

e-science in den Geo- und Umweltwissenschaften

Ansätze und Anforderungen

Lars Bernard · Stephan Mäs

Digitale Geo- und Umweltwissenschaften

Die Geo- und Umweltwissenschaften zählen sicherlich zu den Disziplinen, in denen die Digitalisierung nicht nur Einzug gehalten hat, sondern bereits in weiten Teilen etabliert ist. Forschungsarbeiten in fast allen Bereichen und über den gesamten Forschungszyklus erfolgen in großem Umfang digital bzw. digital unterstützt. Die satelliten- und flugzeuggestützte Fernerkundung der Erdoberfläche ist beispielsweise als Beobachtungsverfahren bereits sehr früh vollständig digitalisiert erfolgt und hat die weitere Digitalisierung stimuliert. Heutige Satellitenmissionen liefern Forschern digitale Daten in erheblichem Umfang: Allein das erst seit 2014 operationelle und noch im Aufbau befindliche Europäische Satellitenprogramm Copernicus¹ hat Wissenschaftlern Ende 2017 monatlich rund 300 Terrabytes neue Erdbeobachtungsdaten kostenfrei zur Verfügung gestellt und verzeichnet monatliche Downloads im Umfang vom 2,7 Petabytes² – bei stark ansteigendem Trend der Datenbereitstellungen und -zugriffe. Auch an der Erdoberfläche (*in situ*) werden Umweltbeobachtungen heute in großem Maße direkt digital erfasst und vielfach über Sensornetze den Forschern in (nahe) Echtzeit zur Verfügung gestellt. Neue Applikationen die für Umweltbeobachtungen die Zeitreihendaten unterschiedlicher Sensornetzbetreiber und für unterschiedliche Parameter zusammenführen (Abb. 1), zeigen einerseits eindrucksvoll die Breite und Umfänge der verfügbaren Sensordaten und skizzieren andererseits das künftige Anwendungspotenzial für den integrierten Zugriff auf unterschiedliche Umweltsensordaten.

Seit mehr als 30 Jahren werden Geoinformationssysteme (GIS) als DV-Instrumentarium der Geo- und Umweltwissenschaften eingesetzt, um raum- und zeitvariante Daten zu analysieren. Methoden zur Analyse und Simulation von beispielsweise atmosphärischen, hydrologischen oder sozioökonomischen Prozessen und deren Interaktionen basieren auf verschiedenen Simulationsstrategien (statistisch, numerisch, agentenbasiert, etc.), sind teilweise sehr rechenintensiv oder/und datenintensiv [19, 23]. Räumliche Analysen werden durch Funktionalitäten heutiger GIS umfassend unterstützt (z. B. geometrisch-topologische Analysen, Netzwerkanalysen, morphologische Analysen, geostatistische Verfahren, multikriterielle Bewertungen). Komplexere raum-zeitliche Simulationsanwendungen (z. B. numerische Strömungsmodelle, agentenbasierte Verfahren zur Populationsentwicklung) sind typischerweise lose für die Vorbereitung der Eingangsdaten bzw. Analyse der Ergebnisdaten an GIS gekoppelt.

Auch wenn sich bereits zahlreiche Ansätze für die Etablierung eines geregelten und nachhaltigen Forschungsdatenmanagements einerseits und andererseits für ein – im Sinne von e-science – kollaboratives Arbeiten in den Geo- und Umweltwissenschaften finden, existieren noch viele Herausforderungen. Entlang eines idealisierten und

<https://doi.org/10.1007/s00287-018-01138-x>
© Springer-Verlag Berlin Heidelberg 2018

Lars Bernard · Stephan Mäs
Technische Universität Dresden,
Fakultät Umweltwissenschaften, Professur für Geoinformatik,
Helmholtzstraße 10, 01069 Dresden
E-Mail: lars.bernard@tu-dresden.de

¹ <http://www.copernicus.eu/>.

² <https://scihub.copernicus.eu/reportsandstats/>.

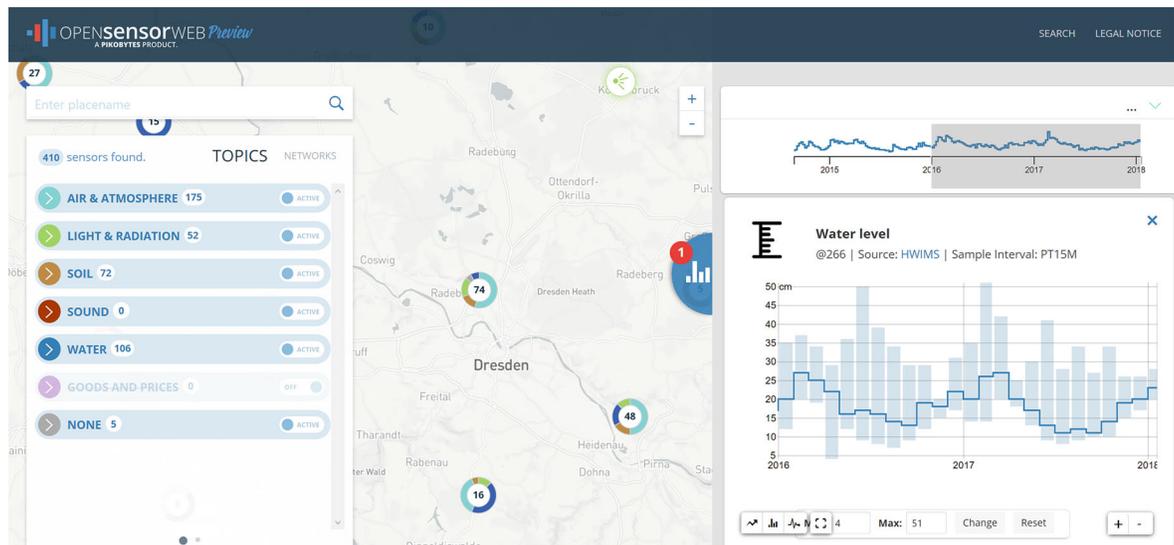


Abb. 1 Beispiel einer Applikation, die einen integrierten und performanten Zugriff auf die Monitoring-Zeitreihen von weltweit rund 400.000 unterschiedlichen Umweltbeobachtungssensoren erlaubt (<https://opensensorweb.de>)

Zusammenfassung

Die Geo- und Umweltwissenschaften arbeiten heute in großem Maße und entlang des gesamten Forschungsprozesses digital: Von der Beobachtung, Erfassung und kontinuierlichen Messung der Umweltphänomene mit unterschiedlicher Sensorik über die Analyse dieser Beobachtungsdaten und Erstellung von Modellen und Simulationen zum Verständnis oder der Vorhersage von Umweltprozessen, der Synthese von Daten verschiedener Quellen zur Beschreibung und Bewertung von Umweltveränderungen bis hin zur Publikation der Ergebnisse als wissenschaftliche Beiträge, Empfehlungen oder interaktive Visualisierungen. Allerdings gibt es viele Brüche in den Arbeitsprozessen und vielfache Hemmnisse für ein kollaboratives, auch interdisziplinäres wissenschaftliches Arbeiten im Sinne von e-science. Hier skizziert der Beitrag aktuelle Entwicklungsstände, derzeitige Forschungsansätze und verbleibende Herausforderungen für Informatik und Geoinformatik.

vereinfachten Forschungszyklus (*Datenakquise und -integration – Analyse und Modellierung – Ergebnispublikation*) versucht dieser Beitrag schlaglichtartig für die Aspekte Datenaustausch sowie Austausch

von Analyse- und Modellierungsmethoden, einige aktuelle Ansätze und Forschungsfragen für die (Geo-)Informatik in diesem Kontext darzustellen.

Austausch, Exploration und Integration von Geo- und Umweltdaten

Anbieter (und Nutzer) von Geo- und Umweltdaten sind, neben der Wissenschaft selbst, in hohem Maße die Verwaltung und kommerzielle Anbieter, aber auch freiwillige und bürgerwissenschaftliche Initiativen [2]. Hier haben speziell die Aarhus-Konvention der UNEC³ zur öffentlichen Bereitstellung von Umweltinformation und die INSPIRE EU-Direktive⁴ zum Aufbau einer Europäischen Informationsinfrastruktur für raumbezogene Umweltinformationen dazu geführt, dass heute zahlreiche und vielfältige behördliche Umweltdaten online und immer häufiger frei verfügbar sind. Auch die Zahl der aus Forschungskampagnen verfügbaren Daten steigt beständig: Das Verzeichnis der Forschungsdatenrepositorien re3data.org listet für das Schlagwort *environment* mehr als 200 Repositorien auf, von denen wiederum allein das Repository PANGAEA⁵ aktuell mehr als 370.000 Datensätze aus mehr als 12 Mio. Beobachtungen enthält. Gleichzeitig finden sich an vielen Stellen Initiativen, Anstöße und

³ <http://www.unec.org/env/pp/welcome.html>.

⁴ <https://inspire.ec.europa.eu/>.

⁵ <https://www.pangaea.de/>.

Abstract

Today's work in geosciences and environmental sciences is to a great extent digital, all along the research cycle. Starting from observations and continuous measurements of environmental phenomena using various sensor techniques, this holds along the subsequent data analysis and development of models and simulation runs to understand and forecast environmental processes, over the synthesis of data from various sources to describe and assess environmental changes towards the publication of the results as scientific papers, recommendations or interactive visualizations. However, there are several discontinuities in the related working processes as well as obstacles towards collaborative, also interdisciplinary, scientific work as envisioned by e-science. In this context, the paper discusses the status, related research approaches and remaining challenges for Computer Science and Geoinformatics.

best practices – nicht nur für das gesamte Wissenschaftssystem [17, 29], sondern auch konkret für die Geo- und Umweltwissenschaften [3, 11] – zur weiteren Entwicklung des Forschungsdatenmanagements und zur Etablierung des Austauschs und der Nachnutzung von Forschungsdaten.

Datenexploration

Dank hinreichend konsolidierter und etablierter Metadatenstandards sind bereitgestellte amtliche Daten und die in den Repositorien verfügbaren Forschungsdaten in der Regel derart beschrieben, dass eine erfolgreiche Datenrecherche auch über mehrere, verteilte Kataloge möglich ist. Aktuelle Arbeiten zeigen, wie diese Metadatenstandards darüber hinaus genutzt werden können, um für wissenschaftliche Datenrecherchen nicht nur die raumzeitliche und thematische Abdeckung, sondern auch die Entstehungsgeschichte der Daten (*provenance*) zu beschreiben und auf den Daten basierende relevante wissenschaftliche Beiträge zu referenzieren (Abb. 2).

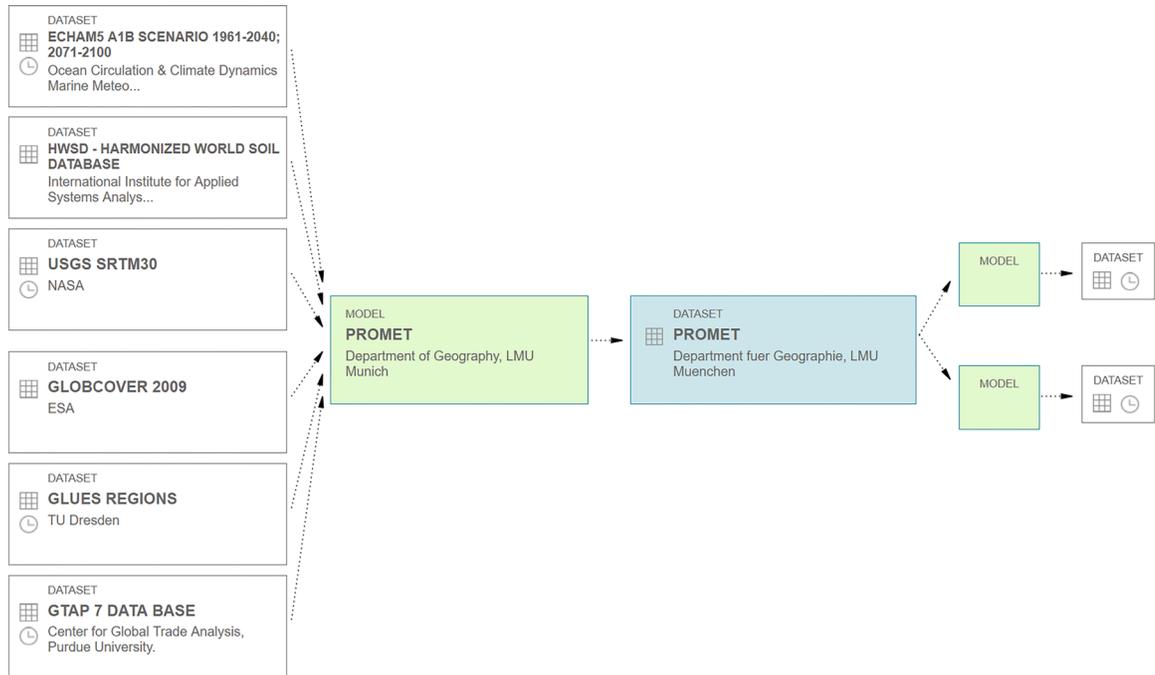
Allerdings finden sich für diese von Wissenschaftlern als sehr nützlich eingestuft und positiv aufgenommenen Anwendungen in der Regel in den aktuellen Metadatenbeständen nicht ausreichend

Inhalte. Die Eignung bzw. die Limitierungen der recherchierten Daten bezüglich der eigenen Anwendungszwecke können daher selten befriedigend beurteilt werden. Das Problem scheint hier weniger das Fehlen von Metadatenstandards – derer gibt es sicher ausreichend viele, die für die verschiedenen Anwendungszwecke einsetzbar sind. Es fehlt hingegen einerseits an vielen Stellen sicher noch ein echter Kulturwandel in Form von Erkenntnis und Willen, Daten zu publizieren und auch umfassende und hilfreiche Metainformationen bereitzustellen. Andererseits sind die manuellen Annotationsaufwände zur Erstellung von Metadaten noch deutlich zu hoch und es mangelt in den datenführenden Systemen an *smarten* Ansätzen zur weitgehend automatisierten Ableitung und Nachführung der benötigten Metadaten.

Hier setzen Arbeiten an, die darauf zielen das Forschungsdatenmanagement von der Erfassung und Simulation über die Forschungsdatenpublikation, Datenqualifikation bis zur Bereitstellung von qualifizierten Daten möglichst umfassend und bruchfrei zu unterstützen. Erste prototypische Lösungen [5] bieten Plattformen für den internen Datenaustausch in verteilten Forschungsteams, die aus einer Modellierung der Analyse- und Datenflüsse entsprechende Datenstrukturen automatisiert ableiten, die Integration externer Umweltdatendienste erlauben und Werkzeuge zur Qualifizierung der Forschungsergebnisse sowie für die Publikation als standardisierte Webdienste anbieten. Eine solche durchgängige Unterstützung des Forschungsdatenmanagements erlaubt es, Metadaten weitgehend automatisiert aus den verschiedenen Annotationen und durchgeführten Prozessierungen zu ermitteln und zu aktualisieren. Die Erstellung der Metadaten und Qualifizierung der Daten zu publizierbaren Ergebnissen ist so nicht mehr von der Datenbearbeitung entkoppelt und nicht länger leidige – gern vergessene – Pflichtaufgabe am Ende eines Forschungsprojekts, sondern ein von Beginn an mitlaufender Prozess.

Datenintegration

Die in den verfügbaren Daten beschriebenen Phänomene reichen über zahlreiche räumliche, zeitliche und thematische Skalen. Qualitäten, Strukturen und Semantiken der erhobenen, genutzten und abgeleiteten Daten differieren ebenso wie die Nutzungsbedingungen. Die für die Geo- und Um-



LINEAGE

PROCESS DESCRIPTION

RATIONALE not set
 DESCRIPTION Simulation of potential yield on representative sampling units in a Glues Region for 16 Glues-Plants

PROCESS PARAMS

TIME OF EXECUTION 2012-01-01T00:00:00
 SOFTWARE REFERENCE not defined

PROCESSOR OF DATA Department of Geography, LMU Munich

PUBLICATION

1. PUBLICATION

TITLE OF PUBLICATION
 Mauser, W., Bach, H. (2009):
 PROMET - Large scale
 distributed hydrological
 modelling to study the impact
 of climate change on the
 water flows of mountain
 watersheds, Journal of
 Hydrology.
 PUBLISHING DATE
 2009-07-22
 CITATION
<http://www.sciencedirect.com/science/article/pii/S0022169409004478>

Abb. 2 Beispiel der Darstellung der Entstehungsgeschichte und Verwendung von Umweltdaten unter Nutzung standardisierter Metadaten [11]

weltforschung betrachteten Daten können im Sinne von *volume* und *variety* sicherlich als *Big Data* gelten, in der Forschung beispielsweise zu Naturkatastrophen geht es oft auch um *velocity* im Kontext der Datenerfassung, -übertragung und -analyse. Zur Schaffung von Interoperabilität und zur Harmonisierung der Datenmodelle für Geo- und Umweltdaten finden sich gesetzliche Rahmenbedingung, internationale Normen und zahlreiche De-facto-Standards, die sich in unterschiedlicher

Weise und verschieden weitreichend durchgesetzt haben [2].

Aber auch wenn neben den existierenden Regelungen und Standards die Verortung in Raum und Zeit zunächst einmal einen Referenzrahmen liefern, stellt die Zusammenführung von Daten unterschiedlicher Quellen immer noch erhebliche Herausforderungen. In Forschungsanwendungen werden solche Datenfusionen beispielsweise notwendig, um räumlich oder zeitlich benachbarte

Daten zu einem Datensatz mit einer möglichst vollständigen Abdeckung zusammenzuführen, die Datensätze für ein Untersuchungsgebiet thematisch anzureichern, Daten zu aktualisieren und um Veränderungen, Unterschiede oder mögliche Fehler zu erkennen. Die Aufwendungen für die notwendigen Datenkuratierungen, Schematransformationen und Datenfusionierung sind heute meist erheblich. Die Prozesse dazu sind oft nicht oder nur begrenzt automatisiert. Diese Problematik adressieren aktuelle Forschungsarbeiten, die zeigen, wie unter Verwendung von automatisiert abgeleiteten geometrischen, topologischen und semantischen Ähnlichkeitsmaßen Vorschläge für Fusionspfade zwischen zwei oder mehreren Datensätzen abgeleitet werden können. Solche Fusionspfade können dann beispielsweise in Linked-Data-Strukturen abgelegt und so für weitere Abfragen zugänglich gemacht werden [25, 26]. Die explizierte Bereitstellung der Ähnlichkeitsmaße und Fusionspfade schafft insbesondere die Möglichkeit, ad-hoc und transparent Daten für eigene Anwendungszwecke zu integrieren. Weitere aktuelle Forschungsarbeiten zeigen, wie weit aus Fernerkundungsdaten automatisiert Informationen zu Objekten der Erdoberfläche abgeleitet werden können [8]. Die bis dato wenig genutzten Verfahren des maschinellen Lernens finden in Anbetracht der heute verfügbaren großen Menge geeigneter *Lern-daten* zur Objekterkennung in Fernerkundungsdaten eine immer maßgeblichere Rolle.

Künftige Forschungsarbeiten sollten hier auf eine weitere Zusammenführung der unterschiedlichen Verfahren zielen. So gilt es beispielsweise, die verschiedenen Datenfusionsansätze derart zu kombinieren, dass Attribute eines Geoobjekts automatisiert und etwa unter Einbindung neuer Sensordatenströme ergänzt, aktualisiert bzw. auf Veränderungen untersucht werden können. Darüber hinaus sollten solche Verfahren auch (automatisiert) immer Unsicherheitsmaße und Information zur Datengeschichte und -genese liefern können.

Kollaborative raumzeitliche Analysen und Umweltmodellierungen

Wang et al. [24] haben *CyberGIS* als eine Bezeichnung für die aktuell im Kontext von Cloud-Umgebungen und Big-Data-Anwendungen entstehende neue Generation von GIS eingeführt, die neben die etablierten desktop- und serverbasierten GIS sowie zu den Programmierschnittstellen für

webbasierte Geoinformationsangebote treten sollen. Unter Nutzung von Cyberinfrastructures [15] als Kombination von Hochleistungsrechnern, gemeinsamen großen Datenrepositorien, Visualisierungsumgebungen, hoch performanten Netzumgebungen sollen Domänenexperten und Entwicklern solche CyberGIS als geeignete Umgebungen bereitgestellt werden, um raum-zeitliche Analysen auch für sehr große – bis dato nicht effizient nutzbare – Datenmengen bzw. Berechnungsumfänge durchführen zu können. Beispiele sind hier Analysen auf den immer umfangreicheren Fernerkundungsdaten oder die zusammenhängende Berechnung von hydrologischen Parametern in hoher zeitlicher und räumlicher Auflösung für große Untersuchungsgebiete und in (nahe) Echtzeit für zeitkritische Hochwasserwarnungen. Insbesondere erlauben CyberGIS, nicht nur komplexere und umfangreichere Analyseberechnungen durchzuführen, sondern dafür auch unter Einsatz etwa von Monte-Carlo-Ansätzen in nahe Echtzeit Unsicherheitsmaße abzuleiten, für die bisherige Systeme nicht ausreichend Rechenleistung aufbieten können [7]. Gerade für Vorhersagesysteme und entscheidungsunterstützende Systeme ist die Bereitstellung solcher Unsicherheitsmaße eine sehr wichtige und wertvolle Erweiterung. Um Hochleistungsrechner und die vernetzten Rechenressourcen effizient einzusetzen, sollten die Anwendungen parallelisierbar sein. Teilweise lassen sich Algorithmen bereits durch eine (räumliche) Zerlegung der Eingangsdaten einfach auf verteilte Ressourcen aufteilen – etwa bei der Anwendung lokaler wirkender Operatoren. Für komplexere dynamische Simulationen stellen sich jedoch umfangreichere und zum Teil nicht triviale Anforderungen zur Parallelisierbarkeit der Algorithmen.

Zur weiteren Entwicklung von CyberGIS stellen sich viele Forschungsfragen. Diese betreffen generelle Themen wie beispielsweise Cybersecurity mit dem Ziel der Sicherstellung von Informations- und Systemsicherheit sowie Lizenz- und Datenschutz in den verteilten Anwendungen. Dabei gibt es in Anwendungsbereichen wie der Umweltmedizin und der Auswertung von Bewegungsprofilen oder sozioökonomischen Daten aktuell häufig einen Zielkonflikt zwischen dem Schutz personenbezogener Daten und effizienten und umfassenden wissenschaftlichen Analysen, der sich derzeit nur selten zufriedenstellen lösen lässt. Eine zukünftige Formalisierung von (ge-

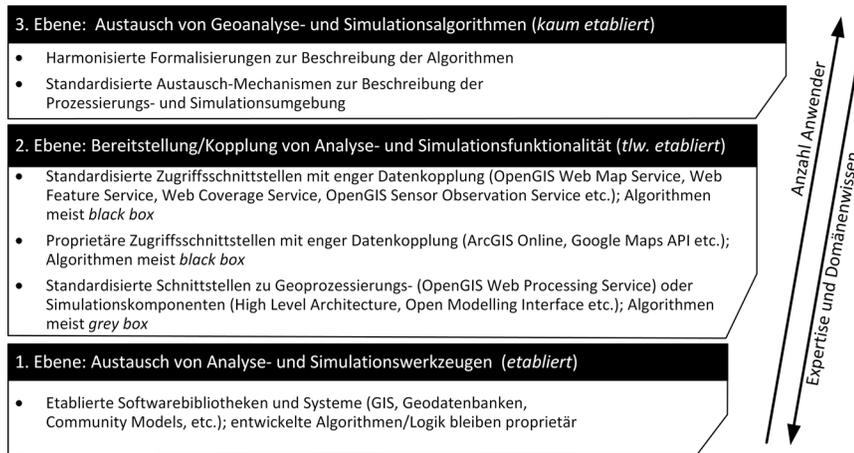


Abb. 3
Interoperabilitätsebenen für den Austausch von Geoanalyse- und Umweltsimulationsalgorithmen.
 (Verändert nach [2])

setzlichen) Datenschutzbestimmungen derart, dass diese als abrufbare und prüfbare Regeln für Analysesoftware bereitstehen, kann hier ein Weg sein, um eine automatisierte Durchsetzung eines Datenschutzes zu ermöglichen und die Anwendungseffizienz nicht über Gebühr einzuschränken.

Die Schaffung von Interoperabilität für verteilte GIS allgemein und im speziellen Fall CyberGIS ist ein zentraler Aspekt zahlreicher Ansätze und Forschungsarbeiten in der Geoinformatik. Ziel ist dabei, Grundlagen zu schaffen, die es erlauben, Simulations- und Geoprozessierungsalgorithmen auszutauschen, vielfältig für eigene Arbeitsabläufe und über idealerweise Domänengrenzen hinweg zu kombinieren und systemunabhängig und transparent zu beschreiben. Grob lassen sich hier nach Grad der erreichten Interoperabilität und Grad der Etablierung drei Ebenen unterscheiden (Abb. 3).

Die erste Ebene umfasst den traditionell und auch heute bei Weitem überwiegenden Austausch von Analyse- und Simulationswerkzeugen über die gemeinsame Entwicklung und Nutzung von meist als Open Source zur Verfügung gestellten Softwarebibliotheken (z. B. ArcGIS, GRASS GIS, PostGIS, QGIS) oder Community Models (z. B. Community Earth System Model⁶, OpenGeoSys⁷). Allerdings beschränkt sich die Nutzung – und der Austausch – dieser Systeme meist auf Nutzer mit einem vertieften Fachwissen und weit überdurchschnittlicher Softwareentwicklungskompetenz. Diese Systeme werden in der Regel nicht fachübergreifend genutzt und die

verwendeten Algorithmen sind zwar idealerweise in Dokumentationen beschrieben, aber dennoch proprietär.

Eine zweite Ebene machen Ansätze zur Bereitstellung und Kopplung von Analyse- und Simulationsfunktionalität aus. Webbasierte Dienst-schnittstellen erleichtern die Nutzung, auch über Domänengrenzen hinweg und helfen, größere Anwendergruppen zu erreichen. Zahlreiche etablierte Angebote stellen Umweltdaten und darauf aufsetzende und eng an die Daten gekoppelte Visualisierungs- und Abfragefunktionen über standardisierte Schnittstellen (z. B. OpenGIS Web Map Service, Web Feature Service, OpenGIS Web Coverage Service; vgl. [2]) oder proprietäre Schnittstellen zur Verfügung (z. B. Google Maps API, ArcGIS Online). Die Nutzung der Analysefunktionalität für eigene Daten ist jedoch nur begrenzt oder überhaupt nicht möglich. Über diese Schnittstellen werden auch Ergebnisse aus (vorab ausgeführten) Simulationsläufen verfügbar gemacht (z. B. des International Panel on Climate Change, IPCC⁸) oder ein robustes und in der Regel vorparametrisiertes Modell erlaubt eine Ad-hoc-Ausführung, etwa zur Abschätzung der Luftverschmutzung unter Nutzung aktueller Wetter- und Luftschadstoffdaten [20, 28]. Weniger etabliert, aber dennoch in einigen Ansätzen zu finden, sind Nutzungen standardisierter Schnittstellen, um eine Kopplung unterschiedlicher Geoprozessierungsfunktionalitäten und/oder Simulationskomponenten auch in loser Kopplung an die eingehenden Daten zu ermöglichen. Der sehr generische Schnittstellen-

⁶ <http://www.cesm.ucar.edu/>.

⁷ <http://www.opengeosys.org/>.

⁸ <http://www.ipcc-data.org/>.

standard des *OpenGIS Web Processing Service* erlaubt es, Geoprozessierungsfunktionalität – z. B. zur raum-zeitlichen Verschneidung von Umweltdaten oder zur räumlichen Interpolation von punkthaft erhobenen Messwerten – als Webdienst bereitzustellen [2]. Die Funktionalität wird dabei im Sinne einer *grey box* gekapselt, die Spezifika des Algorithmus werden also freitextlich, nicht jedoch formal beschrieben [12]. Für die Kopplung von Umweltsimulationskomponenten existieren etwa mit der *High Level Architecture* [9] oder dem *OGC Open Modelling Interface*⁹ Ansätze zur Harmonisierung und Standardisierung der Schnittstellen. Für die Integration von Simulationskomponenten, speziell über Grenzen der Teildisziplinen hinweg, konnte sich jedoch bisher kein Standard etablieren. Die Entwicklungsaufwände für integrierte Umweltmodelle (z. B. zur gekoppelten Simulation von Szenarien zur sozio-ökonomischen Entwicklung und Effekte auf Atmosphäre, Ozean und Vegetation) die idealerweise auch Rückkopplungen berücksichtigen, bleiben daher sehr hoch und die heutigen integrierten Modelle sind eher monolithisch und damit Erweiterungen recht aufwendig [23].

Die dritte Interoperabilitätsebene zielt auf den einfachen Austausch von Analyse- und Simulationsalgorithmen zwischen Systemen bzw. Umweltforschern unterschiedlicher Disziplinen. Hier finden sich viele verschiedene Ansätze auf ganz unterschiedlichen Formalisierungs- bzw. Abstraktionsebenen, die sich zwischen teilweise generischen skriptbasierten Ansätzen [10] bis hin zu Formalisierungen einer Geoprozessierungsalgebra [6, 18] bewegen. Sie haben jedoch selten einen praxisrelevanten Reifegrad erreicht und finden kaum Anwendung in umweltwissenschaftlichen Projekten. Einzig Teilkonzepte, wie die Egenhofer-Operatoren [4] oder die von Tomlin entwickelte Map Algebra [22] haben sich etabliert und sind vielfach in heutigen Datenbanksystemen und GIS zu finden. Hier bleibt es Aufgabe der Forschung, zu prüfen, wie weit realisierbar die Vision von Beschreibungsmöglichkeiten für Algorithmen zur raum-zeitlichen Analyse und Modellierung ist, um zu ermöglichen, dass (a) diese Algorithmen im Sinne eines *moving code* problemlos zwischen verschiedenen Anwendungsumgebungen ausgetauscht werden können [13], (b) daraus komplexere Ana-

lysen leicht in ausführbare Workflows orchestriert werden können [6], (c) Wissenschaftlern zu ihren Analysefragestellungen passende Algorithmen und Operatoren vorgeschlagen werden [21, 27] und (d) durchgeführte Analyseschritte transparent und für andere Forscher nachvollziehbar gemacht werden können. Alternativ dazu stehen pragmatische Ansätze zur Schaffung von Datenmediatoren, die unter Nutzung gemeinsamer Ontologien und von webbasierten Diensteschnittstellen eine Integration von unterschiedlichen Simulationsmodellen zwar im Wesentlichen auf den Datenaustausch beschränken, jedoch eine effizient nutzbare, und mit geringerem Aufwand realisierbare Lösung skizzieren [1].

Herausforderungen und Fazit

Mit dem Ziel des *reproducible research* sind immer häufiger und in vielen Disziplinen wissenschaftliche Publikation von Forschungsergebnissen an die Publikation der zugehörigen Forschungsdaten und/oder der entwickelten Algorithmen und Software gekoppelt. Für die in den Umweltwissenschaften meist als Open Source erfolgenden Softwarepublikationen haben sich verschiedene Plattformen (SourceForge, GitHub etc.) etabliert. Auch für die Datenpublikationen gibt es etablierte Repositorien (z. B. PANGAEA), in denen Datensätze durch einen Digital Object Identifier (DOI) eindeutig referenzierbar sowie durch standardisierte Metadatenangaben und pragmatische Vorgaben zu Datenstrukturen recherchierbar gemacht werden. Es fehlt derzeit jedoch an einheitlichen Vorstellungen bzw. Absprachen zu Aspekten wie Qualitätssicherung, Versionierungen, Lizenzierung oder langzeitlicher Archivierung von Forschungsergebnissen der Umweltwissenschaften als Daten und Software.

Es finden sich generelle Anforderungen von Fördermittelgebern zum Forschungsdatenmanagement und verschiedene auch fachspezifische Empfehlungspapiere und Ansätze [3, 14], die hier jedoch jeweils nur Teilaspekte adressieren. Hier sind weitere gemeinsame Arbeiten von Fachgesellschaften in den Umweltwissenschaften, Forschungseinrichtungen und Fördermittelgebern notwendig. Der Rat für Informationsinfrastrukturen (RfII¹⁰) skizziert mit der Nationalen Forschungsdateninfrastruktur (NFDI) einen Rahmen für diese Bemühungen [16].

⁹ <http://www.openmi.org/>.

¹⁰ <http://www.rfii.de>.

Bereits in 2008 wurde die Schwerpunktinitiative *Digitale Information*¹¹ der Allianz der deutschen Wissenschaftsorganisationen zur Koordination und Bündelung von Aktivitäten zur Digitalisierung in der Wissenschaft begonnen. Diese Initiative hat ebenfalls eine Reihe grundlegender Empfehlungen und Leitlinien, beispielweise zum Forschungsdatenmanagement und dem Umgang mit Forschungssoftware publiziert, die wissenschaftlichen Fachgesellschaften und Einrichtungen als Referenz dienen können.

Eine der großen Herausforderungen liegt sicherlich in der Etablierung von allgemein akzeptierten Formen der Qualifizierung und Qualitätsbewertung für zu publizierende Forschungsdaten und Softwarewerkzeuge. Vorstellbar sind hier beispielsweise die Formulierung unterschiedlicher Qualitätslevel, Nutzung von Selbstverpflichtungsverfahren oder Evaluierungsverfahren (*peer reviews*). Weiterhin denkbar sind geeignete moderierte Verfahren zur kontinuierlichen Qualitätsverbesserung und Versionierung der Publikationen. Daran anknüpfen kann ein Monitoring der Nachnutzung von Daten und Werkzeugen, welches dann auch als Grundlage für Einschätzungen zur Relevanz der publizierten Daten und Software, bis hin zur Etablierung von entsprechenden Nutzungs- und Zitierungsindizes dienen kann.

Weitere große Herausforderungen, etwa die Sicherstellung der langfristigen Archivierung der als relevant erachteten Ergebnisse oder die Etablierung und Durchsetzung von Lizenz- und Urheberschutzrechten, erscheinen weniger fachspezifisch und sollten als generelle Querschnittsaufgabe für das künftige Forschungsdatenmanagement adressiert werden.

Um *e-science* in den Umweltwissenschaften vollumfänglich zu etablieren, sind noch einige Anstrengungen notwendig. Neben den in diesem Beitrag angeführten Forschungsaufgaben für die (Geo-)Informatik gehören dazu insbesondere auch die weitere Förderung und Etablierung einer offenen Daten- und Softwarekultur, die auf Kollaboration, Austausch und Transparenz unter Nutzung der entstehenden Plattformen und Systeme zielt.

Wesentliches Element hierfür ist sicherlich die Anpassung der Aus- und Weiterbildung von

Umweltwissenschaftlern. Schwerpunkte sollten dabei nicht allein der Kompetenzerwerb für das Forschungsdatenmanagement sowie zum fachkundigen Umgang mit den Analyse-, Modellierungs- und Kollaborationswerkzeugen sein, die sie sich bereits – allerdings nur vereinzelt – in derzeitigen Curricula und Promotionsprogrammen finden. Sondern es sollte auch eine Sensibilisierung für Qualitätsanforderungen und für rechtliche Aspekte bei der Arbeit mit und Publikation von Forschungsdaten erreicht werden, die heute nicht oder nur sehr rudimentär in Qualifizierungsprogrammen zu finden ist.

Literatur

1. Belete GF, Voinov A, Morales J (2017) Designing the distributed model integration framework – DMIF. *Environ Modell Softw* 94:112–126, <https://doi.org/10.1016/j.envsoft.2017.04.003>
2. Bernard L, Brauner J, Mäs S, Wiemann S (2016) Geodateninfrastrukturen. In: Freedon W, Rummel R (eds) *Handbuch der Geodäsie*: 5 Bde. Springer, Berlin Heidelberg, S 1–32
3. Bertelmann R, Gebauer P, Hasler T, Kirchner I, Peters-Kottig W, Razum M, Recker A, Ulbricht D, van Gassel S (2014) Einstieg ins Forschungsdatenmanagement in den Geowissenschaften. Helmholtz-Zentrum Potsdam – Deutsches GeoForschungszentrum GFZ, Potsdam, <https://doi.org/10.2312/lis.14.01>
4. Egenhofer M (1994) Spatial SQL: A query and presentation language. *IEEE Trans Knowl Data Eng* 6(1):86–95
5. Henzen D, Müller M, Jirka S, Senner I, Kaeseberg T, Zhang J, Bernard L, Krebs P (2016) A scientific data management infrastructure for environmental monitoring and modelling. Paper presented at the 8th International Congress on Environmental Modelling and Software, Toulouse, France
6. Hofer B, Mäs S, Brauner J, Bernard L (2017) Towards a knowledge base to support geoprocessing workflow development. *Intern J Geogr Inf Sci* 31(4):694–716, <https://doi.org/10.1080/13658816.2016.1227441>
7. Hu H, Lin T, Wang S, Rodriguez LF (2017) A cyberGIS approach to uncertainty and sensitivity analysis in biomass supply chain optimization. *Applied Energy* 203:26–40, <https://doi.org/10.1016/j.apenergy.2017.03.107>
8. Hussain M, Chen D, Cheng A, Wei H, Stanley D (2013) Change detection from remotely sensed images: From pixel-based to object-based approaches. *ISPRS J Photogr Remote Sens* 80:91–106, <https://doi.org/10.1016/j.isprsjprs.2013.03.006>
9. IEEE (2011) 1516 IEEE standard for modeling and simulation (M&S) high level architecture (HLA) framework and rules. Institute of Electrical and Electronics Engineers, New York
10. Knox S, Meier P, Yoon J, Harou JJ (2018) A python framework for multi-agent simulation of networked resource systems. *Environ Modell Softw* 103:16–28, <https://doi.org/10.1016/j.envsoft.2018.01.019>
11. Mäs S, Henzen C, Müller M, Bernard L (2014) GLUES GDI – Eine Geodateninfrastruktur für wissenschaftliche Umweltdaten. *gis.SCIENCE* 2014(4):129–137
12. Müller M (2015) Hierarchical profiling of geoprocessing services. *Comput Geosci* 82:68–77, <https://doi.org/10.1016/j.cageo.2015.05.017>
13. Müller M, Bernard L, Brauner J (2010) Moving code in spatial data infrastructures – web service based deployment of geoprocessing algorithms. *Trans GIS* 14(S1):101–118, doi:10.1111/j.1467-9671.2010.01205.x
14. Nüst D, Konkol M, Pebesma E, Kray C, Schutzzeichel M, Przybytzin H, Lorenz J (2017) Opening the publication process with executable research compendia. *D-Lib Mag* 23(1/2), <https://doi.org/10.1045/january2017-nuest>
15. NSF (2007) Cyberinfrastructure vision for 21st century discovery [online]. US National Science Foundation Cyberinfrastructure Council, https://www.nsf.gov/pubs/2007/nsf0728/nsf0728_1.pdf, last access: 29.9.2018
16. Rat für Informationsinfrastrukturen (2016) Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland. Rat für Informationsstrukturen, Göttingen
17. Rat für Informationsinfrastrukturen (2017) Entwicklung von Forschungsdateninfrastrukturen im internationalen Vergleich: Bericht und Anregungen Rat für Informationsinfrastrukturen. urn:nbn:de:101:1-201711084652

¹¹ <https://www.allianzinitiative.de/>.

18. Scheider S, Gräler B, Pebesma E, Stasch C (2016) Modeling spatiotemporal information generation. *Intern J Geogr Inf Sci* 30(10):1980–2008, <https://doi.org/10.1080/13658816.2016.1151520>
19. de Smith MG, Longley P (2015) Geospatial analysis. <http://www.spatialanalysisonline.com>, last access: 19.12.2018
20. Stasch C, Foerster T, Autermann C, Pebesma E (2012) Spatio-temporal aggregation of European air quality observations in the Sensor Web. *Comput Geosci* 47:111–118, <https://doi.org/10.1016/j.cageo.2011.11.008>
21. Stasch C, Scheider S, Pebesma E, Kuhn W (2014) Meaningful spatial prediction and aggregation. *Environ Modell Softw* 51:149–165, <https://doi.org/10.1016/j.envsoft.2013.09.006>
22. Tomlin D (1990) *Geographic information systems and cartographic modelling*. Prentice-Hall, Englewood Cliffs
23. Voinov A, Shugart HH (2013) “Integronsters”, integral and integrated modeling. *Environ Modell Softw* 39:149–158, <http://dx.doi.org/10.1016/j.envsoft.2012.05.014>
24. Wang S, Anselin L, Bhaduri B, Crosby C, Goodchild MF, Liu Y, Nyerges TL (2013) CyberGIS software: a synthetic review and integration roadmap. *Int J Geogr Inf Sci* 27(11):2122–2145, <https://doi.org/10.1080/13658816.2013.776049>
25. Wiemann S (2017) Formalization and web-based implementation of spatial data fusion. *Comp Geosci* 99:107–115, <https://doi.org/10.1016/j.cageo.2016.10.014>
26. Wiemann S, Bernard L (2016) Spatial data fusion in spatial data infrastructures using Linked Data. *Intern J Geogr Inf Sci* 30(4):613–636, doi:10.1080/13658816.2015.1084420
27. Wiemann S, Karrasch P, Bernard L (2018) Ad-hoc combination and analysis of heterogeneous and distributed spatial data for environmental monitoring – design and prototype of a web-based solution. *Intern J Digit Earth* 11(1):79–94, <https://doi.org/10.1080/17538947.2017.1326984>
28. Wiemann SR, Karrasch P, Brauner J, Pech K, Bernard L (2012) Classification-driven air pollution mapping as for environment and health analysis. Paper presented at the iEMSs Sixth Biennial Meeting: International Congress on Environmental Modelling and Software (iEMSs 2012). International Environmental Modelling and Software Society
29. Wissenschaftsrat (2012) Empfehlungen zur Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen in Deutschland bis 2020, <http://www.wissenschaftsrat.de/download/archiv/2359-12.pdf>, letzter Zugriff: 19.12.2018