

Technische Universität Dresden, Professur für Geoinformationssysteme

GLUES GDI – EINE GEODATENINFRASTRUKTUR FÜR WISSENSCHAFTLICHE UMWELTDATEN

Stephan Mäs, Christin Henzen, Matthias Müller, Lars Bernard

Zusammenfassung: In den letzten Jahren gewinnt die Publikation wissenschaftlicher Daten und der damit verbundene Aufbau von Forschungsinfrastrukturen immer mehr an Bedeutung. Demgegenüber gibt es bisher kaum Implementierungen von Geodateninfrastrukturen und deren Komponenten in umweltbezogenen Forschungsinfrastrukturen. Der Beitrag liefert einen Erfahrungsbericht über den Aufbau einer wissenschaftlichen Geodateninfrastruktur als Datenaustauschplattform innerhalb eines BMBF-Forschungsprogramms mit dem Schwerpunkt Landmanagement und Klimawandel.

Die Veröffentlichung wissenschaftlicher Umweltdaten stellt besondere Anforderungen an die Geodateninfrastruktur. Die hier betrachteten Forschungsschwerpunkte sind dabei insbesondere die Beschreibung der Entstehungsgeschichte der Daten, die Unterstützung von Zeitreihendaten und verteilte Geoprocessing als Werkzeug für die wissenschaftliche Arbeit und zur Datenaufbereitung. Wissenschaftliche Umweltdaten und auch die Wissenschaftler als spezielle Nutzergruppe stellen gerade in diesen Bereichen besondere Anforderungen, welche aber von den am Markt verfügbaren Lösungen in aller Regel nicht unterstützt oder zumindest vernachlässigt werden. Deshalb wurden für die hier vorgestellte wissenschaftliche Geodateninfrastruktur eine Reihe von speziell angepassten Werkzeugen und Lösungen entwickelt, unter anderem für die interaktive Visualisierung der Datenhistorie und Zeitreihendaten. Diese Entwicklungen sowie der allgemeine Entwicklungsstand der GDI nach vierjähriger Projektlaufzeit werden im Beitrag vorgestellt und diskutiert.

Schlüsselwörter: Wissenschaftliche Geodateninfrastruktur, SDI, Austausch von Umweltmodellen, e-Science, Cyberinfrastruktur

GLUES GDI – A SPATIAL DATA INFRASTRUCTURE FOR SCIENTIFIC ENVIRONMENTAL DATA

Abstract: Recently, the publication of scientific data and therewith the implementation of research infrastructures in the Web is getting more and more attention. However, spatial data infrastructures respectively their components are hardly used in research infrastructures of environmental data. This paper reports experiences from the implementation of a scientific spatial data infrastructure supporting the data exchange within a land management and climate change research program.

The publication of scientific environmental data has specific requirements on the spatial data infrastructure. The focus of this article is particularly on the description of data provenance, the support of time series data and distributed geoprocessing in support of scientific work and data preparation. Here, scientific environmental data and also the scientists as specific user group have particular requirements, which are usually not completely supported by available tools. Therefore, the introduced scientific spatial data infrastructure contains a set of specific tools, for example, to support an interactive visualization of data provenance and time series data. The article illustrates and discusses these particular implementations of the GLUES GDI as well as the overall stage of development after four years of project duration.

Keywords: Scientific spatial data infrastructure, sharing environmental models, geospatial cyberinfrastructure, e-science, geodata infrastructure

Autoren

Dr.-Ing. Stephan Mäs
Dipl.-Medieninf. Christin Henzen
Dipl.-Geogr. Matthias Müller
Prof. Dr. rer. nat. Lars Bernard

Technische Universität Dresden
Professur für Geoinformationssysteme
Helmholtzstr. 10
01069 Dresden
E: {Stephan.Maes, Christin.Henzen,
Matthias_Mueller, Lars.Bernard}@tu-dresden.de

1 EINLEITUNG

Die Veröffentlichung wissenschaftlicher Daten als zusätzlichen Output neben wissenschaftlichen Publikationen gewinnt in den letzten Jahren mehr und mehr an Bedeutung. Große Forschungsinitiativen wie die „Initiative on Scientific Cyberinfrastructures“ der US National Science Foundation (NSF 2007) oder im europäischen Raum die „European roadmap for research infrastructures“ (Esfri 2008) forcieren den Aufbau von Forschungsinfrastrukturen für den Austausch wissenschaftlicher Daten. Während in Europa meist von e-Science oder Research Infrastructures die Rede ist, hat sich im amerikanischen und asiatischen Raum hierfür der Begriff Cyberinfrastructure etabliert. Die Ziele von Forschungsinfrastrukturen sind dabei unter anderem die verbesserte Dokumentation und Transparenz der Forschungsarbeiten, eine höhere Rentabilität öffentlich geförderter Forschungsinitiativen und die Unterstützung von datenintensiver multidisziplinärer Forschung. Die bessere Verfügbarkeit der Daten stimuliert deren Wiederverwendung und damit die Kollaboration zwischen Wissenschaftlern. Außerdem stehen die Forschungsergebnisse auch für die Öffentlichkeit und für politische Entscheidungsträger zur Verfügung.

Vor diesem Hintergrund ist das Informationsmanagement rund um digitale Forschungsdaten inzwischen selbst zum Forschungsgegenstand geworden. Dabei wird beispielsweise untersucht, was die tatsächlichen Mehrwerte solcher Forschungsinfrastrukturen und damit die motivierenden Faktoren für ihre Benutzung sind, wie die Qualität der Forschungsdaten sichergestellt werden kann, welche Inhalte in den Metadaten beschrieben werden müssen und wie die Langzeitarchivierung von Forschungsdaten sichergestellt werden kann (Kindling 2013).

Die derzeit verfügbaren wissenschaftlichen Dateninfrastrukturen sind in der Regel auf bestimmte Domänen spezialisiert, wie beispielsweise das soziale Netzwerk MyExperiment¹ (Goble et al. 2010) für den Austausch von Prozessabläufen im Bereich der Bioinformatik oder die iPlant Collaborative Cyberinfrastructure² für Pflanzenwissenschaften. Auch im Bereich der Geo- und Umweltwissenschaften gibt es Publikationsplattformen wie PANGAEA (Data Publisher for Earth and Environmental Science)³ und Geoshare⁴ oder die Global Collaboration

Engine (GLOBE)⁵ für die Recherche und den Vergleich von regionalen Fallstudien (Ellis 2012). Die zentralen Funktionalitäten dieser Plattformen sind in der Regel die Recherche in Katalogen, der Zugriff auf Daten und Visualisierungen, die Bereitstellung von Rechenkapazitäten sowie die Interaktion und Kollaboration zwischen Wissenschaftlern.

Obwohl Geodateninfrastrukturen (GDI) ebenfalls die meisten dieser Funktionalitäten unterstützen, gibt es bisher kaum entsprechende Implementierungen als Forschungsinfrastrukturen. Dabei könnten die standardisierten Schnittstellen und Formate den interoperablen Austausch der Daten unterstützen und potenziell die Arbeit der Wissenschaftler bei der Recherche, der Verarbeitung in unterschiedlichen Systemen und der Integration der Daten und Modelle wesentlich vereinfachen (Bernard 2013). Dieser Beitrag liefert einen Erfahrungsbericht über den Aufbau einer wissenschaftlichen GDI als Datenaustauschplattform innerhalb eines BMBF-Forschungsprogramms. Schwerpunktthemen des Forschungsprogramms sind nachhaltiges Landmanagement, Klimawandel und ökosystemare Dienstleistungen.

Im vorgestellten Projekt wurde untersucht, inwieweit sich die vorhandenen GDI-Technologien und Schnittstellen für die Dokumentation und Publikation wissenschaftlicher Daten im Web eignen und ob GDI die wissenschaftliche Arbeit und die Informationsbereitstellung für die relevanten Stakeholder unterstützen können. Forschungsschwerpunkte waren dabei unter anderem Metadaten für die Beschreibung wissenschaftlicher Umweltdaten und dabei insbesondere die Beschreibung der Entstehungsgeschichte der Daten, die Unterstützung von Zeitreihendaten und verteilte Geoprozessierung als Werkzeug für die wissenschaftliche Arbeit und zur Datenaufbereitung. Dabei wurden für diese Bereiche Probleme und Herausforderungen identifiziert und Lösungsmöglichkeiten untersucht.

Im folgenden Kapitel werden zunächst der Projekthintergrund, die Rahmenbedingungen im Forschungsprogramm und dann die konkreten Ziele der Geodateninfrastruktur beschrieben. Die Kapitel 3 und 4 liefern dann einen Überblick über den Entwicklungsstand der wissenschaftlichen GDI nach vierjähriger Projektlaufzeit und über

die, im Vergleich zu anderen Geodateninfrastrukturen, besonderen Inhalte, Werkzeuge und Funktionalitäten.

2 PROJEKTHINTERGRUND: GLUES ALS KOORDINATIONSPROJEKT IM FORSCHUNGSPROGRAMM „NACHHALTIGES LANDMANAGEMENT“

Die Schwerpunkte des BMBF-Forschungsprogramms „Nachhaltiges Landmanagement“⁶ liegen in den Themenfeldern Landmanagement, Treibhausgasemissionen, Klimawandel und Ökosystemdienstleistungen sowie deren Wechselwirkungen und gegenseitigen Abhängigkeiten. Vor dem Hintergrund dieser globalen Herausforderungen entwickeln zwölf Regionalprojekte im Teil A des Forschungsprogramms Beispiellösungen für die nachhaltige Nutzung der Landoberfläche und der natürlichen Ressourcen in global verteilten Untersuchungsgebieten. Die Regionalprojekte sind mit einem Gesamtvolumen über gut 50 Mio. Euro über fünf Jahre gefördert. Zusätzlich dient das interdisziplinär angelegte Projekt „GLUES“ (Global Assessment of Land Use Dynamics on Greenhouse Gas Emissions and Ecosystem Services) der wissenschaftlichen Begleitung, Koordination und Synthese der Regionalprojekte.

GLUES unterstützt die international fachübergreifende Zusammenarbeit in den Forschungsprojekten durch verschiedene Serviceleistungen im Bereich der Kommunikation, Koordination und Vernetzung sowie bei der Integration der Ergebnisse durch eine gemeinsame Datenplattform und durch die Entwicklung von einheitlichen Szenarien in der Landnutzung, dem Klimawandel und sozioökonomischen Veränderungen (Abb. 1). Durch die Unterstützung bei der Kommunikation von Forschungsergebnissen versucht GLUES die Sichtbarkeit der deutschen Landnutzungsforschung in der internationalen Forschungslandschaft und den Wissenstransfer zu den relevanten Stakeholdergruppen und in die Politikprozesse wie beispielsweise unter der CBD (Convention on Biological Diversity⁷) und der UNFCCC (United Nations Framework Convention on Climate Change⁸) zu verbessern. Die entwickelten Szenarien sind mit den Regionalprojekten abgestimmt und stellen konsistente Rahmenbedingungen für alle Projekte bereit. Die Synthese integriert die Ergebnisse der regionalen Forschungsprojekte in Bezug auf die Anforderungen



Abbildung 1: Schwerpunkttätigkeiten von GLUES (Quelle: UFZ, www.nachhaltiges-landmanagement.de)

verschiedener Nutzer- und Stakeholdergruppen (Eppink et al. 2012). Insgesamt arbeiten die Regionalprojekte dadurch weniger isoliert und die Ergebnisse sind besser vergleichbar. Die Übertragbarkeit der Ergebnisse auf die globale Ebene als auch die Bewertung grenzüberschreitender Auswirkungen wird verbessert (Werntze 2013).

Im Fokus dieses Beitrags steht die GLUES-Geodateninfrastruktur (GLUES-GDI), welche für den Austausch der wissenschaftlichen Daten aus Modellrechnungen und Simulationen im Rahmen von GLUES durch die Professur für Geoinformationssysteme an der TU Dresden aufgebaut wird. Damit wird auf technischer Ebene die Zusammenarbeit innerhalb von GLUES und zwischen den regionalen Projekten des BMBF-Programms unterstützt und eine Analyse und Synthese von globalen und regionalen Datensätzen ermöglicht. Weiterhin stellt die GLUES-GDI technische Komponenten für die Außendarstellung des Projekts bereit. Als Austauschplattform für die wissenschaftlichen Daten hat die GLUES-GDI dabei folgende Ziele:

1. Publikation und Austausch der Modelldaten, Analyseergebnisse und Basiszenarien der involvierten Forschungsgruppen.
2. Nahtlose Verknüpfung existierender Datenquellen durch die GDI, beispielsweise für die Berechnung wissenschaftlicher Modelle oder Vergleichsanalysen.
3. Interessenvertreter verschiedenster Bereiche werden durch die Such- und Analysewerkzeuge der GDI dabei unterstützt, Forschungsergebnisse aufzufinden und zu verstehen und diese für die eigenen Planungs- und Managementaktivitäten einzusetzen.

3 ENTWICKLUNGSSTAND DER GLUES-GDI

Nach dem offiziellen Projektstart in 2010 sind nach vierjähriger Projektlaufzeit die wichtigsten Projektergebnisse aus GLUES in der GDI über entsprechende Metadaten beschrieben und über Visualisierungs- und Downloaddienste für die Regionalprojekte und andere Nutzer verfügbar⁹. Der GLUES-Katalog, die zentrale Suchkomponente in der GDI, referenziert knapp 3.000 regist-

rierte Ergebnisdatensätze aus GLUES. Ein Großteil der Datensätze stammt aus Simulationsmodellen und Modellensembles, die eine Vielzahl von Einzeldatensätzen mit unterschiedlichen thematischen Inhalten liefern. Diese Datensätze sind jeweils einzeln im Katalog aufgeführt und durch eine hierarchische Gliederung als Gesamtdatensatz oder Datenserie verknüpft. Um die einzelnen Ergebnisdatensätze und deren Zusammenhänge zu verstehen, sind diese Verlinkungen zwischen den Datensätzen besonders wichtig.

Eine wesentliche Grundidee bei der Konzeption der GDI war die dezentrale Datenhaltung bei den wissenschaftlichen Institutionen, welche die Daten produzieren, und die Bereitstellung der Daten über dezentrale Dienste. Da das Aufsetzen und die Laufendhaltung der Dienste aber ohne entsprechende Fachkenntnisse kaum mit vertretbarem Aufwand machbar sind, wurden hier im Projekt Abstriche gemacht. Die GLUES Partner am Helmholtz-Zentrum für Umweltforschung (UFZ) und am Potsdam-Institut für Klimafolgenforschung (PIK) stellen ihre Daten über eigene Dienste in der GLUES-GDI bereit. Durch die aktuellen e-Science-Entwicklungen ist die persistente Bereitstellung der wissenschaftlichen Daten im Internet für solche Großforschungseinrichtungen interessant und teilweise sind hier auch eigene Entwicklungen und Projekte für die Zukunft geplant (z. B. die Earth System Knowledge Platform¹⁰ für die Helmholtz Gemeinschaft). Für kleinere Forschungsgruppen unter den GLUES-Partnern werden die Daten zentral durch die TU Dresden gehostet.

Von den zwölf Regionalprojekten sind derzeit Metadaten und Daten aus sechs Regionalprojekten in der GLUES-GDI auffindbar. Diese Regionalprojekte haben, im Unterschied zu den verbleibenden sechs, jeweils mindestens einen Partner mit GDI-Know-how im Projektkonsortium und auch entsprechende Arbeitspakete und Aufwände im Projektplan und der Ressourcenplanung vorgesehen. Die meisten dieser Projekte betreiben eigene Geoportale bzw. eigene Kataloge, die über die CSW-Schnittstelle mit dem GLUES-Katalog verbunden sind. Bei der Suche im GLUES-Katalog können diese Kataloge also als externe Ressourcen mit einbezogen werden. Im Vergleich zu GLUES wurden durch die Regionalprojekte bisher relativ wenige Daten

veröffentlicht. Dies lässt sich dadurch erklären, dass die Regionalprojekte teilweise mit einem deutlichem Zeitverzug nach GLUES starteten und deshalb die meisten Projektergebnisse erst zum jetzigen Zeitpunkt vorliegen.

Auch für die verbleibenden sechs Projekte, welche bisher keine Ergebnisse in der GDI veröffentlichen, sollen in Zukunft die wichtigsten Projektergebnisse in der GDI verfügbar sein. Dafür müssen durch bilaterale Kooperationen zwischen den Entwicklern der GLUES-GDI und den jeweiligen Datenproduzenten in den Regionalprojekten jeweils Lösungen gefunden werden.

4 BESONDERHEITEN DER GLUES-GDI

Die Veröffentlichung von wissenschaftlichen Umweltdaten und die Projektkonstellation von GLUES und den Regionalprojekten stellen einige besondere Anforderungen an die Inhalte und insbesondere auch an die Funktionalitäten, welche durch eine wissenschaftliche GDI bereitgestellt werden. Die konkreten Anforderungen an die GDI und ihre Komponenten wurden im Projekt in verschiedenen Nutzerworkshops erarbeitet.

Während der iterativen Implementierung wurde regelmäßig Nutzerfeedback eingeholt und die weiteren Entwicklungsschritte entsprechend angepasst. Auf einige dieser Anforderungen und Besonderheiten und den daraus resultierenden Entwicklungen in der GLUES GDI soll im Folgenden vertieft eingegangen werden.

4.1 ZEITREIHENVISUALISIERUNG

Ein Großteil der in der GLUES-GDI beschriebenen und bereitgestellten Daten sind Ergebnisse verschiedener Klima-, Landnutzungs- und Biodiversitätssimulationen. Die Zeitreihendaten sind, je nach Fokus des Forschungsprojekts, in verschiedenen zeitlichen Auflösungen und für bestimmte Zeitfenster verfügbar. Werkzeuge für eine dementsprechend flexible Visualisierung von Zeitreihendaten existieren nur vereinzelt in kommerziellen oder komplexen Expertensystemen.

Für die Visualisierung und Analyse der Zeitreihendaten in der GLUES-GDI wurde deshalb eine Webanwendung implementiert, die sowohl den Anforderungen der Nutzer als auch Erkenntnissen neuester Standards und Technologien genügt (Hen-

zen et al. 2014). Der leichtgewichtige Webclient ist intuitiv zu bedienen, wurde auf Basis nachhaltiger Open-Source-Technologien entwickelt und implementiert standardbasierte Schnittstellen. Partner aus dem Programm können den Client als Softwarepaket in ihrer eigenen GDI einsetzen und eventuell anpassen oder den Client parametrisiert mit ihren Daten und Diensten über die GLUES-GDI nutzen. Innerhalb von GLUES wurde eine lose Kopplung des Webclients an verschiedene Anwendungen der GLUES-GDI, wie beispielsweise als Visualisierungskomponente für den Metadatenkatalog benötigt.

Abbildung 2 zeigt die Visualisierung von Lufttemperaturen im GLUES-Webclient Time4Maps. Neben dem großen interaktiven Kartenfenster werden Informationen über den dargestellten Kartenlayer sowie eine Legende und Feature Informationen angezeigt. Unterhalb der Karte befinden sich Steuerungselemente für die verfügbare Zeitspanne, zum Beispiel zum Starten einer Animation sowie zum Drucken des Kartenausschnitts, der Zeit- und Feature-Information. Alle dargestellten Informationen werden durch Time4Maps von dem angebu-

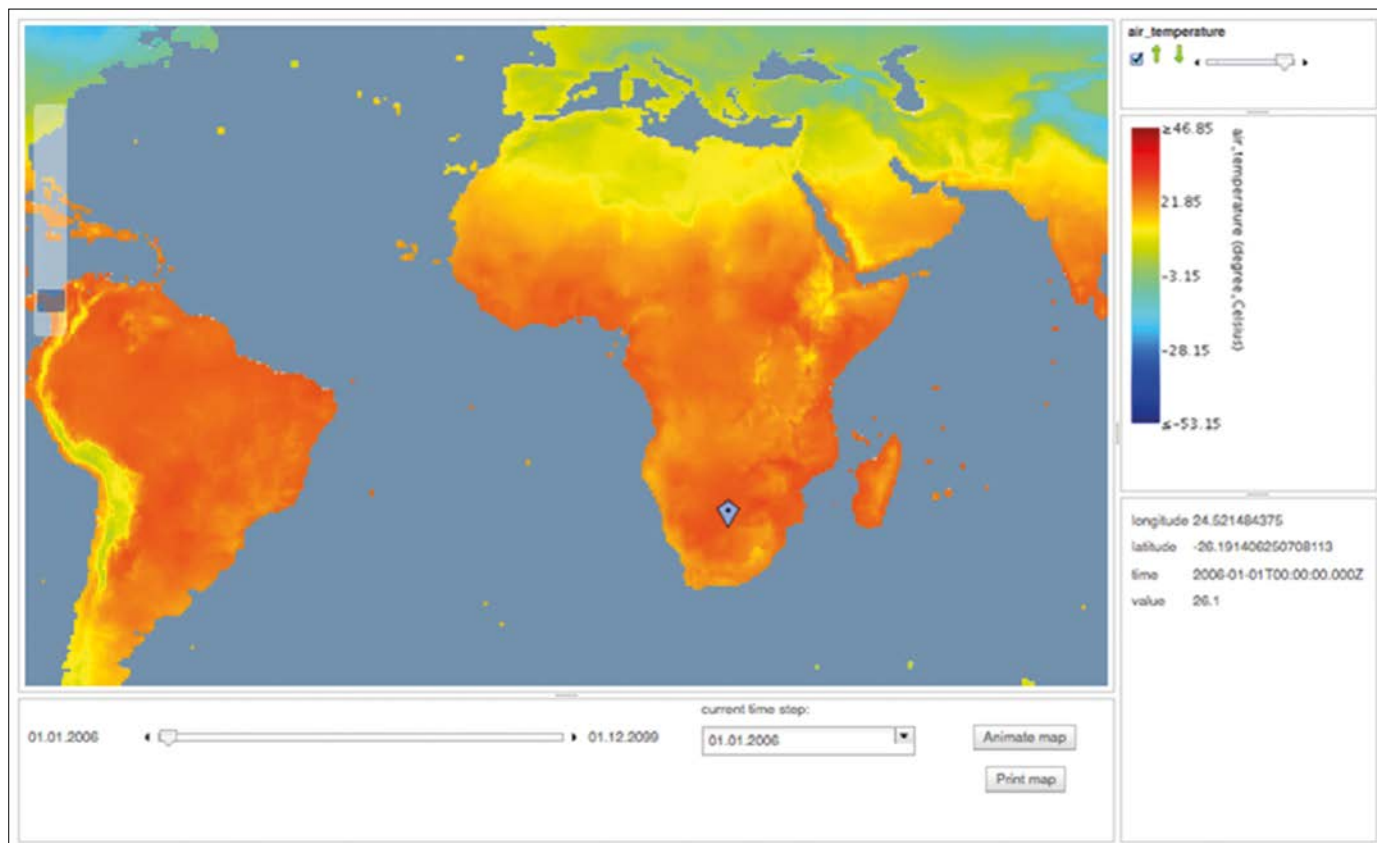


Abbildung 2: Visualisierung der Lufttemperatur für den Zeitraum 2006 bis 2009 in der Anwendung Time4Maps (Screenshot)

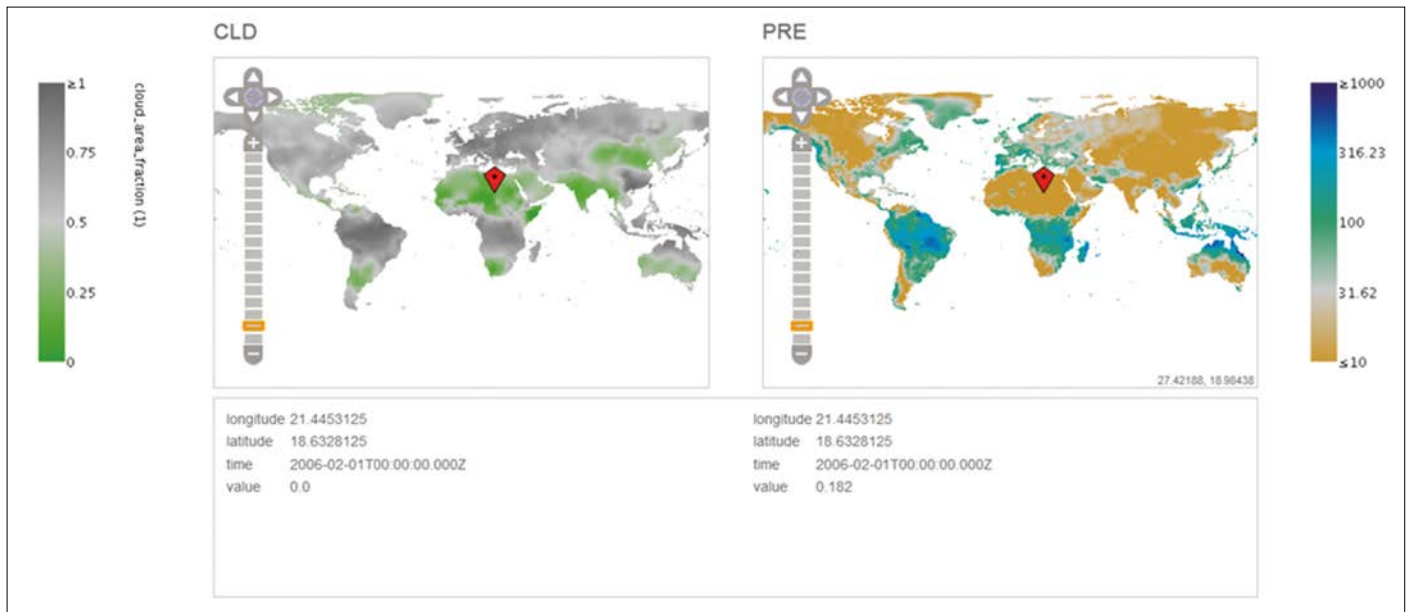


Abbildung 3: Visualisierung von Bewölkung und Niederschlag in der Vergleichsansicht (Screenshot)

denen Darstellungsdienst (WMS) mit den Projektergebnissen abgefragt und automatisch ausgewertet.

Im Rahmen des Programms entsteht eine Vielzahl von Simulationsergebnissen für verschiedenste thematische Schwerpunkte. Eine der Aufgaben von GLUES ist es, die wissenschaftliche Synthese der Projektergebnisse innerhalb des Forschungsprogramms zu koordinieren. Um diese Synthese technisch zu unterstützen, wurde der Webclient Time4Maps um eine Vergleichsansicht erweitert. Die Erweiterung ermöglicht es, zwei thematische Kartenansichten parallel zu visualisieren und für einen definierten Zeitpunkt und Ort Informationen abzurufen und visuell zu vergleichen (Abb. 3).

4.2 BESCHREIBUNG DER ENTSTEHUNGSGESCHICHTE DER DATEN

Die Simulationsmodelle werden von den Wissenschaftlern in der Regel mehrfach mit unterschiedlicher Parametrisierung und Eingangsdatsätzen ausgeführt. Aus diesem Grund sind Informationen zur Entstehungsgeschichte der Simulationsdaten ein besonders wichtiger Bestandteil der beschreibenden Metadaten und zentral für die Beschreibung von wissenschaftlichen Umweltdaten in Forschungsdateninfrastrukturen (He et al. 2014, Anselin et al. 2014). Die ISO Norm 19115 (ISO/TC211) beinhaltet hierfür das Lineage Element zur Beschreibung der Entstehungsprozesse und Eingangsdaten. In den meisten Anwendungen der ISO-Norm wird jedoch auf diese detaillierte

Modellierung verzichtet und die Historie ist bestenfalls als Freitext hinterlegt. Die Metadaten des GLUES-Katalogs beinhalten eine ausführliche Modellierung der Entstehungsgeschichte der Daten, sodass für jeden Datensatz die jeweiligen Simulationsmodelle, die Inputdatsätze und Referenzen zu den entsprechenden wissenschaftlichen Publikationen beschrieben werden können. Jeder Inputdatsatz hat auch einen eigenen Metadatsatz im Katalog, auf den dann in der Historienbeschreibung verlinkt wird. So wird nicht nur die Entstehungsgeschichte abgeleiteter Daten dokumentiert, sondern auch die Weiterverwendung eines Datensatzes in anderen Analyse- und Simulationsprozessen in den Metadaten erfasst.

In den gängigen Katalogoberflächen werden die Metadaten meist tabellarisch abgebildet. Diese sind jedoch wenig nutzerfreundlich und ungeeignet um einen schnellen Überblick über umfangreiche Herkunftsbeschreibungen zu bekommen. Aus diesem Grund wurde in GLUES der Webclient MetaViz für die interaktive Visualisierung der Entstehungsgeschichte und Weiterverwendung eines Datensatzes entwickelt (Abb. 4) (Henzen et al. 2013).

Die Abbildung veranschaulicht die Entstehung des PROMET-Datsatzes (blaue Box) aus dem PROMET-Modell (linke grüne Box) und die dabei verwendeten Inputdaten (linke weiße Boxen). Im rechten Bereich des Graphen wird (hier aus Platzgründen nur vereinfacht) gezeigt, dass es zwei Simulationen gibt, bei denen der PROMET-

Datsatz als Input verwendet wurde. In MetaViz kann der Nutzer diesen Bereich vergrößern und auf die anderen Datsätze fokussieren und sich so interaktiv durch den Entstehungsgraphen navigieren. Solche Visualisierungen veranschaulichen auch den Datenaustausch zwischen Forschungseinrichtungen und damit beispielsweise die Kooperationen zwischen GLUES und den Regionalprojekten. Neben dem Entstehungsgraphen werden in MetaViz für den fokussierten Datsatz weitere Metadaten wie beispielsweise zum Datenproduzenten und die referenzierten Publikationen angezeigt. Die Metadaten werden jeweils direkt aus der Datenbank des GLUES-Katalogs abgerufen.

Die leicht verständliche Visualisierung und das einfache Bedienkonzept von MetaViz bekam von den meisten Wissenschaftlern des Forschungsprogramms ein sehr positives Feedback. In einer weiteren Implementierung wurde gezeigt, wie sich die Visualisierung der Entstehungsgeschichte und die Interaktionskonzepte auch als eine von verschiedenen Sichten auf die Daten bei der Exploration und Suche von Geodaten einsetzen lassen (Henzen et al. 2014).

4.3 REGISTRIES UND VOKABULARE

Simulationsmodelle für den Klimawandel orientieren sich meist an definierten Szenarien, welche grundlegende Annahmen beispielsweise zum Bevölkerungswachstum und ökonomischen Entwicklungen festlegen



Abbildung 4: Visualisierung des Entstehungsgraphen des PROMET-Datensatzes in MetaViz11 (Screenshot)

und zusammenfassen (IPCC SRES 2000). So gibt es beispielsweise Szenarien, die eine starke ökonomische Entwicklung und eine entsprechende Politikausrichtung annehmen. Zum Vergleich stehen demgegenüber Szenarien, die eher auf nachhaltige Entwicklungen im Sozialen- und Umweltbereich fokussieren. Damit werden die Ausgangsparameter der Modelle sinnvoll eingeschränkt und eine gewisse Vereinheitlichung und Vergleichbarkeit der Ergebnisse verschiedener Simulationen erreicht. Um die Synthese der Ergebnisse von GLUES und der Regionalprojekte zu unterstützen, wurden auch für das „Nachhaltige Landmanagement“ Forschungsprogramm Szenarien definiert. Die Szenarienbeschreibungen sind von zentraler Bedeutung bei der Analyse und Interpretation der Ergebnisdaten. Aus diesem Grund werden die Szenariendefinitionen mit ihren Beschreibungen in der GLUES GDI über eine Registry angeboten und in den Metadaten der Datensätze referenziert.

Darüber hinaus verwenden die meisten Wissenschaftler verschiedene thematische Klassifikationen und Vokabulare für ihre Daten, die in der Regel nicht in den Metadaten beschrieben sind. Für die hier relevan-

ten Themenbereiche gibt es nur zum Teil vereinheitlichte oder standardisierte Klassifikationen, die aber häufig aufgrund spezieller Projektanforderungen nicht genutzt werden. Wenn vereinheitlichte Klassifikationen wie beispielsweise das LCCS für Landbedeckung¹² umfangreiche Hierarchien definieren, sollte in den Metadaten dokumentiert werden, auf welcher thematischen Granularität ein Datensatz die Klassifikation benutzt (Mäs et al. 2011). Nicht zuletzt wegen des hohen Aufwands bei der Metadatenerstellung ist dies in der Praxis selten der Fall.

In Zukunft sollen in der Registry der GLUES GDI auch relevante Klassifikationen und Vokabulare verfügbar sein. Ziel ist es nicht nur die verwendete Klassifikation zu referenzieren, sondern auch die jeweilige thematische Granularität zu dokumentieren. Für nicht standardisierte oder modifizierte Klassifikationen werden derzeit auch kollaborative Methoden zur Erstellung der Vokabulare untersucht, wie beispielsweise vorgeschlagen in Kalbasi et al. (2013).

4.4 GEOPROZESSIERUNG

Die Integration von Diensten zur Datenprozessierung in Forschungsinfrastrukturen er-

möglicht interaktive Analysen auf den vorhandenen Datenbeständen. Der Begriff „Geoprozessierung“ hat sich dabei als Oberbegriff für die Verarbeitung (erd-)raumbezogener Daten etabliert. Im Kontext von GDI wird auch Geoprozessierung verstärkt über Webdienste bereitgestellt. In der Literatur werden hier üblicherweise zwei Ansätze unterschieden (Müller et al. 2010). Lose gekoppelte Prozessierungsdienste haben große Ähnlichkeit zu klassischen Tool-sammlungen in GIS. Sie bieten eher allgemeine Schnittstellen und sind damit auf beliebige Datensätze anwendbar, sofern deren Format und Schema von der Prozessschnittstelle unterstützt wird. Eng gekoppelte Funktionen setzen dagegen auf vordefinierten Datenbeständen auf und stellen zugeschnittene Verarbeitungs- und Analysefunktionen bereit. In Geodateninfrastrukturen stellen die gängigen Standards für Geodatendienste (WFS/FES, WCS) einige primitive Verarbeitungsmethoden wie Subsetting oder Pufferung zur Verfügung. Komplexere Funktionen können über Web Processing Services (WPS) angeboten werden.

Die GLUES-GDI bietet unter anderem Zugriff auf ein breites Spektrum an Klimasimulationen. Diese Datensätze enthalten

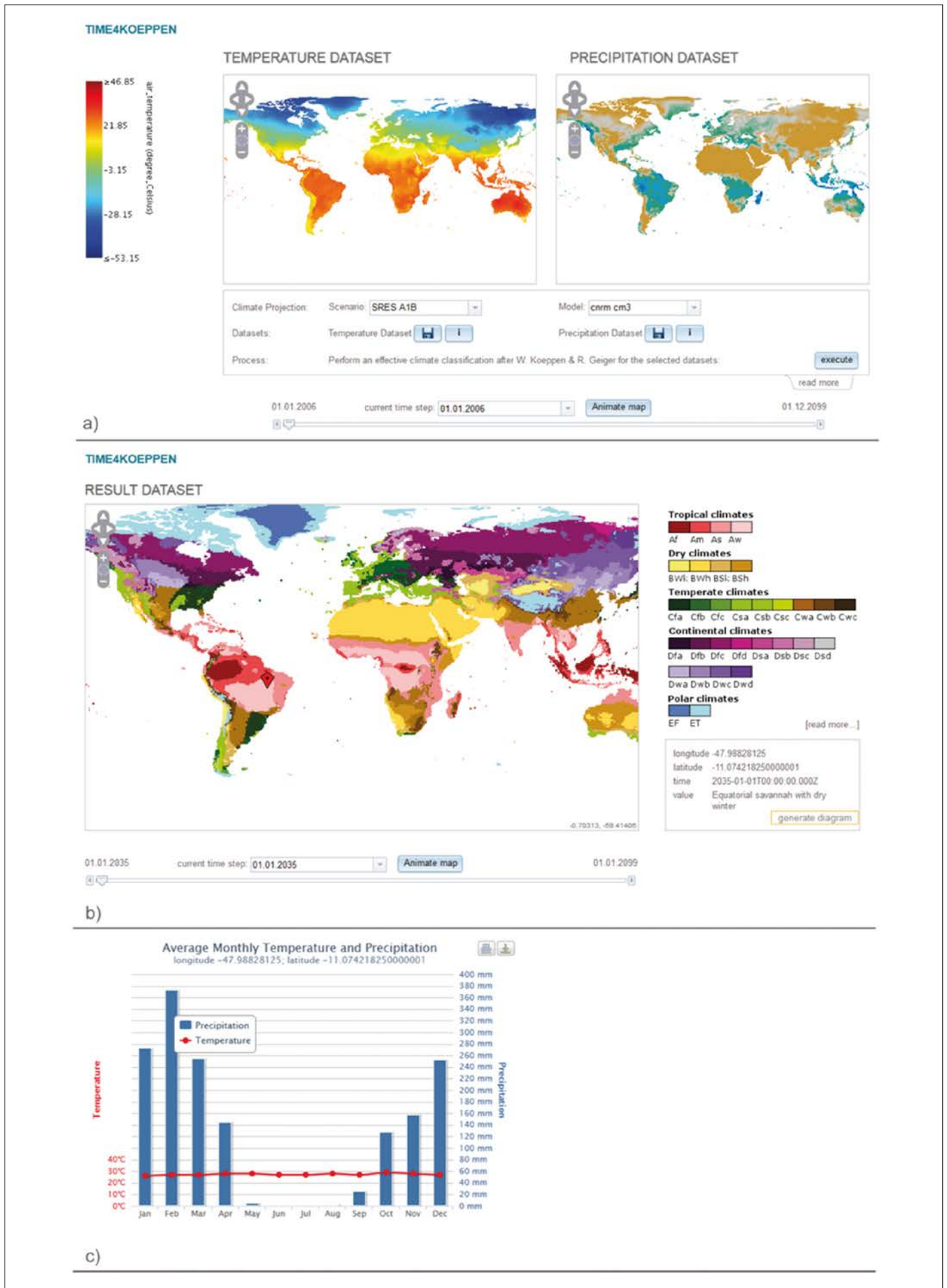


Abbildung 5: Interaktive Klassifikation von Klimadaten: a) Eingangsdaten, b) Klassifikationsergebnis, c) Erzeugung von Klimadiagrammen (Screenshot)

Projektionen zu Klimaelementen, wie Niederschlag, Temperatur und Wolkenbedeckung, die teilweise bis ins Jahr 2100 reichen. Die Simulationen wurden für ein Ensemble gängiger Klimamodelle und verschiedene Szenarien gerechnet (IPCC SRES 2000). Die Visualisierung einzelner Klimaelemente kann die Konsequenzen eines Klimawandels nur unzureichend darstellen. Klimaindizes, die auf historischen und projizierten Daten gerechnet werden, vermitteln ein plastischeres Bild des Weltklimas im Zeitverlauf.

Ein gängiger Index, ursprünglich für historische Daten entwickelt, ist die effektive Klimaklassifikation nach Koeppen & Geiger (Geiger & Pohl 1954, Koeppen 1936). Die Anwendbarkeit auf projizierten Daten wurde von Rubel & Kottek (2010) nachgewiesen. Die Klassifikation nutzt lange Zeitreihen (mindestens 30 aufeinander folgende Jahre) monatscharfer Daten und bewertet das Klima anhand von Temperatur- und Niederschlagsschwellwerten sowie der geographischen Lage. Im Ergebnis sind auf der Erde 31 Klimazonen zu erkennen, welche Rückschlüsse auf Wasserverfügbarkeit und Temperaturverhältnisse zulassen. Die Klassifikation erfolgt auf einem gleitenden Intervall von 30 Jahren, sodass über den kompletten Simulationszeitraum bis 2100 langfristige Trends sichtbar werden.

In der GLUES-GDI wird der Klassifikationsprozess auf drei verschiedene Arten bereitgestellt:

1. Als ad hoc ausführbarer Prozess auf einem WPS.
2. Als interaktiver Client auf einer vordefinierten Menge an Datensätzen.
3. Als wohldefinierter, austauschbarer Code, der von Experten heruntergeladen und in eigene Systeme und Anwendungen eingebettet werden kann.

Der Klassifikationsprozess selbst ist lediglich lose an die Daten gekoppelt und eignet sich grundsätzlich für beliebige Klimasimulationen. Die Eingangsdaten müssen als netCDF-Daten vorliegen, monatscharfe Temperatur- und Niederschlagswerte enthalten und entsprechend den netCDF-CF-Konventionen beschrieben sein. Der Prozess kann also auch aus fremden Daten (andere Klimasimulationen, beliebige räumliche Auflösung, beliebiger Simulations- oder Erfassungszeitraum) einen klassifizierten Datensatz erzeugen.

Der interaktive Client (Abbildungen 5a und 5b) ermöglicht die Anwendung des Klassifikationsprozesses auf schemakonforme Klimadaten, die sich bereits in der GDI befinden. Er enthält die Logik für einen kompletten Klassifikationsworkflow, angefangen bei der Anzeige der Datensatzmetadaten aus dem Katalog, über die Ansteuerung des WPS mit den korrekten Eingangsdaten bis hin zur Zwischenspeicherung und Anzeige der Klassifikationsergebnisse. Zusätzlich werden Funktionen zur Abfrage von Datenpunkten und zur Erzeugung von Klimadiagrammen bereitgestellt (Abbildung 5c).

In einer wissenschaftlichen Geodateninfrastruktur eignen sich derartige Prozessierungsdienste für einfachere Datentransformationen, um die Wissenschaftler bei der Datenvorverarbeitung zu unterstützen sowie für einfache Analysen und Visualisierungen, insbesondere als Entscheidungsunterstützung für Stakeholder.

5 RESÜMEE UND AUSBLICK

Die GLUES-GDI als wissenschaftliche Geodateninfrastruktur hat sich mittlerweile als Austauschplattform innerhalb des BMBF-Forschungsprogramms „Nachhaltiges Landmanagement“ etabliert. Daneben verbessert sie vor allem die Dokumentation und Verfügbarkeit der Projektergebnisse im Web, unterstützt die Wissenschaftler bei der Stakeholderarbeit und erhöht die Sichtbarkeit und Transparenz nach außen. Prinzipiell stehen viele Wissenschaftler der Publikation ihrer Daten noch sehr skeptisch gegenüber. Häufige Gründe dafür sind unter anderem Bedenken wegen möglicher Fehlinterpretationen der Daten durch fachfremde Nutzer, Laien oder in der Presse und der nicht unerhebliche zusätzliche Arbeitsaufwand bei der Datenpublikation. Für diese Nutzergruppen werden deshalb in Zukunft die wichtigsten Forschungsergebnisse aus den Projekten separat in sogenannten „Storylines“ aufbereitet erklärt und direkt auf der Portaloberfläche angeboten. Intuitiv bedienbare und nutzerfreundliche Werkzeuge und Visualisierungen helfen außerdem dabei, die Wissenschaftler selbst von der Datenpublikation zu überzeugen. Letztlich können aber auch schon in künftigen Forschungsausschreibungen Bedingungen an die Datenpublikation gestellt werden, wenn entsprechende Forschungsinfrastrukturen zur Verfügung stehen.

In der verbleibenden Projektzeit von GLUES sind vor allem Arbeiten geplant, welche die Nachhaltigkeit der Entwicklungen nach dem Ende der Förderung sicherstellen. Die zentralen Komponenten wie beispielsweise der Katalog sollen an das UFZ Leipzig migriert und dort nach Möglichkeit in die Infrastruktur der Großforschungseinrichtung integriert werden. Außerdem wird untersucht, welche Inhalte zu anderen internationalen Forschungsinfrastrukturen verlinkt oder übertragen werden können, um so auch international die Sichtbarkeit zu erhöhen. Der Quellcode der für die GLUES GDI entwickelten Werkzeuge wird Open Source unter der Apache Software License Version 2.0 veröffentlicht.¹³

DANKSAGUNG

Die in diesem Beitrag vorgestellten Ergebnisse und Erfahrungen entstammen zum Großteil aus dem GLUES-Projekt, welches im Rahmen des Forschungsprogramms „Nachhaltiges Landmanagement“ vom Bundesministerium für Bildung und Forschung gefördert wird.

¹ www.myexperiment.org

² www.iplantcollaborative.org

³ www.pangaea.de

⁴ <https://geoshareproject.org>

⁵ <http://globe.umbc.edu>

⁶ <http://modul-a.nachhaltiges-landmanagement.de/de/wissenschaftliche-begleitung-glues/>

⁷ www.cbd.int

⁸ <http://newsroom.unfccc.int>

⁹ www.sustainable-landmanagement.net/glues-geoportal

¹⁰ www.eskp.de

¹¹ Dieses Beispiel ist verfügbar unter: <http://geoportal.glues.geo.tu-dresden.de:8080/MetaViz/detail.jsp?id=glues:Imu:metadata:dataset:promet>

¹² http://www.glc.org/sof_1_en.jsp

Literatur

- Anselin, L.; Rey, S. J.; Li, W. (2014): Metadata and provenance for spatial analysis: the case of spatial weights. In: *International Journal of Geographical Information Science*. doi: 10.1080/13658816.2014.917313.
-
- Bernard, L.; Mäs, S.; Müller, M.; Henzen, C.; Brauner, J. (2013): Scientific Geodata Infrastructures: Challenges, Approaches and Directions. In: *International Journal of Digital Earth*. doi:10.1080/17538947.2013.781244.
-
- Ellis, E. C. (2012): The GLOBE Project: accelerating global synthesis of local studies in land change science. *Newsletter of the Global Land Project*, 5-6.
-
- Eppink, F.; Wertz, A.; Mäs, S.; Popp, A.; Seppelt, R. (2012): Land Management and Ecosystem Services: How Collaborative Research Programmes Can Support Better Policies. In: *GAI – Ecological Perspectives for Science and Society*, 21 (1), S. 55-63.
-
- Esfri (2008): European roadmap for research infrastructures. European Strategy Forum on Research Infrastructures. ftp://ftp.cordis.europa.eu/pub/esfri/docs/esfri_roadmap_update_2008.pdf, Zugriff 5/2014.
-
- Geiger, R.; Pohl, W. (1954): Eine neue Wandkarte der Klimagebiete der Erde nach W. Köppens Klassifikation. In: *Erdkunde*, 8, S. 58-61.
-
- Goble, C. A.; Bhagat, J.; Aleksejevs, S.; Cruickshank, D.; Michaelides, D.; Newman, D.; Borcum, M.; Bechhofer, S.; Roos, M.; Li, P.; De Roure, D. (2010): myExperiment: a repository and social network for the sharing of bioinformatics workflows. In: *Nucleic Acids Research*, 38, S. W677-W682.
-
- He, L.; Yue, P.; Di, L.; Zhang, M.; Hu, L. (2014): Adding Geospatial Data Provenance into SDI – A Service-Oriented Approach. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. doi: 10.1109/JSTARS.2014.2340737.
-
- Henzen, C.; Mäs, S.; Bernard, L. (2013): Provenance Information in Geodata Infrastructures. In: *Vandenbroucke, D.; Bucher, B.; Cromptvoets, J. (Eds.): Geographic Information Science at the Heart of Europe, 2013. Lecture Notes in Geoinformation and Cartography*. Springer-Verlag, Berlin/Heidelberg/New York, S. 133-151.
-
- Henzen, C.; Mäs, S.; Müller, M.; Bernard, L.; Tressel, H.; Haase, S. (2014): GeoMetaFacet 2.0 Interaktive nutzerfreundliche Visualisierung von geographischen Metadaten. *Geoinformatik 2014*, Hamburg.
-
- IPCC SRES (2000); Nakićenović, N.; Swart, R. (Ed.): *Special Report on Emissions Scenarios: A special report of Working Group III of the Intergovernmental Panel on Climate Change* (book). Cambridge University Press.
-
- ISO/TC211: ISO 19115-2 – geographic information – metadata – part 2: metadata for imagery and gridded data.
-
- Kalbasi, R.; Janowicz, K.; Reitsma, F.; Boerboom, L.; Alesheikh, A. (2013): Collaborative Ontology Development for the Geosciences. *Transactions in GIS*.
-
- Kindling, M.; Schirnbacher, P. (2013): „Die digitale Forschungswelt“ als Gegenstand der Forschung. In: *Information – Wissenschaft & Praxis*, 64 (2-3), S. 127-136. doi:10.1515/iwp-2013-0017.
-
- Koepfen, W. (1936): Das Geographische System der Klimate. In: *Koepfen, W. & Geiger, R. (Hrsg.): Handbuch der Klimatologie, C1 – C44*. Gebr. Borntraeger, Berlin.
-
- Mäs, S.; Müller, M.; Henzen, C.; Bernard, L. (2011): Linking the Outcomes of Scientific Research: Requirements from the Perspective of Geosciences. In: *Proceedings of the First International Workshop on Linked Science 2011 (IISC2011), CEUR Workshop Proceedings, Vol. 783*. Bonn, Germany, October 24, 2011. <http://www.ceur-ws.org/Vol-783>.
-
- Müller, M.; Bernard, L.; Brauner, J. (2010): Moving Code in Spatial Data Infrastructures – Web Service Based Deployment of Geoprocessing Algorithms. In: *Transactions in GIS*, 14 (S1), S. 101-118.
-
- NSF (2007): Cyberinfrastructure vision for 21st century discovery. US National Science Foundation Cyberinfrastructure Council. <http://www.nsf.gov/pubs/2007/nsf0728/>, Zugriff 5/2014.
-
- Rubel, F.; Kotteck, M. (2010): Observed and projected climate shifts 1901-2100 depicted by world maps of the Köppen-Geiger climate classification. In: *Meteorologische Zeitschrift*, 19, S. 135-141.
-
- Wertz, A.; Eppink, F. V.; Mäs, S.; Popp, A.; Seppelt, R. (2013): Landmanagement und Ökosystemdienstleistungen: Wie in Verbundforschungs-Programmen integrative Landnutzungsstrategien erarbeitet werden können. In: *Berichte Geographie und Landeskunde*, 87 (3), S. 277-293.
-