

wibaklidama
wissensbasiertes Klima-Datenmanagement

Projektbericht B1
Datenmanagement in der Klimaforschung
Akteure, Bedarfe, Practices

Projektkürzel	Wibaklidama	
Typ und Nr.	B1 State-of-the-Art-Bericht	
Arbeitspaket(e)	AP 1-2	
Datum	18.6. 2010	
Dateiname	wibaklidama-B1V01-04.pdf	Version 01.04
Status	Öffentlich	
Autoren	Silke Grossmann	
Bearbeiter	Lars Müller	
Abstract	Beschreibt die Ergebnisse der ersten Phase des Projektes: Ausgangssituationen aufseiten der Einrichtungen, Bedarfe und Data Practices.	
Stichwörter	Klimaforschung, Klimaplattform, State-of-the-Art, Datenmanagement, Data Practices, Informationsbedarf	



Inhaltsverzeichnis

1. Zielsetzung und Vorgehensweise	3
2. Datenmanagement in der Klimaforschung: Data Practices	4
2.1 Big Science – Small Science.....	4
2.2 Organisationelle Verankerung des Datenmanagements.....	5
2.3 Von Projekten zu Services	5
2.4 Wissenschaftliche Informationsinfrastrukturen	6
2.5 Metadaten	6
2.6 Nutzung und Weitergabe von Daten	7
2.7 Open Access und Open Data.....	7
2.8 Forschungs- und Technologietransfer.....	8
3. Fazit: Umgang mit Daten in den Einrichtungen der Klimaplattform	9
4. Bedarf im Bereich Datenmanagement	10
5. Ansätze für Datenmanagement-Services	12

1. Zielsetzung und Vorgehensweise

In der ersten Phase des Projektes sollte die Ausgangssituation vorrangig aufseiten der Einrichtungen der Klimaplattform und des sie umgebenden Umfeldes möglichst umfassend analysiert werden. Ziel war es, einerseits den State-of-the-Art des Datenmanagements im Klimabereich in Berlin-Brandenburg, andererseits die Bedarfe der beteiligten Akteure zu erheben und zu beschreiben. Hierzu wurden verschiedene Instrumente eingesetzt (Abbildung 1).

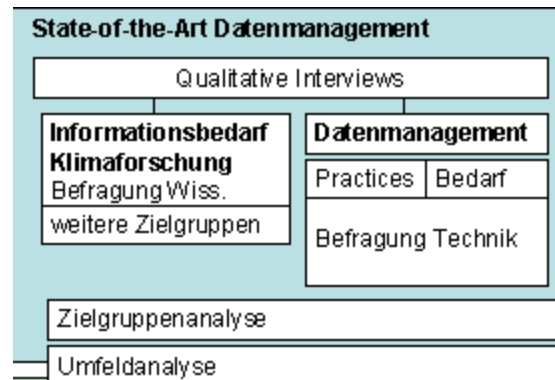


Abbildung 1: Vorgehensweise und Instrumente im Projekt wibaklidama

Ausgehend von dem Kick-Off-Workshop, der bereits erste Eindrücke vom Bedarf der verschiedenen Zielgruppen (Wissenschaftler, Techniker/Bibliothekare, Nutzer in der Politik und Verwaltung) vermittelte, vor allem aber auch die Unterschiedlichkeit dieser deutlich machte, wurden qualitative Interviews mit ausgewählten Vertretern der Einrichtungen der Klimaplattform geführt (insgesamt 10 zwischen Juli und September 2009, sie werden anonym zitiert und wurden zur Unterscheidbarkeit und Referenzierbarkeit mit laufenden Nummern versehen). Interviewt wurden sowohl Wissenschaftler, als auch Personen, die in den Einrichtungen für das Datenmanagement zuständig sind. Darauf aufbauend wurde eine Befragung zum Informationsbedarf in der Klimaforschung konzipiert und zunächst ausschließlich mit Nutzern in der Wissenschaft durchgeführt (Oktober - November 2009). Analog dazu wurde im Dezember 2009 eine Erhebung durchgeführt, die die Beschreibung der Data Practices in den Einrichtungen weiter systematisieren und zudem den Bedarf der „Datenmanager“ feststellen helfen soll.



2. Datenmanagement in der Klimaforschung: Data Practices

Im Folgenden werden die wichtigsten Ergebnisse aus der Institutionenanalyse und den Erhebungen zur Beschreibung des Umgangs mit Daten in den beteiligten Einrichtungen der Klimaplattform zusammengeführt.

Daten spielen in allen Einrichtungen der Klimaplattform eine bedeutende Rolle - sei es, dass Daten in Projekten erzeugt (erhoben) werden, sei es, dass sie als Input für eigene Analysen von Dritten eingeholt und verwendet werden.

2.1 Big Science – Small Science

Sehr deutlich zeigen sich die Unterschiede in der Praxis des Forschungsdatenmanagements zwischen Big Science und Small Science. Hochspezialisierte Forschungsinstitute praktizieren Datenmanagement, wohingegen Hochschulen keine koordiniertes Datenmanagement betreiben.¹ Ein systematisches Datenmanagement ist in der Regel für die mit sehr viel Aufwand erhobenen und qualitätsgesicherten Daten aus großen Projekten gegeben. Dennoch ist davon auszugehen, dass in keiner Einrichtung der Klimaplattform ein vollständiger Überblick über vorhandene Datenbestände existiert. Für Daten, die in kleinen und mittleren Projekten entstehen, existiert ein Datenmanagement nur im Ausnahmefall. Projektdaten sind über die Organisation verstreut, für Außenstehende weitgehend unbekannt und damit auch für eine Nutzung durch Dritte nicht verfügbar. Selbst in Einrichtungen, in denen große (meist disziplinspezifische) Repositories aufgebaut wurden,² fehlt der Überblick über die Daten der so genannten „small science“.

Aus den Interviews mit Vertretern der Einrichtungen der Klimaplattform (Frage 1) stammen die folgenden Aussagen hierzu:

- „die „kleineren“, möglicherweise relevanten Datenbestände aus den Projekten sind in der Regel weit verstreut und „unbekannt.“ Nicht alles ist also verfügbar“³
- „vor allem die kleinen Datenmengen werden oft vernachlässigt (sind nur beim Wissenschaftler gespeichert oder in kleinen Datensammlungen)“⁴
- „mit den kleinräumigen Daten hat die Bibliothek nichts zu tun, man konzentriert sich auf die 'großen' Daten“⁵

¹ Befragung Data Practices, Frage 1.

² Z.B. die „großen“ Daten-Projekte und Repositories wie PANGEA, National Climate Centre, Scientific Drilling Database.

³ Interview Nr. 1

⁴ Interview Nr. 2

⁵ Interview Nr. 2



- „wenn der Post-Doc weg ist, stellt sich die Frage: wer hat die Daten?“⁶

2.2 Organisationelle Verankerung des Datenmanagements

Zentrale Funktion vs. „scientific workspace“: Datenmanagement ist in den untersuchten Einrichtungen ganz unterschiedlich institutionalisiert. Zunächst ist der Wissenschaftler zuständig für seine Daten. Allerdings ist davon auszugehen, dass seine Motivation, Daten zu speichern, verfügbar zu machen oder gar zu archivieren, spätestens mit Ende des Projektes erlischt. Aufgaben des Datenmanagements werden teilweise von zentralen Serviceeinrichtungen (Daten- und Rechenzentrum, wissenschaftliche Infrastrukturen, IT-Einrichtung) übernommen – dies aber eher in technischer Hinsicht. „Wir nehmen sie (die Daten) auf, machen damit aber nichts.“⁷

Unterschiedlich stark ist die dezidierte Benennung und Zuordnung des Aufgabenfeldes „Datenmanagement“ ausgeprägt. Die meisten Einrichtungen verfügen über keine gesonderte Abteilung, wenn überhaupt, dann ist die IT-Abteilung die richtige Anlaufstelle.

2.3 Von Projekten zu Services

In vielen Einrichtungen gelten zum Umgang mit Daten eher allgemeine Grundsätze, wie z. B. die DFG-Empfehlungen zur guten wissenschaftlichen Praxis, aus denen auch die Anforderungen an das Forschungsdatenmanagement abgeleitet werden können.⁸ Das Datenmanagement ist bisher stark projektbezogen.⁹ Wenn ein Projekt endet, wechselt in der Praxis allerdings die Verantwortung für das Projekt. Rohdaten bleiben meist bei den Mitarbeitern. Dennoch werden in allen Institutionen – inkl. der Hochschulen – Daten-Services geplant oder für sinnvoll gehalten.¹⁰ Durchgängig können sich die Wissenschaftler auf IT-Services bei der Speicherung und Sicherung (im Sinne von Backups) ihrer Daten stützen. Eine Archivierung von Forschungsdaten, die über den Charakter von IT-Sicherheit hinausgeht, gibt es bisher kaum.¹¹

Serviceeinrichtungen, die sich mit dem Datenmanagement nachhaltig beschäftigen wollen, müssen weg von den projektbezogenen Lösungen hin zu Services: „Wir haben bis vor zwei Jahren das Datenmanagement immer projektbezogen gesehen und gleich von vornherein Lösungen gesucht und gefunden; daneben gibt es eben die Projekte mit wenigen Daten. Die kommen auf uns zu und wollen Lösungen: das funktioniert auf der Einzel-Projektebene nicht

⁶ Interview Nr. 2b

⁷ Interview Nr. 6

⁸ Befragung Data Practices, Frage 1

⁹ Befragung Data Practices, Frage 3

¹⁰ Befragung Data Practices, Frage 4

¹¹ Befragung Data Practices, Frage 7,8.



mehr. Wir brauchen 'Services'¹². Diese Aussagen korrespondieren mit der Auffassung von James M. Caruthers von der Purdue University, der davon ausgeht, dass große Projekte über die Expertise (und auch die Lösungen) für den Umgang mit Daten verfügen: „Large and very large projects tend to have very particular approaches, and probably represent significant expertise.“¹³ Für die Unterstützung vorwiegend der kleineren Projekte und um den „Ansturm“ künftig überhaupt bewältigen zu können, gilt es, „services that support projects in managing and curating their data better“¹⁴ zu definieren.

2.4 Wissenschaftliche Informationsinfrastrukturen

Bestandteil wissenschaftlicher Infrastrukturen sind Informationssysteme (Daten- und Text-Repositorien, Datenbanken) und das Rechenzentrum. Infrastruktureinrichtung ist auch die Bibliothek mit dem Publikationsportal/-datenbank, ggf. Webdiensten und webbasierten Forschungsplattformen. Alle erbringen Leistungen für die Unterstützung der Forschung. Datenmanagement ist als Funktion und Kompetenz überwiegend auf der wissenschaftlichen, projektbezogenen Ebene angesiedelt (mit Ausnahmen, z. B. GFZ, AIP).

Die D-GRID-Initiative¹⁵ ist ein Beispiel für ein nationales Bemühen um den Aufbau digitaler Informations-Infrastrukturen. Zu den schon bestehenden Infrastrukturen (z. B. in Deutschland) gehören u. a. die digitalen Repositorien, die meist federführend von den Bibliotheken aufgebaut wurden und in erster Linie für Publikationen bzw. Textdokumente gedacht sind.

2.5 Metadaten

In der Regel gibt es in den untersuchten Einrichtungen keine Vorschriften dazu, welche Daten dokumentiert werden müssen. Einigkeit besteht darüber, dass Metadaten für die Dokumentation und auch den Zugriff unerlässlich sind. Es gibt verschiedene Initiativen und Standards. Metadaten sind die Voraussetzung dafür, die Forschungsdaten zu einem späteren Zeitpunkt interpretierbar zu machen. Allerdings stellt der Aufwand der Erstellung eine große Barriere dar. Die Metadaten können nicht vollständig automatisch generiert werden und Wissenschaftler verfügen nicht über die Kapazitäten bzw. sind nicht willens, ihre Daten zu beschreiben. Dieser Umstand macht eine Verfügbarkeit von Daten zur Nachnutzung schwierig, wenn nicht gar unmöglich. Möglicherweise ist daher der Rückgriff auf interpretierte und aufbereitete Ergebnisse (Paper, Studien etc.) der einfachere Weg, Datensets zur Nachnutzung aufzufinden.

¹² Interview Nr. 2b

¹³ <http://digitalcuration.blogspot.com/2008/11/project-data-life-course.html>

¹⁴ <http://digitalcuration.blogspot.com/2008/11/project-data-life-course.html>

¹⁵ <http://www.d-grid.de/>



Aus den Interviews stammen die folgende Aussage hierzu: „Wir sehen uns nicht als Anbieter, weil wir keine Daten haben/produzieren und weil wir keine Metadaten haben bzw. automatisch generieren können. Was geht, ist so etwas wie die Deutschlandstudie, also Dinge, die irgendwie schon aufbereitet sind (das sind Interpretationen, Ergebnisse, Papers).“¹⁶

2.6 Nutzung und Weitergabe von Daten

Ownership of Data: Daten sind mindestens innerhalb der Einrichtung frei verfügbar. Die Weitergabe erfolgt dann aber meist nur auf Anfrage z. B. im Rahmen wissenschaftlicher Kooperationen an bekannte Leute. Die Verantwortung für das Datenmanagement liegt i.d.R. bei den Wissenschaftlern. So wird Zugang und Nutzung auch auf Projektleiterebene geregelt.¹⁷ Gegen eine freie Verfügbarmachung steht aufseiten der Wissenschaftler die Gefahr der Fehlinterpretation und des Datenmissbrauchs und auf der anderen Seite fehlendes Vertrauen in die Verlässlichkeit fremder Daten. Zugangsregelungen scheinen also unerlässlich.

Datenpublikation: Daten bilden die Grundlage für Publikationen und werden – wenn überhaupt – als elektronisches Supplement veröffentlicht. Die Publikation von Daten findet statt, ist aber noch nicht die Regel. Die (separate) Datenpublikation wird allgemein als adäquates Mittel angesehen.

Die Publikation ist oft auch das Mittel, um an Daten anderer Kollegen heranzukommen. Da dieses oft auf ein (bekanntes) Netzwerk beschränkt bleibt, kann hierbei einiges verloren gehen.

Portale: es gibt Portale/Projekte, die auch über Data Policies verfügen. Neben den großen Data Repositories gibt es auch fachspezifische Datenbanken. Viele Maßnahmen gehen von der Wissenschaft selber aus. Verzeichnisse in den Einrichtungen müssen mit dem Problem der Datenflut umgehen: nicht alle Daten müssen in einen „Verbundkatalog“.

Zugang zu Daten: Die Weitergabe von Daten muss im Dialog erfolgen. Der Zugang kann nur über den Bearbeiter erfolgen, der hat die aktuellen Daten und kann Fragen, die bei der Nutzung eigentlich immer entstehen, beantworten.

Nutzungsregelungen müssen aus den Kulturen heraus, die herrschen, formuliert werden.

2.7 Open Access und Open Data

Die Rahmenbedingungen sind da, aber keiner macht mit. Um diesem Problem zu begegnen, braucht man Strukturen, die das Mitmachen erzwingen (z. B. die elektronische Registrierung als Pflicht, Grundlage einer Evaluation): Was nicht veröffentlicht bzw. gemeldet ist, existiert nicht. „Wir brauchen Informationen über den wissenschaftlichen Output – und das muss von den Wissenschaftlern selber erhoben werden. Wir haben dazu ein System eingeführt, wel-

¹⁶ Interview Nr. 4

¹⁷ vgl. Ergebnisse der „Data-Practices-Befragung“, Fragen 5, 7, 9,10



ches von der Leitung unterstützt wird und somit über Druck funktioniert: alles was nicht im System ist, gibt es nicht (und kann damit auch nicht für eine Evaluation herangezogen werden).¹⁸ Es sprechen einige Gründe dagegen, Daten frei zur Verfügung zu stellen, obwohl Open-Access-Erklärungen breite Unterstützung finden. Die Gefahr der Missinterpretation wird hier angeführt und ein fehlendes Anreizsystem.

2.8 Forschungs- und Technologietransfer

Der Forschungs- und Technologietransfer (auch: Angewandte Forschung und Technologietransfer) ist in der Regel separat organisiert als Aufgabe/Abteilung – mit dem Ziel des Transfers der Know-hows in eine wirtschaftliche Anwendung (Technologietransfer). Diese „Stelle“ ist per Definition die zentrale Schnittstelle für Forschungsk Kooperationen. Zu den Aufgaben gehörten die Initiierung und Betreuung von Kontakten zwischen der Forschungseinrichtung/Universität und ihrem wirtschaftlichen und gesellschaftlichem Umfeld:

- Unterstützung von Wissenschaftlern bei der Anbahnung von Projekten
- Akquise von Transferideen / -vorhaben / Verwertungsideen
- über Forschungsgebiete/Schwerpunkte (mit Ansprechpersonen)
- z.B. in Form einer Forschungsdatenbank, die über Forschungsprojekte informiert (vgl. <http://www.fh-oow.de/forschung/index.php?id=402>)
- Vermittlung von Kontakten zwischen Hochschule und Wirtschaft
- Unterstützung der Nutzbarmachung der Forschungsergebnisse
- Forschungsnetzwerke (hochschulübergreifend, siehe z. B. <http://www.fh-oow.de/forschung/index.php?id=359>)
- Publikationen¹⁹

Klassische „Instrumente“ zur Förderung des Forschungs- und Technologietransfers sind z.B. Forschungs- und Technologiedatenbanken, in denen die Hochschulen ihre Angebote zugänglich machen, sowie der Aufbau von Forschergruppen, Clustern und Forschernetzwerken (Wissensverbände, Wissenschaftsallianzen). Schwerpunkt des Technologietransfers ist in der Regel die Verwertung von Forschungsergebnissen.

siehe z. B. auch: Kompetenzlandkarte <http://www.lifecompetence.eu/>

¹⁸ Interview Nr. 10

¹⁹ Vgl. <http://computerphilologie.tu-darmstadt.de/jg07/aschmeff.html>



3. Fazit: Umgang mit Daten in den Einrichtungen der Klimaplattform

In den meisten Einrichtungen fehlt noch die ausreichende Sensibilisierung für die Bedeutung des Themas und die Notwendigkeit einer systematischen Auseinandersetzung damit – und das, obwohl die Bedeutung von Daten eigentlich schon unbestritten ist. Selbst die aktive Unterstützung der Wissenschaftler scheint noch sehr in den Anfängen begriffen, kaum eine Einrichtung bietet aktive Dienste für Daten an (vgl. Abschnitt Data Services). In der Regeln fehlen Vorschriften zum Umgang mit Daten, also z. B. Richtlinien, welche Daten dokumentiert werden müssen (Selection Policies).



4. Bedarf im Bereich Datenmanagement

„Metadaten und Standards“ sowie „Langzeitarchivierung“ wurden auf die Frage nach dem Fortbildungsinteresse am häufigsten genannt. Sie werden gefolgt von Data-Policies, Datenanalyse und Publikation. Somit zeigt sich, dass die dringendsten Probleme im Bereich der fachlichen und organisatorischen Fragen liegen und die technischen Probleme (z.B. Systeme zum Betrieb eines Repositoriums oder rechtliche Aspekte) leichter lösbar erscheinen.²⁰

Die folgenden Äußerungen, die zum Teil aus einer anonymen Kartenabfrage auf dem Kick-Off-Meeting entstammen, zum Teil in persönlichen Gesprächen geäußert wurden, helfen, den Bedarf näher zu bestimmen. Gebraucht wird aus Sicht der Datenmanagement-Experten:

- der Austausch von Know-how
- ein Überblick über Vorhandenes
- kooperative Services
- eine Lösung, die Strategie, technische Werkzeuge und Kultur zusammenbringt

Datenmanagement muss „in die Arbeitsabläufe integriert werden; wir brauchen Strukturen, die das Mitmachen erzwingen“.²¹

Natürlich haben auch Wissenschaftler Bedarf in der Verbesserung ihres Datenmanagements, allerdings wurde dieses weniger als Unterstützung konkreter datenbezogener Aufgaben (z.B. Planung der Datenerhebung etc. geäußert), als vielmehr in der Such- bzw. Vermittlungsfunktion bei der Suche nach konkreten Anfragen auf spezielle Fachfragen. Der Bedarf aus wissenschaftlicher Sicht wurde in einer separaten Befragung, die anonym durchgeführt wurde, erhoben; im Folgenden sind auch hieraus einige einschlägige Bedarfsäußerungen wiedergegeben:

- Metainformationsverzeichnis: Daten über Daten – wo ist etwas zu finden?
- Hilfe bei der Datenorganisation – bessere Aufbereitung der eigenen Daten
- Überblick von Werkzeugen für das Datenmanagement
- Verdichtung von Daten zu Wissen
- Sicherstellung der Verlässlichkeit der Daten

“Ein Traum wäre es, wenn man nur das schicken müsste → allerdings ist das ein Widerspruch, denn: die Beschreibung/Erläuterung muss sicher der Wissenschaftler machen. Das

²⁰ Befragung Data Practices, Frage 12, 13.

²¹ Interview Nr. 2b.



ist aber ein zeitlicher Aufwand, den keiner leisten will. Das wird nur gemacht, wenn man getreten wird.“²²

²²

Interview Nr. 2b.



5. Ansätze für Datenmanagement-Services

Eine moderne wissenschaftliche Infrastruktur sollte auch dezidierte Services und Werkzeuge für die Unterstützung des Wissenschaftlers beim Umgang mit Daten umfassen. Wie bereits dargestellt, ist das aber bisher noch wenig systematisch umgesetzt. Die Rolle der Bibliothek in diesem Prozess ist bisher sehr gering: 'Die Bibliothek hat mit den Daten nichts zu tun und kann das auch nicht leisten'. Als Begründung hierfür wird der fehlende fachwissenschaftliche Hintergrund angeführt. Wird die Bibliothek einbezogen, dann vor allem zur Unterstützung im Bereich Metadaten und in der Koordination der Datenerfassung. „Die Primärdaten machen wir alleine – die Bibliotheksleute sind dafür auch nicht entsprechend ausgebildet.“

Hinweise darauf, wie Data Services aussehen könnten, geben folgende Aussagen, die sowohl im Kick-Off-Meeting zum Projekt als auch in persönlichen Gesprächen gesammelt wurden:

- Unterstützung in der Aufbereitung meiner Daten
- es geht nicht nur um Metadaten, viele Informationen sind Kontextinformationen
- nicht alle Daten sind relevant
- Anwendung informationswissenschaftliche Methoden
- es muß eine zentrale Verantwortung – weg vom Wissenschaftler – geben. Dieser liefert den Final Draft, den Rest machen die Experten, vgl. dem Prozeß bei Publikationen (daran kann man sich orientieren) (für die Persistent Domain)
- in Ableitung der institutionellen Repositories braucht man Data Repositories
- Beratung im Umgang mit konkreten Daten: Tipps und Tricks

Folgende Tabelle gibt eine Übersicht über die Aufgaben einer zentralen Daten-Service-Abteilung:

Erhebung von Daten
Datenvalidierung
Speicherung / Aufbewahrung von Daten
Dokumentation/Beschreibung
Weitergabe von Daten auf Anfrage (intern/extern)
Aufbau einer zentralen Datenbank/Data Repository
Nutzung/Analyse / Bearbeitung der Daten
Publikation (=> Artikel, Aufsätze)

Tabelle 1: Sammlung datenbezogener Aufgaben