for active involvement of the participants in all NFDI4Earth Task Areas. Furthermore, we develop and **implement the communication strategy and support the consortium's activities on community building** (Action 2).

We will **develop the (future) operation model of NFDI4Earth** – depending on the legal and organisational framework of the overall NFDI – and establish the NFDI4Earth organisational structure for the NFDI4Earth consortium bodies as well as the interfaces to the NFDI bodies (Action 3). It will be pivotal to ensure an NFDI4Earth organisational structure and governance that can be easily embedded into the overall NFDI and interact with the NFDI Consortia Assembly, the NFDI Scientific Senate and the NFDI Directorate. M4.1 will coordinate and support in particular M3.3 to moderate the cross-links with other NFDI consortia on the NFDI Commons, M3.4 in the international embedding and M2.2 on joint NFDI capacity building.

Actions

Action 1 Establishing the NFDI4Earth Coordination Office

Being a core element of NFDI4Earth, the Coordination Office (at TUDD) manages the initiative specifically by **coordinating and controlling the finances** within the NFDI4Earth consortium

– implementing DFG rules as well as the decision procedures in NFDI4Earth for allocating centrally managed funds to the participants for specific developments of NFDI4Earth. It also acts as the back office of the steering group and the advisory board.

The office monitors and guides project progress, technical and financial reporting and supports the TAs and Co-Applicants, respectively. It consolidates the project results and provides quality assurance of NFDI4Earth outputs such recommendations, publications and white papers. Annual reports will address the overall development of NFDI4Earth and will in particular feature success measures (see chapter 2.2)

The NFDI4Earth Coordination Office will be hosted at TUDD and is expected to become part of the new Lehmann Centre at TUDD. **TUDD has committed to support the coordination office** by not only providing office space, but also providing guarantees for permanent staff positions. The Saxon State Ministry for Higher Education, Research and the Arts (SMWK) has also committed support for the NFDI4Earth work at TUDD. This support guarantees the potential for long-term operation of the NFDI4Earth Coordination Office. A managing director heads the NFDI4Earth Coordination Office and supports the Spokesperson as well as the Co-Spokespersons of the initiative.

Action 2: Communication, Outreach and support of Community Building

The Coordination Office acts as the **NFDI4Earth point-of-contact for internal and external communication**. This action realizes and manages communication and exchange within the consortium (by conferences, workshops, meetings, virtual collaboration platforms, etc.) to harmonize the project results and outcomes as well as with the (scientific) community, policy and the general public. We plan the first conference for November 2020 (see www.nfdi4earth.de). We will put a special focus on support for NFDI4Earth Interest Groups, as an important pillar of community building. In terms of outreach activities, we will provide several information channels (website, (social-) media, presentations, stands at conferences such as EGU General Assemblies and IGU congresses and fairs, etc.) to inform and engage with students, researchers, information infrastructure and software providers and the wider public about the status, progress and collaboration opportunities of NFDI4Earth. Communication and outreach will be embedded in a communication strategy which will be established immediately following the start of the project.

Action 3: Towards a sustainable NFDI4Earth Governance and Operation Model in the NFDI

In the preparation phase for this proposal, the NFDI4Earth consortium drafted and agreed upon an organizational structure (see chapter 3.4). In accordance with DFG rules, this structure will be further detailed and manifested in the **NFDI4Earth consortium agreement**. These bylaws include a clear definition of members' roles, rights and duties as well as responsibilities. The agreement must be acknowledged by the applicant, co-applicants and participants.

NFDI4Earth will become an active part of the (future) NFDI Verein, the umbrella structure for all NFDI consortia (see chapters 3.2, 3.4 and 3.5). We will contribute towards the development of this overall structure and integrate NFDI4Earth into it.

This will be followed by joint development **towards an elaborated model and agreements to ensure the operation of NFDI and NFDI4Earth in the long term**. This action will elaborate from the NFDI4Earth perspective – and based on a moderated dialogue within the NFDI4Earth community – on ways and instruments towards NFDI4Earth sustainability. This links tightly with M3.3 and supports communication with and within the NFDI and the provision of joint work plans to achieve the NFDI commons (see chapter 3.2). We will also address possible synergies from joining and consolidating RDM offerings. Furthermore, we will recommend possible adaptations of funding regimes for universities and research institutions running infrastructure services for NFDI4Earth and the NFDI to secure their long-term operation. Starting with a white paper, we will develop towards a commonly agreed model to operate and govern NFDI4Earth within NFDI. In close cooperation with M1.3, M2.1, M3.1, M3.3 and M4.1, we will in particular address responsibilities and agreements for the long-term operation of NFDI4Earth core components.

Deliverables and Milestones

Mile-stone	Delive-rable	Type*	Description	Due end of
MS4.1.1	-	S	NFDI4Earth Coordination Office in Operation	Q4 2021
-	D4.1.1	R	NFDI4Earth Consortium Agreement (according to DFG Rules)	Q1 2022
-	D4.1.2 D4.1.3	R	NFDI4Earth Communication Strategy (Revision in Q3 2024)	Q2 2022 Q3 2024
MS4.1.2	-	S	Advisory Board established	Q3 2022
-	D4.1.4 D4.1.5 D4.1.6 D4.1.7	R	NFDI4Earth Annual Report including impact assessment	Q4 2022 Q4 2023 Q4 2024 Q4 2025
-	D4.1.8	R	White Paper on the NFDI4Earth/NFDI long term Operation Model	Q3 2024
MS4.1.3	-	S	Coordination and submission of proposal for 2nd funding phase	Q3 2025
-	D4.1.9	R	Agreed NFDI4Earth long term operation (linked into NFDI)	Q1 2026
-	D4.1.10	R,S	Business plan for long term operation of NFDI4Earth Coordination Office	Q1 2026

* R-Report, S-Service

5.4.2 Measure 4.2 (M 4.2): Towards a Cultural Change in ESS Research Data Management

Contributors: TUDD (lead), FUB, GFZ

Links to M1.1, M1.3, M1.4, M2.4, M3.3, M3.4

Objectives

Values, behaviour patterns and habits are difficult to change, because people love to keep what they are used to. This makes cultural change very challenging. The availability of user-friendly technical infrastructure, meaningful services as well as up-to-date educational opportunities are all important elements in NFDI4Earth for advancing cultural change in the ESS research community towards FAIR and Open research data management. However, for more comprehensive success, other specifically non-technical aspects need to be addressed.

One key to motivating NFDI4Earth researchers and supporting staff for cultural change is to secure that the leadership of NFDI4Earth-related research institutions and research organisations strongly support the NFDI4Earth spirit and actions at stake. **To achieve this, an overarching statement, the *NFDI4Earth FAIRness and Openness Commitment* will be developed and promoted as the common vision of NFDI4Earth, to be endorsed by the institutions and organisations involved and related to NFDI4Earth.** Moreover, this common agreement on the basic principles of FAIR and Open research data management will create a sense of identity for all actors in NFDI4Earth and will build confidence and trust in this national enterprise. Additionally, this agreement may become a useful instrument for research funders as it provides commonly accepted and manifested documentation that the signatory institutions adhere to best practices in ESS research data management (Action 1).

Promoting cultural change in the ESS must involve ESS academic societies. Traditionally, they are key players in shaping scientific and ethical standards within the scientific discipline they represent. ESS academic societies will be targeted in order to support them in adopting FAIR and Open practices (Action 2). Progress in cultural change towards FAIR and Open practices is also dependent on the broader research framework, on issues and incentives beyond individual disciplines. Therefore, it is crucial that **NFDI4Earth is active in working with future NFDI-cross-cutting topic groups** (Action 3). **Reporting and exchanging experiences on cultural change towards FAIR and Open practices** will demonstrate benefits of this change, but will also identify dead ends or critical paths. The work in M4.2 includes engaging with community networks as for example ABC/J, Geo.X, DAM and Water Science Alliance to address FAIR and Open practices in their work and integrate their education/dissemination capacities.

Action 2 will be carried out in close cooperation with the Specialized Information Service Geosciences (Fachinformationsdienst Geowissenschaften, FID GEO), which is located at GFZ Potsdam and Göttingen University. The FID GEO, in turn, closely collaborates with the associated FID Montan and FID Karten and thus ensures broad coverage.

Actions

Action 1: NFDI4Earth FAIRness and Openness Commitment

M4.2 will moderate and coordinate the process towards the common vision of the NFDI4Earth, the *NFDI4Earth FAIRness and Openness Commitment* and will develop and detail these general principles. All current and future NFDI4Earth members as well as all researchers in ESS will be encouraged to acknowledge and endorse this commitment.

The core conception of the common vision is to firmly associate FAIR general principles with the principles of openness, i.e. of sharing open data and open source software, and to implement a no-fee policy for the services offered by NFD4Earth to the community. This concept is a vital part of the NFDI4Earth operating model (see chapter 3.5). The commitment will have a clear definition of all included criteria and their applicability. Several issues to be addressed in the statement are

detailed in chapter 4.3, where the implementation of FAIR principles as part of the NFDI4Earth RDM Strategy is described. Legal aspects, for example, will be explored in close cooperation with M3.3 and also with other NFDI consortia, since they are already identified as major topic cross-cutting all NFDI consortia (see the Leipzig-Berlin Declaration, Bierwirth et al.). Embargo periods will be addressed as well, as they are current national and international practice to guarantee the 'first publishing right'. We will crosslink with M3.3 and M3.4 on related national and international developments and agreements and with M2.4 on linking governmental data to NFDI4Earth. The NFDI4Earth commitment will also address adherence to quality criteria for data and software publications, pursuing a careful balance between the necessity of quality assurance versus practicability and adequate allocation of time to conduct quality checks and implement curation. Practically, this will be supported with procedures and tools by M2.1, M2.2 and M4.3.

To strive for the best balance between FAIRness/openness principles and the feasibility for achieving these goals is understood as the major prerequisite for broad acceptance of the NFDI4Earth FAIRness and Openness Commitment in the ESS community. Therefore, we will establish (online) community-consultations to allow an open commenting process for the draft commitment and update the commitment in the course of the NFDI4Earth's lifetime.

Action 2 Engaging with academic societies and publishers

Academic societies are important stakeholders in the science system and drivers of cultural change in that they develop a shared understanding among their members and by defining binding standards, which they establish within their specialist communities. Currently, the issue of FAIR and Open data is, with very few exceptions, absent in German geoscience societies´ mission statements, journal policies, and activities (see e.g. Hübner, 2020). Action 2 stimulates and supports German geoscience academic societies in promoting and implementing FAIR and Open data principles as an integral part of their mission and their activities.

Action 2 will guide the boards of the societies, their members as well as representatives for scientific publishing towards FAIR and Open data principles and support the societies´ actions in this regard. This may include amending their mission statements and signing the NFDI4Earth FAIRness and Openness Commitment, the COPDESS statement of commitment, engagement in the development of community standards, infrastructure, tools and services to enable open and FAIR data practices, regular education and outreach to their communities regarding these principles and practices, and/or promoting open and FAIR data activities as important criteria for awards and honours.

The vast majority of geoscience academic societies publishes journals. Action 2 puts special emphasis on improving FAIRness and openness in research data sharing of the societies' journals, including policies and author instructions. We will guide societies in particular in evaluating policies of their journals and will provide support for the implementation and/or

improvement of data policies, to meet current FAIR standards and promote openness in data publication.

Action 3 Rewarding systems: status and recommendations

Cultural change towards FAIR and Open practices in the ESS will additionally require changing the incentives for researchers on a broader scale, beyond disciplinary borders (e.g. acknowledging FAIR and Open data activities in publication records, funding schemes, professional careers, rewarding mentoring and encouraging others in developing their FAIR and Open data capabilities). We will map incentive mechanisms already in place at NFDI4Earth co-applicant and participant institutions, at relevant funders, at geoscience academic societies as well as at related communities like the Water Science Alliance, DAM or the Meteorological Society. We will closely cooperate with M3.3 and support common NFDI activities, like future NFDI cross-cutting topic groups, addressing these aspects.

Action 4 Monitoring

Monitoring and analysing cultural change towards FAIR and Open practices within NFDI4Earth and beyond will demonstrate the benefits of this change, but will also identify dead ends or critical paths. Here, we will closely collaborate especially with TA 1 to gather experience and success stories from pilots, incubator Projects, and the education and academy activities, the ESS community networks, and with future NFDI cross-cutting topic groups. Also, the impact of other non-technical activities such as the NFDI4Earth FAIRness and Openness Commitment (M4.2), the NFDI4Earth Label (M3.2) and implementation of and compliance with publisher data polices (M4.2) will be evaluated. The report will be part of the NFDI4Earth‘s Living Handbook (see TA 3, M3.3).

Deliverables and Milestones

Mile-stone	Delive-rable	Type*	Description	Due end of
MS4.2.1	-	S	Community consultations on draft NFDI4Earth FAIRness and Openness Commitment; Consultation for 2nd Version in 2025	Q4 2022 Q2 2026
-	D4.2.1, D4.2.2	R	NFDI4Earth FAIRness and Openness Commitment / 2nd Version in Q2/2026	Q2 2023 Q2 2026
-	D4.2.3, D4.2.4	R	Report FAIR and Open RDM in ESS academic Societies and Journals / 2nd Version in Q3/2026	Q4 2022 Q3 2026
-	D4.2.5	R	NFDI4Earth White Paper: Implementing Science Rewarding Systems for FAIR and Open RDM in ESS	Q4 2024
-	D4.2.6	R	Report and Scientific Publications on Cultural change towards FAIR and Open practices in the ESS	Q3 2026

* R-Report, S-Service

5.4.3 Measure 4.3 (M 4.3): Central Support Services for the federated NFDI4Earth Community

Contributors: TUDD (lead), SGN, HSBO, KIT

Links to M1.1, M1.2, M1.3, M2.1, M2.2, M3.1, M3.2

Objectives

The focus of M4.3 is on the technical implementation and provision of the services developed in TA1, TA2 and TA3 for the NFDI4Earth community. The content of the services (e.g. the knowledge in the Knowledge Hub) is provided by other measures. This requires close cooperation with a number of other measures regarding the conceptualization, development and operation which will be detailed in the description of the actions below. NFDI4Earth will run four backbone services to support the NFDI4Earth community and its tasks:

- OneStop4All as the central access point to the NFDI4Earth virtual research environment, i.e. all content and services provided (see M2.1)
- The Knowledge Hub to collect, structure, analyse and retrieve the collective knowledge of NFDI4Earth as well as its public representation, the Living Handbook. The hub includes information about all NFDI4Earth services, NFDI4Earth architecture, standards, research commons, education etc. (see M1.3, M2.4, M3.1, M3.2, M3.3)
- Virtual Research Environments for the development, piloting and operation of NFDI4Earth software and services
- A ticket system for the User Support Network (see M2.2)

Actions

Action 1: Developing the Technical Base of the OneStop4All

OneStop4All will function as the central access point to the NFDI4Earth virtual research environment, i.e. all content and services provided by NFDI4Earth (see other TAs). It will link all distributed community portals, data and software repositories, computing and other services provided by NFDI4Earth members and the ESS community, user support, the available collaboration environments (TA2), education and learning materials (EduHub, TA1), access to the content of the Knowledge Hub and the Living Handbook (TA3), documentation and contact information. **OneStop4All will not only be the central entry point for all users from different ESS domains but also for other – national and international – scientists, who require spatial reference data or visualizations**, for instance. It will also be used to publish calls for pilots and incubators (see TA1), the calls for interest groups, and to promote their results. It will act as an information hub, e.g. to spread information about conferences, workshops, webinars and other events organized under the umbrella of NFDI4Earth but also to provide information about other relevant events at the national and international level. We envision OneStop4All as

the virtual NFDI4Earth point of contact, which will guide users according to their specific ESS RDM and Data Science requirements and capabilities.

In an initial step, we will identify the requirements and design the concept for OneStop4All. This will be done in a close cooperation with the other TAs that contribute to the structure and content of OneStop4All. The major partners are: M2.1 for the overall concept of the interaction of the users with OneStop4All, M2.2 for integration of (virtual) user support, M3.1 for integration of the Knowledge Hub, M1.1 for pilot integration, M1.3 for education and training material and formats (EduHub). **The very different user requirements must be reflected by the structure and functionalities of OneStop4All. Therefore, a user-oriented guidance concept will be developed aiming at an intuitive and effective use (look-and-feel, navigation, seamless user experience …) and including procedures and roles in the portal**.

Based on the concept and the requirements, we will evaluate appropriate software solutions (such as HubZero, Vaadin, Django) and implement and operate OneStop4All. OneStop4All won't be a monolithic block, but a central entry point organizing the requests to the various microsites representing different NFDI4Earth components. Single sign-on authentication and authorization will be included based on DFN-AII and other user required solutions.

Action 2: Realizing the NFDI4Earth Knowledge Hub and the Living Handbook

The **NFDI4Earth Knowledge Hub will become the central instrument to collect and structure the information and knowledge of the distributed competence network NFDI4Earth**. In this measure, we will develop and operate the technical base of the Knowledge Hub and its public interface/representation, the Living Handbook. The content of the Knowledge Hub will be provided by the cooperating measures:

- M3.1 – knowledge about the components, services interactions, workflows of the NFDI4Earth architecture; a service catalogue with information about all NFDI4Earth services such as descriptions, standards used, specific user communities and their services, etc.
- M3.2 – knowledge about standards relevant for the ESS community and NFDI4Earth
- M2.2 – knowledge for user support such as answers to questions, cross-references to best suited services for specific requests, documentation and more
- M1.3 – knowledge about education and training like curricula, teaching materials, teaching forms

On the other hand, knowledge must be linked and retrieved for different purposes: to provide knowledge directly to users via the web portal, to find information, to answer questions in user support, to document NFDI4Earth findings and developments such as the architecture or teaching strategies, to track internal documentation and procedures and more.

This shows that the content-related as well as technical development of the Knowledge Hub must involve close cooperation amongst the technical development team of this measure with the

content suppliers of other measures. In addition, we will cooperate with experts on knowledge management and retrieval such as ScaDS.AI Dresden/Leipzig and DKRZ data management.

At first, we will design a concept for the Knowledge Hub in a common effort of all involved parties based on the requirements of the various usage scenarios. As we expect this conceptualization phase and the first implementation to last at least a year, we will start in parallel with a simple solution (e.g. a CMS with a database or file-based knowledge management) to support initial OneStop4All operation. Based on the concept and its requirements, we will evaluate different solutions (like ontology and graph based concepts) and software tools. The actual implementation and further development of the knowledge base will be continuously evaluated and tested intensively in cooperation with different users and their usage scenarios. The results of the evaluation, changes needed and new requirements resulting from this step will be fed back into system design and implementation.

Action 3: Provision of Virtualization Services for Development and Operation

The development of high-quality services and software is an integral part and major goal of NFDI4Earth (see also action 5). We will **provide an environment for service developers (such as the pilots in M1.1, advancing tools in M2.5, software quality procedures in M4.4) to support the development, testing and evaluation of their products**, e.g. regarding interoperability, standard conformity or for data quality procedures. This environment will be a virtualization infrastructure and will be hosted at TUDD-ZIH. A general procedure for an access model will be defined and implemented with the AAI infrastructure as well as a user model for use of the resources. We will define procedures and workflows for facilitated use of the infrastructure and provide appropriate tools. An example is the TUDD-ZIH Research Cloud, which allows users a web-based setup of a virtual environment. The virtualization infrastructure will also be used to operate our central services such as the Knowledge Hub and NFDI4Earth's OneStop4All. The action also supports the development of hosting strategies for services, operated by NFDI4Earth participants

Action 4: Ticketing System for User Support

M2.2 will establish the multi-level user support structure for NFDI4Earth. In this action, we will **provide them with an infrastructure supporting the distributed nature of the User Support Network**. The central part will be a ticket system to manage user requests, their processing and responses. We will use the ticket system of TUDD-ZIH (OTRS), which has been in use at TUDD for 14 years. We will extend it with the NFDI4Earth support processes (e.g. for incident models for incident management, ticket classification, answer templates) and provide training (tutorials, documentation) for NFDI4Earth support staff. Privacy issues of the clients according to General Data Protection Regulation will be taken into account.

Action 5: Support and Coordination of Software Developments

This **cross-cutting action coordinates the implementation of procedures and tools to ensure the quality and sustainability of the software developed within NFDI4Earth**. The action will support software developers in producing open source software that can be used or extended by other scientists or other interested persons. This support includes means and best practices for software management (e.g. continuous integration and task management for development, automatic unit and functionality tests, community building, licensing, software publishing). In the GI domain, but also mainstream IT, there are numerous open source frameworks available that are relevant for the software developments of the NFDI4Earth community (see chapter 4). The action will provide an overview of the relevant frameworks and develop a cooperation strategy with existing GI open source software frameworks and platforms. Another goal of this action is coordination of the organizational, methodological and technical aspects of software development to ensure a high level of software quality. This requires close cooperation with OneStop4All (M2.1), Knowledge Hub (M3.1, M3.2), the Pilots (M2.1) on the general Open Source Strategy in NFDI (M3.3) and the international networks on ESS related to open source (M3.4).

Action 6: Providing a technical Platform for the NFDI4Earth EduHub

The **technical platform serving the management and provision of education and learning materials developed and provided within the EduHub** (M1.3) will be implemented in collaboration with TA1 and operated at TUDD. The platform gives open access to all the of the education materials provided by NFDI4Earth. Further functionalities are conceptualized in M1.3.

Deliverables and Milestones

Mile-stone	Delive-rable	Type*	Description	Due end of
MS4.3.1	-	S	Virtualization Sandbox Services available	Q4 2022
-	D4.3.1	R	OneStop4All - Requirements and Concept Design	Q4 2022
MS4.3.2	-	S	Ticketing System - 1st Version online	Q3 2022
MS4.3.3	-	S	OneStop4All - 1st Version with basic information functions online	Q4 2022
-	D4.3.2	R	NFDI4Earth Knowledge Hub - Concept	Q3 2022
MS4.3.4 MS4.3.5	-	S	Technical Platform NFDI4Earth EduHub - 1st Version online; 2nd Version online	Q4 2022 Q4 2023
-	D4.3.3	R	Cooperation strategy with existing GI Open Source Software Frameworks and Platforms	Q4 2022
-	D4.3.4 D4.3.5	R	OneStop4All – Implementation Strategy, Evaluation and Selection of Software Solutions / Revision in Q4/2025	Q2 2023 Q4 2025
MS4.3.6 MS4.3.7	-	S	NFDI4Earth Knowledge Hub - 1st Version as basic Implementation; 2nd Version allowing for continuous enriching and updating	Q4 2023 Q4 2024
-	D4.3.6	R	White Paper: Best Practices organizational, methodological and technical aspects of GI Open Source Software Development, incl. recommendations for Open Source Frameworks	Q1 2024
MS4.3.8 MS4.3.9	-	S	OneStop4All – Iterations: 2nd Version Implementing D4.3.4.; 3rd Version Implementing D 4.3.5	Q2 2024 Q4 2025

* R-Report, S-Service

5.5 Bibliography and list of references

References (bold indicates from members of the consortium)

AtMoDat, 2019. Atmosphären-Modelldaten: Datenqualität, Kurationskriterien und DOI-Branding (AtMoDat). https://www.atmodat.de/ .

Barker, M., Wilkinson, R., Treloar, A., 2019. The Australian Research Data Commons. CODATA,vol. 18.

Bernard, L., Mäs, S., Müller, M., Henzen, C., Brauner, J., 2014. Scientific geodata infrastructures: challenges, approaches and directions. International Journal of Digital Earth,vol. 7, 613–633.

Bierwirth, M., Glöckner, F.O., Grimm, C., Schimmler, S., Boehm, F., Busse, C., Degkwitz, A., Koepler, O., Neuroth, H. Leipzig-Berlin-Erklärung zu NFDI-Querschnittsthemen der Infrastrukturentwicklung.

BMJV, 2019. Gesetz über den Zugang zu digitalen Geodaten. http://www.gesetze-im-internet.de/geozg/ .

BMWI, 2019. Geologiedatengesetz. Referentenentwurf des Bundesministeriums für Wirtschaft und Energie. https://www.bmwi.de/Redaktion/DE/Artikel/Service/Gesetzesvorhaben/geologiedatengesetz.html .

Bodesheim, P., Jung, M., Gans, F., Mahecha, M.D., Reichstein, M., 2018. Upscaled diurnal cycles of land–atmosphere fluxes: a new global half-hourly data product. Earth Syst. Sci. Data,vol. 10, 1327–1365.

Buttlar, J. v., Zscheischler, J., Mahecha, M.D., 2014. An extended approach for spatiotemporal gapfilling: dealing with large and systematic gaps in geoscientific datasets. Nonlin. Processes Geophys.,vol. 21, 203–215.

Cayoglu, U., Schröter, J., Meyer, J., Streit, A., Braesicke, P., 2018. A modular software framework for compression of structured climate data. In: Banaei-Kashani, F., Hoel, E., Güting, R.H., Tamassia, R., Xiong, L. (eds.), Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems - SIGSPATIAL '18. ACM Press, New York, New York, USA, pp. 556–559.

COPDESS, 2019a. Commitment Statement in the Earth, Space, and Environmental Sciences. https://copdess.org/enabling-fair-data-project/commitment-statement-in-the-earth-space-and-environmental-sciences/ .

COPDESS, 2019b. Enabling FAIR Data Project. https://copdess.org/enabling-fair-data-project/ .

DataCite e.V., 2019. DataCite Metadata Schema. https://schema.datacite.org/ .

DCAT-AP, 2019. DCAT-AP. https://www.dcat-ap.de/ .

Dillo, I., Doorn, P., 2014. The Front Office–Back Office Model: Supporting Research Data Management in the Netherlands. IJDC,vol. 9, 39–46.

Directive (EU) 2019/1024, 2019. 32019L1024 - EN. Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information. https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1561563110433&uri=CELEX:32019L1024 .

DKRZ, 2019. ESGF Node at DKRZ. https://esgf-data.dkrz.de/projects/esgf-dkrz/ .

Downs, R.R., Lenhardt, C.W., Robinson, E., Ethan, D., Weber, N., 2015. Community Recommendations for Sustainable Scientific Software. Journal of Open Research Software,vol. 3, 4.

EIONET, 2019. GEneral Multilingual Environmental Thesaurus (GEMET). https://www.eionet.europa.eu/gemet/en/about/ .

ENVRIplus Project Office, 2019a. ENVRI-FAIR. ENVironmental Research Infrastructures building Fair services Accessible for society, Innovation and Research. http://envri.eu/envri-fair/ .

ENVRIplus Project Office, 2019b. Virtual ENVRI community platform. http://envri.eu/ .

European Commission, 2019a. ERIC Landscape. Members of the European Reasearch Infrastructure Consortium (ERIC). https://ec.europa.eu/info/research-and-innovation/strategy/european-research-infrastructures/eric/eric-landscape_en .

European Commission, 2019b. Infrastructure for spatial information in Europe (INSPIRE). INSPIRE Knowledge Base. https://inspire.ec.europa.eu/ .

European Commission, 2020. Open Science. https://ec.europa.eu/research/openscience .

Ferrari, T., Scardaci, D., Andreozzi, S., 2018. The Open Science Commons for the European Research Area. In: Mathieu, P.-P., Aubrecht, C. (eds.), Earth Observation Open Science and Innovation. Springer International Publishing, Cham, pp. 43–67.

FGDC, 2019. Executive Orders. https://www.fgdc.gov/policyandplanning/executive_order .

GeoKur, 2019. GeoKur Project. https://geokur.geo.tu-dresden.de/index.php .

GeoNames, 2019. GeoNames. http://www.geonames.org/ .

GEOSS. Global Earth Observation System of Systems. https://www.earthobservations.org/geoss .

GFZ Data Services, 2019. GFZ Data Services. Metadata Portal. http://dataservices.gfz-potsdam.de/portal/ .

Glöckner, F.O., Diepenbroek, M., Felden, J., Overmann, J., Bonn, A., Gemeinholzer, B., Güntsch, A., König-Ries, B., Seeger, B., Pollex-Krüger, A., Fluck, J., Pigeot, I., Kirsten Toralf, Mühlhaus, T., Wolf, C., Heinrich, U., Steinbeck, C., Koepler, O., Stegle, O., Weimann, J., Schörner-Sadenius, T., Gutt, C., Stahl, F., Wagemann, K., Schrade, T., Schmitt, R., Eberl, C., Gauterin, F., Schultz, M., Bernard, L., 2019. Berlin Declaration on NFDI Cross-Cutting Topics. Zenodo.

Haase, P., Tonkin, J.D., Stoll, S., Burkhard, B., Frenzel, M., Geijzendorffer, I.R., Häuser, C., Klotz, S., Kühn, I., McDowell, W.H., Mirtl, M., Müller, F., Musche, M., Penner, J., Zacharias, S., Schmeller, D.S., 2018. The next generation of site-based long-term ecological monitoring: Linking essential biodiversity variables and ecosystem integrity. The Science of the total environment,vol. 613-614, 1376–1384.

Henzen, C., 2018. Building a Framework of Usability Patterns for Web Applications in Spatial Data Infrastructures. IJGI,vol. 7, 446.

Higman, R., Teperek, M., Kingsley, D., 2017. Creating a Community of Data Champions.

Hoffmann, A., Penner, J., Vohland, K., Cramer, W., Doubleday, R., Henle, K., Kõljalg, U., Kühn, I., Kunin, W., Negro, J.J., Penev, L., Rodríguez, C., Saarenmaa, H., Schmeller, D., Stoev, P., Sutherland, W., Ó Tuama, É., Wetzel, F., Häuser, C.L., 2014. The need for an integrated biodiversity policy support process – Building the European contribution to a global Biodiversity Observation Network (EU BON). NC,vol. 6, 49–65.

Horstmann, W., Nurnberger, A., Shearer, K., Wolski, M., 2017. Addressing the Gaps: Recommendations for Supporting the Long Tail of Research Data.

Hübner, A., 2020. Earth science and biodiversity journals can improve support for data publication.

IGSN, 2019. Technical Documentation of the International Geo Sample Number (IGSN). Metadata. https://igsn.github.io/metadata/ .

ISO, 2014. ISO 19115-1:2014. Geographic information — Metadata — Part 1: Fundamentals. https://www.iso.org/standard/53798.html .

ISO, 2017. ISO15836-1:2017. Information and documentation — The Dublin Core metadata element set — Part 1: Core elements. https://www.iso.org/standard/71339.html .

Kahn, R., Wilensky, R., 2006. A framework for distributed digital object services. Int J Digit Libr,vol. 6, 115–123.

Karam, N., Khiat, A., Algergawy, A., Sattler, M., Weiland, C., Schmidt, M., 2020. Matching biodiversity and ecology ontologies: challenges and evaluation results. The Knowledge Engineering Review,vol. 35.

Kaspar, F., Niermann, D., Borsche, M., Fiedler, S., Keller, J., Potthast, R., Rösch, T., Spangehl, T., Tinz, B., 2020. Regional atmospheric reanalysis activities at Deutscher Wetterdienst: review of evaluation results and application examples with a focus on renewable energy. Adv. Sci. Res.,vol. 17, 115–128.

Koppe, R., Gerchow, P., Macario, A., Haas, A., Schafer-Neth, C., Pfeiffenberger, H., 2015 - 2015. O2A: A generic framework for enabling the flow of sensor observations to archives and publications, OCEANS 2015 - Genova. IEEE, pp. 1–6.

Lannom, L.L., Wittenburg, P. Global Digital Object Cloud (DOC) - A Guiding Vision,vol. 2016.

Lemmens, R., By, R. de, Gould, M., Wytzisk, A., Granell, C., van Oosterom, P., 2007. Enhancing Geo-Service Chaining through Deep Service Descriptions. Transactions in GIS,vol. 11, 849–871.

Mahecha, M.D., Gans, F., Brandt, G., Christiansen, R., Cornell, S.E., Fomferra, N., Kraemer, G., Peters, J., Bodesheim, P., Camps-Valls, G., Donges, J.F., Dorigo, W., Estupinan-Suarez, L.M., Gutierrez-Velez, V.H., Gutwin, M., Jung, M., Londoño, M.C., Miralles, D.G., Papastefanou, P., Reichstein, M., 2020. Earth system data cubes unravel global multivariate dynamics. Earth Syst. Dynam.,vol. 11, 201–234.

Mirtl, M., T Borer, E., Djukic, I., Forsius, M., Haubold, H., Hugo, W., Jourdan, J., Lindenmayer, D., McDowell, W.H., Muraoka, H., Orenstein, D.E., Pauw, J.C., Peterseil, J., Shibata, H., Wohner, C., Yu, X., Haase, P., 2018. Genesis, goals and achievements of Long-Term Ecological Research at the global scale: A critical review of ILTER and future directions. The Science of the total environment,vol. 626, 1439–1462.

Mollenhauer, H., Kasner, M., Haase, P., Peterseil, J., Wohner, C., Frenzel, M., Mirtl, M., Schima, R., Bumberger, J., Zacharias, S., 2018. Long-term environmental monitoring infrastructures in Europe: observations, measurements, scales, and socio-ecological representativeness. The Science of the total environment,vol. 624, 968–978.

Müller, M., Bernard, L., Kadner, D., 2013. Moving code – Sharing geoprocessing logic on the Web. ISPRS Journal of Photogrammetry and Remote Sensing,vol. 83, 193–203.

NASA, 2019. Global Change Master Directory (GCMD). Keywords. https://earthdata.nasa.gov/earth-observation-data/find-data/gcmd/gcmd-keywords .

NFDI4Earth. NFDI4Earth: https://www.nfdi4earth.de/ .

PANGAEA, 2019. PANGAEA. Data Publisher for Earth & Environmental Science. https://www.pangaea.de/ .

re3data.org, 2019. re3data.org. Registry of Research Data Repositories. https://www.re3data.org/ .

Research Data Alliance FAIR Data Maturity Model Working Group, 2020. FAIR Data Maturity Model: specification and guidelines - draft.

RfII, 2019. Digital competencies – urgently needed! Recommendations on career and training prospects for the scientific labour market. Rat für Informationsinfrastrukturen, Göttingen, http://www.rfii.de/?wpdmdl=4015 .

RfII, 2020. The Data Quality Challenge. Recommendations for Sustainable Research in the Digital Turn. Rat für Informationsinfrastrukturen, Göttingen, http://www.rfii.de/?wpdmdl=4203 .

Rönsdorf, C., Mason, P., Holmes, J., Gerber, U., Streilein, A., Bos, M., Shaon, A., Naumann, K., Kirstein, M., Samuelsson, G., Rantala, M., Kvarteig, S., Adams, L., Svennewall, J., Stößel, W., 2016. GI+100: Long Term Preservation of Digital Geographic Information — 16 Fundamental Principles Agreed by National Mapping Agencies and State Archives. IJDC,vol. 11, 156.

ScaDS, 2019. Competence Center for Scalable Data Services and Solutions (ScaDS). https://www.scads.de/en/ .

Schlichtkrull, M., Kipf, T.N., Bloem, P., van den Berg, R., Titov, I., Welling, M. Modeling Relational Data with Graph Convolutional Networks,vol. 10843, 593–607.

Schmidt, B., Wytzisk, A., 2019. Nachhaltigkeitsaspekte in der Softwareentwicklung. In: Englert, M., Ternès, A. (eds.), Nachhaltiges Management,vol. 14. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 651–668.

Schultz, M.G., Apweiler, S., Vogelsang, J., Hagemeier, B., Kleinert, F., Mallmann, D. A Web Service Architecture for Objective Station Classification Purposes, IEEE 14th International Conference 29.10.2018 - 01.11.2018, pp. 283–284.

Schwardmann, U., 2020. Digital Objects – FAIR Digital Objects: Which Services Are Required? Data Science Journal,vol. 19.

Smaili, F.Z., Gao, X., Hoehndorf, R., 2019. OPA2Vec: combining formal and informal content of biomedical ontologies to improve similarity-based prediction. Bioinformatics (Oxford, England),vol. 35, 2133–2140.

Smedt, K. de, Koureas, D., Wittenburg, P., 2020. FAIR Digital Objects for Science: From Data Pieces to Actionable Knowledge Units. Publications,vol. 8, 21.

Stocker, M., Darroch, L., Krahl, R., Habermann, T., Devaraju, A., Schwardmann, U., D'Onofrio, C., Häggström, I., 2020. Persistent Identification of Instruments. Data Science Journal,vol. 19.

Stocker, M., Paasonen, P., Fiebig, M., Zaidan, M.A., Hardisty, A., 2018. Curating Scientific Information in Knowledge Infrastructures. Data Science Journal,vol. 17.

Strawn, G., 2019. Open Science, Business Analytics, and FAIR Digital Objects. IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC), 658–663.

Thessen, A.E., Woodburn, M., Koureas, D., Paul, D., Conlon, M., Shorthouse, D.P., Ramdeen, S., 2019. Proper Attribution for Curation and Maintenance of Research Collections: Metadata Recommendations of the RDA/TDWG Working Group. CODATA,vol. 18, 54.

Wiemann, S., Bernard, L., 2016. Spatial data fusion in Spatial Data Infrastructures using Linked Data. International Journal of Geographical Information Science,vol. 30, 613–636.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, Luiz Bonino, Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., Hoen, P.A.C. 't, Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Scientific data,vol. 3, 160018.

Wilkinson, M.D., Sansone, S.-A., Schultes, E., Doorn, P., Bonino da Silva Santos, L.O., Dumontier, M., 2018. A design framework and exemplar metrics for FAIRness. Scientific data,vol. 5, 180118.

5.6 List of Information Infrastructures and Data Repositories, Software Repositories (developed and operated by members of the consortium)

Infrastructure / repository / project / analytics framework	Name(s) of the maintaining organization(s)	Weblink
O2A - SENSOR (Sensor observations to archives)	AWI	http://sensor.awi.de
O2A DASHBOARD	AWI	https://dashboard.awi.de
GITLAB	AWI	https://gitlab.awi.de/
PANGAEA	AWI, MARUM	https://pangaea.de/
GDI-BGR	BGR	https://produktcenter.bgr.de
Earth System Data Lab (ESDL)	Brockmann Consult GmbH, MPIBGC	https://www.earthsystemdatalab.net/
re3data	DataCite, GFZ, KIT	https://repositoryfinder.datacite.org/
World Data Center for Climate (WDCC)	DKRZ	http://cera.wdc-climate.de
DKRZ	DKRZ	http://www.dkrz.de
DKRZ User Portal (LTA DOKU)	DKRZ	https://www.dkrz.de/up/services/data-management/lta-docu-1
CMIP Data Pool (DKRZ CDP)	DKRZ	https://www.dkrz.de/up/services/data-management/cmip-data-pool
Data Distribution Centre of the Intergovernmental Panel on Climate Change (IPCC-DDC)	DKRZ and others	https://www.ipcc-data.org/
ENES Climate Analytics Lab (ECASLAB)	DKRZ, CMCC	https://ecaslab.dkrz.de/home.html
ERA5 data pool	DKRZ, ECMWF	https://www.dkrz.de/up/de-services/de-data-management/de-projects_cooperations/de-era
B2SHARE	DKRZ, EUDAT	https://b2share.eudat.eu/
B2HANDLE	DKRZ, EUDAT	https://eudat.eu/catalogue/B2HANDLE
Data for Marine Science & Offshore Industry (coastDat)	DKRZ, HZH	https://cera-www.dkrz.de/WDCC/ui/cerasearch/project?acronym=coastDat
WDC-RSAT	DLR	http://wdc.dlr.de/
German Satellite Data Archive (D-SDA)	DLR	http://eoweb.dlr.de
Earth Observation Center (EOC) Geoservice	DLR	https://geoservice.dlr.de/
Copernicus Data and Exploitation Platform - Germany (CODE-DE)	DLR	https://code-de.org/

Infrastructure / repository / project / analytics framework	Name(s) of the maintaining organization(s)	Weblink
Alpine Environmental Data Analysis Center (AlpEnDAC)	DLR, LRZ, U Augsburg, bifa mbH, UFS GmbH	http://www.alpendac.eu
The Bacterial Diversity Metadatabase (BacDive)	DSMZ	https://bacdive.dsmz.de/
The Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures (DSMZ)	DSMZ	https://www.dsmz.de/
EUMETSAT Climate-Monitoring Satellite-Application Facility (CMSAF)	DWD	https://www.cmsaf.eu/
Climate Data Center (CDC)	DWD	https://cdc.dwd.de/catalogue/srv/en/main.home
Repository for Research Sites & Datasets (DEIMS-SDR)	eLTER	http://www.lter-europe.net/carousel/deims-sdr
Hi-Knowledge	FU Berlin, IGB	https://hi-knowledge.org/
Geoportal	GDI-DE	https://www.geoportal.de/
Ocean Science Information System (OSIS)	GEOMAR	https://www.geomar.de/en/research/research-data/
GEOMAR (Near-) Realtime-Data from Glider and Moorings	GEOMAR	https://data.geomar.de/realtime/html/index.html
Meteorite Information Database (MetBase)	GeoPlatform	https://metbase.org
Centre for High-Performance Scientific Computing in Terrestrial Systems (HPSC TerrSys)	Geoverbund ABC/J	www.hpsc-terrsys.de
Laboratory for Clouds and Precipitation Exploration (CPEX-LAB)	Geoverbund ABC/J	www.cpex-lab.de
Geo.X Laboratory Infrastructure Search (LI@Geo.X)	Geo.X	https://www.geo-x.net/geox-laboratory-infrastructure-search/
German Federation for Biological Data (GFBio)	GFBio e.V.	https://www.gfbio.org
Research Data Repository of GFZ Data Services (GFZ Data Services)	GFZ	http://dataservices.gfz-potsdam.de/portal/
GFZ Data and Research Infrastructures (RI@GFZ)	GFZ	http://dataservices.gfz-potsdam.de/mesi/index.html
GEOFON	GFZ	https://geofon.gfz-potsdam.de/
World Stress Map (WSM)	GFZ	http://www.world-stress-map.org/
ICDP Drilling Information Systems	GFZ	https://www.icdp-online.org/
Gravity Recovery and Climate Experiment (GRACE)	GFZ	https://isdc.gfz-potsdam.de/grace-isdc/
Gravity Recovery and Climate Experiment Follow-on (GRACE-FO)	GFZ	https://isdc.gfz-potsdam.de/grace-fo-isdc/
CHAMP data	GFZ	https://isdc.gfz-potsdam.de/champ-isdc/

Infrastructure / repository / project / analytics framework	Name(s) of the maintaining organization(s)	Weblink
Indices of Global Geomagnetic Activity (kp-index)	GFZ	https://isdc.gfz-potsdam.de/kp-index/
IAG - ICGEM International Center for Global Earth Models	GFZ	http://icgem.gfz-potsdam.de/home
IAG - IGETS International Geodynamics and Earth Tide Service	GFZ	http://isdc.gfz-potsdam.de/igets-data-base/
IAG - International GNSS Service (IGS) - Analysis Centre	GFZ	https://www.gfz-potsdam.de/en/section/space-geodetic-techniques/projects/igs/
Geophysical Instrument Pool Potsdam (GIPP)	GFZ	http://gipp.gfz-potsdam.de/webapp/projects
European Plate Observing System (EPOS)	GFZ	https://www.epos-ip.org/
European Integrated Data Archive (EIDA)	GFZ and others	http://www.orfeus-eu.org/data/eida/
International Laser Ranging Service (ILRS)	GFZ and others	https://ilrs.cddis.eosdis.nasa.gov/
International VLBI Service for Geodesy and Astrometry (IVS)	GFZ and others	https://ivscc.gsfc.nasa.gov/
International DORIS Service (IDS)	GFZ and others	https://ids-doris.org/
Implementation Organization of the International Geo Sample Number e.V. (IGSN)	GFZ and others	https://igsn.github.io/
GravIS, the Gravity Information Service	GFZ, AWI, TUDD	http://gravis.gfz-potsdam.de/home
Earth System Knowledge Platform (ESKP)	Helmholtz Association, GFZ	https://www.eskp.de/en/home/
Coastal Observing System for Northern and Arctic Seas (COSYNA)	HZH	http://codm.hzg.de/codm/
Marine geoportal of HZH (coastMap)	HZH	https://www.coastmap.org
Climate Service Center Germany (GERICS)	HZH	https://www.climate-service-center.de
IGB Freshwater Research and Environmental Database (FRED)	IGB	https://fred.igb-berlin.de/
Freshwater Information Plattform (FIP)	IGB	http://www.freshwaterplatform.eu/
Monitor of settelement and open space development (IOER-Monitor)	IÖR	https://www.ioer-monitor.de/
Oceanographic Database Search with Interactive Navigation (ODIN 2)	IOW	https://odin2.io-warnemuende.de
Integrated Carbon Observation System (ICOS)	JÜLICH	https://www.icos-cp.eu/
TOAR Database of Surface Ozone	JÜLICH	https://join.fz-juelich.de
TERENO	JÜLICH	https://tereno.net

Infrastructure / repository / project / analytics framework	Name(s) of the maintaining organization(s)	Weblink
In-service Aircraft for a Global Observing System (IAGOS)	JÜLICH	https://www.iagos.org/
Jülich Research on Exascale Cluster Architectures (JURECA)	JÜLICH	https://www.fz-juelich.de/ias/jsc/EN/Expertise/Supercomputers/JURECA/JURECA_node.html
B2ACCESS	JÜLICH, EUDAT	https://www.eudat.eu/services/b2access
Steinbruch Centre for Computing (SCC)	KIT	https://www.scc.kit.edu/index.php
Geoportal Geopotential German North Sea (GPDN)	LBEG, BGR, BSH	https://www.gpdn.de/gpdn/cardomap3.aspx
Geophysics Information System (FIS GP-Explorer)	LIAG	https://www.fis-geophysik.de
Geothermal Information System (GeotIS)	LIAG	https://www.geotis.de
Data Portal for German Marine Research	Marine Network for Integrated Data Access (MaNIDA)	http://manida.awi.de
BGI Data Portal	MPIBGC	https://www.bgc-jena.mpg.de/geodb/
Geochemistry of Rocks of the Oceans and Continents (GeoRoc)	MPIC	http://georoc.mpch-mainz.gwdg.de/georoc/
Geological and Environmental Reference Materials (GeoRem)	MPIC	http://georem.mpch-mainz.gwdg.de
Climate Data Operator (CDO)	MPIM	https://code.mpimet.mpg.de/projects/cdo/
Earth System Grid Federation (ESGF)	PCMDI, DKRZ, BADC, IPSL, and others	https://esgf.llnl.gov
Sea Ice Data Portal	REKLIM, U Bremen, AWI	https://www.meereisportal.de/en/
Edaphobase - The database for distribution and ecology of soil organisms	Senckenberg	https://portal.edaphobase.org
Centre for Information Services and High Performance Computing (ZIH)	TUDD	https://tu-dresden.de/zih
Generic Research Data Infrastructure (GeRDI)	TUDD/ZIH, ZBW Kiel, LRZ, CAU Kiel, DFN	https://www.gerdi-project.eu

List of international Infrastructures relevant for NFDI4Earth

Infrastructure / repository / project / analytics framework	Name(s) of the maintaining organization(s)	Weblink
UK Polar Data Centre	BAS	https://www.bas.ac.uk/data/uk-pdc/

Infrastructure / repository / project / analytics framework	Name(s) of the maintaining organization(s)	Weblink
British Oceanographic Data Centre (BODC)	BODC	https://www.bodc.ac.uk/data/numerical_model_data/
Oceanographic Data Center (DOD)	BSH	https://www.bsh.de/DE/DATEN/Ozeanographisches_Datenzentrum/ozeanographisches_datenzentrum_node.html
GeoSeaPortal	BSH	https://www.geoseaportal.de/
Climate Data Store (CDS)	ECMWF	https://cds.climate.copernicus.eu/
European Environmental Agency (EEA)	EEA	http://www.eea.europa.eu
ESA Earth Online Data	ESA	https://earth.esa.int/web/guest/data-access
GEOSS Portal	ESA and others	http://www.geoportal.org
EUDAT - B2FIND	EU	http://b2find.eudat.eu
European Marine Observation and Data Network (EMODnet Biology)	EU	http://www.emodnet-biology.eu/portal/
Australia Ocean Data Network (AODN)	IMOS	https://portal.aodn.org.au/
Planet OS	Intertrust	https://data.planetos.com/
IRIS Seismic Data Access	IRIS	http://ds.iris.edu/data/access/
Japan Agency for Marine-Earth Science and Technology (JAMSTEC)	JAMSTEC	http://www.jamstec.go.jp/e/database/
Intergrated Microbial Genomes and Microbiomes (IMG/M)	JGI	https://img.jgi.doe.gov/cgi-bin/w/main.cgi
The INSPIRE Geoportal	JRC	http://inspire-geoportal.ec.europa.eu/
KNMI Climate Explorer	KNMI	https://climexp.knmi.nl/start.cgi
Marine Geoscience Data System (MGDS)	Lamont	http://www.marine-geo.org/index.php
Interdisciplinary Earth Data Alliance (IEDA)	Lamont	https://www.iedadata.org
National Snow and Ice Data Center (NSIDC)	NASA	https://nsidc.org
Earthdata	NASA	https://earthdata.nasa.gov/
CDS - Climate Model Data Services	NASA	https://cds.nccs.nasa.gov/
NASA's Earth Observing System Data and Information System (EOSDIS)	NASA	https://earthdata.nasa.gov/about/daacs
Climate Data Guide	NCAR	https://climatedataguide.ucar.edu/
National Center for Biotechnology Information (NCBI)	NCBI	https://www.ncbi.nlm.nih.gov/
NIH Genetic Sequence Database (GenBank)	NCBI	https://www.ncbi.nlm.nih.gov/genbank/
National Centers for Environmental Information (NCEI)	NOAA	https://www.ncei.noaa.gov/

Infrastructure / repository / project / analytics framework	Name(s) of the maintaining organization(s)	Weblink
Earth Impact Database (EID)	PASSC	http://www.passc.net/EarthImpactDatabase/New%20website_05-2018/Index.html
JASMIN Scientific Data Analysis Environment	RAL, U Leeds, U Bristol, U Reading	http://www.jasmin.ac.uk
Earth Reference Data and Models (EarthRef)	SCRIPPS	https://earthref.org
International Data Citation Initiative (Data Cite)	TIB	https://datacite.org/
Environmental Systems Science Data Infrastructure for a Virtual Ecosystem (ESS-DIVE)	DOE	https://ess-dive.lbl.gov/
EarthCube	UCAR	https://www.earthcube.org
Repository for Archiving and Managing Diverse Data (RAMADDA)	UCAR	http://atm.ucar.edu/repository
University of Wyoming Atmospheric Soundings	U Wyoming	http://weather.uwyo.edu/upperair/sounding.html
USGS Earth Explorer	USGS	https://earthexplorer.usgs.gov/
ClimateWatch	WRI	https://www.climatewatchdata.org/