

**Digital Earth Zwischenbericht - Teil Sachbericht für 2019
Zwischenbericht zu ZT-0025**



Förderkennzeichen:	ZT-0025
Vorhabenbezeichnung:	Digital Earth
Laufzeit des Vorhabens:	01.06.2018 - 31.05.2021
Berichtszeitraum:	01.01.2019 - 31.12.2019

1. Aufzählung der wichtigsten wissenschaftlich-technischen Ergebnisse und anderer wesentlicher Ereignisse unter Bezugnahme der entsprechenden Workpackages.

In *Digital Earth* wurden vier übergreifende geowissenschaftliche Fragen definiert, die auf innovative Ansätze für die Erforschung des Erdsystems abzielen und Studien über alle Schnittstellen hinweg integrieren. Umgesetzt wird dies anhand von zwei beispielhaften Show Cases, die siloübergreifend (Disziplin, Daten, Erdkompartiment) gestaltet sind. Generell liegt der Schwerpunkt auf SMART Monitoring Designs und einem Data Exploration Framework, für die gemeinsame Arbeitsabläufe und Methoden durch eine nachhaltige Zusammenarbeit weiterentwickelt werden.

Die Show Cases sind A) die Analyse von Flutereignissen entlang der Elbe in enger Zusammenarbeit mit *MOSES* und (B) die Quantifizierung von Methanflüssen in/aus der Nordsee. Der „Fluten Show Case“ betrachtet die Entstehung von Fluten entlang der Elbe (z.B. durch Starkregen), deren Verlauf und Folgen auf Umwelt und Mensch. Im „Methan Show Case“ steht die Nordsee als Methan-Quelle im Fokus; Methan ist nach Kohlendioxid das zweitwichtigste Treibhausgas.

Fluten Show Case: Fluten sind kompartiment-übergreifende Ereignisse. Sie zeigen komplexe Verkettungen und ihre kaskadierenden Effekte haben vielfältige Auswirkungen. Es wurden exemplarische Fragestellungen formuliert, die mit Hilfe von Workflows und Data Science Methoden besser und schneller untersucht und beantwortet werden:

- Warum sind manche Hochwasserereignisse extreme Ereignisse?
- Wie sind die Auswirkungen eines Hochwassers?
- Wie verhalten sich in Zukunft fluten-relevante Klimavariablen, wie z.B. extreme Niederschläge?

Um diese komplexen Prozesse und Zusammenhänge besser zu verstehen, wurde ein *Flood Event Explorer* konzipiert und in Teilen bereits implementiert. Er soll es Wissenschaftlern ermöglichen, Hochwasserereignisse umfassend zu bewerten und dabei die volle Prozesskette zu betrachten. Dazu zählen: Vorbedingungen im Einzugsgebiet, Abflussbildung, Überschwemmung, Hochwasserschäden und Auswirkungen im betroffenen Ästuar. Dazu wurden verschiedene Workflows definiert, geeignete Datenanalysemethoden identifiziert und bewertet und interaktive Visualisierungen zur Datenexploration entwickelt (Details siehe **Task 2.1, 2.2, 2.3**).

Methan Show Case: Neben Kohlenstoffdioxid ist Methan ein wichtiges Treibhausgas. Es trägt etwa zu einem Fünftel des menschengemachten Treibhauseffekts bei. Daher ist eine genaue Beobachtung der Methanemissionen sinnvoll. Im Rahmen von diesem Show Case wird ein Workflow entwickelt, der die Methanemissionen der Nordsee im Jahr 2018 genauer analysieren soll. Zu den wissenschaftlichen Zielen zählen:

- das Abschätzen der absoluten Methanemissionen aus der Nordsee,
- die Bestimmung des anthropogenen Anteils an den Emissionen (z.B. durch die Erdöl- und Erdgasgewinnung in der Nordsee),
- der Vergleich von Schätzungen basierend auf atmosphärischen und ozeanographischen Daten und
- eine Analyse der Variabilität der Emissionen, sowohl räumlich als auch zeitlich.

Dazu werden sowohl atmosphärische als auch ozeanographische Daten verwendet. Eine Herausforderung ist das Zusammenführen von Daten aus verschiedenen Quellen und auf verschiedenen

Skalen. Hierzu sollen die notwendigen Workflows und Tools entwickelt werden (Details siehe **Task 2.1, 2.3., 2.4**).

WP1 SMART Monitoring Framework

SMART Monitoring Framework befasst sich zu einem mit der Weiterentwicklung und Standardisierung des Datenflusses vom Sensor bis zur Datenanalyse und zu anderen mit der Digitalisierung von Überwachungsstrategien durch die Verwendung von Methoden aus WP2. Dies geschieht, um beispielsweise Beobachtungslücken zu erkennen und die Sensoranordnungen zu verfeinern, um so eine bessere und statistisch robuste Datenerhebung und Extrapolation zu ermöglichen. Es werden Möglichkeiten zur Einbeziehung opportunistischer Sensoren, d.h. von Geräten, die für andere Zwecke betrieben werden, aber nützliche Beobachtungen liefern, untersucht. Dieses WP erarbeitet des Weiteren Workflows für die Bereitstellung qualitätskontrollierter/gesicherter (nahezu) Echtzeitdaten (QC/QA-Methoden aus WP2), die als Grundlage z.B. für eine schnelle Datenexploration mit dem Ziel der Ereigniserkennung durch visuelle und rechnergestützte Methoden (WP2) für eine adaptive Überwachung verarbeitet werden müssen. WP1 definiert die Anforderungen der naturwissenschaftlichen Gemeinschaft an die datenwissenschaftlichen Aufgaben von WP2. Um Workflows in der Realität testen und gegebenenfalls verbessern zu können, ist WP1 sehr eng in die *MOSES* Feldkampagnen der Wirkungskette „Hydrologische Extreme“ integriert.

Task 1.1 Match monitoring/observation approaches with cross-compartmental needs

Task 1.1.1 List of observation parameters/proxie

In diesem Task sollten Beobachtungsparameter gemäß ihrer räumlichen und zeitlichen Repräsentativität generisch zusammengefasst werden, um somit die durch Sensoren zur Verfügung gestellten Daten besser einschätzen zu können. Folgende Punkte müssen dabei charakterisiert werden: Measurement Principle, Device Manufacturer, Measurement Requirements, Data, Quality Assessment, Calibration, Metadata, Measurements. Abbildung 1 gibt einen genaueren Überblick dazu.

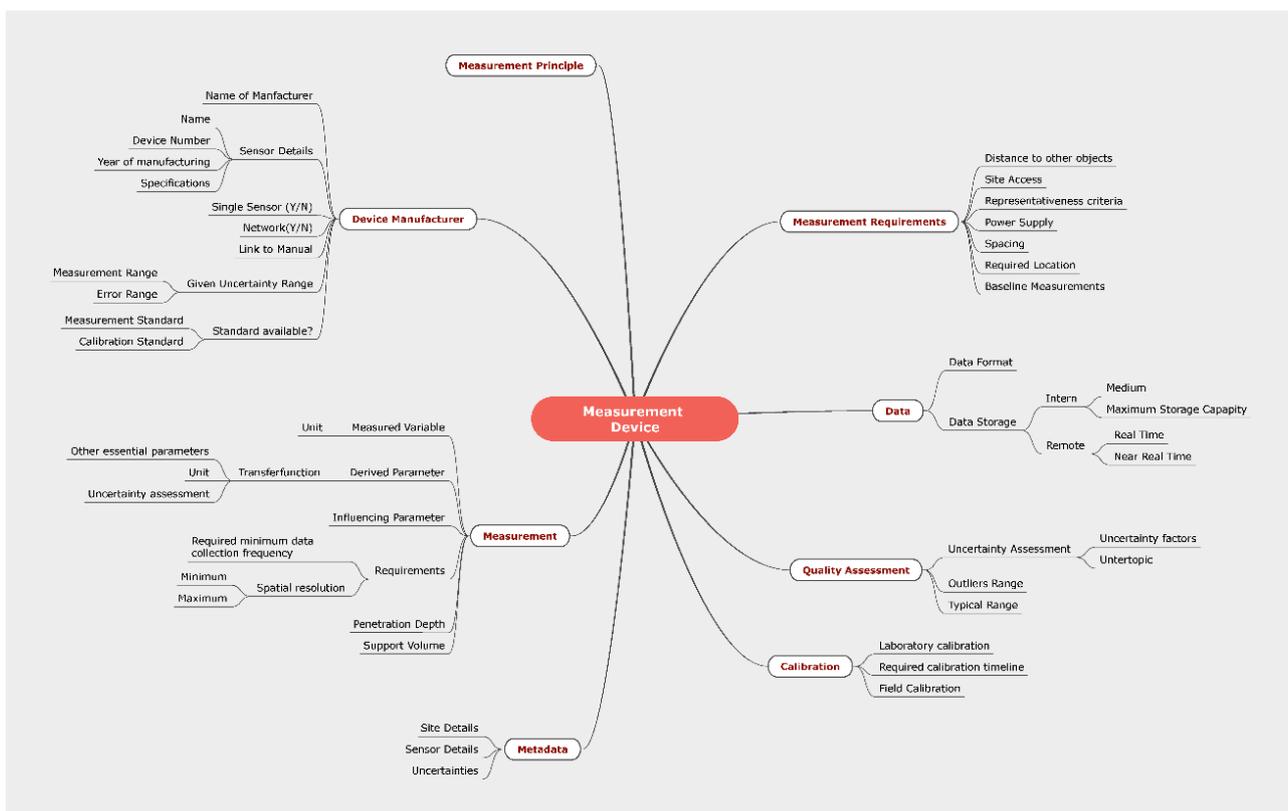


Abbildung 1: Abfragekriterien

Diese identifizierten Abfragepunkte sind als Beispiel zu sehen und dienen dazu, wichtige Informationen über ein Messgerät zu identifizieren und können so dazu beitragen:

- entsprechend der wissenschaftlichen Aufgabenstellung repräsentative Messschemata aufzustellen. So ist beispielsweise die Kenntnis der Eindringtiefe und des Einflussbereiches (Footprint) unerlässlich um Werte hoch zu skalieren. Ein hierarchisches Monitoringkonzept berücksichtigt diese, indem z.B. Messtechniken mit unterschiedlichen, aber ineinander übergreifenden Einflussbereichen miteinander kombiniert und beispielsweise durch Korrelationsanalysen Abhängigkeiten untersucht werden können.
- passende Messstrategien hinsichtlich zeitlich und räumlicher Auflösung und deren Footprint zu identifizieren. Dabei ist es wichtig das wissenschaftliche Ziel zu betrachten und der Aufwand-Nutzen sowie die Machbarkeit abzuwägen.
- automatisierte Datenverarbeitung zu ermöglichen und mögliche Schwachstellen (wie geräteinternes Pre-Processing) zu erkennen. Dabei werden auch die möglichen gerätespezifischen Messgrenzen miterfasst und dienen in der automatisierten QA/QC Analyse der physikalischen Plausibilitätsprüfung.
- mit Hilfe des Gesamtmessfehler die Zuverlässigkeit der Daten einzuschätzen.
- eine Einordnung der Messgeräte hinsichtlich ausgewählter Kriterien zu ermöglichen.
- Messkampagnen mit Hilfe der Aussagen zu Footprint, Auflösung und Messanforderungen zu planen.
- Metadatenabfragen zu standardisieren.

Ebenso ist die Kenntnis von Betriebsbereichen, Messbereichen und deren zeitliche Auflösung wichtig um automatisierte QA/QC Kriterien anzuwenden. Die herausgearbeiteten Punkte sind bei dem im Rahmen des Projektes *MOSES* erarbeiteten Datenmanagementtool eingeflossen und unterstützen ebenso die Neuausrichtung eines Sensorverwaltungstools welches durch das UFZ und GFZ entwickelt wird.

Task 1.1.2 Standard Operation Procedures (SOPs) for benchmarking of measurement

WP1 fokussierte seine bisherigen Arbeiten hauptsächlich auf Show Case A. Es ist klar, dass bei diesem Show Case, die räumliche und zeitliche Auflösung bei der Überwachung von Kompartiment übergreifenden Flutereignissen und dem resultierenden Oberflächenabfluss eine wichtige Rolle spielt. Die *MOSES* Kampagnen haben gezeigt, dass auslösende Starkniederschläge wenige Minuten bis Stunden andauern und lokal stark begrenzt sind. Dies kann wiederum eine schnelle und intensive Abflussreaktion aus einem Einzugsgebiet nach sich ziehen. Diese komplexen Verkettungen und ihre kaskadierenden Effekte machen eine Vorhersage sehr schwierig. Da Starkregenereignisse meist an bestimmte meteorologische Zirkulationsmuster und -abläufe gebunden sind, ermöglichten die kontinuierlichen atmosphärischen Beobachtungen, *Intensiv Operation Period's* innerhalb der 2019-Kampagne die Zeitpunkte an denen intensive Messungen stattgefunden haben, zu definieren. Um den hydrologischen Einfluss dieser Starkregenereignisse einschätzen zu können, wurden Bodenfeuchte, Abflüsse und Wasserstände an ausgewählten Punkten gemessen. Als Grundlage für ein Benchmarking von unterschiedlichsten Messungen wurden Informationen als unerlässlich herausgearbeitet. Diese sind in Tabelle 1 und Abbildung 2 dargestellt. Dabei ist die Qualitätssicherung der Daten als auch die Einschätzung der Datenverlässlichkeit bei Messungen auf unterschiedlichsten Skalen sehr wichtig, um die Bereitstellung der ermittelten Forschungsdaten an Dritte gerecht zu werden.

Um die Datenverlässlichkeit richtig einschätzen zu können und ein Benchmarking der Messungen zu ermöglichen, sind unterschiedlichste Anforderungen an Referenz- oder Vergleichsmessungen zu erfüllen. Während bei großskaligen Messkampagnen meist Vergleichsmessungen z.B. Satellitendaten mit bodengestützten Messungen verwendet werden, sind bei mesoskaligen und kleinskaligen Messkampagnen Sensorvergleiche über eine definierte Zeit am Ende und Anfang an einem definierten Ort notwendig.

Tabelle 1: Übersicht über notwendige Informationen, die als Basis für ein Benchmarking von verschiedenen Messdaten stehen.

Skale	Messungen	Datenverarbeitung
<i>Punkt</i>	<ul style="list-style-type: none"> • Footprintanalyse • Messfehlern (z.B. Instrumentenfehler) • Unsicherheiten (z.B. bei Proxi-Transferfunktionen) • Gerätedrift → Referenzmessungen an einem definierten Ort mit definierter Zeitdauer 	<ul style="list-style-type: none"> • temporale Variabilität • Ersetzen von Nan-Werten mit Hilfe geeigneter Data Imputation Methods • Drift- Korrekturfunktionen mit Hilfe der Referenzmessungen
<i>Kleinräumige</i>	<ul style="list-style-type: none"> • räumlicher Variabilität → Mittelwerte und Varianz • Vergleichbarkeit der Sensoren → Referenzmessungen an einem definierten Ort mit definierter Zeitdauer 	<ul style="list-style-type: none"> • räumliche Korrelationen innerhalb von Messgrößen → Kovarianzfunktionen (Semivariogramme) • Pedotransferfunktionen: Ableitung von Größen mit Hilfe von einfach bestimmbar Größen
<i>Großräumige</i>	<ul style="list-style-type: none"> • Vergleichsmessungen kleinskalig/grossskalig 	<ul style="list-style-type: none"> • Validation der Flächenmessdaten mit Interpolationsalgorithmen • Überprüfung der Korrelationen zu anderen Größen (z.B. Bodenfeuchte → Bodentemperatur)

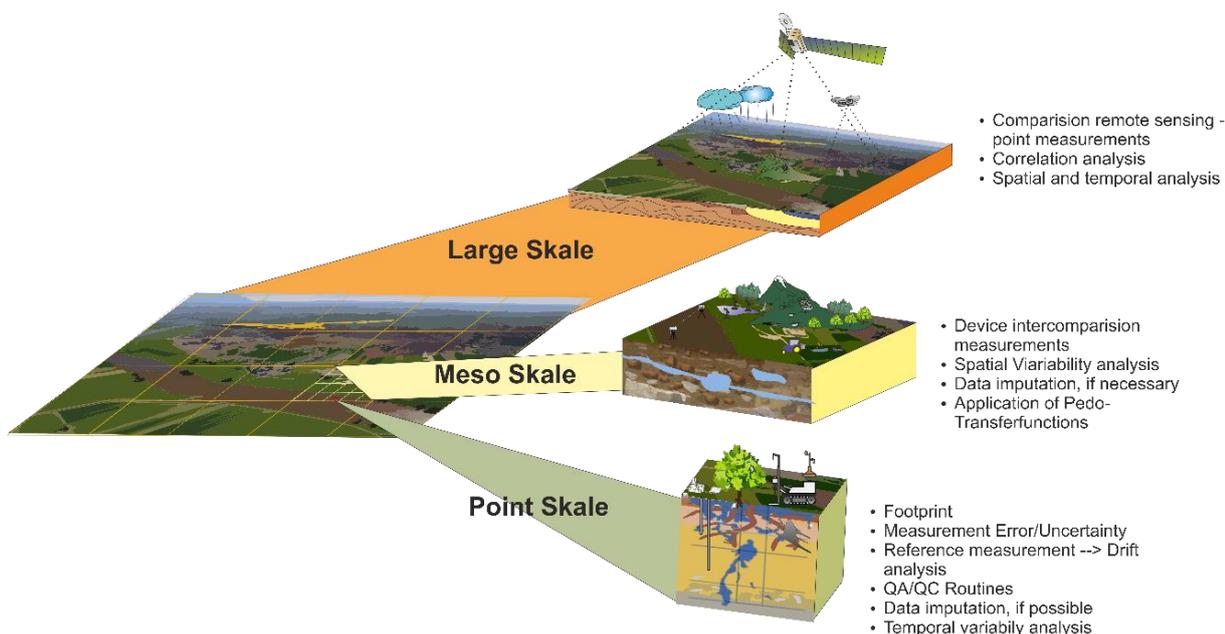


Abbildung 2: Anforderungen bei den verschiedenen Skalen

Die im Rahmen vom *MOSES* durchgeführten Messkampagnen z.B. innerhalb der Wirkungskette „Hydrologische Extreme“ zeigten, dass solche Vergleichsmessungen fest im Workflow integriert werden müssen. Im Rahmen der *MOSES* Aktivitäten wurden umfangreiche Sensorvergleiche zur Messung physikalischer, chemischer und biologischer Wasserqualitätsparameter durchgeführt. Die dadurch erkannten Probleme sollen durch neue Ideen, z.B. Etablierung von zwei Referenzsonden in den Häfen und bei den Schiffsfahrten durch die Mitnahme eines Sensors für Referenzmessung behoben werden. Ebenso wird an der Definition von notwendigen Metadaten-Informationen hinsichtlich der Benchmark-Messungen gearbeitet.

Ein weiteres Ziel war es, Empfehlungen zur Anpassung der aktuellen Systeme/Praktiken für die Erhöhung der Zuverlässigkeit von Daten zu erstellen und die Unsicherheiten dieser zu quantifizieren. Hierfür wurde der Begriff *SMART-Monitoring* im Rahmen des *Digital Earth* Projekts eingeführt und wie folgt definiert: gemessene Variablen und ihre abgeleiteten Parameter müssen für eine nachhaltige Nutzung als Daten spezifisch (**S**pecific), messbar (**M**easurable), akzeptiert (**A**ccepted), relevant (**R**elevant) und nachvollziehbar (**T**rackable) -SMART- sein. *SMART-Monitoring* heisst Überwachung des Datenflusses von einzelnen Sensoren zu Datenbanken unter Einbeziehung automatisierter (z.B. maschineller Lernprozesse) und echtzeitnaher interaktiver Datenanalysen/Exploration.

Um eine gemeinsame Datenanalyse zu ermöglichen spielt beim *SMART-Monitoring* der Umgang mit Metadaten eine entscheidende Rolle. Die Anwendung von standardisierten Schritten innerhalb dieses Flusses ist unerlässlich, um eine höhere Datenzuverlässigkeit zu erreichen. Folgende Tasks wurden definiert, welche im Rahmen der bisherigen Arbeit als besonders wesentlich im Umgang mit Metadaten identifiziert wurden:

- standardisierter Inhalt
- definierte Daten oder Formate und standardisierte Datenformattransformation
- standardisierte Routinen für Q/A-Routinen mit dem Ziel eine identische Datenkennzeichnung zu gewährleisten; standardisierte Routinen sollten den bestehenden Standardarbeitsanweisungen (SOPs), z.B. aus ICOS, entsprechen
- hierarchische Datenspeicherstruktur um relevante Hilfsdaten, Kalibrierprotokolle und Daten aus Gerätevergleichsexperimenten zu kombinieren
- standardisierte Verarbeitungsroutinen bei der Anwendung von statistischen Verfahren oder Proxy-Transfer-Funktionen
- standardisierte Berichterstattungsroutinen, um sicherzustellen, dass alle Schritte genau beschrieben und nachvollziehbar sind und dass alle Benutzer die Datenqualität der von den Proxy-Übertragungsfunktionen abgeleiteten Parameter beurteilen können
- standardisierte Visualisierungswerkzeuge

Deliverables und Milestones

Deliverables



D1.1.1 List of observation parameters/proxies

M6

Milestones

Keine entsprechend des Projektplans

Task 1.2 Improve Data-Flow Framework from Sensor to Data-Exploration

Innerhalb des *Forschungsbereiches Erde & Umwelt* besitzt die **Observation to Archive** Architektur des AWIs (O2A) eine Vorreiterrolle. Seine Infrastruktur zur Metadaten-Aufnahme, Datenarchivierung sowie deren Visualisierung steht allen beteiligten HGF Zentren innerhalb von *MOSES* und *Digital Earth* zur Verfügung. Im Berichtszeitraum wurden Schulungen bezüglich der Anlage von Sensoren im O2A angeboten. Durch die intensive Einbindung von WP1 in die *MOSES* Kampagnen wurden Punkte in den Metadaten identifiziert, deren Informationen bisher nicht vollständig oder gar nicht hinterlegt sind.

Task 1.2.1 Data and data products (e.g. acquisition of CML attenuation for countrywide precipitation estimation) provided

Satelliten-Fernerkundung bietet die Möglichkeit, kontinuierlich und global Informationen über bestimmte Eigenschaften der Erdoberfläche zu sammeln. Je nach wissenschaftlicher Fragestellung werden Daten zur satellitengestützten Erdbeobachtung von unterschiedlichsten Satellitensystemen (optisch oder Radarsysteme (SAR)) genutzt. Um eine umfassende Betrachtung zu ermöglichen, wurde sich im Folgenden auf Satellitendaten bezüglich der im Projekt gesuchten Größe Bodenfeuchte fokussiert. Bodenfeuchtheitsdaten können mit Hilfe der satellitengestützten Messung von Mikrowellen

(1mm – 1m Wellenlänge), die von der Erdoberfläche reflektiert oder abgestrahlt werden, ermittelt werden. Dabei hängt die Intensität des gemessenen Signals von der Wassermenge im Boden ab. Diese Mikrowellen durchdringen die Vegetation und Boden bis zu einer gewissen Tiefe. Dabei nimmt die Eindringtiefe mit der Wellenlänge des Messsignals zu. Bereits seit 33 Jahre gibt es passive und aktive Satellitenbeobachtungen hinsichtlich der Bodenfeuchtigkeit (z.B. SMOS, SMAP, ERS-1 and ERS-2, MetOp-A,B,C und Coriolis WindSat). Um eine Übersicht der weltweiten Aktivitäten zu erstellen wurden wichtige Betriebsparameter (z. B. Frequenz und Sensororientierung der Fernerkundung) hinsichtlich folgender Parameter definiert und in einer Tabelle zusammengefasst. Dabei wurden am Beispiel der im Projekt gesuchten Größe *Bodenfeuchte* folgende Parameter zur Charakterisierung von Satellitenbeobachtungen identifiziert:

· Parameter	· Sensing depth [m]
· Unit	· Revisit Time [days]
· Measured Parameter	· Accuracy
· Derived Parameter	· Spatial Resolution [m] (Product)
· Satellite	· Temporal Resolution (Product)
· Measurement Principle	· Time period of downloadable data
· Sensor description	· Uncertainty Assessment (Product)
· Operating Agency	· Link (Product)
· Operation Timeline	· Download Program available e.g. in Python (Product)
· Measurement Frequency	· References
· Polarization Direction Resolution [km]	

Diese umfassende Liste mit essentiellen Informationen steht allen Wissenschaftlern, die mit Remote Sensing Daten in *MOSES* und *Digital Earth* arbeiten zu Verfügung und wird regelmäßig aktualisiert und kann für andere Größen erweitert werden. Bezüglich der Größe *Bodenfeuchte* ist aus der in der Tabelle erarbeiteten Zusammenfassung zu erkennen, dass *Bodenfeuchte* zum einen bei passiven Verfahren aus der „Brightness Temperature“, und bei aktiven Verfahren aus dem „Backscatter“ abgeleitet wird. Die Eindringtiefe dieser satellitengestützten Verfahren liegt im Durchschnitt bei 5cm. Dies zeigt, dass eine Kombination anderen Messungen wie zum Beispiel mit mesoskaligen *Cosmic-Ray* Messungen, die eine Eindringtiefe von ca. 50cm erreichen, unerlässlich ist.

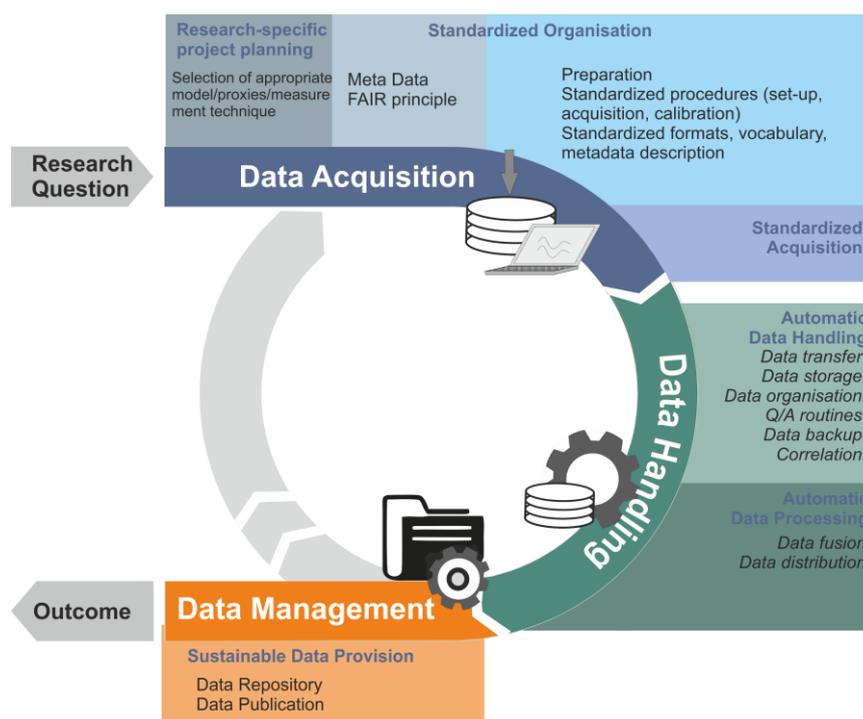


Abbildung 3: Datenfluss vom Sensor in die Datenbank (cc: Uta Ködel)

Eine weitere Aufgabe war es eine Standardisierung und Automatisierung des Datenflusses vom Sensor in die Explorationsumgebung vorzunehmen und des Weiteren erforderliche Metadateninformationen (z.B. Datenerfassungsparameter oder lokale Einstellungen) zu definieren. Der Datenfluss vom Sensor in die Datenbank kann in drei Schritte eingeteilt werden: (i) die Datenakquisition, (ii) das Datenhandling und (iii) das Datenmanagement (Abbildung 3). Besonders in den ersten beiden Schritten sind standardisierte Vorgehensweisen unerlässlich, um verlässliche Daten generieren zu können.

Die *FAIR*-Prinzipien bieten eine Anleitung zur Verbesserung der Auffindbarkeit, Zugänglichkeit, Interoperabilität und Wiederverwendbarkeit aller digitalen Ressourcen. Es ist wesentlich, maschinell bearbeitbare Prozesse in Dateninfrastrukturen zu unterstützen, um die Ressourcen für Maschinen und damit auch für Menschen auffindbar zu machen. Diese *FAIR*-Prinzipien zielen darauf ab, den größtmöglichen Nutzen aus Forschungsdaten zu ermöglichen und erfordern keine wissenschaftliche Datenqualität. Insbesondere für Datenanalyseaufgaben müssen alle Arten von Daten mit einem umfassenden Metadaten-Satz beschrieben werden.

Tabelle 2: Zusammenstellung der notwendigen Informationen zu Metadaten entsprechend dem *FAIR*-Prinzip.

Kategorie	Metadaten
Sensor-Sensor-Vergleich	<ul style="list-style-type: none"> • Informationen über das Vergleichsexperiment (Wo, Wann, Dauer, Bedingungen, welche Geräte) • Sind Referenzpunkte verfügbar? • Umfassende Bewertung der Unsicherheiten
Anwendung der Proxy-Transfer-Funktionen	<ul style="list-style-type: none"> • Beschreibung der Ausgabe (Einheit, zeitliche und räumliche Auflösung) • Notwendige Hilfsdaten (Beschreibung, Auflösung, Unsicherheitsbewertung) • Unsicherheitsabschätzung (z.B. Fehlerfortpflanzungsanalyse) mit allen relevanten Informationen
Zuverlässige Karten	<ul style="list-style-type: none"> • Daten (Anbieter, Zeitpunkt der Aufzeichnung, Quelle der aufgezeichneten Daten und der damit verbundenen Metadaten, Lizenzierung, Lebensmitteldruck, räumliche Datenabdeckung, Dichte der Beobachtungen) • Kartenproduktion (Zeit, Hersteller, Kartenquelle, Projektion, Koordinatensystem, Erklärung der Produktionsschritte/Techniken, Lizenz und Verfügbarkeit) • Genauigkeit (z.B. Position und Techniken) • Unsicherheit

Eine falsche Interpretation und Verwendung dieser Daten können in einer gemeinsamen Interpretation zu falschen wissenschaftlichen Schlussfolgerungen führen. Deshalb muss die Datenqualität ausreichend beschrieben sein. Im Rahmen von *Digital Earth* wurden drei relevante Kategorien betrachtet, mit dem Ziel, die Datenverlässlichkeit dieser drei Kategorien für eine bessere gemeinsame Nutzung der Daten gemäß den *FAIR*-Prinzipien zu erhöhen. Dabei wurden Metadaten identifiziert, die für jede dieser Kategorie unerlässlich sind (Tabelle 2).

Die Eingangsparameter in das *O2A* wurden hinsichtlich der Vollständigkeit analysiert und eine Liste mit notwendigen Informationen in Tabelle 3 zusammengestellt. Diese Informationen dienen als Basis um zukünftige Messkampagnen adäquat planen und um die so generierten Daten vergleichen zu können.

Tabelle 3: Übersicht Informationskatalog Eingangsparameter für O2A.

Punkte	Abfrage	AWI O2A	Beschreibung	Notwendig für
Hersteller des Gerätes		O2A: Manufacturer (Overview)	Es gibt Geräte mit gleicher Funktionsweise von unterschiedlichen Hersteller	Datenanalyse
Anschaffungsjahr	Pull down Menue (Jahr)		Es gibt Versionen von Geräten, daher wichtig	Datenanalyse
Upload Funktionen	Paper/ Foto/ Manual	O2A: Ressources	Manual, Fotos und relevante Paper sollten hinterlegt werden und bei Bedarf von anderen einsehbar sein	Datenanalyse
Zeitstempel (Koordinierte Weltzeit (UTC)?	Y/N (Differenz zu UTC)	In O2A Festlegung, alle Zeiten UTC	Abfrage, bei gemeinsamen Kampagnen wichtig um eine Vergleichbarkeit zu erzielen, ebenso ist es bei manchen Geräten nicht möglich Zeit zu ändern	Datenanalyse
Positionsdaten (Latitude, Longitude), evtl. Höhe	Auswahl geräteinterne Erfassung/ anderes Geräte (was, evtl. Abweichung); Angabe in welchem Format (Dezimal Degree/anderes)	In O2A als Parameter hinterlegt	Abfrage, werden geräteinterne Positionsdaten erfasst oder Verweis auf anderes Instrument welches die gleichen Postionsdaten erfasst, da evtl. Abweichung angeben	Q/A Analyse (Überprüfung des Standortes), Datenanalyse
Geforderte Standortanforderung einer Messung	Mindestplatz/Zufahrt (z.B. mit PKW, LKW); Funknetzabdeckung/Stromversorgung (s.u.); Geländetopographie/Notwendigkeit einer Abgrenzung (Zaun) - Ausgabe und Angabemöglichkeiten anderer Informationen		Zur späteren Kampagnenplanung (z.B. man fügt bei neuen Kampagnen die entsprechenden Geräte hinzu und verfügt so über alle wichtigen Informationen)	Kampagnenplanung
Stromversorgung	(autark /intern/extern; z.B. Batterie/Solar) /Netzversorgung /anderes), evtl. notwendige Leistung?		Stromversorgung ist ein wichtiger Punkt bei der Standortwahl	Kampagnenplanung als auch zur Auswertung (Q/A)
Unterbrochene Stromversorgung	definierter Wert als Rohdatenfile (Y /N)			Q/A Auswertung, Datenanalyse
Datentransfer of raw data	vor Kampagne (manuell/automatisch)/ während Kampagne (real time/ near real time); Angabe Dateiformat (z.B. csv, tab sep .txt, ncd)	In O2A Ingest (after expedition by scientist/central transmission, continuously real time etc.)		Kampagnenplanung und Datenanalyse
Zusätzliche Erfassung wichtiger Einflussparameter	Y (gekoppelt/ anderes Gerät) /N	in O2A als Parameter	evtl. (automatische) Verknüpfung mit anderen Sensoren	Kampagnenplanung, Datenanalyse, Ableitung von Parametern
Angabe der Messunsicherheiten/Ranges	Gerätespezifische Messunsicherheit (in SI Einheit oder % des Messwertes)/ Operational Range (in welchem Bereich sind Messwerte zu erwarten)	In O2A Parameter Properties (Manufacturer Range, Operational Range, Thresholds)		automatische Q/A Analyse
Angabe der Messrate	Taktung (s, min, h)/ geräteinterne automatische Mittelung (Y/N)			Datenanalyse, Q/A Analyse
Angabe der Eindringtiefe/ Footprint/	defined penetration depth (Y (Angabe)/N/ not relevat) / defined footprint (Y (Angabe)/N/ not relevant)	In O2A Parameter Properties (Measurement resolution)	Eingabe der gerätespezifischen Parameter evt. Upload-Funktion zu wichtigen Paper	Datenanalyse, Upscaling
Möglichkeit eines Messwertdrifts	Y (bekannt vor Messung/ Analyse nach Messung möglich)/N	In O2A Parameter Properties (genaue Einschätzung vor der Kampagne?)	wie ein Drift aussieht kann manchmal erst nach Abschluss eingeschätzt werden, evtl. gibt es schon SOP	Datenanalyse; Q/A Analyse
Abgeleitete Parameter	Parameter mit Einheit/Formel/zusätzliche genutzte Inputparameter	O2A (in Planung): Parameterdescriptor ; wie wurde der Parameter berechnet, umgewandelt, etc. primär oder abgeleiteter Parameter?	manche Messvariablen wie z.B. Spektren sind nur durch Berechnungen für andere verständlich, meist existieren verschiedene Transferfunktionen, daher Angabe der genutzten Formel und zusätzlicher Parameter notwendig, nachvollziehbare Verlinkung mit Rohdaten notwendig	Datenanalyse und zum Bereitstellungs von Daten laut FAIR
Notwendige Kalibration des Messgerätes vor Messeinsatz	Laborkalibration (Y(SOP, Zeit, Daten wo)/N) /Feldkalibration (Y (bei Kampagne, SOP, Daten wo) /N)	O2A: Actions und resources	das standardmäßig eine Werkskalibrierung durchgeführt wird ist anzunehmen (diese Daten sollten auch irgendwo hinterlegt werden); Abfrage ob zusätzliche Kalibration wichtig und Informationen	Datenanalyse
Geplante Vergleichsmessung	Y (Angabe Lokation/definierte Zeit/ welche Geräte)/N	O2A: Actions und resources	Bei größeren Kampagnen mit verschiedenen Messgeräten gleiche Parameter messen; Vergleichsmessungen an einem gemeinsamen Ort über eine vorher definierte Zeit sind essentiell; weitere Informationen sollten nach der Kampagne eingetragen werden (Pflicht!); die Daten der verschiedenen Geräte sollten für diesen definierten Zeitraum abrufbar sein	Datenanalyse

Task 1.2.3 Final set of concepts, standards and algorithms for data flow from sensor to archive compiled and documented

Diese Überprüfung erfolgt im Rahmen der MOSES Kampagnen.

D1.2.1 Data and data products (e.g. acquisition of CML attenuation for countrywide precipitation estimation) provided

Abdämpfungsdaten von knapp 4000, über ganz Deutschland verteilten, kommerziellen Richtfunkstrecken (CMLs) werden mithilfe eines ‚Simple Network Management Protocols‘ (SNMP) erfasst. Einmal pro Minute werden momentane TX und RX Werte aller 4000 CMLs im privaten Netzwerk des Betreibers Ericsson über deren IP-Adresse abgefragt und auf einen Server in diesem privaten Netzwerk abgelegt. Von dort werden die Daten in Echtzeit auf einen Server des KITs übertragen. Die Latenzzeit beträgt dabei circa 20 Sekunden. Diese Erfassung läuft seit August 2017 mit einer Frequenz von einer Minute. Die Software hierfür ist frei zugänglich (<https://github.com/cchwala/pySNMPdaq>). Aus den so gewonnen Daten erfolgt eine Abschätzung der gemittelten Regenrate entlang jedes CMLs.

D1.2.2 Pre-operational, documented Data-Flow Framework system established

Das Data-Flow Framework O2A wurde als ein Netzwerk aus mehreren Quellen, Institutionen und Abteilungen eingerichtet. Digital Earth nutzt O2A als einen Dienst, der Anwendungen und zusammengesetzte Dienste für Kollegen aller Helmholtz-Zentren und deren Anwendungen bereitstellt.

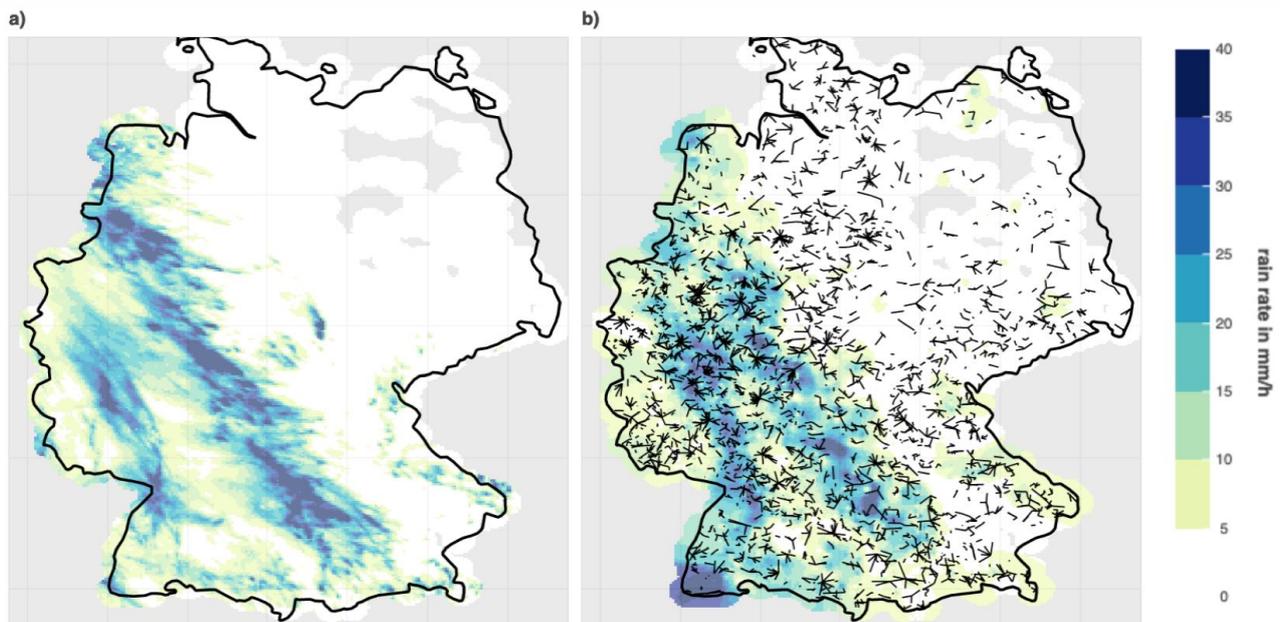


Abbildung 4: Akkumulierte Niederschlagssumme über 48 Stunden von 12.05.2018 - 14.05.2018 für a) RADOLAN-RW (operationelles Radarprodukt des Deutschen Wetterdienstes) und b) von CMLs abgeleitete Regenraten (linien-integrierte Regenraten wurden mittels inverser Distanzen Interpolation interpoliert. Regionen die weiter als 30 km von einem CML entfernt sind wurden ausgegraut. (cc: Maximilian Graf)

O2A ist in Betrieb, wird aber kontinuierlich weiterentwickelt und konzentriert sich momentan auf die Verwaltung heterogener Quellen durch die Sensor-Webanwendung. Im Herbst (5.12.2019) fand am AWI ein Workshop zur Schulung der Sensor-Webanwendung und der Dateneinspeisung in die echtzeitnahe Datenbank des O2A statt.

Deliverables und Milestones

Deliverables



D1.2.1 Data and data products (e.g. acquisition of CML attenuation for countrywide precipitation estimation) provided

M12

Milestones



M1.2.1 Initial set of concepts, standards and algorithms analysed and selected

M11

Task 1.3 Close data gaps and optimize data acquisition for adaptive hierarchical monitoring

Task 1.3.1 Guidelines implementation of hierarchical monitoring concept

Es ist unerlässlich, mathematische und statistische Tools zu verwenden, um zuverlässige Stichprobenpunkte, Korrelationsfunktionen oder Wechselbeziehungen zwischen Modelldaten, vorhandener Überwachungsdaten und anderer Hilfsdaten zu identifizieren. Dies ist z.B. Notwendig um Upscaling-Funktionen abzuleiten. Dabei tragen Methoden des SMART-Monitorings dazu bei, Sensoreinstellungen und Überwachungsstrategien in Zeit und Raum durch iterative Rückkopplung besser anzupassen, um auch diese Daten zu liefern, die für Modellierungsläufe oder Validierung benötigt werden. Folgende zeitlichen und räumlichen Anforderungen sind hinsichtlich eines umfassenden Monitorings wichtig:

Zeitliche Überwachung:

- kontinuierliche *baseline*-Messungen
- automatische Identifizierung von Events (Messintervall ober-/unterhalb bestimmter Thresholds)

Räumliche Überwachung:

- Datenanalyse von vorhandenen Mess-, Zusatz-, und Modelldaten zur Identifizierung von repräsentativen Messlokalitionen
- Kopplung von unterschiedlichen Footprints der Messungen zur Gewährleistung von der Messung von gemeinsamen
- Identifizierung repräsentativer Probenahmepunkte durch Fuzzy C-Means Cluster Algorithmus

Diesbezüglich bietet die Vielfältigkeit der zur Analyse verfügbaren Daten (Fernerkundungsdaten, Modelldaten) neue Möglichkeiten, z.B. für die Abschätzung zeitlicher Bodenfeuchtigkeitsvariationen. Dabei ist klar, dass die räumliche Überwachung der Bodenfeuchte in kleinen Einzugsgebieten (0,1-1 km²) eine Herausforderung darstellt, und herkömmliche *in-situ* Bodenfeuchtemessungen nach wie vor unverzichtbar sind. Hierfür ist die Erstellung von technisch und wirtschaftlich machbaren Messdesigns bei gleichzeitigem maximalem Informationsgehalt essentiell. Abbildung 5 zeigt einen Workflow für die Identifizierung von repräsentativen Probenahmepunkten.

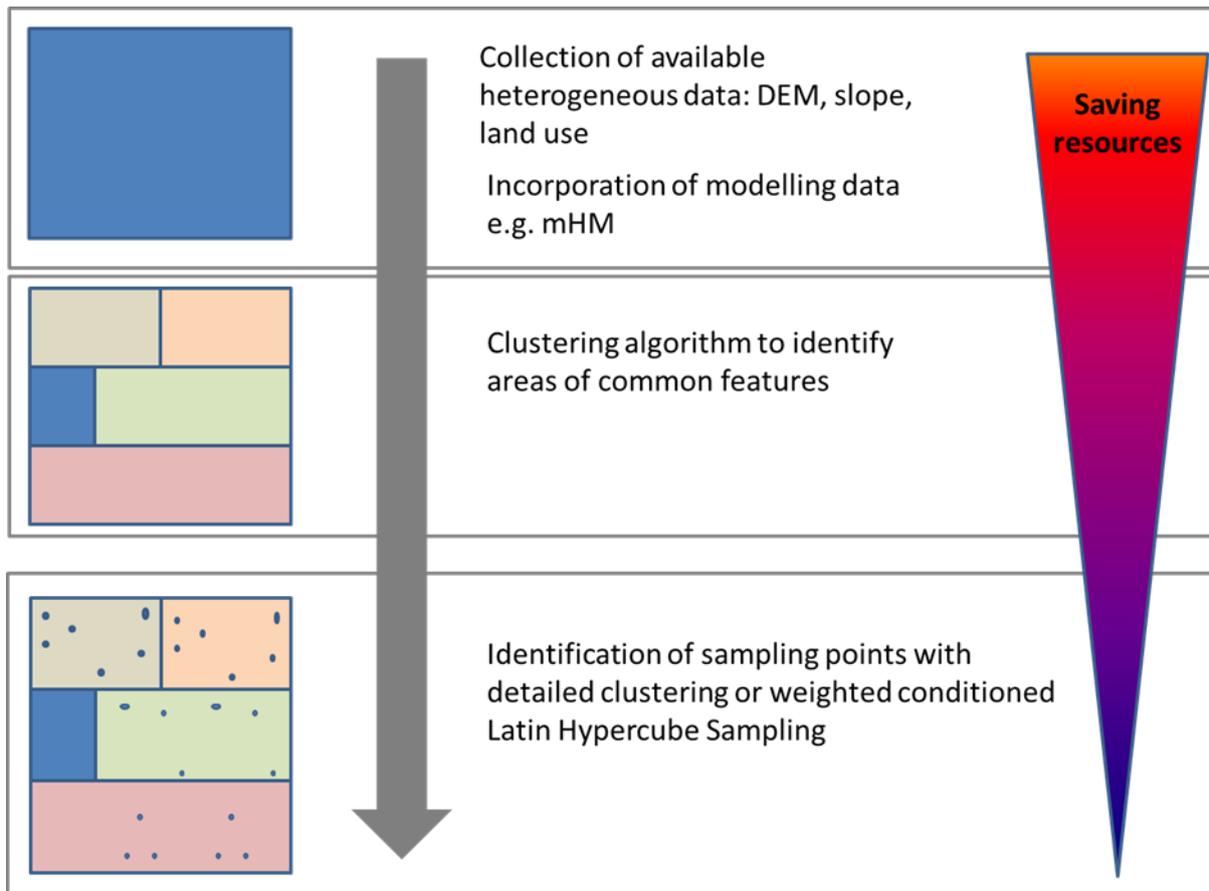


Abbildung 5: Workflow für die Auswahl von repräsentativer Probenahmepunkte

Zu den Herausforderungen bei der Erstellung solcher strategischer Probenahmepläne für *in-situ* Bodenfeuchtemessungen gehören: (i) Wie kann *a priori* eine angemessene Anzahl repräsentativer Messstandorte definiert werden?; und (ii) Wie kann unter Verwendung von limitierten Punktmessungen und weitere vorhandene Zusatzinformationen regional die Bodenfeuchtemuster für ein Zielgebiet so abschätzt werden, dass die abgeleiteten räumlichen Informationen aus Fernerkundungsdaten oder Modellergebnissen vergleichbar sind?

Im Rahmen der Arbeiten in *Digital Earth* wurde die *Fuzzy C-Means* (FCM) Cluster-Algorithmus nach Paasche et al., 2006 angewendet, um *a priori* eine begrenzte Anzahl von repräsentativen Probenahmepunkten zur Überwachung der Bodenfeuchtigkeitsdynamik in einem kleinen Einzugsgebiet zu identifizieren. Die FCM-Clustering-Technik kann angewendet werden, um ein Einzugsgebiet, beispielsweise basierend auf einer Reihe von Proxies für bodenfeuchtigkeitssteuernde Faktoren, in eine bestimmte Anzahl von Clustern zu unterteilen.

Das Ergebnis der Anwendung des FCM zeigt Abbildung 6. Dabei wurden auf der einen Seite Gebietsinformationen wie das Höhenmodell, die Neigung als auch Landnutzungsdaten und auf der anderen Seite Modelldaten des mesoscaligen hydrologischen (mHM)-Modells verwendet.

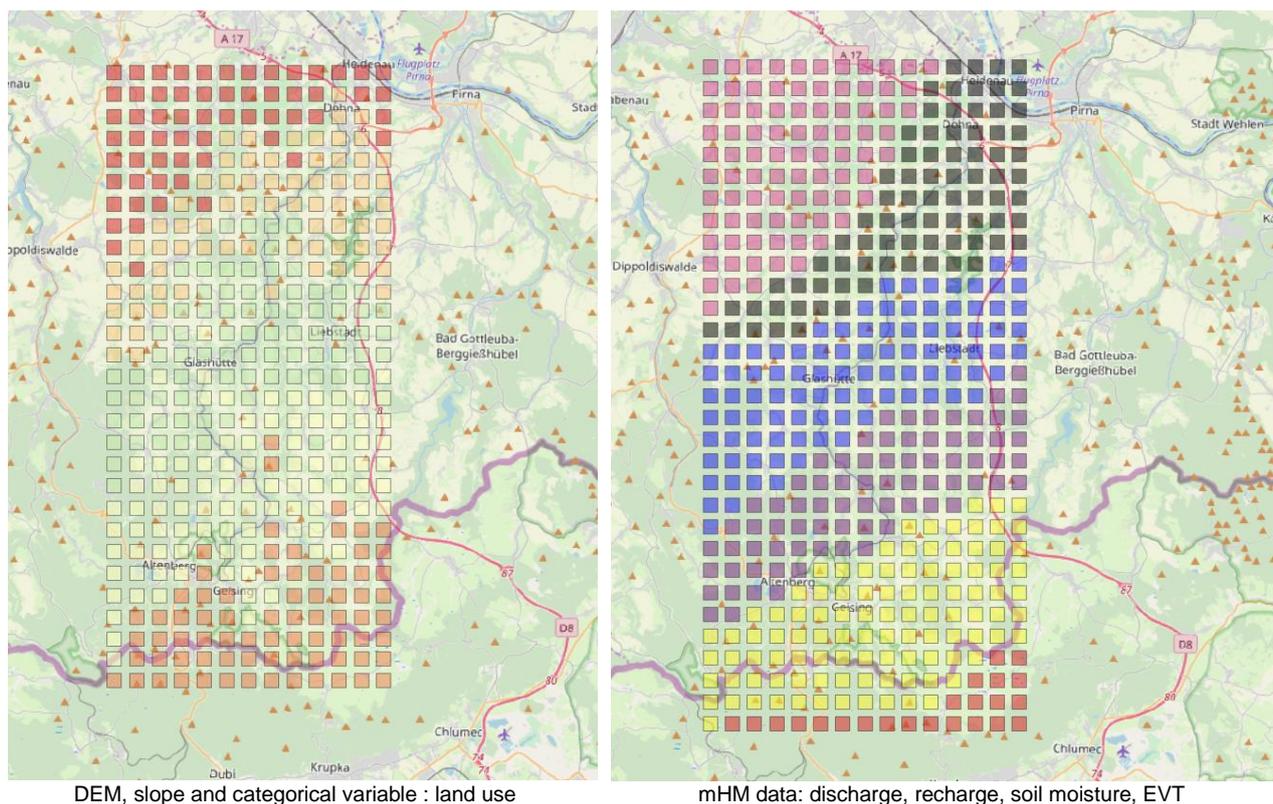


Abbildung 6: Gebieteinteilung (Cluster) nach Anwendung des *Fuzzy-C-Means Clustering* im Gebiet des Müglitztal's

Im Rahmen eines *Bridging Postdocs* wird dieses Verfahren weiter vorangetrieben und beispielhaft aus der meereslithologischen Kartierung und der Detektion von Massivsulfid am Meeresboden (SMS), räumlich erfasste Bathymetriedaten und gleichzeitigen magnetischen Anomalie-Daten miteinander im dem Algorithmus kombiniert, um eine räumliche Segmentierung des Meeresbodens zu erhalten. Die Anwendung dieses Algorithmus ermöglicht die Identifikation potenziell unbekannter Standorte und die Optimierung von Beobachtungsstrategien.

In *Digital Earth* wird ein integratives, hierarchisches Monitoringkonzept angewendet, in dem Methoden und Technologien aus verschiedenen Disziplinen entweder kombiniert oder komplementär eingesetzt werden (siehe Abb. 7). Dieses Konzept ermöglicht eine konsistente Erfassung auf großen räumlichen Skalen ohne die räumliche und zeitliche Auflösung zu beeinträchtigen. Große Flächen werden meist mit *Remote Sensing* Verfahren erkundet, welche aber nur auf eine begrenzte Eindringtiefe basieren und die Messdaten oft indirekt oder mehrdeutig durch sogenannte *Proxy-Transfer* Funktionen mit dem Zielparameter (z.B. Bodenfeuchte) verknüpft sind. Der Vorteil eines hierarchischen Monitoringkonzeptes liegt in der Datenintegration/-verknüpfung. Das bedeutet, dass zu klären ist, inwieweit Punktmessungen mit Messungen auf der meso- bzw. großen Skala verknüpft werden können. Solche Verknüpfungen können mit Hilfe von sogenannten *footprints*, der Auflösung und Unsicherheiten sowie der Referenzpunkte analysiert werden. In Abbildung 8 ist der Workflow eines hierarchischen Monitoringkonzeptes dargestellt. Durch *in situ* Analysen können diese Konzepte angepasst werden. Für dieses Konzept ist das Zusammentragen von sämtlichen verfügbaren Daten und die Identifizierung bestmöglicher Messstandorte durch z.B. oben beschriebene *Fuzzy-C-Means Clustering* unerlässlich.

Für die Entwicklung und Erprobung von hierarchischen und modellbasierten Monitoringkonzepten wurde das MOSES-Testgebiet Müglitztal als Referenzgebiet ausgewählt. Zunächst erfolgte das Zusammentragen verschiedenster bereits verfügbarer Daten, sowie deren Aufbereitung als Grundlage für die Erprobung von im Rahmen von *Digital Earth* entwickelten Tools zur Datenanalyse und zum Messdesign. Danach wurden verschiedene Messungen in unterschiedlichen Skalen durchgeführt. So wurden zum Beispiel Punktmessungen für die Bodenfeuchte und mesoskalige Messungen mit dem *Cosmic Ray* Verfahren durchgeführt. Die Ergebnisse beider Messverfahren zeigen eine hohe Korrelation und besitzen daher ein sehr hohes Potential miteinander kombiniert werden zu können.

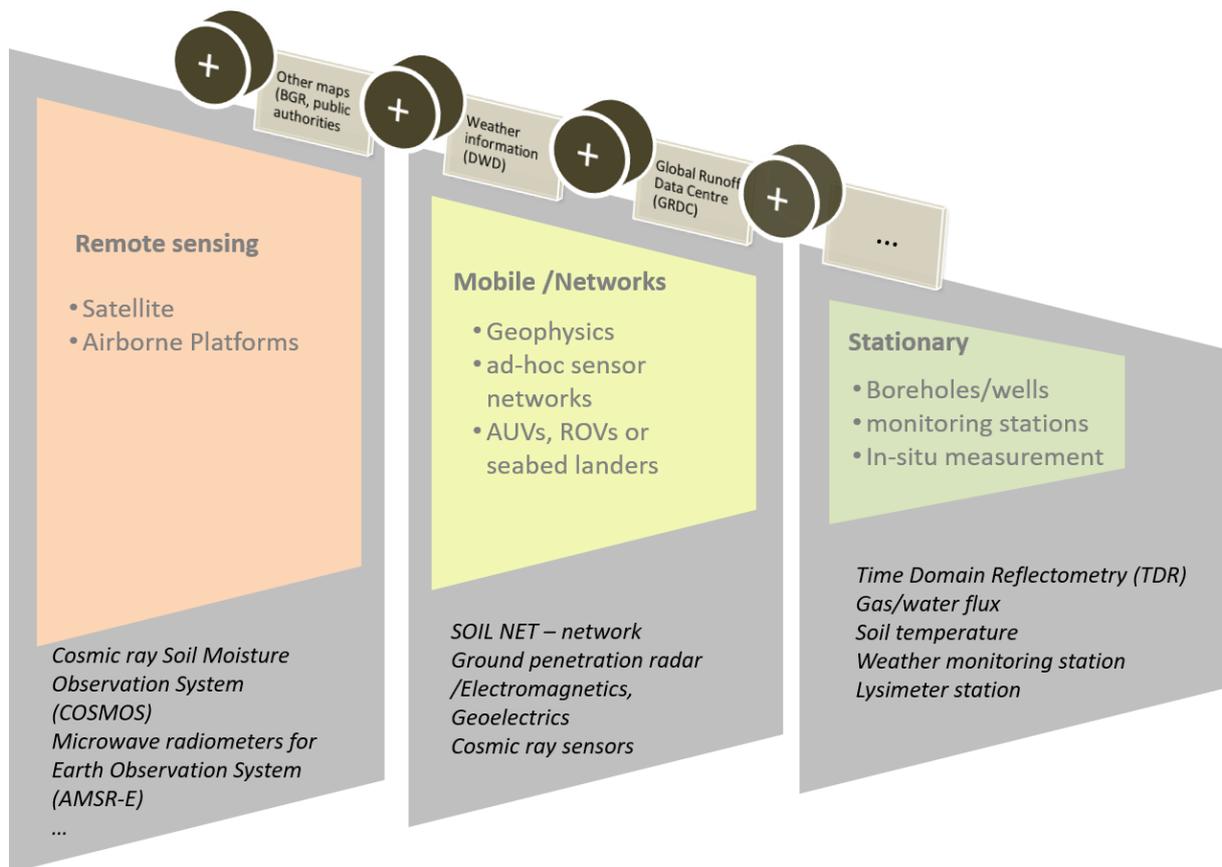


Abbildung 7: Beispiele für Messmethoden eines hierarchischen Monitoringkonzeptes (cc: Uta Ködel)

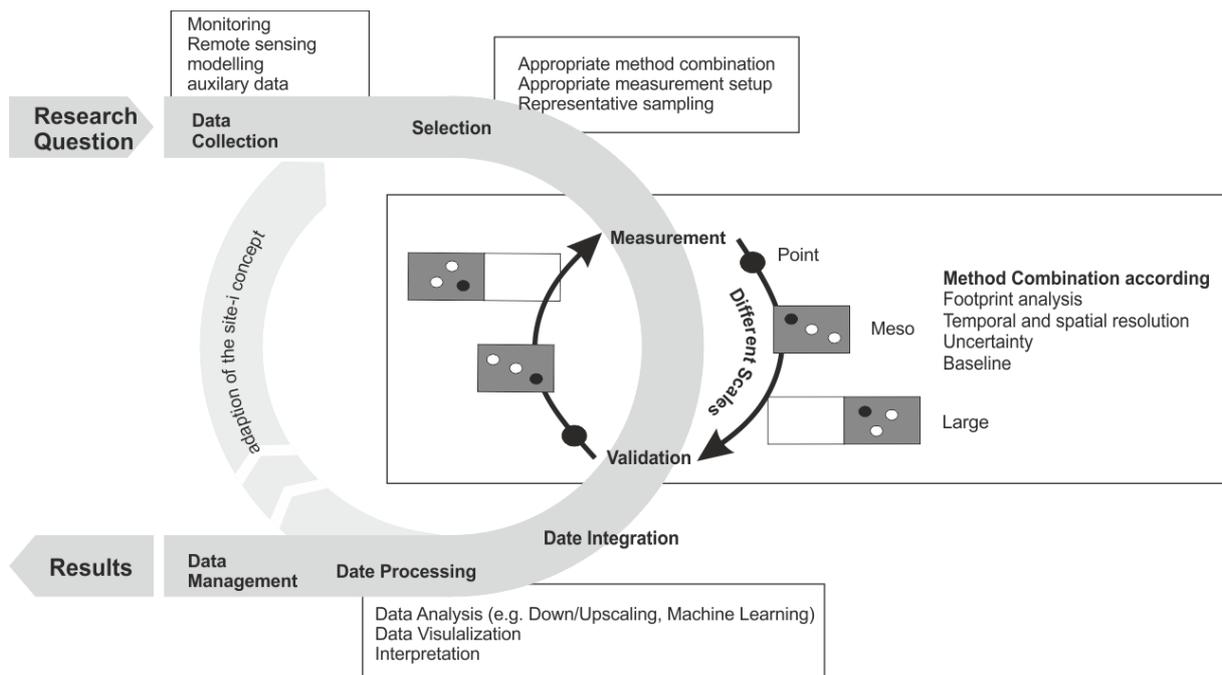


Abbildung 8: Ablauf eines hierarchischen Monitoring Konzeptes (cc: Uta Ködel)

Deliverables und Milestones

Deliverables

Keine entsprechend des Projektplans.

Milestones



M1.3.1 Workshop about challenges and opportunities for hierarchical monitoring concepts (finalized)

M6

Theoretischer Teil wurde während des 1. Interim Meetings im Dezember 2018 absolviert. Mit dem praktischen Teil während des Jahrestreffens im Mai 2019 in Novy Kostel wurde der Workshop abgeschlossen.

WP2 Data Exploration Framework

Mit dem *Data Exploration Framework* sollen Wissenschaftler disziplinübergreifende Datenanalysen mit *state-of-the-art* und neuen *Data Science* Verfahren durchführen können, um ein datengetriebenes Verständnis über große, skalen- und domainübergreifende Fragestellungen wie auch kompartment- und domainspezifische Fragestellungen zu entwickeln. Im Berichtszeitraum wurden dafür moderne computerbasierte Datenanalyseverfahren aus Statistik, Künstlicher Intelligenz und interaktiver 2D bis 4D+ Visualisierung evaluiert, angepasst und angewendet. Außerdem wurde das Konzept von Workflows von der konzeptionellen wissenschaftlichen Analyseketten bis hin zur technischen Umsetzung in digitale Workflows eingeführt und in ersten Teilen umgesetzt. Auch wurden gemeinsame Standards für die komponentenbasierte und kollaborative Software-Entwicklung festgelegt und implementiert. Für das Datenmanagement wurden keine eigenen Entwicklungen durchgeführt. Es wird als sinnvoller erachtet, mit bereits existierenden Initiativen und Infrastrukturen zu kooperieren und deren Ergebnisse mitzugestalten, auf *Digital Earth* anzupassen und zu nutzen.

Als exemplarische wissenschaftliche Fragestellungen für die Entwicklung des *Data Exploration Framework* wurden zwei Show Cases definiert, für die zum einen Workflow-Konzepte entwickelt und Methoden getestet werden, und die zum anderen zur Evaluation der Umsetzbarkeit, Akzeptanz und Nützlichkeit von *Digital Earth* Ergebnissen dienen.

Task 2.1 Discuss workflows, practices, demands and solutions

Für die Fragestellungen aus den Show Cases wurden Workflows systematisch in Form von wissenschaftlichen Analyseketten beschrieben und diese in ersten Schritten in digitale Workflows überführt. Für die technische Realisierung des *Data Exploration Frameworks* wurde entsprechend des in **Task 3.4** entwickelten Softwarekonzepts, ein plattformunabhängiger, modularisierter Ansatz gewählt. Dies soll die Wiederverwendbarkeit und Kombinierbarkeit von Softwarekomponenten ermöglichen.

Für den *Fluten Show Case* wurden folgende fünf Workflows definiert; sie werden ganz oder in Teilen in einem *Flood Event Explorer* implementiert (siehe Abb. 9).

1. Similarity of floods: Untersuchung und Bewertung der Ähnlichkeit zwischen einem Flut-Referenzereignis (z.B. aktuell oder historisch) und einer Liste von Vergleichsereignissen (historisch) auf der Grundlage relevanter Hochwasserereignisdaten

In diesem Workflow wurde ein Ansatz entwickelt, um ein beliebiges Hochwasserereignis gegen historische oder zukünftige Ereignisse zu bewerten. Dies schliesst die gesamte Prozesskette von den auslösenden atmosphärischen Bedingungen und hydrologischen Vorbedingungen bis zu den Hochwasserfolgen ein. Der Ansatz umfasst die letzten 70 Jahre und bezieht sich auf einen Hochwasserereigniskatalog für Deutschland. Dieser bietet die Grundlage, um ähnliche Hochwasserereignisse identifizieren und vergleichen zu können, z.B. auf der Basis von hydrologischen Vorbedingungen, Gebietsniederschlägen, Abflussganglinien und Überschwemmungskarten. Der Workflow, als Teil des *Flood Event Explorers*, ermöglicht die Analyse

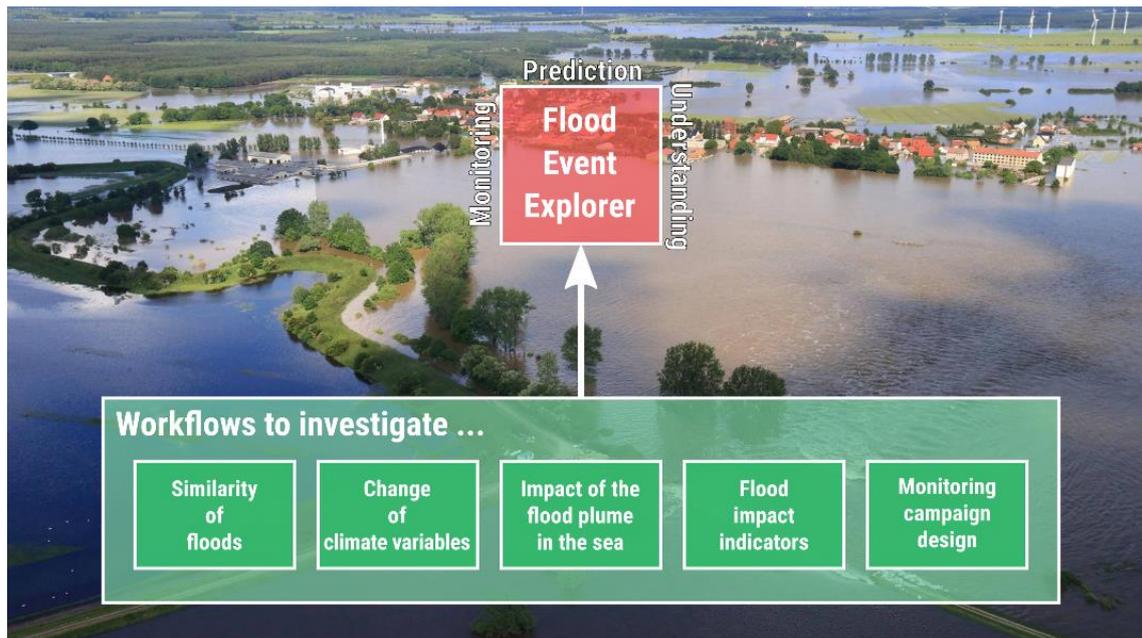


Abbildung 9: Flood Event Explorer - unterstützte Workflows

mehrdimensionaler Hochwassereigenschaften sowie aggregierter Indikatoren (in Raum und Zeit), räumlicher Muster und Zeitreihensignaturen. Die Ähnlichkeit zwischen den Hochwasserereignissen wird mit einer Vielzahl von statistischen Methoden, maschinellen Lernalgorithmen und visueller Datenexploration bewertet (siehe **Task 2.3 und 2.4**).

2. Change of climate variables: Quantifizierung und Bewertung von Veränderungen von Klimavariablen und damit verbundener Ereignisse wie z.B. Fluten oder Dürren durch den Vergleich von Ergebnissen aus verschiedenen Simulationen

Das Ziel dieses Workflows ist die Bewertung von Änderungen relevanter Klimavariablen (Niederschlag, Temperatur, u. a.) und daraus abgeleiteter Ereignisse (z.B. Starkregen). Die allgemeine Bewertung beruht dabei auf der Quantifizierung der Änderungen über klimarelevante Zeiträume auf Basis von Klimamodellsimulationen für historische und zukünftig projizierte Zeiträume. Eine detaillierte Bewertung erfolgt mittels einer Analyse von aus den Modelldaten abgeleiteten Ereignissen. Die Analyse erfolgt über die Auswertung der raum-zeitlichen Signifikanz und Robustheit der abgeleiteten Ereignisse, sowie über deren Eigenschaften mittels statistischer und visueller Methoden (siehe **Task 2.3**). Dieser Workflow wurde bereits als wissenschaftlicher Workflow beschrieben und in einen digitalen Workflow überführt. Umgesetzt wurde bisher die allgemeine Bewertung auf Basis der klimarelevanten Zeiträume. Als Datenbasis wurden zunächst Simulationsdaten des regionalen Klimamodells *REMO* verwendet. Für zeitlich hochaufgelöste Daten wurde ein Testdatensatz in die *AWI rasdaman* Instanz eingespielt. Die verwendeten Analysemethoden wurden dabei gemäß des Architekturmodells als Backend-Module umgesetzt und in den Workflow eingebunden. Erste Deployment-Tests fanden bereits statt. Zurzeit werden die umgesetzten Analysemethoden für definierte Szenarien getestet und weiterentwickelt. In der finalen Version sollen die Daten direkt aus dem *ESGF* Archiv geladen werden können. Dabei soll auf Datenbasis von Ensemble-Simulationen, wie beispielsweise der *EURO-CORDEX* Initiative zugegriffen werden. Die Umsetzung des Ansatzes zur Detailbewertung auf Basis von abgeleiteten Ereignissen erfolgt in 2020.

3. Impact of the flood plume in the sea: Detektion der Elbfahne in der Nordsee während und nach einem Hochwasser (*Anomaly Detection*) zur Quantifizierung von Nähr- und Schadstoffeinträgen und deren langfristige Auswirkung auf das marine Ökosystem

Dieser Workflow ist am Ende einer Flut-Ereigniskette angesiedelt. Exemplarisch werden die Auswirkungen der Elbe-Flut vom Juni 2013 auf das marine Ökosystem der Nordsee untersucht. Ziel ist es, die Elbfahne nach einem Flutereignis in der Nordsee zu detektieren, ihre zeitliche und räumliche Ausbreitung zu untersuchen und die Folgen der durch die Flut in die Nordsee

geschwemmten Nähr- und Schadstoffe abzuschätzen. Aktuell ist es möglich, im digitalen Workflow die Elbfahne mittels Ferrybox-Trajektorien zu detektieren, ihre Ausbreitung durch Modelldaten zu approximieren und durch zusätzliche, ausgewählte Beobachtungsdaten zu verifizieren (siehe **Task 2.3**). In der nächsten Entwicklungsphase sollen weitere Beobachtungsdaten aus einem Datenportal zur Verfügung gestellt und automatisch dem zeitlichen und räumlichen Verlauf der Elbfahne zugeordnet werden. Die Elbfahne wird momentan in den Ferrybox-Trajektorien per Nutzer-Interaktion detektiert und soll um eine Funktion zur automatischen Kennzeichnung auffälliger Gebiete erweitert werden. Zur Quantifizierung der Einflüsse der Elbfahne auf das Ökosystem der Nordsee existiert bereits die Möglichkeit zur Durchführung einer einfachen statistischen Analyse der Unterschiede zwischen Elbfahne und dem umgebenden Meerwasser. Zukünftig wird dies durch die Analyse von weiteren Beobachtungsdaten ergänzt werden. Ziel des Workflows ist es, durch das gezielte Kombinieren verschiedener Datentypen, z.B. Ferrybox-, Satelliten- und Modelldaten aus verteilten Datenquellen, das Ausmaß der Elbfahne und ihre Auswirkungen besser erfassen, analysieren und beobachten zu können.

4. Flood impact indicators: Bewertung der treibenden Faktoren und Auswirkungen aktueller und zukünftiger Hochwasserereignisse durch Indikatoren

Das Ziel dieses Workflows ist es Indikatoren zu entwickeln, mit denen die treibenden Faktoren und Auswirkungen aktueller und zukünftiger Hochwasserereignisse bewertet werden können und somit die Interaktionen im System 'Mensch-Hochwasser' über die Grenzen der Kompartimente 'Klima und Atmosphäre', 'Einzugsgebiet und Flussnetz', und 'Sozio-Ökonomie' hinweg besser zu verstehen. Der Ansatz sieht vor, Daten aus den genannten Kompartimenten zu integrieren und somit neue Einblicke in die Zusammenhänge zu gewinnen. Treibende Faktoren stellen Größen dar, die die Ausprägung und die Auswirkung von Hochwasser beschreiben, beispielsweise Niederschlag, Bodenfeuchte und Schutzmaßnahmen. Die zu entwickelnden Indikatoren repräsentieren einzelne oder aggregierte treibende Faktoren und werden zur Bewertung von Ereignissen verwendet. In der ersten Phase des Workflows werden geeignete Indikatoren identifiziert. Dafür werden Daten aus den verschiedenen Kompartimenten zusammengeführt. Dieser Schritt greift auf Beobachtungsdaten (z.B. Niederschlag, Abflusszeitreihen) und *Hindcast*-Ergebnisse von Simulationsmodellen (z.B. Bodenfeuchte, Überflutungsflächen, Hochwasserschäden) zurück. Als räumliche Abgrenzung dient das Einzugsgebiet der Elbe. Der Betrachtungszeitraum orientiert sich an der Verfügbarkeit von Beobachtungsdaten und dem Auftreten von Hochwasserereignissen in der Vergangenheit. Es stehen Informationen zu Hochwassern seit dem Jahr 1950 zur Verfügung. Beobachtungsdaten umfassen die Aufzeichnungen an den Klima- und Niederschlagsstationen des Deutschen Wetterdienstes, die Pegelaufzeichnungen des Wasser- und Schifffahrtsamtes sowie der hydrologischen Messdienste der Bundesländer. *Hindcast*-Simulationen mit räumlichen und zeitlichen Daten zu Niederschlägen, Bodenfeuchte und Oberflächenabfluss stehen aus dem *MHM* Modell bereit. Diese liegen in einer Auflösung bis zu 1km x 1km als Tageswerte vor. Weitere *Hindcast*-Simulationen liegen aus Evaluationssimulationen des Regionalmodells (*REMO*) für den Zeitraum von 1979 bis 2012, angetrieben mit *ERA*-Interim Daten, als Tageswerte vor. Ein wichtiger Schritt ist die Zusammenführung der genannten Daten und die Analyse, ob für die Simulationsdaten von *MHM* und *REMO* ein Zusammenhang mit dem Auftreten, der Ausprägung und den Auswirkungen von Hochwassern hergestellt werden kann. Abhängig von den Erkenntnissen aus Phase 1 werden in der zweiten Phase aktuelle und zukünftige Hochwasser bewertet. Die Anwendung der Indikatoren auf aktuelle Ereignisse ermöglicht eine schnelle Einordnung von zu erwartenden Hochwasserauswirkungen. Die Anwendung auf zukünftige Ereignisse gibt Einblicke in die in der Zukunft zu erwartenden Veränderungen von Hochwasserauswirkungen.

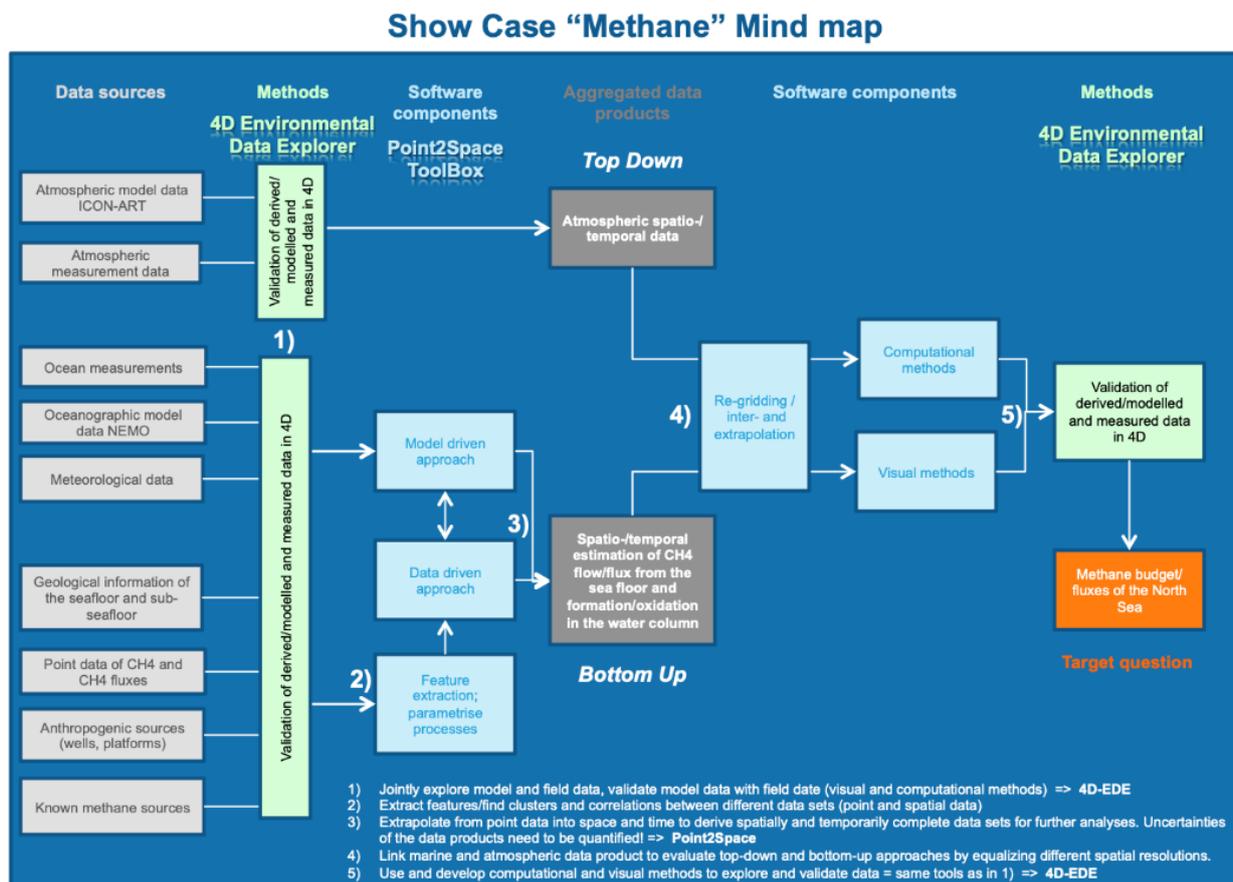
5. Monitoring campaign design: Ermittlung geeigneter Zeitpunkte und Orte für Ereignis-getriebene (near realtime, NRT) Datenerhebung z.B. in der Hydrologie

Kern dieses Workflows ist die Entwicklung und Umsetzung eines Ansatzes zur raum-zeitlichen Exploration von Einflussfaktoren zur Ermittlung eines geeigneten Zeitpunktes und Ortes für geowissenschaftliche Messkampagnen. Die Planung von geowissenschaftlichen Messkampagnen berücksichtigt Faktoren wie bereits vorhandene Messdaten, das Vorliegen spezieller

Umweltbedingungen, wie Oberflächenbedeckung, sowie Daten aus Vorhersagen, wie bspw. der zu erwartende Niederschlag. Der entwickelte Workflow soll es ermöglichen, die für eine spezifische Messkampagne relevanten Daten für die räumliche Verteilung zu explorieren und zu bewerten, um einen optimalen Ort für eine zukünftige Messkampagne zu ermitteln. In einem zweiten Schritt soll eine Analyse von projizierten Daten (e.g. Niederschlagsprognose, Verfügbarkeit von Satellitendaten, bspw. zur Nutzung als globale Referenzwerte) die Ermittlung eines geeigneten Zeitpunktes für das Durchführen von Messungen ermöglichen. Die Bewertung der räumlichen Verteilung basiert dabei auf der gewichteten Integration von gefilterten Eingangsdaten, was den Fokus der aktuellen Umsetzung darstellt.

Für den *Methan Show Case* wurde ein Workflow entwickelt, der aus einem ozeanographischen und einem atmosphärischen Teil besteht. Letztere können dann kombiniert werden und zu einer Schätzung der Methanemissionen in der Nordsee führen (siehe Abb. 10). Zur Umsetzung dieses Workflows wurden verschiedene Toolboxen und Methoden definiert:

- **Konventionen für Daten:** Da Daten aus verschiedenen Quellen zusammengeführt werden, werden Regeln für die Formate der Daten benötigt um diesen Prozess zu vereinfachen und Fehler zu vermeiden. Hierzu zählen unter anderem die Verwendung von einheitlichen Variablenamen. Dabei wurde sich an den *CF Conventions* und Datenformaten orientiert.
- **Point2Space Toolbox:** Um atmosphärische und ozeanographische Daten zusammenbringen zu können, muss aus den ozeanographischen Messwerten, die als Punktdaten vorliegen, zunächst ein Rasterdatensatz erzeugt werden. Dazu wird die *Point2Space* Toolbox verwendet. Dabei werden zwei verschiedene Ansätze verfolgt: (A) ein modellbasierter Ansatz und (B) ein datenbasierter Ansatz (siehe **Task 2.4**).



RR/13.02.2020

Abbildung 10: Workflow für den Methan Show Case

- **Interpolation/Re-gridding Toolbox:** Da die Modelldaten nicht auf dem gleichen Raster vorliegen, ist eine Transformation der Daten notwendig. Dazu wurde die Interpolation/Re-gridding Toolbox verwendet. In ihr werden verschiedene Methoden implementiert, die die verschiedenen Datensätze zusammenführen können. Dabei finden sowohl klassische Ansätze wie lineare Interpolation als auch neue Methoden (siehe **Task 2.4**) Verwendung.
- **Visuelle Validierung von Modelldaten durch Messwerte:** Im Rahmen des Workflows werden verschiedene neue Datensätze algorithmisch erzeugt. Außerdem werden Modelldaten verwendet, für die reelle Messwerte vorliegen. Um diese Modelldaten und Messwerte zu vergleichen wurde der *4D Data Viewer* konzipiert, der beide Datentypen gleichzeitig darstellen kann und so visuelle Vergleiche ermöglicht (siehe **Task 2.3**).

Für die Konzeption, Dokumentation und Umsetzung digitaler Workflows setzt *Digital Earth* auf innovative Technologien aus der Informatik. So wird die moderne Softwareentwicklungsplattform (*Gitlab*, siehe Abb. 11) genutzt. Weitere innovative Technologien (*Apache Pulsar*) ermöglichen a) die Entwicklung entkoppelter und leicht wiederverwendbarer Softwaremodule, b) die Integration von bereits existierenden Softwaremodulen (bspw. als Python, R oder Bash Skripte) und c) die beliebige Kopplung von Softwaremodulen (*Method-Chaining*). Die effektive Zusammenarbeit von Fachwissenschaftlern und Informatikern bei der Entwicklung der digitalen Workflows wird ebenfalls durch innovative agile Methoden und Techniken (*Continuous Integration CI*) unterstützt. Die Definition klarer Schnittstellen für Daten und Methoden ermöglichen die Wiederverwendbarkeit von erstellten Softwaremodulen und Datenarchiven.

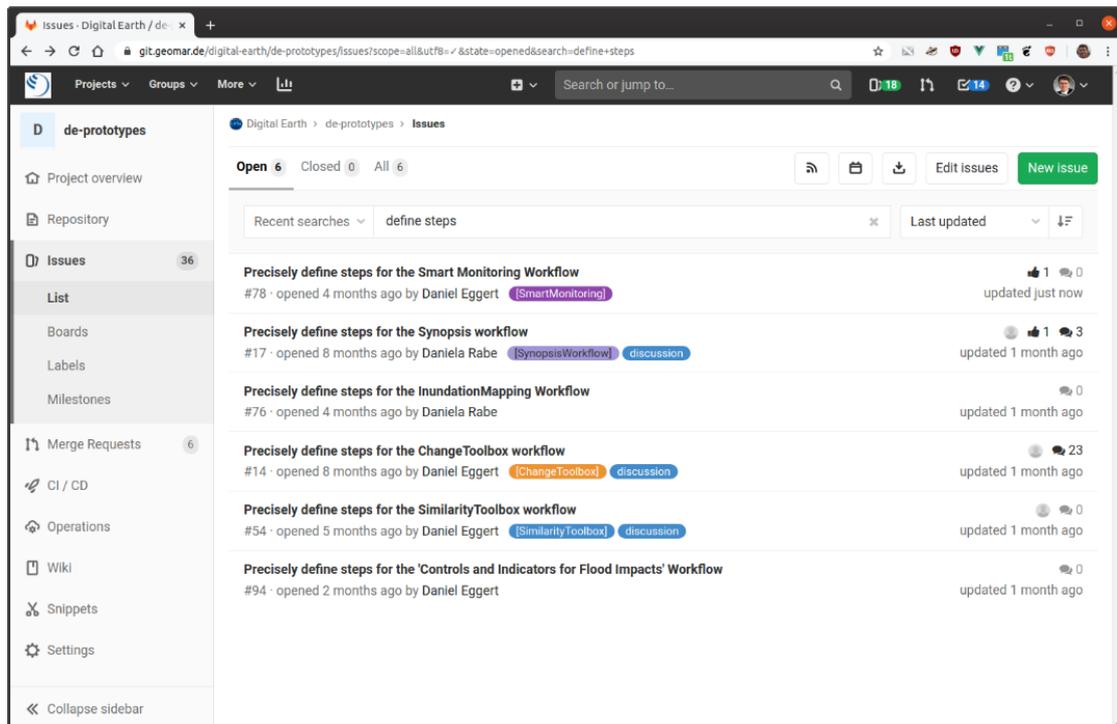


Abbildung 11: Screenshot *Gitlab Issue Tracker*

Deliverables und Milestones

Deliverables

D2.1.1 Model compilation of traditional and enhanced scientific workflows M12

Milestones

M2.1.1 Methodical gaps are identified M11

M2.1.2 Suitable methods are identified M12

Task 2.2 Data provision and quality

Für den datengetriebenen Ansatz von *Digital Earth* und das *Data Exploration Framework* ist es essentiell, eine umfassende Datenbasis und -infrastruktur zu etablieren, die die erforderlichen Daten schnell und einfach zur Verfügung stellt. *Digital Earth* betreibt dazu keine eigenen Entwicklungen, sondern kooperiert mit anderen Initiativen (z.B. NFDI) und baut auf deren Ergebnisse auf (siehe dazu **Task 3.1**). Für die nächste Projektphase in *Digital Earth* ist es geplant, konkrete Kooperationen mit den *NFDI Data Hubs* und Helmholtz Plattformen zu initiieren. Unabhängig von diesem langfristigen Konzept wurde für die Projektphase eine Entwicklungsumgebung geschaffen, bei der die für die Show Cases erforderlichen Daten beim AWI gesammelt und über Standardschnittstellen und Services zur Verfügung gestellt werden. Zukünftig sollen weitere existierende Dateninfrastrukturen genutzt werden.

Die "*List of data products*" gibt einen Überblick über die für die *Digital Earth* Show Cases relevanten und derzeit verfügbaren standardisierten Datenservices und Speicherschnittstellen. Gestartet wurde die Liste mit der Erfassung der Datenprodukte, die bei der Umsetzung der Workflows eingesetzt bzw. exploriert werden sollen. Obgleich ein Hauptaugenmerk auf diesen Datenprodukten lag, wurde und wird die Liste themenübergreifend weiter ausgebaut um eine mögliche hohe Parameterabdeckung zu gewährleisten. Zum aktuellen Zeitpunkt liegen neben *Digital Earth*-Forschungsdaten auch Ergebnisse aus Internet- und Literaturrecherchen zu öffentlich zugänglichen Datenressourcen vor. Gelistet wurden und werden auch die Zugriffsmöglichkeiten, wie beispielsweise Daten-Kataloge mit Bestellmöglichkeit (z.B. ftp-Download) oder auch Web-Dienste, die OGC-standardisierte Datenservices wie z.B. *WebFeatureServices* (WFS) bereitstellen.

Eine große Herausforderung für die Nutzung und Analyse von Daten ist es, die Daten bezüglich ihrer Qualität bewerten zu können. Ein Schwerpunkt gilt hier der Sicherstellung bzw. der Überprüfung von Daten(services) sowie deren Qualitätsbeschreibungen (*Garbage in - Garbage out*). Es wurden verschiedene Ansätze in Kooperation mit anderen Initiativen entwickelt und implementiert:

Die webbasierte Sensoranwendung des *O2A-Systems* wurde zur Unterstützung von Datenberichten und Kennzeichnungsprotokollen verwendet. Innerhalb des *O2A-Systems* können die Datenberichte durch technische Dokumentationen sowie Einsatzparameter beschrieben werden, und diese können über die gesamte Zeit verfolgt werden. Obwohl das System bereits für *Digital Earth* eingesetzt wird, wurde es kontinuierlich weiterentwickelt und um neue Verbesserungen ergänzt. Zum Beispiel ist das derzeit verwendete Schema zur Qualitätskennzeichnung sequentiell und qualitativ, d.h. es beschreibt ein Qualitätsniveau der Daten. Ein neues Kennzeichnungssystem wird derzeit entwickelt, um ein beschreibendes Merkmal aufzunehmen, das technische- und Nutzer-Interpretation umfasst. Um dies zu erreichen, wurde eine Überprüfung verschiedener Kennzeichnungssysteme (z.B. *ARGO*, *GTPPS*, *OOD*) durchgeführt, die die Definition eines codebasierten Systems unterstützen und die erforderliche Verbindung zwischen Nutzerbedürfnissen sowie den technischen Anforderungen herstellt. Qualitätskennzeichen inklusive Fehlerbeschreibung resultieren in einer Aussage zur Analysefähigkeit von Daten und stellen die für die Rückverfolgbarkeit erforderlichen Informationen zur Verfügung. Das in der Entwicklung befindliche Qualitätskategorisierungssystem, wird in Anlehnung an den allgemein verwendeten *UNESCO*-Standard (Zwischenstaatliche Ozeanographische Kommission der *UNESCO* 2013) entwickelt. Die technischen sowie konzeptionellen Aspekte dieser Entwicklungen wurden auf *Digital Earth* Meetings und internationalen Konferenzen zur Diskussion gestellt (z.B. *REKLIM*, 2019).

Parallel wurde die Konzeptionierung eines integrierten Frameworks für eine variable, automatisierte Qualitätskontrolle von Daten eruiert. Es wurde recherchiert, ob Workflow-Konzepte mit integrierten Prozessierungsschritten und abspeicherbaren Zwischenprodukten bereits existieren und ob diese mit ausreichend Metadaten zur Nachvollziehbarkeit von Prozessen beschrieben sind. Dabei überzeugte zunächst das im Rahmen der *CUAHSI Initiative* entwickelte *WaterShare Tool*. Dieses kam bereits in einer ersten Version im Rahmen der Forschungsinfrastruktur *TERENO/eLTER* zur Anwendung. Die Übernahme des Tools ist nach erfolgreichem Abschluss für *Digital Earth* geplant. Darüber hinaus wurde ein Server zur Bereitstellung von *Jupyter* Notebooks in Betrieb genommen, um die Entwicklung von

Software-Bibliotheken zur automatisierten Qualitätskontrolle (*Jupyter* Notebooks bzw. *WaterShare*) fortzuführen.

Für das automatisierte Datenqualitätskontrolle-Tool *AutoQA4Env* wurde eine standardisierte Architektur entwickelt. Diese beinhaltet für mehrere Ebenen Datenqualitätstests: *qc-read* testet beispielsweise die Rohdaten während des Einlesens, *qc-check* führt statistische Prüfungen der finalen Daten durch, *qc-score* bewertet die Resultate von *qc-check* und ordnet Qualitäts-Flags in *qc-flags* zu, die finale Ausgabe wird im abschließenden Schritt *qc-outputs* bewertet (siehe Abb. 12). Die Vorstellung und Diskussion von *AutoQA4Env* erfolgte bei einem *Digital Earth* Workshop Ende Februar 2019. *AutoQA4Env* soll im Verlauf von *Digital Earth* weiterentwickelt und bei den einzelnen Partnern implementiert werden. Das Qualitätskontrolle-Tool ist eine zentrale Komponente im digitalen Workflow für die Datenerfassung und -speicherung, der in den operationellen schon existierenden Forschungsinfrastrukturen *LAGOS* und *TERENO/eLTER* in ersten Teilen umgesetzt wurde. Der automatisierte Workflow umfasst die Datenströme von den einzelnen Sensoren im Feld oder an Bord von Flugzeugen zur Empfängerstation und von dort weiter zu den Instrumenten – Datenbanken, wo die Daten verarbeitet werden. Das hier entwickelte Konzept kann später auf andere Sensorumgebungen von *Digital Earth* Partnern angepasst werden. Basierend auf dem Workflow der Forschungsinfrastrukturen wird ein *Best-Practice* Dokument entwickelt, das allen *Digital Earth* Partnern als Blaupause für die jeweilige Implementierung dienen kann.

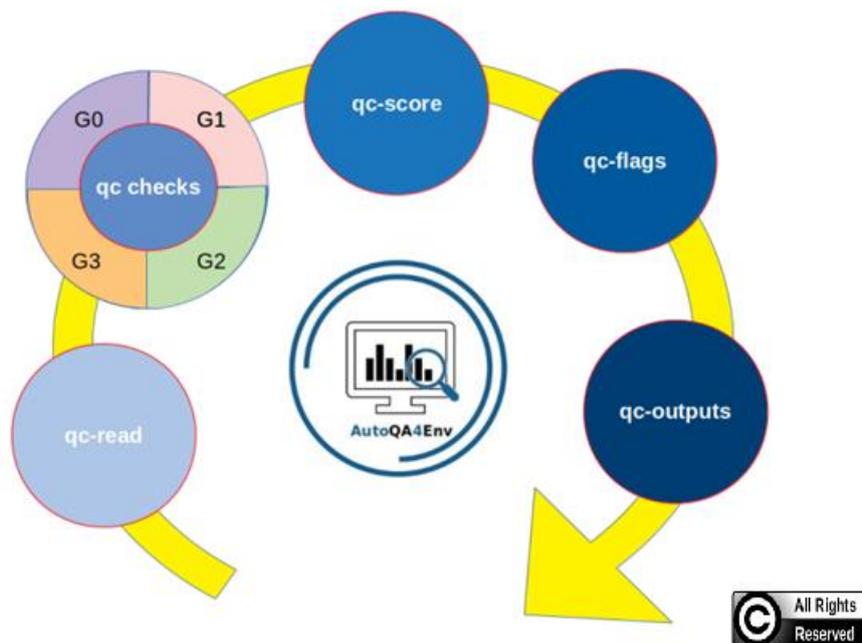


Abbildung 12: Schematischer Workflow von *AutoQA4Env*

Deliverables und Milestones

Deliverables

- 👍 D2.2.1 List of data products M6
- 👍 D2.2.2 Concept of harvesting and data fusion M12
- Die Fertigstellung dieses Konzepts hat sich verzögert, da das Deliverable 2.2.3 Code base for standardized QA/QC methods vorgezogen wurde.
- 👍 D2.2.3 Code base for standardized QA/QC methods M36

Milestones

- 👍 M2.2.1 Definition of standards for data reporting and flagging protocols including traceability and documentation M12

Task 2.3 Visual data exploration

Für die Anforderungen der Show Cases und die in **Task 2.1** definierten Workflows wurden mehrere Datenviewer und interaktive Explorationswerkzeuge entwickelt; auch wurde bestehende Software zur wissenschaftlichen Visualisierung bezüglich der Anforderungen getestet.

In einem ersten Schritt wurden die Anforderungen für die Visualisierungen identifiziert und beschrieben: wofür werden Visualisierungen gebraucht, und welche Datentypen müssen visualisiert werden? Diese Anforderungsanalyse ist die Basis für die Auswahl und Kombination geeigneter Visualisierungsmethoden und -software.

Die wissenschaftliche Arbeit im Bereich Erde und Umwelt verwendet Daten, die auf einer Vielzahl verschiedener Sensoren und Modellen beruht. Diese Daten sind in der Regel heterogen und variieren stark in ihrer zeitlichen und räumlichen Auflösung und Verfügbarkeit. Die Variation reicht von einzelnen Punktdaten bis hin zu Modelldaten, die einen dreidimensionalen Raum abdecken und sich über lange Zeitspannen erstrecken. Ziel ist es, diese Daten gleichzeitig zu analysieren, beispielsweise um Unterschiede zwischen Messdaten und Modelldaten zu finden. Diese Analyse kann sowohl algorithmisch als auch visuell erfolgen. Die Menge der einzelnen Datenquellen, die aus verschiedenen Beobachtungen und Modellen stammen und für unsere Anwendungsfälle relevant sind, ist zu groß, um sie aufzulisten oder um sie in einer Lösung für die visuelle Exploration zu implementieren. Stattdessen wurde eine Datenquellenkonzeption entwickelt, die die Datenquellen nach ihrer zeitlichen und räumlichen Dimension gruppiert. Dieses Konzept deckt die Gesamtheit der möglichen Datentypen ab, die im Projekt verwendet werden:

Die erste Kategorie beschreibt Beobachtungen, die als zeitlich konstant (für nicht-geologische Zeitskalen) betrachtet werden können und daher nicht an eine zeitliche Komponente gebunden sind. In unserem Fall beschränken sich diese auf geologische Karten, die ein Breiten-Längen-Raster mit Geländehöhenwerten, z.B. Bathymetrie oder Topologie, umfassen. Beobachtungen von einem einzigen räumlichen Punkt aus, die ihre Werte im Laufe der Zeit ändern, bilden die nächste Kategorie. Hier kann die räumliche Position konstant sein (z.B. statische Sensoren) oder sich im Laufe der Zeit verändern (z.B. fahrzeugmontierte Sensoren). Die letzte Kategorie entspricht dem Konzept des Datenwürfels: Diese Datenquellen sind an eine zeitliche und drei räumliche Komponenten gebunden und bilden einen oder mehrere Werte über das gesamte 4-dimensionale Gitter ab. Beispiele hierfür in *Digital Earth* sind Daten des atmosphärischen Modells *ICON*-Art und des ozeanographischen Modells *NEMO*. Diese Daten können wie bisher skalare Werte oder Vektoren darstellen. Wenn letzteres der Fall ist, beschreibt ein solcher Würfel ein Vektorfeld.

Auf Basis der definierten Anforderungen, die sich aus den Daten und Nutzungszielen ergeben, wurden folgende Datenviewer und Explorationswerkzeuge entwickelt:

Data Catalog-Viewer (siehe Abb. 13) ist ein Online *WebGIS*-Viewer, der die verfügbaren Daten der "*List of Data Products*" zeigt. Für *Digital Earth* wurde eine Auswahl aktuell zugänglicher bzw. bereitgestellter Daten aus den Show Cases homogenisiert und für die Veröffentlichung als digitaler *WebGIS*-Service vorbereitet. Das Datenspektrum umfasst derzeit öffentlich verfügbare Daten wie Landnutzung und Topographie, Messpunkte mit Informationen zu Datenloggern, und diverser anderen Messungen, sowie die Präsentation von hydrogeologischen Modellierungsergebnissen. Dieser kuratierte *WebGIS*-Service kann in zweierlei Hinsichten verwendet werden: Er visualisiert die bereits im Rahmen des Show Cases aggregierten Daten für alle Beteiligten in *Digital Earth*, zudem macht er für eine interessierte Öffentlichkeit die Ergebnisse aus der Wissenschaft in einer allgemeinverständlichen Form zugänglich¹.

¹ https://maps.awi.de/awimaps/projects/public/?cu=digital_earth_moses#home

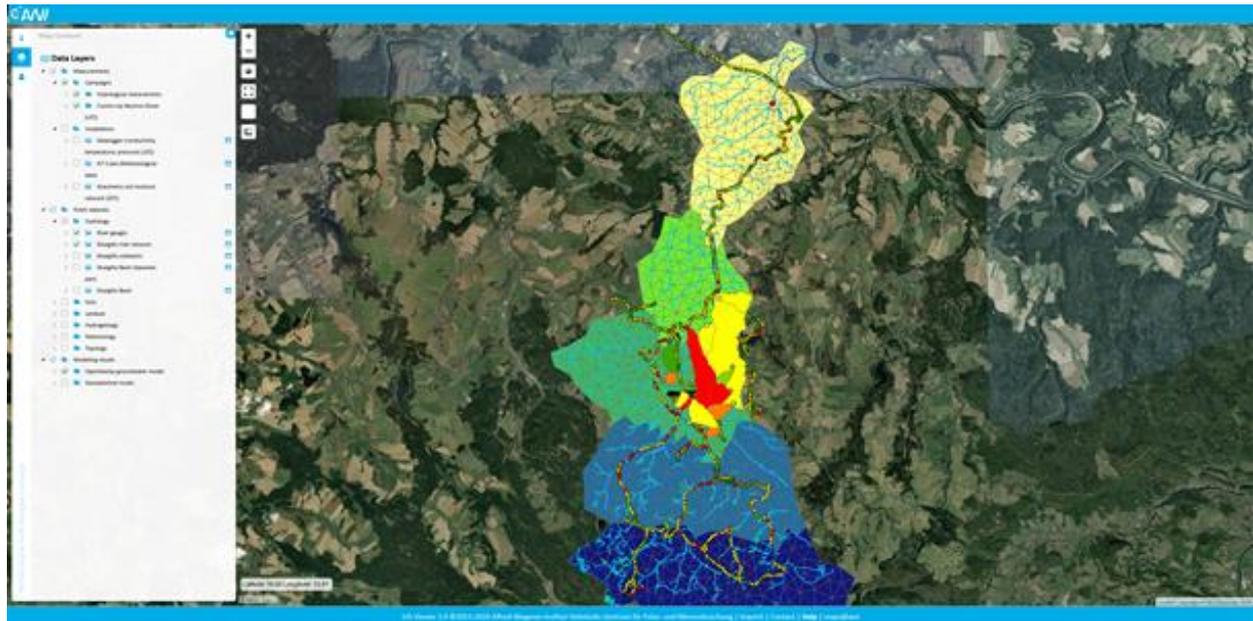


Abbildung 13: Data Catalogue -Viewer

4D Data Viewer (Workflow 6): Das Ziel bei der Entwicklung des *4D-Data Viewer* war die Visualisierung und räumlich-zeitliche Kontextualisierung von heterogenen Datentypen. Es handelt sich um eine portable Web-Anwendung. Die Anwendung kann sowohl lokale Dateien öffnen als auch über das Internet auf andernorts verfügbare Daten zugreifen. Alle Datenquellen werden in einer dreidimensionalen Umgebung dargestellt, die aus einem Breiten-Längen-Raster und einer Höhenkomponente besteht. Daten mit einer zeitlichen Dimension werden mittels einer gemeinsamen Zeitachse synchronisiert. Maus- und Tastatureingaben sowie grafische Elemente ermöglichen die Navigation in Zeit und Raum und über Datenstrukturen hinweg (siehe Abb. 14). Die Visualisierung einzelner Datenquellen wird durch ein zweidimensionales "Heads-up-Display"-Element mit speziellen 2D-Grafiken und Bedienelementen unterstützt.

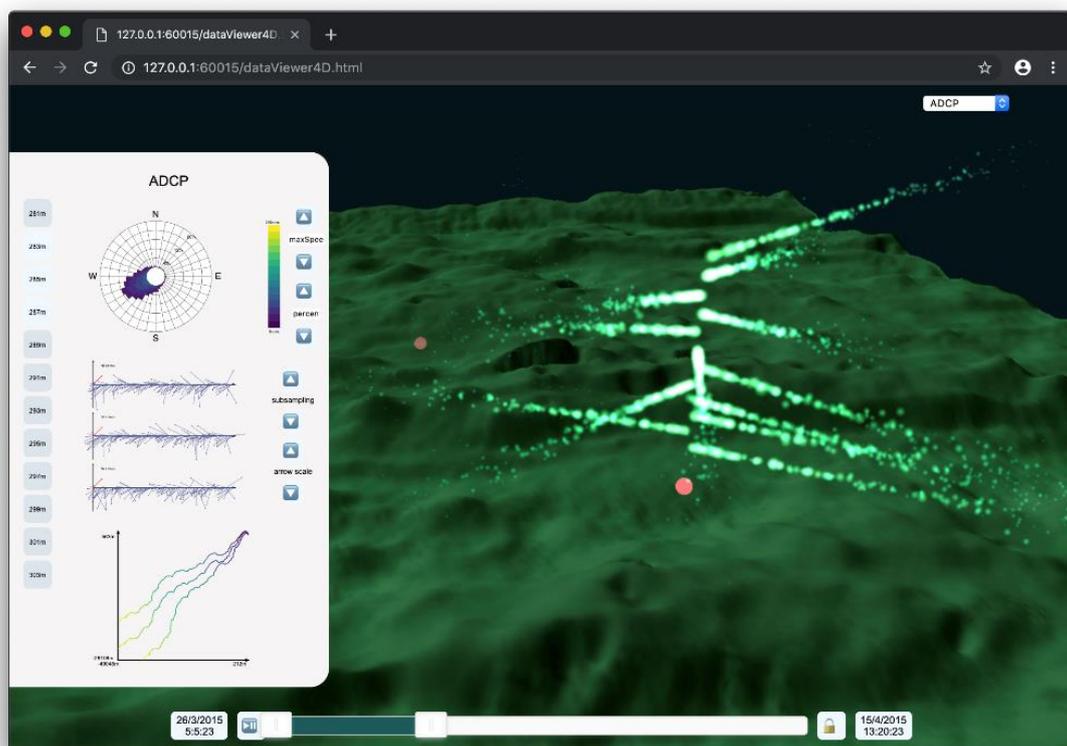


Abbildung 14: 4D Data Viewer

Anomaly Explorer (Workflow 3): Dieses Explorationstool unterstützt Nutzer dabei, Anomalien in Datensätzen zu explorieren und zu identifizieren. Es ermöglicht Nutzern im Workflow 3 die Elbfahne in der Nordsee auf Grund von Merkmalsunterschieden zu identifizieren (siehe Abb. 15). Basis dafür sind Daten aus *Ferrybox*-Trajektorien und zusätzliche Modelldaten, die auf einer Karte visualisiert werden. Die verwendeten Modelldaten, sogenannte Synopsis-Daten, dienen dazu, das zeitliche und räumliche Ausmaß der Elbfahne während oder nach einer Flut zu erfassen. Verschiedene visuelle Marker ermöglichen es, die Position und Ausbreitung der Elbfahne im Meer über bis zu 20 Tage abzuschätzen und mit verfügbaren Beobachtungsdaten abzugleichen. Für Detailuntersuchungen können "regions of interest" (ROI) ausgewählt werden, entweder manuell oder in Zukunft durch eine automatische Anomalien-Detektion in den *Ferrybox*-Daten (siehe Abb. 16). Auch steht eine Funktion zur Verfügung, die statistische Größen der Daten innerhalb und außerhalb der ROI miteinander vergleicht und damit Flutereignisse vergleichbar macht. Darüber hinaus kann die ROI als eigenes Layer auf andere Datenquellen gelegt werden, um die gesuchte Anomalie auch in diesen Daten zu finden und zu analysieren. Zur Identifizierung und Quantifizierung der Elbfahne wird zusätzlich eine einfache statistische Analyse durchgeführt: ein sogenannter *Similarity Score* berechnet die Abweichungen zwischen Datenpunkten inner- und außerhalb der ROI nach Variablen getrennt. Diese Analyse hilft dabei, die Hypothese zu überprüfen, ob es sich bei der in der ROI markierten Anomalie tatsächlich um eine Elbfahne handelt. Darüber hinaus lassen sich so interessante Variablen für die weitere Analyse mit zusätzlichen Beobachtungsdaten bestimmen. Um die Folgen des Flutereignisses für das marine Ökosystem abschätzen zu können, werden zusätzliche Beobachtungsdaten wie Satelliten- und *in-situ* Chlorophyll-Daten in den Viewer geladen. Zurzeit geschieht das manuell via *drag'n'drop*. In Zukunft soll dies dann über eine standardisierte Schnittstelle zu einem Datenportal erfolgen.

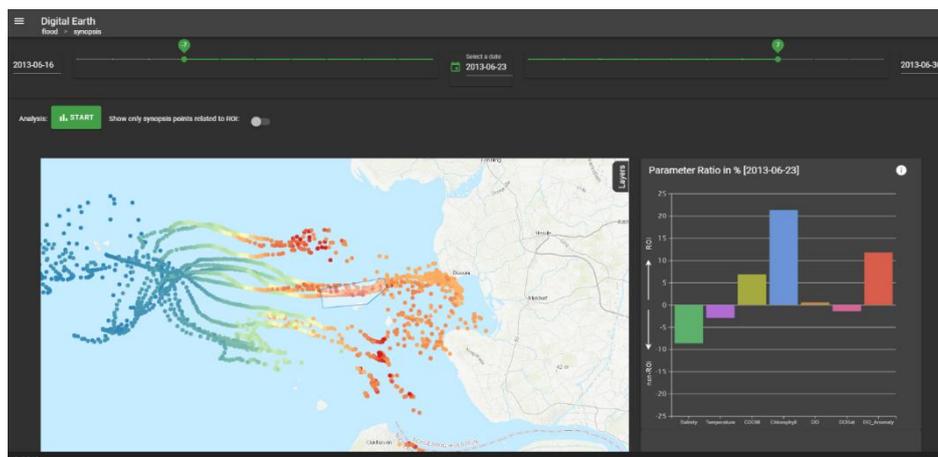


Abbildung 15: Anomaly Explorer

Similarity Explorer (Workflow 1): Der *Similarity Explorer* unterstützt die Untersuchung von Ähnlichkeiten; im Workflow 1 unterstützt er die Vergleichsanalyse verschiedener Flutereignisse. In der ersten Ausbaustufe ermöglicht der *Similarity Explorer* die räumliche und zeitliche Auswahl von Ereignissen für Pegelmessstationen (siehe Abb. 17). Die räumliche Auswahl der zu untersuchenden Pegelmessstation erfolgt über ein Kartenelement. Die Auswahl des relevanten Zeitbereiches erfolgt über eine Zeitbereichssliderkomponente. Die katalogisierten Hochwasserereignisse für die gewählte Messstation und den definierten Zeitbereich werden zusammen mit dem zugrundeliegenden Abfluss in einem Zeitreihendiagramm dargestellt. Die Auswahl der zu vergleichenden Ereignisse erfolgt über interaktive Diagramme durch Selektion der jeweiligen Säulen. Die vom Nutzer ausgewählten Hochwasserereignisse werden anschließend zur weiteren Analyse an ein Backend-Analysenmodul übermittelt.

Das Backend-Analysenmodul wie auch die verwendeten Methoden werden zurzeit entwickelt. Das Analyseergebnis wird abschließend an das visuelle Frontend übermittelt und zur visuellen Exploration für den Nutzer aufbereitet, dies ist ebenfalls Gegenstand der nächsten Projektphase.

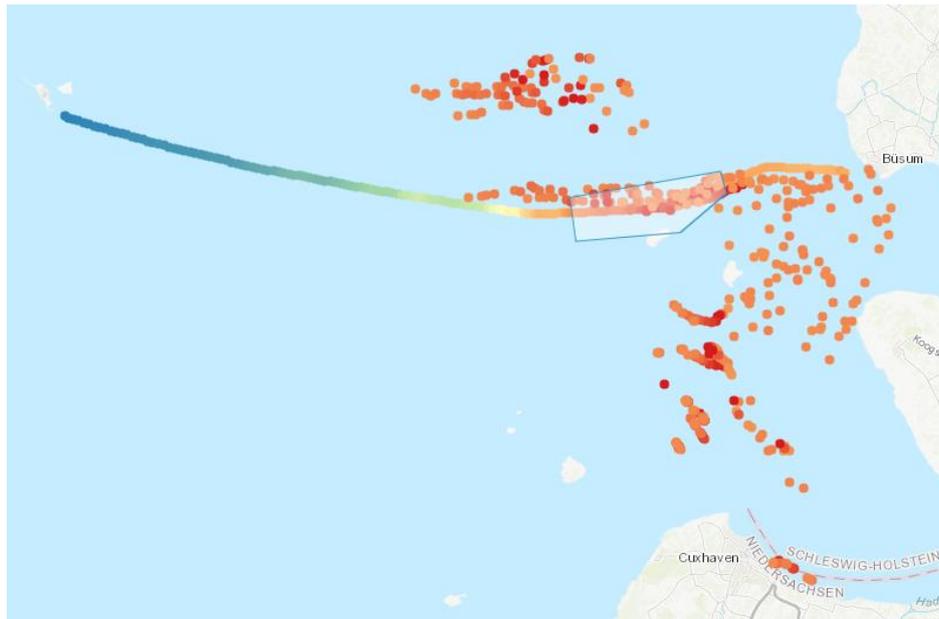


Abbildung 16: Anomaly Explorer mit "region of interest" Auswahl

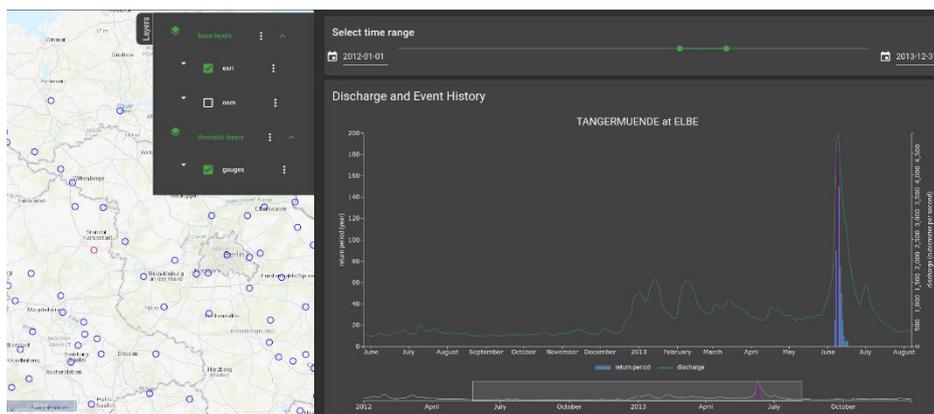


Abbildung 17: Similarity Explorer

Change Explorer (Workflow 2): Der *Change Explorer* ist ein interaktives Tool zur Visualisierung und Vergleich von Änderungen; im Workflow 2 werden simulierte Niederschlagsänderungen aus Klimamodellsimulationen für historische und zukünftige Zeiträume im Vergleich zu einer Referenzperiode für eine benutzerdefinierte Region untersucht (siehe Abb. 18). Für den visuellen Vergleich kann der Nutzer mit dem *Change Explorer* aus mehreren Datensätzen regionaler Klimasimulationen einzelne Simulationen und Variablen des ESGF-Datenarchivs auswählen, sowie Zeitbereich und Mittelungsintervalle und die zu untersuchende Region definieren. Auch können "regions of interest" in einer Karte mittels 'drag and drop' definiert und aus einer großen (mehrere GB) Datenmenge (*netCDF*-Daten) selektiert und anschließend mit verschiedenen Methoden visualisiert werden. Dies erleichtert wesentlich den Zugang zu Klimainformationen. Außerdem wurde damit begonnen, ein Konzept für eine detaillierte Analyse der Veränderungen von Starkniederschlagsereignissen (insbesondere Dauer, Intensität, Häufigkeit) zu entwickeln. Die Entwicklung des *Change Explorers* ist ein iterativer Prozess, wobei die Informatiker den Rahmen für den Explorer schaffen und die Naturwissenschaftler Klimasimulationsdaten in *netCDF* Format und Originalskripte zur Verarbeitung und Darstellung der Klimadaten bereitstellen und die Schritte zur Verbesserung der Funktionalität des interaktiven Werkzeugs definieren. Obwohl eine systematische Bewertung des *Change Explorers* noch aussteht, zeigt sich bereits jetzt, dass ein etablierter Workflow aus der Klimawissenschaft durch *Data Science* Methoden, in diesem Fall der visuellen Datenexploration, verbessert werden kann. Der *Change Explorer* kann auch für den Vergleich von anderen Modellvariablen verwendet werden.

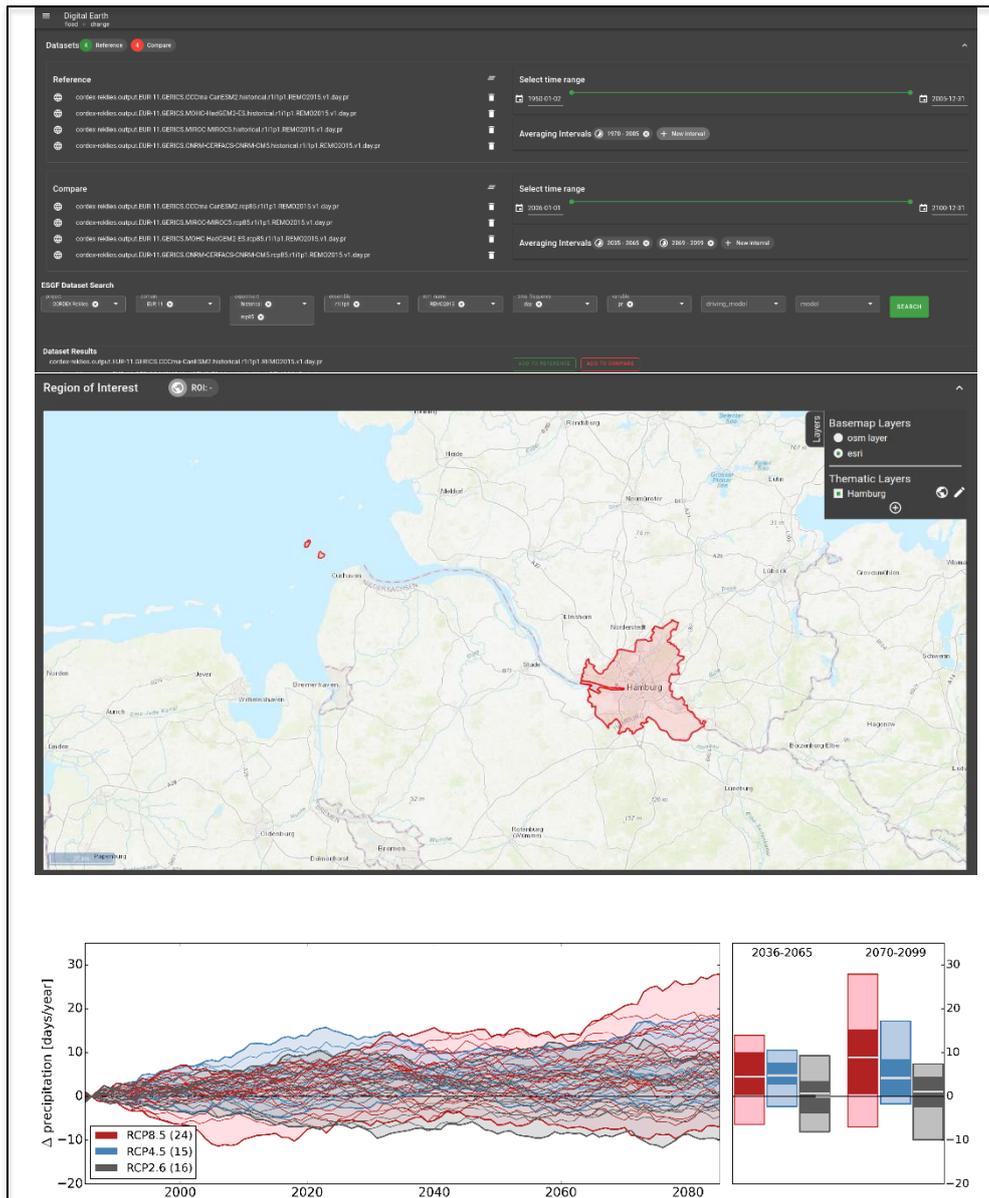


Abbildung 18: Change Explorer

Community Explorer. Der *Community Explorer* ermöglicht die Bewertung der Ähnlichkeiten von Gruppen durch drei Komponenten: (i) Interaktionsnetzwerk, (ii) Clustering und (iii) ‘stacked Barchart’. Diese drei Komponenten werden über multiple *Linked-View-Technik* miteinander verknüpft. So kann der Benutzer durch manuelle Auswahl einige Knoten aus dem Netzwerk auswählen, die dann als Linien im Cluster angezeigt werden (siehe Abb. 19). Der *Community Explorer* wurde in R entwickelt und basiert auf *R-shiny*. Die entwickelten Komponenten des *Community Explorer* stehen im *Digital Earth Data Exploration Framework* zur Verfügung und können in Workflows eingebunden werden.

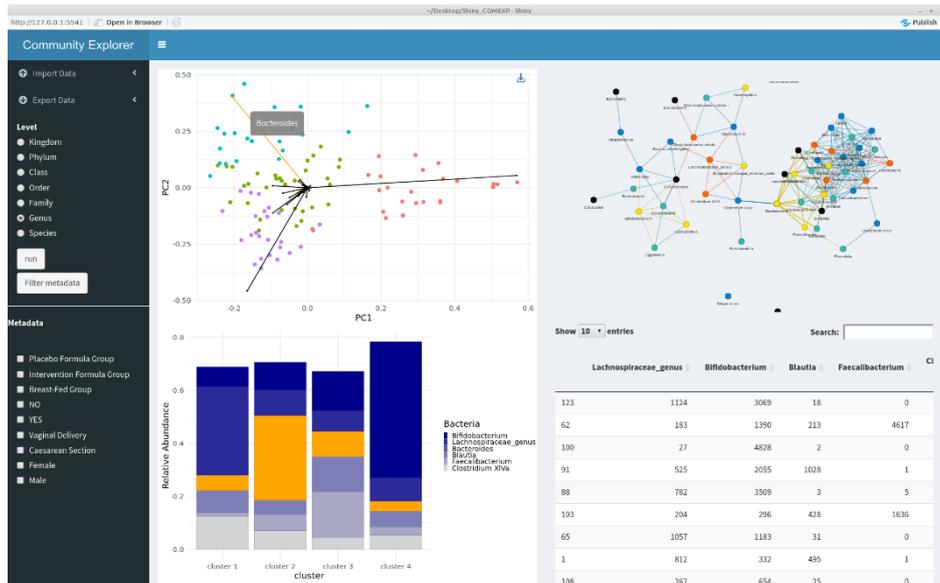


Abbildung 19: Community Explorer

QtAC-Paket ist ein Werkzeug zur Quantifizierung des adaptiven Zyklus für Zeitreihen als Realisierung von *Markov*-Prozessen der Ordnung 1, unabhängig vom zugrunde liegenden System. Das *QtAC*-Paket wurde entwickelt, um Zeitreihen in komplexen Systemen zu verarbeiten (siehe Abb. 20). Das Übertragungsentropienetzwerk ist ein direkter Graph, der verwendet wird, um die Übertragungsentropie als Übertragung von Informationen von einer Komponente zu einer anderen zu einem bestimmten Zeitpunkt zu interpretieren. Die *Adaptive Cycle-Metapher* im 2D- oder 3D-Format soll das Potential, die Resilience und die Connectedness in der Zeitreihe aller Knoten anzeigen. Potential ist die Entropie in den Informationskanälen. Die Resilience ist ein Maß dafür, wie viel Störung erforderlich ist, damit der Graph in eine andere verbundene Komponente zerfällt. Connectedness ist die gegenseitige Information zwischen Zu- und Abfluss von Informationen. Dieses Tool kann auf alle Zeitreihen im allgemeinen Format angewendet werden. Zum Beispiel kann das *QtAC*-Paket im Fluten Show Case, Workflow 1 verwendet werden. Das *QtAC*-Paket steht für Zeitreihenanalysen im *Digital Earth Data Exploration Framework* zur Verfügung und kann in Workflows eingebunden werden.

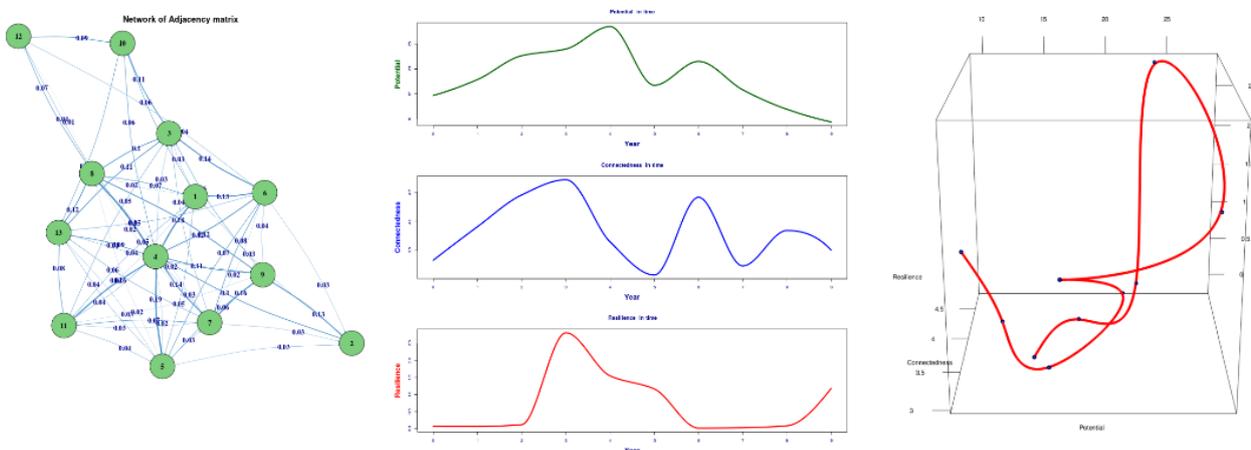


Abbildung 20: *QtAC*-Paket

Deliverables und Milestones

Deliverables

Keine entsprechend des Projektplans

Milestones

- 👍 M2.3.1 User tasks and data types identified and described M6
- 👍 M2.3.2 First visualization concepts developed M18

Task 2.4 Computational/Machine Learning Data Exploration

Für die Fragestellungen aus den Show Cases und wissenschaftlichen Workflows wurden in **Task 2.4** *Machine Learning* Algorithmen für die spezifischen Anforderungen evaluiert und angepasst. Anwendungsfelder waren beispielsweise das Erkennen von Fluss- oder Küstendeichen anhand von digitalen Geländemodellen und Orthofotos und das Erkennen von Munition am Meeresboden aus bathymetrischen Daten. Für die Arbeit in diesen Projekten sind zwei Arten von Daten besonders wichtig: Rasterdaten und Vektordaten. Diese Datentypen werden häufig für geographische Daten verwendet und sind daher weit verbreitet. Ein wichtiger Teil der Arbeit war das Zusammenführen der verschiedenen Datensätze, um Algorithmen auf die Daten anzuwenden. Hierzu wurden Methoden implementiert, die für ähnliche Probleme wiederverwendet oder angepasst werden können. Die zusammengeführten Rasterdaten haben hohe Ähnlichkeit mit Bilddaten (zweidimensional, räumlicher Zusammenhang). Der Fokus der betrachteten Methoden lag daher auf *Deep Learning* und anderen Methoden des maschinellen Lernens, da diese Methoden dort mit großem Erfolg angewendet werden. Die verwendeten Algorithmen hatten das Ziel Ausschnitte zu klassifizieren (bspw. als Munition) und neue Rasterdaten mittels semantischer Segmentierung zu erzeugen (bspw. die Fläche eines Deichs oder Wasserflächen, siehe Abb. 21). Dabei hat sich *Deep Learning* als die präziseste Methode herausgestellt. Die verwendeten neuronalen Netzwerke gehören zu der Klasse der *Convolutional Neural Networks* (für Klassifizierung) und *U-Net* (semantische Segmentierung). Die entwickelten Pipelines lassen sich auf ähnliche Probleme anwenden.

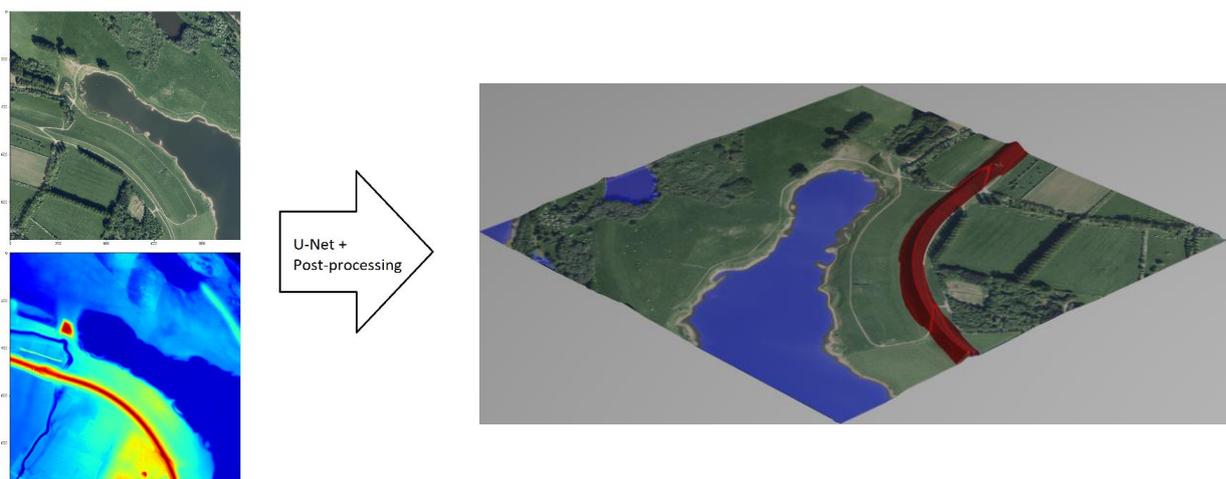


Abbildung 21: Semantische Segmentierung zur Erkennung von Deichen in Luftbildern und digitalen Höhenmodellen.

Als Teil von Show Case Methan wurden die Schwierigkeiten bei der Modellierung der Methankonzentration im Meerwasser mit Methoden zur Inferenz der räumlich-zeitlichen Auflösung bearbeitet. Es gibt nur eine niedrige vierstellige Anzahl an Messwerten in der Nordsee. Gleichzeitig sind viele dieser Messwerte an nah beieinanderliegenden und nicht zufällig ausgewählten Orten. Von diesen Messwerten ein Model mit Werten für die gesamte Nordsee zu entwickeln ist nicht trivial. Es wurde sich hierbei auf die Auswahl sinnvoller Informationskriterien und Validierungsmethoden konzentriert. Einfache Interpolationsmethoden, die sich auf die räumliche Nähe beziehen, haben sich als nicht zielführend herausgestellt, da kleinräumige Effekte nicht abgebildet wurden (diese treten beispielsweise

nah an Bohrlöchern auf, wo eine höhere Konzentration gemessen wird). Um den Einfluss von anthropogenen Strukturen (bspw. Ölplattformen) und Flussmündungen, die eine höhere Methankonzentration annehmen lassen, einzubeziehen, liegt der Fokus aktuell auf Modellierungsansätzen, die nicht primär auf räumliche Effekte setzen, sondern auf einer breiteren Datenbasis fußen. Das Problem der sehr unterschiedlichen raum-zeitlichen Auflösung von Daten mit Datenlücken und ungleichmäßigen Verteilungen existiert auch im Fluten Show Case. Die methodischen Entwicklungen und Erfahrungen aus dem Methan Show Case werden in der nächsten Projektphase auf den Fluten Show Case übertragen.

Methoden zum Füllen von Datenlücken: hierfür wurden Superauflösung, Multi-Grid und Kombinationstechnik identifiziert und befinden sich derzeit in der Entwicklung. Alle diese Methoden haben eine ähnliche Idee und verfolgen ein gemeinsames Prinzip. Die Konzentration liegt hierbei darauf, bereits vorhandene Lösungen mit verschiedenen Auflösungen zu verwenden und diese so zu kombinieren (siehe Abb. 22). Somit wird eine genauere Lösung mit höherer Auflösung erreicht. Es besteht die Möglichkeit, dass in jedem Fall die Multi-Grid-Methode angewendet werden muss, um sicherzustellen, dass beim Übergang von Coarser-Grid-Daten zu feineren Grids keine Informationen verloren gehen. Ziel ist es, diese Methoden zu untersuchen, um die Lücke zwischen *Top-Down*- und *Bottom-Up*-Ansätzen zu schließen.

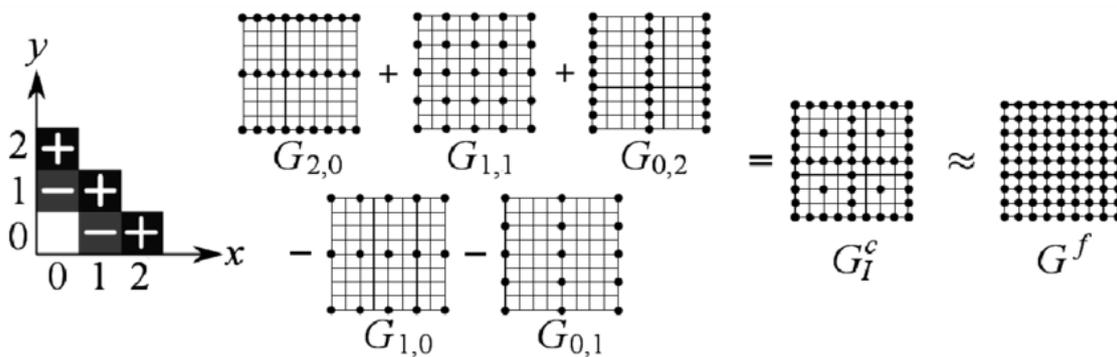


Abbildung 22: Grundlegende Kombinationstechnik-Konzeptillustration in 2D (Ali, Md Mohsin. (2016). High Performance Fault-Tolerant Solution of PDEs using the Sparse Grid Combination Technique. doi: 10.13140/RG.2.2.11378.71369.)

Methoden zur räumlichen Interpolation "Point to Space Problem": Ein "Point to Space Workshop" wurde organisiert, um die Methoden und die Fragestellungen innerhalb des *Point to Space*-Problems zu definieren. Die einfache *Kriging*-Methode (Gaußsche Regression) hat Probleme mit der Glättung und kann die sekundäre Datenquelle nicht so berücksichtigen, dass eine verbesserte Genauigkeit entsteht. Daher wird hier ein Verfahren vorgeschlagen, das auf einer bedingten Simulation basiert. Diese Simulationen haben den Vorteil, dass die in den Daten beobachtete Varianz erhalten bleibt und nicht nur der Mittelwert wie bei der Interpolation realisiert wird. Ihr stochastischer Ansatz verwendet Berechnungen vieler gleich wahrscheinlicher Lösungen (Realisierungen) und ermöglicht, Unsicherheiten zu bewerten und in der Nachbearbeitung zu quantifizieren (siehe Abb. 23). Derzeit werden Implementierungen speziell für den Methan Show Case entwickelt, die sich mit räumlich-zeitlichen ozeanografischen und räumlich-zeitlichen atmosphärischen Datenquellen befassen.

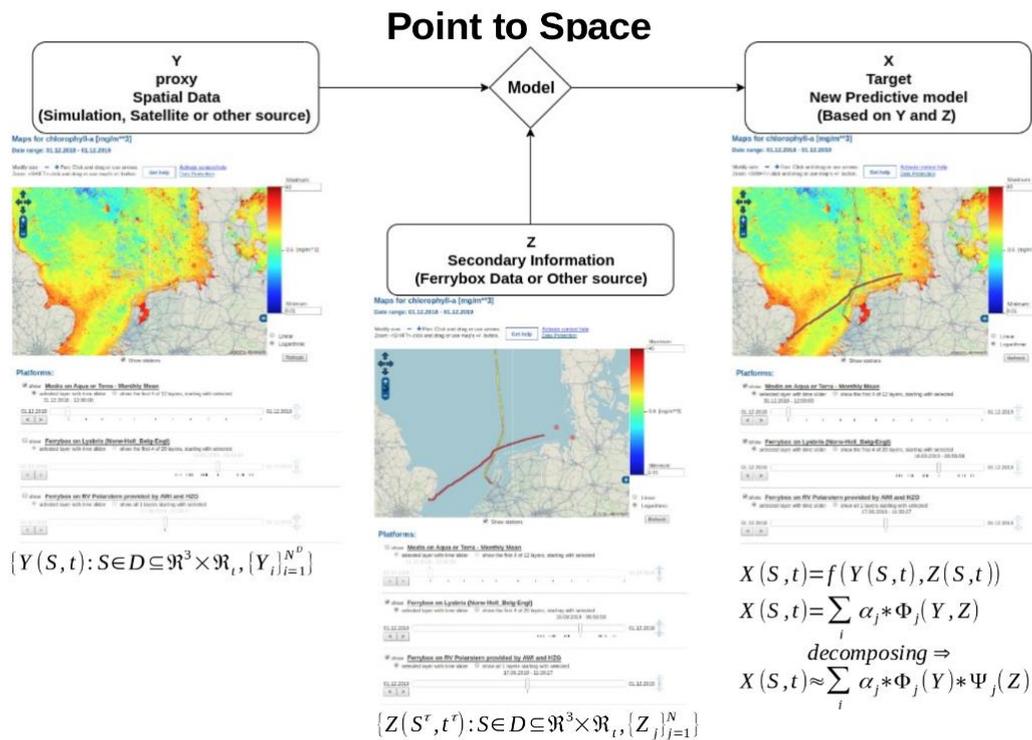


Abbildung 23: Point to Space Problem Ansatz

Auch wurden Methoden identifiziert, um das *Point to Space* Problem durch das Zusammenführen verschiedener Datenquellen anzugehen. Eine erste Bewertung geeigneter Methoden des maschinellen Lernens wurde anhand von Vorversuchen mit Ferrybox- und Satellitendaten sowie CML-Niederschlagsdaten durchgeführt. In Abbildung 24 sind die Ergebnisse mehrerer ML-Algorithmen am Beispiel des Salzgehalts von Ferryboxdaten (siehe dazu Workflow 3) dargestellt. Die ML Verfahren erzeugen ein besseres Ergebnis als Kriging Methoden (siehe Abb. 25).

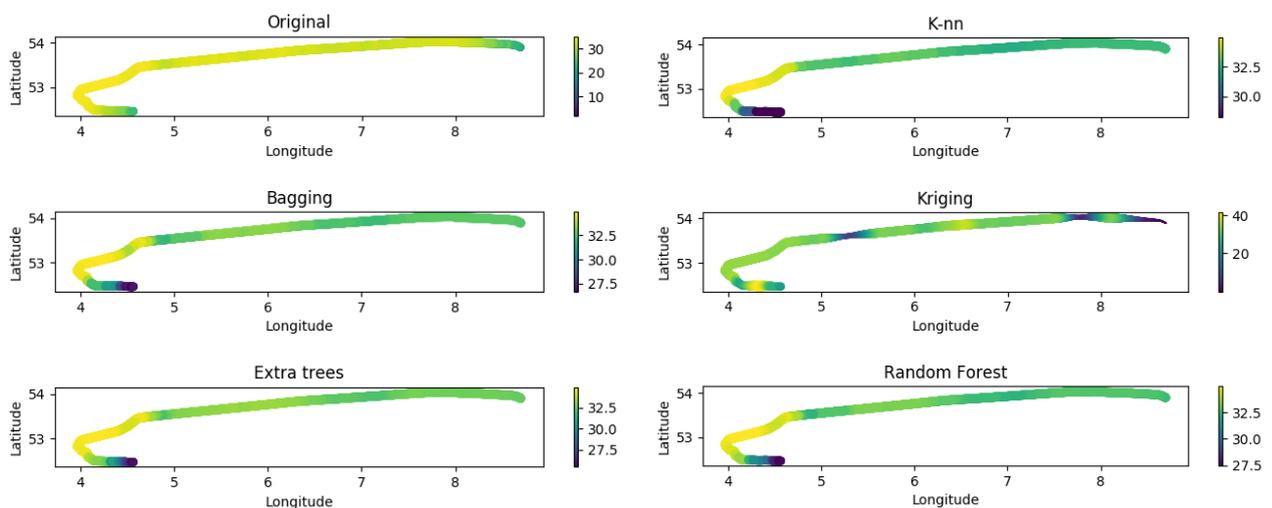


Abbildung 24: Ergebnisse mehrerer ML-Algorithmen am Beispiel des Salzgehalts von Ferryboxdaten

Im Rahmen des seit November 2019 laufenden Bridging Post-Doc Projekts "*Machine Learning methods for assessing causal links in heterogenous data: applied to Climate Change and Health*" (HZG und HMGU) wurde eine Literaturrecherche zum Thema Hitzewellen und Gesundheit durchgeführt, die Datenakquise geplant, und das Projekt wurde von Dr. Peter Hoffmann (HZG/GERICS) auf einer Konferenz zum Thema "*Machine Learning for sustainable development*" in Tokio, Japan vorgestellt.



Abbildung 25: Vergleich zwischen ML-Methoden, beobachteten Daten und normalen Kriging-Methoden

Deliverables and Milestones

Deliverables

D2.4.1 Modules of anomaly detection and imputation M18
 Dieses Deliverable wurde zugunsten der Entwicklung von Methoden zum Füllen von Datenlücken und für die "Point to Space" Problematik zeitlich nach hinten verschoben. Aus den Show Cases hat sich eine hohe Dringlichkeit für die Lösung der beiden methodischen Herausforderungen ergeben, daher wurde diese Anpassung im Projektplan vorgenommen.

Milestones

- M2.4.3 Tools for knowledge extraction identified M6
- M2.4.4 Prototype for algorithms for inference of spatio-temporal resolution identified M12

WP3: Sustainable Collaboration

Task 3.1 Establish structures and strategies for joint data driven science within E&E and other partners

Eine übergreifende Governance-Struktur wurde durch das Einsetzen einer *Arbeitsgruppe Digitalisierung* im Forschungsbereich *Erde und Umwelt* etabliert. Diese implementiert in strukturierter Weise die Entwicklung notwendiger Infrastrukturen und Dienste für datengetriebene Forschungsansätze, wie z.B. in *Digital Earth*. Zentral dafür ist die Einrichtung des *Erde und Umwelt DataHubs*, der eine kollaborative Initiative darstellt, um das Datenmanagement an den Zentren auf einen gehobenen Stand zu bringen. Das Ziel ist, virtuelle Forschungsumgebungen auf Basis offener und *FAIR*ere Daten, thematische Viewer (Web-Portale) und Daten-Dienste wie DataCubes aufzustellen, um hochperformante Datenzugriffe und -analysen im Verbund der Zentren zu ermöglichen. Prinzipien der *Digital Earth* Governance-Struktur wurden in die AG Digitalisierung übernommen, insbesondere die Kommunikationswege in das Managementboard des Forschungsbereichs, was zu einem strategischen Berichtswesen für *Digital Earth* beiträgt.

In Anlehnung an die *Digital Earth* Seed Groups wurden im DataHub Arbeitskreis zu den Querschnittsthemen Metadaten, verteilte Datenanalysen, Datenflüsse und Sensormanagement eingerichtet. Diese werden die Harmonisierung von Diensten und Standards auf synergistischer Basis herstellen, was die Implementierung von *Digital Earth* Datenprodukten und Diensten im nächsten Forschungsprogramm unterstützen wird. Sehr relevant für diese Entwicklungen ist der geplante WP3

Stakeholder Networking Workshop (Frühsommer 2020), der *Digital Earth* Produkte sowie Beziehungen und Rollen der Partner klären wird (siehe **Task 3.2**, unten).

Digital Earth hat seine Stellung im kommenden Forschungsprogramm “*Changing Earth - Sustaining our Future*” in Form einer *Cross-Topic Activity* (CTA) eingenommen und ist als solches im Abschnitt 2.2.4 “*Towards Smart Monitoring and Integrated Data Exploration of the Earth System*” beschrieben. Insgesamt wurden 19,5 FTEs für *Digital Earth* über alle Zentren alloziert. Diese sind in zahlreichen Topics des Programms eingebettet. Es ist dort explizit festgehalten, dass *Digital Earth* den Rahmen für strategische Entwicklungen der Data Science Aktivitäten im Forschungsbereich setzt.

Deliverables und Milestones

Deliverables



D3.1.3 Performance indicators established for running Digital Earth in PoF-IV

M12

Milestones

Keine entsprechend des Projektplans

Task 3.2 Develop collaborative framework for operating technical platforms

Als technisches Plattformkonzept für kollaboratives Arbeiten wurde das *O2A*-System am AWI für Nutzer in *Digital Earth* geöffnet, insbesondere für Bodenfeuchte-Sensoren und Datenmanagement (UFZ), und für Software-Entwicklung zur visuellen Datenexploration (GFZ). Ein cloud-basierender Ansatz wird über den sogenannten Marketplace virtueller Anwendungsdienste bereitgestellt, so z.B. Jupyter-Notebooks, Rasterdaten-Manager, sowie NRT-Datenbanken. Diese Dienste werden auf der *Helmholtz Data Federation* (HDF), aufgebaut und betrieben am AWI, bereitgestellt.

Das Konzept der Technischen Plattform für *Digital Earth* wird aus den Services und Standards, wie sie in den Show Cases definiert und genutzt werden, konsolidiert. Es wird auf den Entwicklungen der Querschnittsthemen im *DataHub EuU* basieren, und auf Standards der geowissenschaftlichen Community, wie *OGC* und Services für georeferenzierte Daten. Es ist wichtig hervorzuheben, dass der synergistische, kollaborative Betrieb von Daten-Diensten, also für multiple Nutzer-Kontexte, klar definierte Rollen und Workflows erfordert, die der Nutzer selbst bedient. Dies ist essenziell, um eine Konformität mit Qualitäts- und Format-Standards der Dateninhalte auf der Technischen Plattform zu gewährleisten. Dieses Prinzip der geteilten Verantwortung (Bereitstellung von Diensten und Inhalten) ist möglicherweise das wichtigste tragende Element der Nachhaltigkeit, wie es in *Digital Earth* gesehen wird.

Deliverables und Milestones

Deliverables



D3.2.2 Technical platform concept, including data and service standards

M9

Milestones

Keine entsprechend des Projektplans

Task 3.3 Foster innovation for societal benefits

In **D4.1.1** wurden mehrere Aktivitäten von Partnern in *Digital Earth* berichtet, die den Transfer von Daten und Informationen der Forschung über das Erdsystem an die Gesellschaft und potenzielle Nutzer unterstützen. Mehrere Zentren stellen solche Daten und Informationen aktiv für Parteien außerhalb der wissenschaftlichen Bereiche zur Verfügung. Darüber hinaus wurden Dokumentationen und Anleitungen entwickelt mit dem Ziel die Nutzbarkeit von Daten zu verbessern.

Innerhalb WP4 und WP3 wird das *TransFAIR*-Konzept entwickelt, das die Aufnahme von Daten und Informationen durch andere Nutzer und Praktiker verbessern soll. Es wird eine Checkliste bereitstellen, um Forschungsdaten und Informationen so zu entwickeln und bereitzustellen, dass sie auf die Gesellschaft übertragen werden können und besser geeignet sind, von Praktikern aufgenommen zu werden. Ansätze zur Verbesserung des Datenzugangs, der Methoden der *Computational Data Exploration* (KI und maschinelles Lernen) und des Datentransfers wurden ebenfalls mit Forschern und Interessenvertretern diskutiert und vorgestellt, unter anderem auf der *REKLIM*-Konferenz, dem

Helmholtz Data Science Symposium, und bei Veranstaltungen in Italien (*ESA Living Earth Symposium*) und in Japan (KI-Konferenz).

Auf der REKLIM-Konferenz wurde ein Poster mit zwei Beispielen für die Entwicklung von Klimadiensten im Rahmen des *Digital Earth*-Projekts vorgestellt, die sich auf die Bedarfe der Nutzer nach Visualisierung komplexer und großer Klimadatenätze beziehen. Diese prototypischen Beispiele sind der Event Tracker zur Identifizierung von Starkniederschlagsereignissen in großen Beobachtungs- und Modelldatenätzen und die Change Tool Box, die Veränderungen von Niederschlag in großen Klimamodell-Datenätzen visualisiert.

Task 3.4 Software architecture-concept for *Digital Earth* and beyond

Im Rahmen des Arbeitspakets WP2 wurden Anforderungen für die Daten-getriebene Forschung über mehrere Kompartimente erhoben. Basierend auf diesen Anforderungen wurde ein Rahmen entwickelt, welcher eine hierarchische Softwarearchitektur definiert. Diese Architektur definiert Ebenen, die funktional zusammengehörige Komponenten und Pakete gruppieren und definiert die benötigten technischen Strukturen. In der obersten Hierarchiestufe werden im Wesentlichen drei Aspekte unterschieden: *Front-end*, *Back-end* und Datenservices. Das *Front-end* umfasst alle visuellen Komponenten für eine visuelle Datenexploration im Rahmen der umgesetzten digitalen wissenschaftlichen Workflows. Die entwickelten und angewendeten rechnergestützten Ansätze und Methoden wurden als *Back-end* Module umgesetzt. Die Verknüpfung der Module und Komponenten aus *Front-* und *Back-end* erfolgt über ein modernes *Message Broker* System, welches die unabhängige Entwicklung und Nutzbarmachung in einer Vielzahl von Programmiersprachen und Umgebungen ermöglicht. Zu Test- und Evaluierungszwecken wurde eine den Anforderungen entsprechende Umgebung eingerichtet und die entwickelten Komponenten ausgerollt. Als *Message Broker* System kommt *Apache Pulsar* zum Einsatz. Der Zugriff auf die zu prozessierenden und zu visualisierenden Daten erfolgt primär über standardisierte Schnittstellen, beispielsweise *OGC Services* (*Rasdaman* und *maps.awi.de*, *AWI*) und *Thredds/Opendap* Protokoll (*HZG/GERICS* und *GEOMAR*).

WP4 Evaluation of Success

Task 4.1 Requirement identification for scientific use and evaluation criteria definition

Im vergangenen Jahr wurden die wichtigsten Kriterien und Indikatoren für die Bewertung des Projekterfolgs gemeinsam festgelegt. Es handelte sich um eine Querschnittsaktivität, die sich über die **Tasks 4.1, 4.2** und **4.3** erstreckte. Es wurde eine Liste von Kriterien und Indikatoren erarbeitet, die die folgenden Themen abdeckt:

- (1) Anforderungen an die Datenwissenschaft,
- (2) Anforderungen an den wissenschaftlichen Fortschritt und
- (3) die Benutzerfreundlichkeit der Arbeitsabläufe.

Daraus resultierte ein Dokumententwurf, welcher als Grundlage für die erste Umfrage in *Digital Earth* diente und allen Partnern des *Digital Earth*-Projekts vorgelegt wurde. Mit diesem Dokument wurden folgende Milestones abgedeckt: **M4.1.1** *First set of criteria and indicators*, **M4.2.1** *Set up of the monitoring scheme for deliverable* und **M4.3.1** *Initial set up of the survey/ evaluation scheme for deliverable*. In Tabelle 4 sind die Kriterien und zusätzliche Unterkriterien identifiziert, welche in der Umfrage verwendet wurden. Ziel war es, die ursprüngliche Liste von Kriterien und Indikatoren zu testen und diese für die abschließende Bewertung am Ende des *Digital Earth*-Projekts weiterzuentwickeln, um Input für die Deliverables **D4.2.1** und **D4.3.1** zu liefern, die auf den abschließenden Bewertungsberichten von *Digital Earth* (Teil A bzw. B) aufbauen.

Die erste Umfrage wurde im Zeitraum vom 5. März bis 15. Juni 2019 mit der *eSurvey*-Software durchgeführt. Es gab 54 Einzelrückmeldungen, was ein Rücklauf von etwa 50% der versandten Einladungen entspricht und deshalb als aussagekräftige Bilanz zu bewerten ist. Die wichtigsten Ergebnisse der Umfrage sind unter **Task 4.3** aufgeführt.

Tabelle 4: Kriterien für die Bewertung des Erfolgs von Digital Earth.

Kriterien	Kapazitäten für die Datenwissenschaft	Wissenschaftlicher Fortschritt	Verwendbarkeit der Ergebnisse
Unterkriterien	<ul style="list-style-type: none"> • Kapazitäten für Umsetzung von Data Science • Herausforderungen • Daten, Infrastruktur, Modelle • datenwissenschaftliche Methoden • Datenexplorations-Werkzeuge • Projektzusammenarbeit und -management • SMART-Monitoring • wissenschaftliche Workflows 	<ul style="list-style-type: none"> • wissenschaftliche Ziele • Ziele des Forschungsprozesses 	<ul style="list-style-type: none"> • Auffindbarkeit • Zugänglichkeit • Interoperabilität • Wiederverwendbarkeit • Unsicherheit (<i>FAIR plus</i>) • Verwendbarkeit der Ergebnisse über die Forschung hinaus

Alle Ergebnisse werden in **D4.1.1** dokumentiert und sind via *Helmholtz-Netz* für alle *Digital Earth* Mitgliedern einsehbar. Die Ergebnisse der Umfrage, die Interpretation und Auswertung werden außerdem in einem Kapitel des *Digital Earth* Buches vorgestellt, welches in der Reihe *Springer Briefs* veröffentlicht werden soll. Ein Abstract unter Beteiligung aller WP4- Partner wurde hierfür eingereicht. Außerdem wurde eine Bewerbung für den *Digital Leader Award*² stellvertretend für das gesamte *Digital Earth* Konsortium und unter Beteiligung aller Zentren vorbereitet. Der Antrag dokumentiert exemplarische Erfolgsgeschichten, welche im Rahmen von *Digital Earth* zu verzeichnen sind. Es werden hier *SMART-Monitoring* Ansätze aufgezeigt, die es ermöglichen auf Basis einer integrierten real-time Datenbasis und eines digitalen Workflows die Datenerhebung im Feld zu verbessern. Außerdem wurde der *Data Exploration Framework* vorgestellt, mit dessen Hilfe Wissenschaftler siloübergreifende Datenanalysen mit *state-of-the-art* und neuen *Data Science* Verfahren durchführen können.

Für **D4.1.2** *Refined set of criteria and indicators* wurde ein Arbeitsversion erstellt, welche seit dem 3. Dezember 2019 auf einer gemeinsamen Arbeits-Plattform vorliegt und im Laufe des Jahres 2020 als Arbeitsdokument weiterentwickelt werden soll. Diese Kriterien und Indikatoren werden die Grundlage für die abschließende Umfrage im Jahr 2021 bilden.

Darüber hinaus wurde beschlossen, dass im Rahmen des 2nd Interims Meetings im Januar 2020 ein Workshop abgehalten wird, der zum Ziel hat, den Erfolg der Zusammenarbeit im Projekt *Digital Earth* intern zu evaluieren. Die Ergebnisse dieses Workshops werden Gegenstand im nächsten Jahresbericht werden.

Deliverables and Milestones

Deliverables

-  D4.1.1 Inventory of categorised requirements M9
-  D4.1.2 Report with refined set of criteria and indicators M18

Milestones

-  M4.1.1 First set of criteria and indicators M12

Task 4.2 Usability assessment of workflows, methods, and software

Für diese Aufgabe wurden in der ersten Umfrage (siehe **D4.1.1**) mehrere Kriterien für die Anwendbarkeit von Workflows, Methoden und Software entwickelt. Diese Kriterien basieren weitgehend auf den *FAIR*-Kriterien für die Nutzung wissenschaftlicher Daten. Die einzelnen Kriterien sind in Tabelle 4 aufgeführt und wurden gemeinsam mit allen in *Digital Earth* beteiligten Zentren entwickelt.

² <https://www.digital-leader-award.de/>

Die Ergebnisse der ersten Umfrage können auch für in der Entwicklung des Software-Architekturkonzepts von *Digital Earth* verwendet werden (**Task 3.4**). Die Kriterien und Indikatoren die in der abschließenden Umfrage verwendet werden sollen, werden entsprechend der erarbeiteten Bedürfnisse weiter ausgearbeitet (**D4.1.2** - Entwurf) und stetig weiterentwickelt. Sie bilden die Grundlage für **D4.2.1** (Vorlage im Januar 2021 - **M32**).

Task 4.3 Assessing the success of the project for gaining new scientific insight

Für diese Aufgabe wurden in der ersten Erhebung mehrere Kriterien zur Bewertung des Projekts zum wissenschaftlichen Erfolg entwickelt (siehe **D4.1.1**). Die einzelnen Kriterien sind in Tabelle 4 aufgeführt. Die Entwicklung dieser Kriterien wurde mit Beiträgen von allen in *Digital Earth* beteiligten Zentren vorgenommen.

Diese Kriterien werden in einem verfeinerten Satz von Kriterien und Indikatoren (**D4.1.2** - Entwurf), der in der abschließenden Umfrage am Projektende verwendet werden soll, weiterentwickelt und bilden die Grundlage für **D4.3.1** (Vorlage im Januar 2021).

Die wichtigsten Ergebnisse der ersten Umfrage, die **Tasks 4.1, 4.2** und **4.3** abdeckt, sind:

- Im Rahmen des *Digital Earth*-Projekts wurden seit Beginn mehrere Kooperationen zwischen den verschiedenen Zentren aufgebaut. Der Wunsch nach einem weiteren Ausbau der Zusammenarbeit ist ein wichtiges Feedback der Umfrageteilnehmer. Die Umfrage zeigt außerdem, dass eine besonders starke Nachfrage an Kooperationen mit Zentren besteht, die über außerordentliche datenwissenschaftliche Kompetenzen verfügen.
- *Digital Earth* hat zudem Fortschritte bezüglich der interdisziplinären Ansätze gemacht, die sich in der Zusammenarbeit zwischen verschiedenen Bereichen der Erd- und Umweltwissenschaften zeigen, aber auch die Verknüpfung von Fachkenntnissen die Datenwissenschaften betreffend, was im Besonderen die visuelle Datenexploration sowie die *Computational Data Exploration*, einschließlich künstlicher Intelligenz, z.B. maschinelles Lernen betrifft. Zu Beginn des Projekts verfügten viele Zentren nur über eine begrenzte Erfahrung mit einigen dieser Methoden. *Digital Earth* setzte hier an und bot die Basis für den Austausch zwischen Natur- und Datenwissenschaftlern. Dabei ist anzumerken, dass das Projekt *Digital Earth* ohne die starke Basis an Infrastrukturen und Datenbanken nicht realisierbar wäre. *Digital Earth* baut dabei auf mehreren komplementären Maßnahmen auf, die sich auf Feldbeobachtungen und Dateninfrastruktur in den einzelnen Helmholtz-Zentren sowie auf gezielte Kooperationsprojekte wie *TERENO* und *MOSES* konzentrieren und profitiert somit von diesen Strukturen.
- Im Hinblick auf die wissenschaftliche Praxis wurde über mehrere Aspekte von *FAIR Science* berichtet. Viele Zentren stellen ihre Daten und ihren Code offen zur Verfügung, aber in vielen Fällen fehlen gemeinsame Lizenzen und Richtlinien für die Veröffentlichung und weiterer Nutzung.
- Wissenschaftliche - und Beobachtungsdaten, Werkzeuge und Informationen werden auch außerhalb des akademischen Bereichs zur Verfügung gestellt. Dies wird durch die Veröffentlichung von Richtlinien für die Nutzung, die Anpassung an spezifische Anwendungen und die Qualitätssicherung ergänzt.
- Im Hinblick auf den Projekterfolg berichten die Forscher über mehrere wichtige Indikatoren: (1) Die Forscher streben eine bessere Integration und Zusammenarbeit zwischen den Disziplinen innerhalb des Forschungsbereichs *Erde und Umwelt* an; (2) *Digital Earth* soll dazu beitragen, die Nutzbarkeit von Daten, Informationen und Arbeitsabläufen zu erhöhen; und (3) gemeinsame wissenschaftliche Veröffentlichungen, neue Forschungsideen und neue datenwissenschaftliche Aktivitäten werden als wichtige Erfolgszeichen für *Digital Earth* angesehen.

Am ursprünglichen Forschungsplan für das WP4 gibt es keine Änderungen. Alle Ziele für die vorangegangene Periode sind erreicht worden (Milestones und Deliverables). D4.1.2 liegt in einer Entwurfsfassung vor und bleibt ein Arbeitsdokument für die zweite Umfrage, die Anfang 2021 stattfinden wird.

WP5 Project Management

Die *Digital Earth* Project Management Unit (PMU) am GEOMAR koordiniert und führt mit Unterstützung der Steering Group und anderen Leitungsgremien projektbezogene Aufgaben durch. Projektbericht für 2018. Ausserdem wurde *Digital Earth* im PoF-IV Antrag als *Cross-Topic Activities* (CTA) implementiert. Des Weiteren wurde der zentrale Antrag zur Aufstockung von *Digital Earth* Mitteln ab Mitte 2019 organisiert und mit der Helmholtz- Geschäftsstelle abgestimmt.

Task 5.1 Project coordination and administration

Die Projektaufgaben umfassten die Zusammenstellung des Projektbericht für 2018 sowie die Darstellung von *Digital Earth* im PoF-IV Antrag als *Cross-Topic Activities* (CTA). Des Weiteren wurde der zentrale Antrag zur Aufstockung von *Digital Earth* Mitteln ab Mitte 2019 organisiert und mit der Helmholtz-Geschäftsstelle abgestimmt. Die *Digital Earth* PMU co-organsierte des Weiteren ein gemeinsames Treffen der Steering Boards von ESM, MOSES und *Digital Earth* im Januar 2019 in Berlin. Dieses Treffen, wie viele andere Aktivitäten sind zwar nicht explizit Bestandteil des Arbeitsprogrammes von *Digital Earth*, wurden aber entsprechend der strategischen Bedarfe kurzfristig implementiert.

Die *Digital Earth* PMU, organisierte zusammen mit Kollegen des UFZ das 1. *Digital Earth* Jahrestreffen (**D5.1.1**), welches im böhmischen Novy Kostel, Tschechien, vom 19.-24.05.2019 unter der Beteiligung von mehr als 55 Teilnehmern stattfand. Das Programm beinhaltete neben den Berichten zu den einzelnen Arbeitspaketen, *Key Notes* zu Themen wie “*Machine Learning*“, “*Uncertainty Assessment in Geosciences*“ und “*Inter- and transdisciplinary Research*“ auch explizite Break-Out Sessions zu den *Digital Earth* Seed Groups und Show Cases und eine gemeinsame Exkursion in den in der Nähe gelegenen Eger Graben. Ein weiterer Schwerpunkt des Treffens in Novy Kostel war die Auswahl der Bridging Postdoc-Kandidaten. Dieses Treffen wird von allen Teilnehmern als ein Meilenstein für *Digital Earth* betrachtet, da es den direkten Dialog zwischen den WissenschaftlerInnen forcierte und die Zusammenarbeit über die wissenschaftlichen Silogrenzen hinaus belebte.

Es fanden regelmäßig alle 2 Monate Treffen des *Digital Earth* Steering Boards per Videokonferenz statt.

Die *Digital Earth* PMU übernahm die Leitung diverser Aktivitäten wie:

- Koordination der Beteiligung von *Digital Earth* an diversen Konferenzen (e.g. REKLIM, 2019; EGU 2020)
- Koordination und Einreichung des Antrages zum *Digital Leader Award* 2020
- Erstellen von Informationsmaterialien
- Beschaffung und Etablierung der Arbeitsplattform *Confluence* für *Digital Earth* und *MOSES* Nutzer
- Veröffentlichung des monatlichen Newsletters ³, der an alle in *Digital Earth* Mitglieder versendet wird

Deliverables und Milestones

Deliverables



D5.1.1 Coordination and planning of yearly project meetings

M12

Milestones

Keine entsprechend des Arbeitsplanes

Task 5.2 Bridging Post-Docs

Die Ausschreibung zur Einreichung von Bridging Postdoc Projekten wurde im Januar 2019 geschaltet und über alle *Digital Earth* Zentren kommuniziert und auf der Projektseite ausgeschrieben. Innerhalb der Bewerbungsfrist gingen 20 Bewerbungen ein. Davon wurden durch das *Digital Earth* Bridging PostDoc – Auswahlkomitee 11 Kandidaten zu Vorstellungsgesprächen am 22.05.2019 eingeladen (**D5.2.1**). Die Vorstellungsgespräche wurden im Rahmen des 1. *Digital Earth* Jahrestreffen in Novy Kostel, Tschechien abgehalten (**M5.2.1**). Es ist zu betonen, dass die Qualität der vorgestellten Projekte sowohl ein sehr hohes

³ <https://www.digitalearth-hgf.de/de/newsletter>

wissenschaftliches Niveau als auch einen starken Bezug zu den Inhalten von *Digital Earth* hatten. Aufgrund des limitierten finanziellen Budgets musste dennoch eine Auswahl getroffen werden. Das Auswahlkomitee identifizierte sechs Kandidaten (**D5.2.1**), die ab Juli 2019 nach und nach über *Digital Earth* eingestellt wurden. Leider gab es trotz vorheriger Absprache bei der Einstellung eines Kandidaten an einem Zentrum strukturell-administrative interne Probleme, die sich trotz stetiger Unterstützung durch die betroffenen Kollegen und der *Digital Earth* PMU über 5 Monate hinzogen und letztendlich dazu führte, dass die Einstellung nicht umgesetzt werden konnte. Der betroffene PostDoc konnte dann mit einer 5-monatigen Verspätung ab Januar 2020 über das HZG seine Stelle in *Digital Earth* antreten. Die damit verkürzte Laufzeit seines Vertrages hat zur Folge, dass die Fertigstellung seines Projekts bis zum offiziellen Ende von *Digital Earth* im Mai 2021 unwahrscheinlich ist.

In Anbetracht der Tatsache, dass es sich bei der Idee der Bridging Postdocs um ein Profil handelt, welches als eine wichtige aktive Säule in der Zusammenarbeit zwischen den Zentren auch zukünftig in PoF-IV Anwendung finden soll, hoffen wir, dass die zentrumsinternen administrativen Strukturen zeitnah für die Umsetzung derartiger gemeinsamer Projekte optimiert werden können.

Deliverables und Milestones

Deliverables

 D5.2.1 Evaluation of Post-Doc applications and final selection of candidates M12

Milestones

 M5.2.1 Advertise Post-Doc positions and organize Symposium as part of the first annual meeting M10

Referenzen:

Paasche, H., Tronicke, J., Holliger, K., Green, A.G., Maurer, H., 2006. Integration of diverse physical-property models: Subsurface zonation and petrophysical parameter estimation 816 based on fuzzy c-means cluster analyses. *Geophysics* 71, H33–H44. 817 doi:10.1190/1.2192927

Präsentationen, Poster, Papers

- Bouwer, L., Nam, C., Sieck, K., Rechid, D. (24.09.2019). Large climate datasets to support the development of Climate Services: Examples from Digital Earth. Poster, REKLIM Second International Conference “Our Climate, Our Future”, Berlin, Germany.
- Hoffmann, P., Bouwer, L., Nam, C., Pfeifer, S., Rechid, D., Reinhart, V., Jacob, D. (24.10.2019). How can Machine Learning algorithms be used to develop innovative climate service products? Presentation, AI for SDGs - How Can AI Help Solve Environmental Challenges? Japanese - German - French Conference, Tokyo, Japan.
- Nam, C., Sieck, K., Bouwer, L., Rechid, D. (17.05.2019). Developing Climate Services by exploiting large datasets: examples from the Digital Earth Project. Poster, ESA Living Planet Symposium, Milan, Italy.
- Nam, C., Bouwer, L., Sieck, K., Rechid, D. (20.06.2019). Developing Climate Services by exploiting large datasets: examples from the Digital Earth Project. Poster, 4th Helmholtz Data Science Symposium, Geesthacht, Germany.
- Schäfer-Neth, C., Haas, A., Fischer, P., Koppe, R., Gerchow, P., Frickenhaus, S. (24.09.2019): Elbe river flood and draught scenarios - the MOSES and Digital Earth initiatives. Poster, REKLIM Second International Conference “Our Climate, Our Future”, Berlin, Germany.
- Silva, B., Koppe, R., Haas, A., Schäfer-Neth, C., Fischer, P., Immoor, S., Gerchow, P., Fritsch, B. and Frickenhaus, S. (24.9.2019): Automatic data quality control for understanding extreme climate event, 2nd International REKLIM Conference, Berlin, 23 September 2019 - 26 September 2019.
- Sieck, K., Nam, C., Bouwer, L.M., Rechid, D., Jacob, D. (submitted). Weather extremes over Europe under 1.5°C and 2.0°C global warming from HAPPI regional climate ensemble simulations. *Earth System Dynamics Discussions*, <https://doi.org/10.5194/esd-2020-4>
- Wichert, V. (30.01.2019). Time series in the Digital Earth Project. Poster, Time Series Analysis Conference and Training School 28-31 January 2019, Tromsø, Norway.

2. Vergleich des Stands des Vorhabens mit der ursprünglichen (bzw. mit Zustimmung des HGF geänderten) Arbeits-, Zeit- und Kostenplanung.

Im Jahr 2019 konnten die geplanten Arbeiten fast vollumfänglich so wie beantragt/nach Zustimmung mit der HGF, abgearbeitet werden. Aufgrund administrativer Probleme bei der Einstellung der Bridging PostDocs, kam es allerdings zu Verzögerungen bei deren Einstellung und daher zu einem Verzug bei der Abarbeitung der in den individuellen Projekten gesetzten Zielen. Dies hat aber keine direkten Auswirkungen auf die im Projektantrag gesetzten Ziele/Zwischenziele.

3. Haben sich die Aussichten für die Erreichung der Ziele des Vorhabens innerhalb des angegebenen Berichtszeitraums gegenüber dem ursprünglichen Antrag geändert (Begründung)?

Die Aussichten zum Erreichen der Ziele haben sich bezüglich der Antragsphase eher zum Positiven verändert. Die allgemein starke Zunahme des Interesses an Data Driven Science sowohl im Rahmen der HGF (Inkubator Projekte wie HAICU, HIP, HIDA) als auch darüber hinaus an den Zentren selbst und den Partneruniversitäten, haben der Thematik einen extra Schub verliehen. International ist die Thematik ebenfalls hochaktuell, sodass es neben der direkten Förderung durch das Projekt *Digital Earth* zu weiteren Initiativen bezüglich der *Digital Earth* Thematik kommt. Zu nennen ist hier auch das Thema Data Viewer des DataHubs für Erde & Umwelt, zu dem es thematisch eine große Schnittmenge gibt. *Digital Earth* Partnern sind maßgeblich auch in koordinierender Stellung am DataHub und ihren drei Unter-Hubs Mare, Terra, Atmo, beteiligt.

4. Sind inzwischen von dritter Seite FE-Ergebnisse bekannt geworden, die für die Durchführung des Vorhabens relevant sind (auch Darstellung der aktuellen Informationsrecherchen nach Nr. 6.1 NKBF 98)?

Die Thematik der *Data Driven Science* wird weltweit stark vorangetrieben. Im Detail gibt es zahlreiche Ergebnisse insbesondere im Bereich der Anwendung von KI im naturwissenschaftlichen Bereich. Die explizite Beschäftigung und Suche nach Anwendungsbeispielen von KI in den Erdwissenschaften bringt wöchentlich neue Erkenntnisse Dritter zu Tage. Dies ist in der sehr dynamischen Entwicklung auch zu erwarten. *Digital Earth* versucht durch seine Experten in den verschiedenen Zentren diese Ergebnisse zu bündeln und der gesamten Community zur Verfügung zu stellen.

5. Sind oder werden Änderungen in der Zielsetzung notwendig?

Die Zielsetzung muss nicht verändert werden.

6. Jährliche Fortschreibung des Verwertungsplans. Diese soll, soweit im Einzelfall zutreffend, Angaben zu folgenden Punkten enthalten (Geschäftsgeheimnisse des ZE brauchen nicht offenbart zu werden):

Nicht zutreffend.