



Projektbericht B2

Umfeldanalyse zum Klimadaten-Management

Projektkürzel	Wibaklidama	
Typ und Nr.	B2, Umfeldanalyse	
Arbeitspaket(e)	AP 3	
Datum	27.4.2010	
Dateiname	wibaklidama-B2-V02-02.pdf	Version 02.02
Status	Öffentlich	
Autoren	Lars Müller	
Bearbeiter	Silke Grossmann	
Abstract	Auf Basis der Auswertung einer Internet- und Literaturrecherche wird das Umfeld des Klimaforschungsdatenmanagements analysiert und das Projekt Wibaklidama im Kontext vorausgehender wie aktueller Entwicklungen und Projekte verortet.	
Stichwörter	Forschungsdateninfrastruktur, Klimadaten-Management	



LAND
BRANDENBURG

Ministerium für Wissenschaft,
Forschung und Kultur



Inhaltsverzeichnis

1	Forschungsdaten im Kontext der Wissenschaftsentwicklung	3
1.1	Wissenschaftspraxis	3
1.2	Wissenschaftssteuerung und Infrastrukturbildung	4
2	Datenprojekte im weiteren Umfeld	5
2.3	„Meilensteine“ und „Leuchttürme“ des Forschungsdatenmanagements	5
2.3.1	Trends- und Tendenzen der Entwicklungen	5
2.3.2	Auswahl relevanter Pilot- und Modellprojekte.....	6
2.4	Entwicklung einer Forschungsdaten-Infrastruktur	8
2.4.3	Praktische Ansätze zur Schaffung einer Forschungsdateninfrastruktur	8
2.4.4	Auswahl relevanter Dateninfrastruktur-Projekte	10
2.5	Projekte der Klimaplattform-Teilnehmer.....	11
2.5.5	Datenprojekte der Klimaplattform-Partner	12
3	Zusammenfassung und Ergebnisse	13



1 Forschungsdaten im Kontext der Wissenschaftsentwicklung

Seit ca. 15 Jahren gibt es intensive Bemühungen, das Thema „Forschungsdaten“ systematisch zu entwickeln.

Bestehende Ansätze werden von zwei Seiten vorangetrieben: die strukturellen Notwendigkeiten datenintensiver Wissenschaftszweige und die Qualitätsanforderungen an Wissenschaften in digitalen Umgebungen im Allgemeinen.

1.1 Wissenschaftspraxis

Die Notwendigkeit der systematischen „Data Curation“ offenbarte sich zuerst in den datenintensiven Wissenschaften. Dazu gehört insbesondere auch die Klimaforschung. Mit dem exponentiellen Anwachsen der zu speichernden und zu verarbeitenden Datenmengen seit den 1980er Jahren wurde es erforderlich, Informationssysteme zu entwickeln, die diese Daten langfristig speichern und für weitere Nutzung zugänglich halten.¹ Insbesondere für Messdaten, die nur einmalig erhoben werden können, besteht langfristiger Bedarf wiederholt darauf zuzugreifen.²

Das Kernproblem besteht dabei weniger in der technischen Speicherung der Daten, als in der langfristigen Interpretierbarkeit der Datensets. „The logical availability of the data is rapidly degrading due to a lack of structural knowledge about the data and a lack of information describing the data (metadata).“³ Deshalb stammen zentrale Entwicklungen zu Datenmanagement und -publikation aus dem Umfeld der datenintensiven Wissenschaften, weil sie den dynamischen Lebenszyklus von Forschungsdaten handhabbar machen müssen. Obwohl in den vergangenen zehn Jahren große Fortschritte in Entwicklung und Anwendung bspw. von digitaler Langzeitarchivierung

¹ Vgl.: Michael Diepenbroek/M. Futterer/Hannes Grobe/H. Miller/Manfred Reinke/R. Sieger: PANGAEA information system for glaciological data management, in: Annals of Glaciology 27 (1998), S. 655–660, <http://hdl.handle.net/10013/epic.11222.d001>, [S. 2.]

² Vgl. Klump, Jens (2009): Digitale Forschungsdaten. Online verfügbar unter http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_292.pdf, S. 111.

³ Michael Diepenbroek/M. Futterer/Hannes Grobe/H. Miller/Manfred Reinke/R. Sieger: PANGAEA information system for glaciological data management, in: Annals of Glaciology 27 (1998), S. 655–660, <http://hdl.handle.net/10013/epic.11222.d001>, [S. 2.]



und Grid-Technologien erzielt wurden, hat sich am Grundproblem wenig geändert: die Anpassung der Technik an die sozialen Strukturen der Forscher-Communities und deren inhaltliche Anforderungen an die technischen Systeme.⁴

1.2 Wissenschaftssteuerung und Infrastrukturbildung

Mit dem Ziel, wissenschaftliches Fehlverhalten zu verhindern bzw. aufdecken zu können, hat die DFG im Jahr 1998 „Vorschläge zur Sicherung guter wissenschaftlicher Praxis“ verfasst.⁵ Darin wird u.a. die Aufbewahrung von Primärdaten für zehn Jahre empfohlen, damit die publizierten wissenschaftliche Ergebnisse nachvollzogen und überprüft werden können. Die DFG-Empfehlungen nennen bei der Speichertechnik noch Diskette oder CD-ROM.⁶ Im Januar 2009 hat der zuständige DFG-Ausschuss ausdrückliche „Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten“ veröffentlicht. Darin heißt es:

„In jedem Fall sollten die erhobenen Daten nach Abschluss der Forschungen öffentlich zugänglich und frei verfügbar sein. Dieses ist die wesentliche Voraussetzung dafür, dass Daten im Rahmen neuer Fragestellungen wieder genutzt werden können sowie dafür, dass im Falle von Zweifeln an der Publikation die Daten für die Überprüfung der publizierten Ergebnisse herangezogen werden können.“⁷

Die Forderung der DFG wird unabhängig von Datenart oder Wissenschaftsdisziplin erhoben und zielt auf die Schaffung nachhaltiger Strukturen, die aus den Fachdisziplinen heraus entwickelt werden sollen. Dabei sollen bereits praktizierte und erprobte Verfahren berücksichtigt werden.⁸

⁴ Vgl. Klump, Jens (2009): Digitale Forschungsdaten. Online verfügbar unter http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_292.pdf. S. 109.

⁵ DFG: Vorschläge zur Sicherung guter wissenschaftlicher Praxis. Empfehlungen der Kommission „Selbstkontrolle in der Wissenschaft“. Denkschrift. Weinheim 1998, S. 2., vgl. auch: Jens Klump: Digitale Forschungsdaten. http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_292.pdf S. 104.

⁶ DFG: Vorschläge zur Sicherung guter wissenschaftlicher Praxis. Empfehlungen der Kommission „Selbstkontrolle in der Wissenschaft“. Denkschrift. Weinheim 1998, S. 13.

⁷ DFG: "Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten." Bonn 2009. S. 2.

⁸ Vgl.: Ebenda, S. 2.



Um die Empfehlung in elektronischen Arbeitsumgebungen praktisch umzusetzen, ist faktisch die Entwicklung von Informationssystemen notwendig, die Forschungsdaten zuverlässig speichern, referenzieren und zugänglich machen.

Zurzeit befindet sich die Entwicklung in einer Phase, in der bestehende Systeme zu einer Infrastruktur ausgebaut werden. Wissenschaftspolitische Organisationen haben die Forderung nach freiem Zugang zu Forschungsdaten aus öffentlich finanzierten Projekten erhoben und formuliert.⁹ Die entwickelte Praxis und die ausgereiften Strukturen datenintensiver Wissenschaften sollen nun für weitere Wissenschaftsfelder nutzbar gemacht werden.¹⁰

2 Datenprojekte im weiteren Umfeld

2.3 „Meilensteine“ und „Leuchttürme“ des Forschungsdatenmanagements

2.3.1 Trends- und Tendenzen der Entwicklungen

Datenpublikationssysteme werden seit 10 bis 15 Jahren systematisch ausgebaut. Es sind auf Projekt- oder Institutionenbasis zahlreiche Einzellösungen entstanden. Insofern kann man von einem großen Erfahrungsschatz in Bezug auf Forschungsdatenverwaltung und Publikation ausgehen, der aber breit gestreut ist.

Im vergangenen Jahrzehnt hat eine Fokus-Verschiebung stattgefunden. Zunächst lag er auf digitaler Speicherung/Sicherung, stark ausdifferenziert nach Disziplinen und Teildisziplinen mit der zentralen Aufgabe, Projektergebnisse online zu dokumentieren. Es hat nun eine Schwerpunktverlagerung stattgefunden hin zu Interoperabilität, Integration von heterogenem Material unter gemeinsame Schemata und insbesondere auch die Entwicklung von Userinterfaces zum unmittelbaren Onlinezugriff auf Metadaten und Primärdaten. Derzeit aktuell sind Fragen großangelegter Vernetzung und des Infrastrukturausbaus.

⁹ Vgl.: Arbeitsgruppe Elektronisches Publizieren: Positionspapier Forschungsdaten.
<http://edoc.hu-berlin.de/series/dini-schriften/2009-10/PDF/10.pdf>, S. 6.

¹⁰ Ebenda, S. 6.



Den Modellprojekten gemeinsam ist neben den technischen Aspekten ein starker Fokus auf Metadaten und Standards. Ohne standardisierte interoperable Metadatenstrukturen können Forschungsdaten nicht über das Netz geteilt und nachgenutzt werden. Die eindeutige Referenzierbarkeit über das System der Digital Object Identifier (DOI)¹¹ scheint sich in den Datenrepositorien als faktischer Standard durchzusetzen. Praktische Anwendung findet die unmittelbare Verlinkung über DOI von Publikationen zu Datensets beispielsweise im klassischen institutionellen Publikationsserver des Alfred Wegner Instituts, ePIC (electronic Publication Information Center).¹² Markante Icons zeigen hier im Kontext der Bibliografischen Angaben die Existenz publizierter Daten an und verlinken auf diese. Eindeutig vorherrschendes Geschäftsmodell ist „Open Access“ zu den publizierten Daten.¹³

Neuerdings werden Datensets auch im Peer-Review-Verfahren veröffentlicht. Damit bereichert ein neuer Publikationstyp die wissenschaftliche Informationslandschaft. In Analogie zu herkömmlichen akademischen Zeitschriften werden Datensets publiziert und erläutert. Den publizierenden Wissenschaftlern verschaffen sie Renommee und stellen zugleich eine zuverlässige und leicht zugängliche Datenquelle für andere dar.

2.3.2 Auswahl relevanter Pilot- und Modellprojekte

Pangaea

<http://www.pangaea.de/>

Bereits 1998 wurde im Umfeld der deutschen Klimaforschung der Bedarf nach einem Informationssystem formuliert: "The goal was the implementation of an information system which is able to give an overview of the sampling material available, to store paleoclimatic data of any kind in a consistent form, to provide all metadata, necessary for the understanding of the data, and to make these data easy accessible to the scientific community."¹⁴ Somit wurden die wesentlichen Anforderungen für Datenrepositorien schon für den Prototyp PANGAEA genannt.¹⁵

¹¹ <http://www.doi.org/>

¹² <http://epic.awi.de/epic/>, letzter Zugriff: 5.1.2010

¹³ Inwieweit es sich dabei um wirkliche Open-Data-Konzepte handelt – im Sinne nicht nur von Gratis-Zugang, sondern von auch freier Verwendung – wurde nicht im Detail untersucht.

¹⁴ Diepenbroek et al. 1998 (wie Anm. 1), [S. 3.]

¹⁵ Vgl.: Ebenda [S. 4.] 1) Daten müssen in Standardformaten gespeichert werden 2) Retrieval- und Visualisierungswerkzeuge müssen mit den Datenrepositorien eng verknüpft sein 3) Datenrepositi-



World Data Center for Climate (WDCC) / CERA

<http://cera-www.dkrz.de/CERA/>

Im WDCC werden Klimadaten gespeichert und auf verschiedenen Wegen für die Forschung zur Verfügung gestellt: Vorortnutzung, CD-ROM, Internet. Das zentrale Konzept bedingt, dass keine Rohdaten gespeichert werden. Der Schwerpunkt liegt auf Klimamodellen und vergleichbaren Daten.¹⁶ Bemerkenswert an diesem Ansatz ist die internationale Koordination und Kooperation bei der zentralen Datenhaltung und die Entwicklung eines Metadatenmodells für die Speicherung heterogener Datensets der Klimaforschung.¹⁷ Das WDCC ist Teil des World Data Center Systems. Dieses in die 1950er Jahre zurückreichende Netzwerk koordiniert und sichert die Langzeitverfügbarkeit von Daten für die internationale Forschung.¹⁸

GRADE (Geospatial Repository for Academic Deposit and Extraction) und ShareGeo

<http://edina.ac.uk/projects/grade/>

Das Ziel vom 2007 abgeschlossenen britischen Projekt GRADE war die Schaffung einer nachhaltigen Infrastruktur zur Weiterverwendung von Geodaten. Besonderer Fokus lag im Projekt auf dem Ineinandergreifen von kulturellen und technischen Lösungen.¹⁹ Die Projekterkenntnisse sind in den Aufbau von der Nachweisdatenbank ShareGeo geflossen, ein „klassisches“ Web-2.0-Tool zum Teilen von Geodaten.

PhysNet

<http://physnet.uni-oldenburg.de/PhysNet/>

PhysNet ist ein Informationssystem über Einrichtungen und Publikationen der Physik. Primärdaten werden hier nicht berücksichtigt. Für Wibaklidama ist PhysNet jedoch aufgrund der Datenaufbereitung relevant. Wegweisend ist das RDF-basierte Metada-

torien sollen als Publikations- und Nachweissysteme fungieren 4) Datensets sollen mit neuen Publikationen verknüpft werden 5) Datenqualität muss prüfbar sein, Datenplausibilität muss automatisch geprüft werden 6) Daten sollen in verschiedene Stadien gespeichert werden können (Primär-, Sekundär-, Tertiärdaten) 7) Urheberrechte der Datenbesitzer müssen geklärt und berücksichtigt werden

¹⁶ <http://www.mad.zmaw.de/wdc-for-climate/>, letzter Zugriff: 6.1.2010

¹⁷ Toussaint, Lautenschlager, Reinke: "CERA-2: Ein Raumbezogenes Daten- und Metadatenmodell" in: Kramer & Hosenfeld (Hrsg.): Heterogene aktive Umweltdatenbanken, Workshop Vilm 1998, Metropolis-Verlag Marburg 1999.

¹⁸ http://www.world-data-system.org/front_content.php?idcat=356 ,
http://www.icsu.org/1_icsuinscience/PDF/Flyer_WDS_web.pdf

¹⁹ <http://www.edina.ac.uk/projects/grade/index.html>, letzter Zugriff: 6.1.2010



tenschema, über das Themen, Dokumente, Institutionen und geographische Verortung miteinander in Beziehungen gesetzt werden.

Earth Systems Science Data (ESSD)

<http://earth-system-science-data.net/>

Seit 2009 erscheint die Zeitschrift "Earth Systems Science Data (ESSD)". Ihr Anliegen ist es, "to publish data according to the conventional fashion of publishing articles, applying the established principles of quality assessment through peer-review to datasets."²⁰ Die Aufsätze sollen Reviewer und Leser in die Lage versetzen, Datensets ohne Aufwand zu verstehen und zu nutzen. Neben der Prüfung auf Plausibilität der Daten soll der Peer-Review-Prozess u.a. sicherstellen, dass auf die Daten Open Access zugegriffen werden kann, sie mit standardisierten Metadaten beschrieben sind und sie sich auf zuverlässigen Archivservern befinden.²¹

Mit diesem Verfahren erproben die Herausgeber eine Zwischenform der Datenpublikation. Es besteht zwar der Anspruch in der Zeitschrift selbst möglichst viele Daten zu publizieren, faktisch handelt es sich jedoch um „klassische Aufsätze“ (mit den dazugehörigen umfangreichen Zitierungen von Sekundärliteratur), die jedoch den besonderen Gegenstand „Datensets“ haben. Die online publizierten teils sehr großen vollständigen Datenbestände werden ihrerseits über einen DOI referenziert.

2.4 Entwicklung einer Forschungsdaten-Infrastruktur

2.4.3 Praktische Ansätze zur Schaffung einer Forschungsdateninfrastruktur

In den bereits erwähnten „Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten“ wird die dafür nötige „Schaffung der erforderlichen organisatorischen Voraussetzungen und Strukturen“ gefordert.²² Die „Emp-

²⁰ http://www.earth-system-science-data.net/general_information/about_this_journal.html, letzter Zugriff: 4.1.2010

²¹ Vgl. Ebenda.

²² DFG (2009): Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten. Online verfügbar unter



fehlungen“ betrachten die Wissenschaftsdisziplinen als tragende Säule der Datenspeicherung und –bereitstellung. Wie gezeigt, haben sich frühe Datenprojekte aus bestimmtem Bedarf von Fachdisziplinen entwickelt. Dennoch gibt es darüber hinausgehende Ansätze für fachübergreifende Infrastrukturen.

Die DINI-Arbeitsgruppe „Elektronisches Publizieren“ hat im Frühjahr 2009 Positionen zur Schaffung von Strukturen des Forschungsdatenmanagements formuliert und damit das konzeptionelle Feld abgesteckt. Fünf Themenfelder, die in diesem Prozess von zentraler Bedeutung sind, wurden identifiziert: 1) Disziplinarität, 2) Organisation, 3) Zugang, 4) Qualität, 5) Technik.²³ Auch hier steht die Ausrichtung an den Fachdisziplinen an erster Stelle, jedoch findet eine zunehmende Vernetzung von Einzelprojekten auch fachübergreifend statt. Praktische Umsetzungsprobleme, deren Lösung derzeit entwickelt und erprobt werden, beziehen sich im Wesentlichen auf folgende Aspekte²⁴:

- Zitierbarkeit von Forschungsdaten
- Vernetzung von Publikationen mit den Datenquellen
- automatische Erzeugung von Metadaten
- Akzeptanz von Datenpublikationen in den Fachcommunities
- Adaption von Publikationsserver-Lösungen für Datenpublikation
- Ausbildung in Betrieb und Nutzung von Forschungsdatenrepositorien
- Standards und Formate
- Vernetzung von Repositorien

Derzeit werden Standards und Qualitätskriterien entwickelt, die fachübergreifende Gültigkeit besitzen. Beispielsweise existieren weitgehend ausgereifte Konzepte zur Zitierbarkeit von Forschungsdaten und wie beschrieben gibt es vielversprechende Ansätze, die Publikation von Forschungsdaten zu formalisieren und aufzuwerten. Durch solche Entwicklungen werden die Grundsteine für eine umfassende fachübergreifende Infrastruktur gelegt. Dabei ist die Entwicklung von Technik und tragfähigen

http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/veroeffentlichungen/dokumentationen/download/ua_inf_empfehlungen_200901.pdf. S. 2.

²³ Arbeitsgruppe Elektronisches Publizieren 2009 (wie Anm. 9), S. 7f.

²⁴ Vgl. auch: Gibbs, Harry (2007): DISC-UK DataShare:DISC-UK. State-of-the-Art Review. Online verfügbar unter <http://www.disc-uk.org/docs/state-of-the-art-review.pdf>, S. 5f.



Konzepten gegenüber Verbreitung und Akzeptanz der Lösungen oft nur die kleinere Aufgabe.

2.4.4 Auswahl relevanter Dateninfrastruktur-Projekte

PARSE.Insight

<http://www.parse-insight.eu/>

Auf EU-Ebene ist das Projekt PARSE.Insight zu beachten. Dort liegt der Fokus auf Schaffung einer „Science Data Infrastructure“ hinsichtlich Technik, Organisation und Finanzierung.²⁵ Hier wird mit dem Bild der Inselösungen gearbeitet. Diese isolierte Praxis soll durch gezielte Maßnahmen zu einer nachhaltigen Infrastruktur verwoben werden.

STD-DOI

<http://www.std-doi.de/>

Um die Nachnutzung von Forschungsdaten zu ermöglichen, braucht es auch bestimmte Kulturtechniken im wissenschaftlichen Umgang mit diesen Daten. In diesem Sinne wurde im Projekt STD-DOI das Ziel verfolgt, „to make primary scientific data citeable as publications.“²⁶ In diesem Kontext wurde DataCite entwickelt.

DataCite

<http://www.datacite.org/>

Auf der Basis eines DFG-Projekts wurde die Technische Informationsbibliothek und Universitätsbibliothek Hannover (TIB/UB) die weltweit erste DOI-Registrierungsagentur für wissenschaftliche Daten. Die Registrierung erfolgt in Kooperation mit "Publikationsagenten". Diese bleiben für Qualitätssicherung und Speicherung der Daten in ihren lokalen Systemen verantwortlich. Die Registrierungsagentur vergibt die DOIs und speichert die Metadaten.²⁷

²⁵ PARSE.Insight (2009): Road Map. Deliverable D2.1. Herausgegeben von PARSE.Insight. Online verfügbar unter http://www.parse-insight.eu/downloads/PARSE-Insight_D2-1_DraftRoadmap_v1-1_final.pdf, zuletzt geprüft am 7.12.2009. S. 5.

²⁶ STD-DOI Publication and Citation of Scientific Primary Data. http://www.icdp-online.org/contenido/std-doi/front_content.php

²⁷ Jan Brase/Jens Klump: Zitierfähige Datensätze: Primärdaten-Management durch DOI. Wissenschaftskommunikation der Zukunft, in: Schriften des FZ Jülich - Reihe Bibliothek (2007)



Kooperative Langzeitarchivierung für Wissenschaftsstandorte (KolaWiss)

<http://kolawiss.uni-goettingen.de/>

Im Projekt KolaWiss werden Konzepte und Geschäftsmodelle zur Langzeitarchivierung insbesondere auch von Forschungsdaten entwickelt. Das Göttinger Projekt verfolgt dabei einen Ansatz, der auf lokaler Ebene interdisziplinär ausgerichtet ist. Im Projekt wurden besonders auch technische Aspekte und Fragen von Standardformaten und -daten behandelt.²⁸

Digital Repository Infrastructure Vision for European Research (DRIVER)

<http://www.driver-community.eu/>

Ziel des EU-Projekts „DRIVER“ war die Vernetzung von bestehenden Open-Access-Repositorien. Im Kontext von „Enhanced Publications“ sind Forschungsdaten ausdrücklich auch als Gegenstand des Projekts genannt.²⁹ Neben organisatorischer Vernetzung wurde eine Suchmaschine erstellt, die eine zentrale und Repositorienübergreifende Suche ermöglicht.

The Data Seal of Approval (DSA)

<http://www.datasealofapproval.org/>

Von einem Zusammenschluss führender Einrichtungen des Datenmanagements wurde das „Data Seal of Approval“ entwickelt. Das Qualitätssiegel kann an Datenrepositorien vergeben werden, die einen Katalog von sechzehn Anforderungen erfüllen, anhand derer der Langzeiterhalt und -zugriff auf Forschungsdaten gesichert ist.³⁰

2.5 Projekte der Klimaplattform-Teilnehmer

Datenprojekte in dem hier untersuchten Sinne wurden nur in den reinen Forschungseinrichtungen (wie Fraunhofer- oder Leibniz-Institute) identifiziert. Ein hohes Maß an internationaler Vernetzung (im Sinne von Dateninfrastrukturen) lässt sich dort ebenfalls beobachten. Klar zeigt sich ein deutlicher Unterschied zwischen diesen großen

²⁸ <http://kolawiss.uni-goettingen.de/?q=de/node/5>, Zugriff: 17.1.2010

²⁹ <http://www.driver-repository.eu/Enhanced-Publications.html>, letzter Zugriff: 7.1.2010

³⁰ Vgl.: http://www.datasealofapproval.org/sites/default/files/DSA_informationfolder_web.pdf



und fachlich hoch spezialisierten Einrichtungen und den Hochschulen. Bei letzteren stehen oft sehr spezifische, lokal ausgerichtete Projekte im Vordergrund. Daten als Gegenstand oder Publikationsform eigener Arbeit wurden dagegen nicht entdeckt, obwohl vermutlich auch dort viel Datenmaterial anfällt.

Eine Zwischenform der Datenpublikation stellt die Praxis dar, auf Internetseiten von Forschungsprojekten zugehörige Forschungsdaten zu publizieren. Diese sind damit zugänglich, aber nicht im Sinne von zuverlässigen Publikationsservern archiviert und auch nicht systematisch mit standardisierten Metadaten erschlossen.

Groß angelegte „Data Center“ werden von einigen Klimaplattform-Partnern selbst betrieben oder unter ihrer Beteiligung unterhalten. Diese sind auch bei den Berlin-Brandenburger Einrichtungen disziplinär bzw. auch nach Teildisziplinen differenziert ausgerichtet. Zugleich sind sie hoch entwickelt und ermöglichen Erfassung nach Metadatenstandards, Online-Recherche, Visualisierung und Download von Datensets.

Diese Beobachtungen decken sich auch mit den Antworten aus der Data-Practices-Erhebung. Es verläuft eine klare Trennlinie in der Datenpraxis zwischen den Forschungsinstituten und den Hochschulen.

2.5.5 Datenprojekte der Klimaplattform-Partner

Da an dieser Stelle die Institutionen im Fokus stehen, werden hier nicht einzelne Projekte aufgelistet, sondern es werden Projekttypen und Datenservices genannt, die unter Regie oder Beteiligung der Klimaplattform-Partner stehen. Der Themenbereich ist denkbar breit gefächert und reicht von Satellitenmissionen ohne unmittelbaren Regionalbezug zu Messdaten eines einzelnen Brandenburger Sees.

Messinstrumente

Unter Beteiligung des AIP werden Großgeräte, wie bspw. Satelliten oder Teleskope betrieben. Die damit erfassten Daten werden über Internetportale zugänglich gemacht und präsentiert, allerdings sind die Daten nicht prinzipiell öffentlich abrufbar.



Datendokumentation zu Einzelprojekten

Daten konkreter Forschungsprojekte werden online und live dokumentiert, bspw. aktuelle Daten vom Großen Müggelsee.³¹

Repositorien-Projekte und Datenzentrum

Zu speziellen Teildisziplinen werden Datenrepositorien unterhalten. Beispielsweise Scientific Drilling Database³² oder das „Information System and Data Center for geoscientific data“.

Dateninfrastrukturen

Einige Einrichtungen weisen institutionalisierte Infrastrukturen für Forschungsdaten auf. So unterhält das ZALF die „Abteilung Landschaftsinformationssysteme“ als eigenständigen Forschungsbereich. Sie hat die Aufgabe, „adäquate Bereitstellung qualitativ hochwertiger, zentral verwalteter Daten und Informationen für die komplexe, inter- wie transdisziplinäre Systemforschung und Modellierung, sowie einen effektiven Wissenstransfer zu Politik, Forschung und Öffentlichkeit“³³ zu sichern. Vergleichbare Infrastruktureinrichtungen werden auch vom PIK und GFZ unterhalten.

3 Zusammenfassung und Ergebnisse

Die treibenden Kräfte für die Entwicklung des Forschungsdatenmanagements sind fachliche Notwendigkeiten nach einer Langzeitverfügbarkeit von Daten und wissenschaftspolitische Weichenstellungen wie z.B. DFG-Empfehlungen zur guten wissenschaftlichen Praxis.

Es existieren hochentwickelte Insellösungen zum Forschungsdatenmanagement. Ebenso gibt es zahlreiche Ansätze, die das Thema Forschungsdatenmanagement und Publikationen systematisieren und im Sinne von Infrastrukturbildung Leitlinien und Forderungen formulieren. Abgeleitet werden diese Positions- und Forderungs-

³¹ http://www.igb-berlin.de/abt2/ms/ms_dat.shtml

³² http://www.icdp-online.org/contenido/lakedb/front_content.php?idcat=235

³³ http://www.zalf.de/home_zalf/service/service/informationssysteme/index.html, letzter Zugriff:

14.1.2010



papiere i.d.R. aus der Auswertung bestehender Praxis und der Analyse existierender Projekte und Systeme zum Forschungsdatenmanagement.

Eine wichtige Rolle in diesem Prozess spielen hierbei auch einige Mitglieder der Klimaplattform. Das Engagement in diesem Feld hängt allerdings stark von Größe und internationaler Vernetzung der Einrichtungen ab. Die Forschungsinstitute der Helmholtz- oder Leibniz-Gesellschaft haben eine Vorreiterrolle, wohingegen die Hochschulen über keine relevanten Forschungsdatenprojekte verfügen.

Verortung von Wibaklidama

Wibaklidama findet im Raum Berlin-Brandenburg mit der Zielgruppe der Klimaplattform-Partner eine reizvolle Forschungslandschaft und den Bedarf nach Ausbau einer Forschungsdateninfrastruktur vor. Hier spiegelt sich die Gesamtsituation. Es existiert keine regionale Dateninfrastruktur, die über eng gefasste Teildisziplinen und Einrichtungen substantiell hinaus reicht. Dort jedoch, wo die fachlichen Infrastrukturen als „Insellösungen“ bestehen, sind sie weit entwickelt und beispielgebend, allerdings auf nationale und internationale Vernetzung ausgerichtet und ohne Regionalbezug.

Wibaklidama reiht sich mit seiner weitgefassten Ausrichtung auf „Klimaforschung“ unter dem Aspekt Infrastrukturentwicklung in aktuelle vielversprechende Projekte ein. Der disziplinäre Kontext sollte im weiteren Projektverlauf auf jeden Fall eine Bezugsgröße bleiben, denn die Orientierung an Fachdisziplinen ist Voraussetzung bei der Entwicklung von Dateninfrastrukturen. Dennoch müssen durch vergleichbar strukturierte Metadaten Übergänge zwischen Teildisziplinen und Disziplinen ermöglicht werden. Insbesondere die „small science“ muss an bestehende Lösungsansätze angebunden werden. Eine technik- und disziplinübergreifende Lösung ist bisher die Ausnahme. Konzeptuell wegweisend ist hier das DRIVER-Projekt, das das Auffinden von Daten über verteilte lokale Systeme und Disziplinen ermöglichen soll.

Hochschulen und Institute, die auf dem Feld des Forschungsdatenmanagements über keine eigenen Entwicklungskapazitäten verfügen, sollten in die Lage versetzt werden, Konzepte zu adaptieren und technische Lösungen nachzunutzen. Zentrales



Problem wird dabei vermutlich der Spagat zwischen disziplinären Spezialisierungsansprüchen und der strukturellen Notwendigkeit von zentralen Lösungen sein.

Bei einer koordinierten Entwicklung bestehen jedoch gute Chancen, dass auch auf regionaler Ebene eine funktionierende Infrastruktur (im Sinne von Kooperation, Interoperabilität, Anwendung) entsteht.