

## Interdisziplinäres Symposium

### Umgang mit Forschungsdaten: Gängige Praxis und neue Anstöße

- Abstracts -

#### **Roland Bertelmann**

Deutsches GeoForschungsZentrum – GFZ

#### ***Herausforderungen und Fragestellungen***

Die Herausforderungen und Fragestellungen des Umgangs mit Forschungsdaten werden anhand eines Modells zum Lebenszyklus von Forschungsdaten kurz umrissen und führen in das Themenspektrum des Symposiums ein.

#### **Matthias Steinmetz**

Leibniz-Institut für Astrophysik Potsdam (AIP)

#### ***Datengrids und virtuelles Observatorium: mehr als ein google für Sterne!***

Sowie das Internet unser tägliches Leben beeinflusst hat, so ändern die modernen Informationstechnologien auch die Art und Weise wie Astronomen und Astrophysiker arbeiten. Moderne Großteleskope, umfassende Himmelsdurchmusterungen mit Teleskopen und Satelliten sowie Simulationen mit Höchstleistungsrechnern liefern Datenmengen, die nicht mehr mit den traditionellen Arbeitsplatzrechnern zu beherrschen sind. Ausgeklügelte Suchstrategien in diesen Datenbanken erlauben die schnelle Entdeckung und Identifizierung von Himmelsereignissen, wie z.B. einer Supernova. Robotische Teleskope ermöglichen es, automatisch und kurzfristig darauf zu reagieren und so wichtige Erkenntnisse aus der Frühphase dieser Objekte zu gewinnen. Die Erträge dieser gigantischen und systematischen Durchmusterungen werden in »Virtuellen Observatorien« zusammengeführt. Sie bieten einen panchromatischen Blick ins All und ermöglichen es auch dem Laien, sich aktiv an der Auswertung der Daten zu beteiligen.

#### **Gert G. Wagner**

Rat für Sozial- und Wirtschaftsdaten (RatSWD)

#### ***Forschungsdatenmanagement: Erfahrungen und verallgemeinerte Erkenntnisse aus dem Bereich der Sozial-, Verhaltens- & Wirtschaftsdaten***

Insbesondere die Sozial- und Wirtschaftswissenschaften, zunehmend aber auch die Verhaltenswissenschaften (Psychologie) haben seit Jahrzehnten Erfahrungen mit der verteilten Nutzung von Forschungsprimärdaten (erhoben insbesondere auf der Ebene von Personen, Privathaushalten, Betrieben oder Unternehmen) Mikrodaten). Die Erfahrungen zeigen, dass die Beherrschung von Datenschutzproblemen möglich ist und viele Mikrodaten der amtlichen Statistik und der forschungsbasierten Dateninfrastruktur sind weltweit mehr oder weniger leicht zugänglich. In Deutschland werden Forschungsdaten- und Datenservicezentren institutionell gefördert. Gleichwohl gibt es nach wie vor – und aufgrund der stärkeren Konkurrenz zwischen Forschungseinrichtungen und Forschern sogar zunehmend – das Problem, dass Datenarchivierung, Dokumentation und ein nutzerfreundlicher Zugang zu Forschungsprimärdaten von der akademischen Community zu wenig gewürdigt werden. Dadurch leidet die Bereitschaft sehr gute Forschungsdaten zu produzieren und zugänglich zu machen. Disziplinen übergreifende neuere Entwicklungen wie eindeutige Kennungen für Daten und Forscher sowie Standards für die Zitation von Daten lassen Abhilfe erwarten. Eine darüberhinausgehende Zusammenarbeit bei der Entwicklung von Archivierungs-, Dokumentations- und Zugang-Standards dürfte sehr hilfreich sein..

### **Jens Klump**

Deutsches GeoForschungsZentrum – GFZ

#### ***Forschungsdaten und Forschungsdateninfrastrukturen – Aktuelle Initiativen***

Die Entwicklung des Internets und die rasanten Fortschritte in der Informationstechnologie stellen uns vor neue Herausforderungen im Umgang mit Forschungsdaten. Technisches Datenmanagement allein reicht nicht mehr aus, auch die Ansprüche an den Zugang und an die Nutzbarkeit von Forschungsdaten sind gestiegen. Wissenschaftsorganisationen, Forschungsförderer, Projekte und Forscher sehen sich hier vor vielfältige Herausforderungen gestellt, denen mit einer Reihe von Initiativen und Förderprogrammen auf nationaler und internationaler Ebene begegnet wird. In der aktuellen Förderung der DFG für Forschungsdateninfrastrukturen sind Antragsteller aus der Region Berlin/Brandenburg besonders stark vertreten. Dieser Vortrag gibt einen kurzen Überblick über aktuelle Initiativen und Projekte.

### **Roland Bertelmann**

Deutsches GeoForschungsZentrum – GFZ

#### ***Auswertung des Fragebogens: Rahmenbedingungen des Umgangs mit Forschungsdaten in der Region***

Zur Vorbereitung des Symposiums wurden Institutionen aus allen Forschungsdisziplinen aus der Region Berlin Brandenburg gebeten, einen kurzen Fragebogen auszufüllen. Der Fragebogen orientierte sich an Grundfragen zur Erstellung von Datenmanagementplänen. Die vorgestellten Ergebnisse der Umfrage geben einen interdisziplinären Überblick zum Stand des Umgangs mit Forschungsdaten, beleuchten Stärken und identifizieren potentielle Lücken.

### **Uwe Heinrich**

Leibniz-Zentrum für Agrarlandschaftsforschung - ZALF

#### **Forschungsdatenmanagement - Erfahrungen aus der Landschaftsforschung**

Zur Erfüllung seines Auftrages beschreitet das ZALF den Weg einer inter- und transdisziplinären integrativen Landschaftsforschung. Eine Voraussetzung für die erfolgreiche Umsetzung ist eine Forschungsdateninfrastruktur. Im Landschaftsinformationssystem des ZALF werden die zentral verwalteten Daten durch über 3000 Metadatensätze beschrieben und über direkten Zugriff auf die Daten oder über Webdienste, je nach Nutzeranforderungen und Zugriffsrechten, für die Aufgaben der Landschaftsforschung zur Verfügung gestellt. Mit der Publikation von Primärdaten und der Nutzung von Webdiensten zur Darstellung von Szenarien und Studien werden neue Wege der Bereitstellung von Forschungsdaten und -ergebnissen der Landschaftsforschung erprobt.

### **Michael Flechsig**

Potsdam-Institut für Klimafolgenforschung (PIK)

#### ***Die Modellierungsstrategie des PIK als Beitrag zu einem effizienten Datenmanagement***

Die wichtigsten methodischen Ansätze der Klimafolgenforschung sind System- und Szenarienanalyse, quantitative und qualitative Modellierung, Computersimulation und Datenintegration. Vor dem Hintergrund, dass am PIK eine ganze Reihe unterschiedlicher Simulationsmodelle entwickelt, adaptiert und genutzt werden, wurde im letzten Jahr am Institut eine Modellierungsstrategie diskutiert und diese gegenwärtig implementiert. Sie basiert auf anerkannten Regeln im Sinne von „best scientific practices“ für Entwicklung, Pflege und Management solcher Modelle und adressiert in diesem Kontext auch eine nachhaltige Modell-, Experiment- und Ergebnisdokumentation. Die Modellierungsstrategie bietet einen flexiblen Rahmen mit klaren Zuständigkeiten für den Umgang mit Forschungsdaten und dabei speziell Modelloutput. Während die Organisation des Datenmanagements in der Verantwortung der einzelnen Modellierungsgruppen und Forschungsbereiche liegt, gibt es zentrale Dienste für Datenarchivierung, -annotation und Metadatenverwaltung. Für die Datenpublikation werden u.a. verteilte Ansätze wie das Datengrid C3Grid von Instituten der Erdsystemforschung genutzt.

**Ulrike Lucke**

Universität Potsdam

***Was Forschung von Lehre lernen kann: Metadaten-Management im E-Learning***

Vernetzung birgt Potential für Kooperation -- das gilt nicht nur in der Forschung. Die nahezu beliebig breite Distribution und Akquisition von Daten das Internet bringt in vielen Domänen das Thema Metadaten auf den Plan. Daten über Daten sind der Schlüssel zum Zugriff auf diese Daten, wie Bibliothekare seit Jahrtausenden wissen. Doch sie sind nur sinnvoll nutzbar, wenn sie vollständig und korrekt sind. Im Vortrag werden Mechanismen aus dem E-Learning vorgestellt, die dieses Problem der Qualität von Metadaten erfolgreich behandeln, und auf das Management von Forschungsdaten übertragen.

**David Lazarus**Museum für Naturkunde - Leibniz-Institut für Evolutions- und Biodiversitätsforschung  
an der Humboldt-Universität zu Berlin***Whole organism research data - from initial samples to global syntheses***

Whole organism data (taxonomic names and data associated to them) are central to many areas of biologic, geologic, environmental and even medical research. Taxonomic names are tags for scientific hypotheses of population structure, linked to reference type specimens stored in museums, and are 'key fields' for information exchange. Building global internet catalogs of standardised taxa names and definitions (text, images) is a central activity for many taxonomists. Associated data, such as occurrences of taxa used in studies of environmental or evolutionary change, are only partially managed. Some data is retained in archives associated with publications, and a few databases (both modern and paleontologic) synthesise published global occurrence information. Archives need better taxonomic data management abilities, while global databases need long-term institutional support.

**Jens Klump**

Deutsches GeoForschungsZentrum – GFZ

***Konzepte und Beispiele für die Publikation und Langzeitverfügbarkeit von Forschungsdaten***

Der überwiegende Teil der Forschungsdaten entsteht heute digital, dennoch bestehen nur in wenigen Fällen Abläufe und Werkzeuge mit denen die Erhaltung und Bereitstellung von Forschungsdaten unterstützt wird. Die schnell wachsende Menge an Forschungsdaten stellt uns vor neue Herausforderungen, diese Daten zugänglich, nachnutzbar und über lange Zeiträume verfügbar zu machen. Wie kann diesen Herausforderungen begegnet werden? In diesem Vortrag werden Konzepte vorgestellt, wie Daten verfügbar gemacht werden können, welche organisatorischen und technischen Modelle dafür entwickelt wurden und wie die Kosten für diese Dienste abgeschätzt werden können; die vorgestellten Konzepte werden mit Beispielen illustriert.

**Harry Enke**

Leibniz-Institut für Astrophysik Potsdam (AIP)

***Data Management und Data Mining von kosmologischen Simulationen***

Hochauflösende kosmologische Simulation bestehen aus Rohdaten von hunderten Terabyte. Verschiedene Zeitschritte werden in der Regel an vielen HPC gerechnet, um die hierzu notwendigen Millionen CPU Stunden zu bekommen. Die wissenschaftliche Analyse der Rohdaten erfordert die Kollaboration vieler Wissenschaftler, die durch die über die HPC verteilten Daten erheblich behindert wird. Für das Datenmanagement einer solchen Kollaboration ist CoSiMM (Collaborative Simulation Metadata Management) in Zusammenarbeit mit dem CLUES-Projekt entwickelt worden. Der Zugang zu den Ergebnissen der wissenschaftlichen Analyse der Simulationsdaten, Halo-Katalogen und Merger-Trees und anderen Daten wird durch die Nutzung von Datenbanken und SQL effizienter, und es ist möglich, grösseren Gruppen von interessierten Wissenschaftlern Zugang zur Nutzung der Daten über Webinterfaces zu ermöglichen. Es werden die Arbeiten des AIP e-Science Teams zu den Projekten MultiDark und CoSimm/CLUES vorgestellt.

### **Stefanie Schumacher**

Alfred Wegener-Institut für Polar- und Meeresforschung – AWI

#### ***Datenarchivierung und Datenpublikation mit PANGAEA***

PANGAEA - Data Publisher for Earth and Environmental Science ist ein digitales Bibliothekssystem für die Daten-Archivierung und -Veröffentlichung aus der Erdsystemforschung und den Umweltwissenschaften. Das System wird durch das Alfred-Wegener-Institut für Polar und Meeresforschung, Bremerhaven und das Zentrum für Marine Umweltwissenschaften, Bremen getragen. Beide Institute garantieren eine Langzeit-Verfügbarkeit der Datenbibliothek und ihren Inhalt. Wissenschaftliche Daten werden georeferenziert in einer relationalen Datenbank abgelegt. Alle Daten sind frei verfügbar (Open Access) und werden über Webdienste im Internet zur Verfügung gestellt. Zu jedem Datensatz gehört eine Datensatzbeschreibung, die ein bibliographisches Zitat beinhaltet und die dauerhaft über einen Digital Object Identifier (DOI) verfügbar ist. Datensätze können als eigenständige Daten-Publikation oder als Supplement zu einer Publikation veröffentlicht werden. PANGAEA und Elsevier haben seit 2009 eine Kooperation, PANGAEA Daten sind direkt mit dem entsprechenden Elsevier-Artikel verlinkt.

### **Stephan Büttner**

Fachhochschule Potsdam (FHP)

#### ***Das Handbuch Forschungsdatenmanagement***

Zur diesjährigen Frankfurter Buchmesse erscheint das "Handbuch Forschungsdatenmanagement" als Hybrid-Publikation. In diesem Handbuch werden zentrale Aspekte des Forschungsdatenmanagements aus informationswissenschaftlicher und anwendungsbezogener Perspektive disziplinübergreifend von zahlreichen ausgewiesenen Expertinnen und Experten behandelt. Das Handbuch gliedert sich in drei Teile. Im ersten Teil erfolgt eine allgemeine Einführung in das Themengebiet. Grundlegende Begriffe werden erläutert, die Entwicklung und Bedeutung des Themas in den Wissenschaften werden dargestellt. Der zweite Teil widmet sich der Praxis des Forschungsdatenmanagements. Einzelne Problemfelder werden aufrissen und praktisch-technische Lösungsansätze und Empfehlungen dargestellt. Schwerpunkt des dritten Teils sind die strategisch-konzeptionellen Entwicklungen des Forschungsdatenmanagements, die auf die Entwicklung von Dateninfrastrukturen zielen und auch Konsequenzen für die Aus- und Weiterbildung erfordern.