

Nationale Forschungsdateninfrastruktur für die Erdsystemwissenschaften

Tag der Forschungsdaten 2021

Andreas Wytzisk-Arens (Hochschule Bochum, Fachbereich Geodäsie)

Fachbereich Geodäsie @ Hochschule Bochum

- Fachgebiete Geoinformatik und Vermessung
- Mitglied des Instituts f
 ür Angewandte KI und Data Science Ruhr (AKIS)
- Forschungsschwerpunkte, u.a.
 - (Geo-) Informationsinfrastrukturen
 - Spatial Data Science
 - Interaktive Geovisualisierung
 - Nachhaltige Softwareentwicklung
 - Optische 3D Mess- und Auswertetechnik
 - Kinematische Messtechnik / Multisensorsysteme
 - Gebäude- und Oberflächenmonitoring



Motivation

Rasches Wachstum und zunehmende Heterogenität der Forschungsdatenbestände stellt die (Geo-) Wissenschaftscommunity vor große Herausforderungen

- Ineffiziente Nutzung verfügbarer Daten aufgrund eines hohen Fragmentierungsgrades
- Hohe Kosten durch komplexe (oder mehrfache) Datenerfassung und -integration
- Eingeschränkter Erkenntnisgewinn, wenn relevante Daten unbekannt bleiben oder für externe Forscher nicht zugänglich sind
- Daten werden unbrauchbar, wenn das Wissen über die Daten verloren geht (z. B. wenn Mitarbeiter die Organisation verlassen oder sich die Verantwortlichkeiten ändern)

"Der digitale Wandel hat Forschungsdaten und Forschungsmethoden erfasst, er verändert die Wissenschaft grundlegend."

RfII-Positionspapier Leistung aus Vielfalt (2016)

Motivation

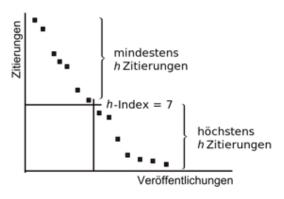
FDM zentrales Thema der Forschungscommmunity

- Good Practices, Richtlinien und Technologien zur Akquise, Bewertung, Analyse, Austausch und Publikation von Forschungsdaten und Forschungssoftware
- Forschungsdaten und –software als Grundlage,
 Zwischenprodukt und Endprodukt von Forschung
- Effizienzsteigerung und Reproduzierbarkeit datengetriebener Forschung (für Forschende, Labs, Communities)
- Enabler für neue (interdisziplinäre) Forschung

- Typically inhouse Research Questions Experiment Design Collect or Archive discover ata & method Peer review Integrate Publish results Analyze (data & methods) Validate Describe Typically in a ommon infrastructure
- → Sicherstellung der Austauschbarkeit von Forschungsergebnissen
- → Enabler für einen kulturellen Wandel hin zu offener und reproduzierbarer Wissenschaft

Forschungsdatenmanagement – generelle Herausforderungen

- Reputationsmechanismen
 - Aktuell zählt das paper nicht die Daten/Softwarepublikation
- Kuratierung und Qualitätssicherung
- Datenschutz
- Langzeit-Archivierung für Software und Daten
- Forschungsdatenmanagement als Standardbestandteil in der Ausbildung
- Support
- Souveränität und Nachhaltigkeit



https://de.wikipedia.org/wiki/H-Index



Cases of Cholera Victims in London (John Snow 1854)

Forschungsdatenmanagement – Status Technologien

Es existieren zahlreiche Plattformen und Softwaresysteme, die (wissenschaftliche) Datenpublikation, Datenaustausch, Analyse und den Systemaustausch unterstützen...

...tlw. Open Source und tlw. etabliert bei IT-affinen Wissenschaftlern.







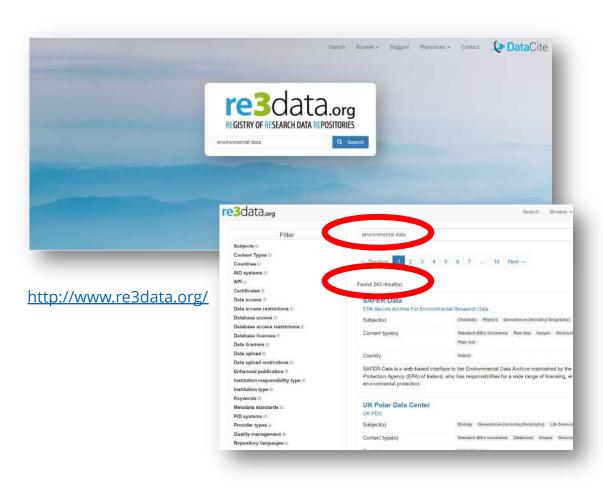






Forschungsdatenmanagement – Status Plattformen

In einigen Domänen existieren (etablierte) Plattformen für die Publikation von Forschungsdaten...





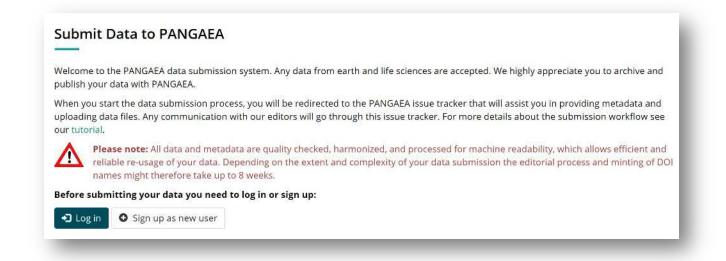
Forschungsdatenmanagement - Status Initiativen

Best Practices und Standards existieren etwa für ...

- Zur Angabe von Datenzitationen
- Qualitätskriterien für Repositorien
- Kriterien f
 ür Publikation offener Daten







Forschungsdatenmanagement – Status Initiativen

Zahlreiche international und nationale Initiativen zum Thema Forschungsdatenmanagement...













Rat für Informationsstrukturen (RfII)

SCHARNIERFUNKTION ZWISCHEN WISSENSCHAFT UND POLITIK

Der Rat für Informationsinfrastrukturen

- > berät als Sachverständigengremium Politik und Wissenschaft in strategischen Zukunftsfragen der digitalen Wissenschaft;
- > will Kooperation und Abstimmung bestehender Aktivitäten befördern;
- > identifiziert Synergiepotenziale im Wissenschaftssystem und will Doppelförderung vermeiden helfen;
- > benennt neue Handlungsfelder, die durch technischen und kulturellen Wandel entstehen;
- > will Selbstorganisationsformen in der Wissenschaft stärken, um wissenschaftsgetriebene Entwicklungen zu befördern;
- > gewährleistet die Rückkopplung mit internationalen Debatten.



http://www.rfii.de/

- 2014 durch die Gemeinsame Wissenschaftskonferenz (GWK) eingerichtet
- 24 Ratsmitglieder

(8 Wissenschaftliche Nutzer, 8 Einrichtungen, 4 Bund und Länder, 4 Öffentliches Leben)



Rat für Informationsstrukturen (RfII)

Empfehlungen von 2016 an wissenschaftliche Einrichtungen u.a.:

- übergreifende Dienste, vereinheitlichende Datenbeschreibungen und universelle Zugänglichkeit für eine Nationale Forschungsdateninfrastruktur (NFDI) zu entwickeln bzw. sicherzustellen
- als Infrastrukturanbieter durch Methodenkompetenz und überzeugende Services
 Anreize für die Nutzung eines Forschungsdatenmanagements und der NFDI zu setzen
- Für das Forschungsdatenmanagement **qualifiziertes Personal** durch attraktive Ausgestaltung neuer Berufswege in der Forschung halten

• ...

http://www.rfii.de/?wpdmdl=1998

Der Prozess zur Nationalen Forschungsdateninfrastruktur

Empfehlung zum Aufbau einer Nationalen Forschungsdateninfrastruktur (NFDI; RfII 2017)

- Umsetzung der FAIR-Prinzipien (Findable, Accessible, Interoperable, and Re-usable)
- Nachhaltig (nicht nur Projektlösungen)
- Schrittweise unter Berücksichtigung der unterschiedlichen digitalen Kulturen in den Disziplinen
- International eingebunden
- Getrieben durch die wissenschaftlichen Nutzer
- Durch fachlich geeignete Konsortien aus Nutzern und Anbietern



Diskussionsimpulse des RfII zur NFDI: http://www.rfii.de/de/themen/ bzw. http://www.rfii.de/?wpdmdl=2269

Der Prozess zur Nationalen Forschungsdateninfrastruktur

"...mit einer nationalen Forschungsdaten-Infrastruktur wissenschaftliche Datenbestände systematisieren und einen nachhaltigen Zugang sicherstellen. (S. 33)"

Koalitionsvertrag, März 2018

https://www.bundesregierung.de/Content/DE/StatischeSeiten/Breg/koalitionsvertrag-inhaltsverzeichnis.html



picture alliance / dpa

Fördermaßnahmen aus dem Bund-Länder Programm (11/2018)

- Finanzierung über 10 Jahre mit 90 Mio € p.a.
- Förderung eines NFDI Direktorats (→ Karlsruhe...)
- Förderung von bis zu 30 Konsortien in 3 Ausschreibungsrunden

https://www.gwk-bonn.de/fileadmin/Redaktion/Dokumente/Papers/NFDI.pdf

NFDI: Geförderte Konsortien – Stand 10/2021

Seit 10/2020

- <u>DataPLANT</u>: Daten in Pflanzenforschung
- 2. GHGA: Deutsches Humangenom-Phenom-Archiv
- KonsortSWD: Konsortium für die Sozial-, Bildungs-, Verhaltens- und Wirtschaftswissenschaften
- 4. <u>NFDI4Biodiversity</u>: Biodiversität, Ökologie und Umweltdaten
- NFDI4Cat: NFDI für Wissenschaften rund um Katalyse
- 6. NFDI4Chem: Fachkonsortium Chemie für die NFDI
- 7. NFDI4Culture: Konsortium für Forschungsdaten zu materiellem und immateriellem kulturellem Erben
- 8. NFDI4Health: NFDI für personenbezogene Gesundheitsdaten
- 9. <u>NFDI4Ing</u>: NFDI für die Ingenieurswissenschaften

Ab 10/2021

- 10. BERD@NFDI: NFDI für Betriebswirtschaftslehre, Volkswirtschaftslehre und verwandte Daten
- 11. DAPHNE4NFDI: Daten aus Photonen- und Neutronen-Experimenten
- 12. FAIRmat: FAIRe Dateninfrastruktur für die Physik der kondensierten Materie und die chemische Physik von Feststoffen
- 13. MaRDI: Mathematische Forschungsdateninitiative
- 14. NFDI4DataScience: NFDI für Datenwissenschaften und Künstliche Intelligenz
- 15. NFDI4Earth: NFDI-Konsortium für Erdsystemforschung
- 16. NFDI4Microbiota: NFDI für Mikrobiota-Forschung
- 17. NFDI-MatWerk: Nationale Forschungsdateninfrastruktur für Materialwissenschaft und Werkstofftechnik
- 18. PUNCH4NFDI: Teilchen, Universum, Kerne und Hadronen für die NFDI
- 19. Text+: Sprach- und textbasierte Forschungsdateninfrastruktur

Erdsystemwissenschaften (ESS)



- Geosphäre, Atmosphäre, Biosphäre, Hydrosphäre, Kryosphäre, Anthroposphäre und all ihre Interaktionen
- Von lokalen Prozessen zu globalen Herausforderungen e.g. Erdsystemdynamik, Naturgefahren, Klimawandel, Umweltverschmutzung, Wasserknappheit, Landnutzungsveränderungen, Rohstoffmangel
- Beobachtung, Messung, Modellierung, Analyse,
 Vorhersage des Erdsystems und seiner Veränderungen
- Nutzung von raumzeitvarianten Daten als gemeinsame Referenz internationalen und interdisziplinären Settings



Wo starten wir in den ESS?

NFDI₄Earth

- Daten aus vielfältigsten Quellen
- Oftmals anspruchsvolle Modelle und datenintensive Anwendungen
- Grundsätzlich positive Haltung bzgl.
 Offenheit, FAIR und kollaborativen Entwicklungsmethoden
 - Nicht zuletzt vor dem Hintergrund des langjährigen Aufbaus von Geodateninfrastrukturen (GDI-DE, INSPIRE)
- Etablierte Datenrepositorien
- Zahlreiche Interoperabilitätsstandards



https://cera-www.dkrz.de/



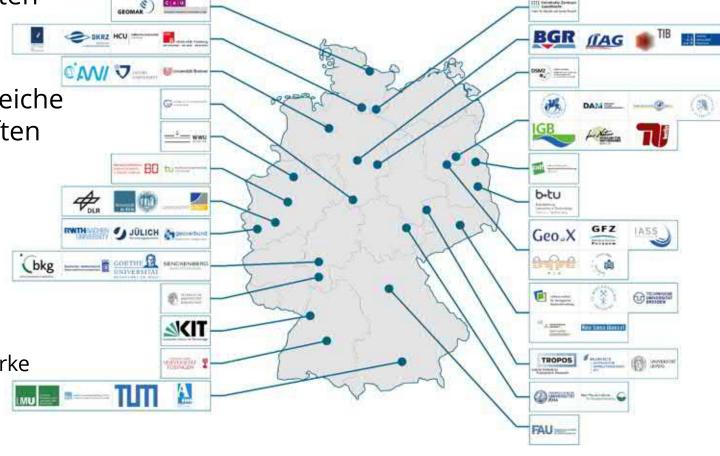
NFDI4Earth - Wer sind wir?



 2018 als offenes Konsortium und als Zweig der NFDI für Erdsystemwissenschaften gegründet

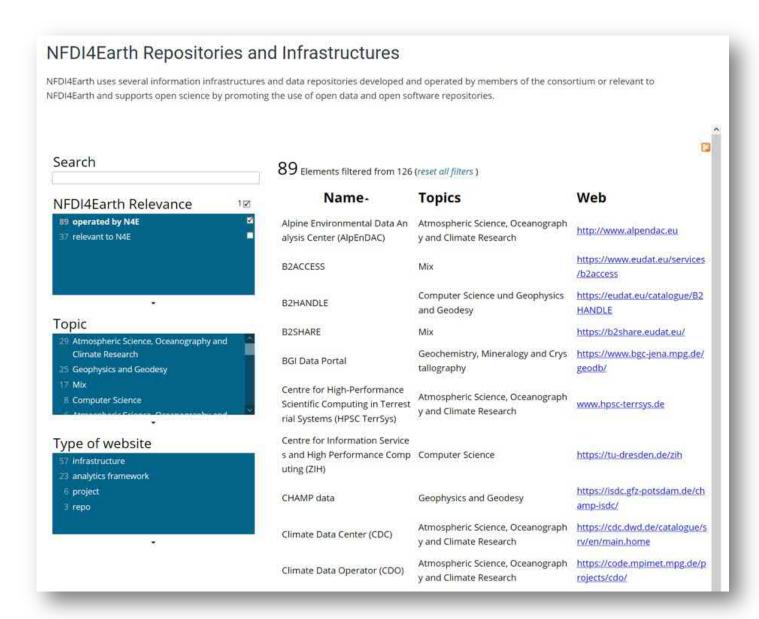
 Aktuell 59 Organisationen, die zahlreiche Facetten der Erdsystemwissenschaften abdecken

- Universitäten
- Forschungsorganisationen
- Infrastrukturbetreiber
- Institutionen aus Politik und Verwaltung
- Wissenschaftliche Verbände und Netzwerke
- https://www.nfdi4earth.de



NFDI4Earth - existierende Repositorien





 NFDI4Earth Partners betreiben ca. 90 Forschungsdatenrepositorien

https://www.nfdi4earth.de/aboutus/repositories-and-infrastructures

Unsere Herausforderungen



- Große Anzahl heterogener Datendienste (>> 100)
 und Aktivitäten im Bereich ESS, nur wenige nachhaltig
- Datenheterogenität
 Qualität, Level der Kuratierung, Lizenzen, Semantik, Datenumfang,
 Wiederholungen, Maßstab, raum-zeitliche Auflösung, ...
- Unterschiedliche Datenkulturen und Geschwindigkeiten bzgl. FAIR - Umsetzungen Von (Nahe-) Echtzeitdatenströmen zu "long tail" Daten; RDM Experten und Novizen
- Unvollständiger Support entlang des Data Lifecycle
- Fehlende Plattformen und Tools für gemeinsame und kollaborative Interpretation heterogener, dezentraler Daten



Zentrale Ziele

NFDI₄Earth

- One Community Approach für ein nachhaltiges offenes und FAIRes Forschungsdatenmanagement in den Erdsystemwissenschaften (ESS)
- Community getriebene agile Kopplung existierender und Entwicklung neuer innovativer Plattformen zur Datenintegration und kollaborative Datenanalyse
- Qualifizierung von Menschen, Daten, Werkzeugen und Diensten als Basis für FAIRes FDM und dessen langfristiges Überleben
- OneStop4All und User Support Network für FDM in des ESS als integraler Bestandteil der NFDI
- Key-driver für den Aufbau und den Betrieb der NFDI

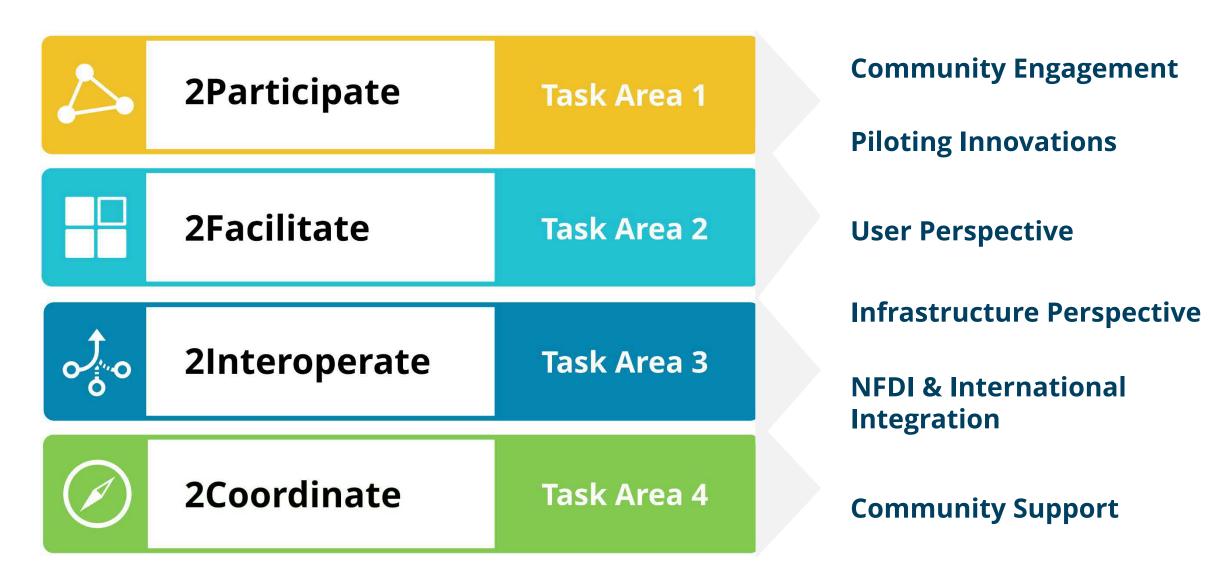






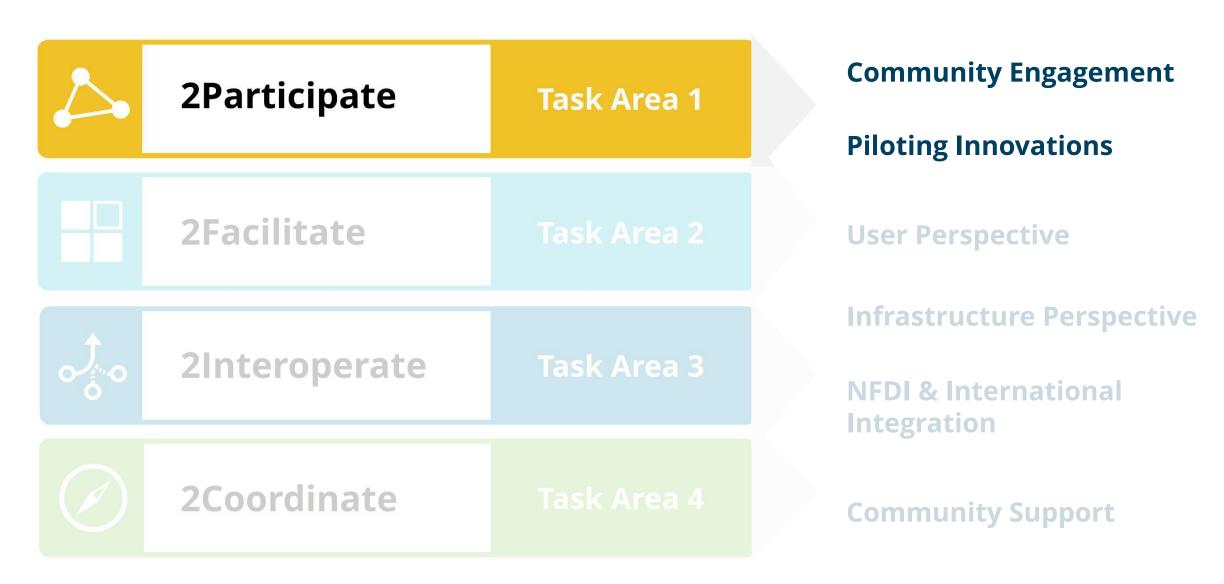
NFDI4Earth – Strategie 2021-26





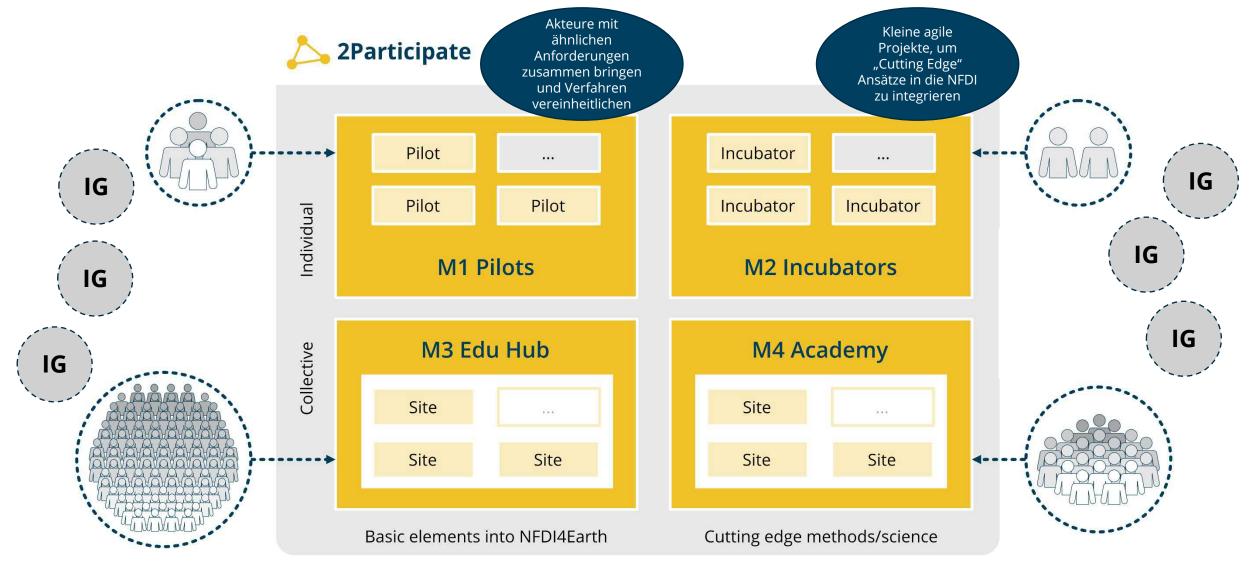
NFDI4Earth – Strategie 2021-26





Community Engagement, Innovationen fördern

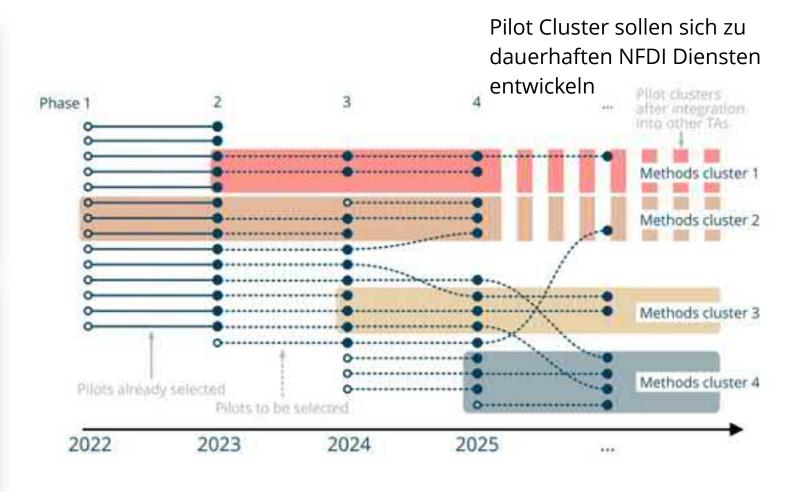




Pilots: von Individuen zu Clustern ...







https://www.nfdi4earth.de/participate/get-involved-by-pilots

Umfangreicher offener Review-Prozess, Monitoring während der Durchführung; zukünftige Pilot Calls können bedarfsgerecht formuliert werden, wenn Gaps identifiziert wurden

Beispiel: Pilot Cluster Data Cubes



Statistical Learning on data cubes.

Winkler et al.



Model evaluation in data cubes.

Eyring et al.



EO4Glaciers data cube. Braun et al.



Herausforderungen:

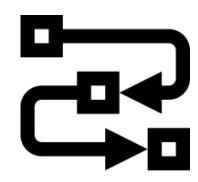
- Heterogene Datenmodelle zusammenführen
- FAIR Prinzipien umsetzen
- Zukunftsorientierte Implementierungen (e.g. cloud-ready)

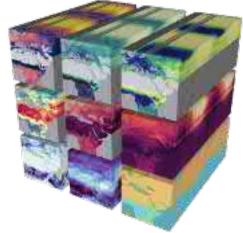
Socioeconomic data cubes.

Kraemer et al.



Processing and Visualization.Unnithan et al.





Aus- und Weiterbildung auf allen Ebenen





Edu Hub: Kompetenzentwicklung in FDM und Spatiotemporal Data Literacy

 Zielgruppen: Praktiker, Studierende, Postdocs, Senior Researcher, Infrastrukturbetreiber

 Aufbau und Bereitstellung umfangreicher "ready-to-use" Open Educational Resources

 Entwicklung neuer Qualifikationsprofile (z.B. research data curator)

 Aufbau eines Netzwerks von "Education and Training Sites"

Academy: Verprobung und Entwicklung neuer Ansätze

- Vernetzung junger Forscher:innen im Kontext daten-getriebener Anwendungen und Forschung
 - Entwicklung einer Nutzer- und Multiplikatorenbasis
- Aufbau von eines Netzwerks Academy Sites



Status



- 10/2020 wurde ein Projektantrag zum Aufbau der NFDI4Earth bei der DFG eingereicht (Laufzeit 5 + 5 Jahre, Fördervolumen ~20 Mio. € für 2021-26)
- 07/2021 Gemeinsame Wissenschaftskonferenz beschließt Förderung (allerdings knapp 30% Kürzung)
- Etablierung der Konsortialverträge
- 10/2021 initialer Projektstart

Thank You!



https://www.nfdi4earth.de/