

An der Schwelle zum 4. Paradigma

Datenmanagement in der Klimaplattform

Lars Müller



Themen

- Das 4. Paradigma
- Forschungsdomäne Klimawandel – Projekt Wibaklidama
- Data Practices in der Klimaplattform
- Brücken zu einer Forschungsdateninfrastruktur
- Praktische Ansätze zum Ausbau der Dateninfrastruktur



Das 4. Wissenschafts-Paradigma

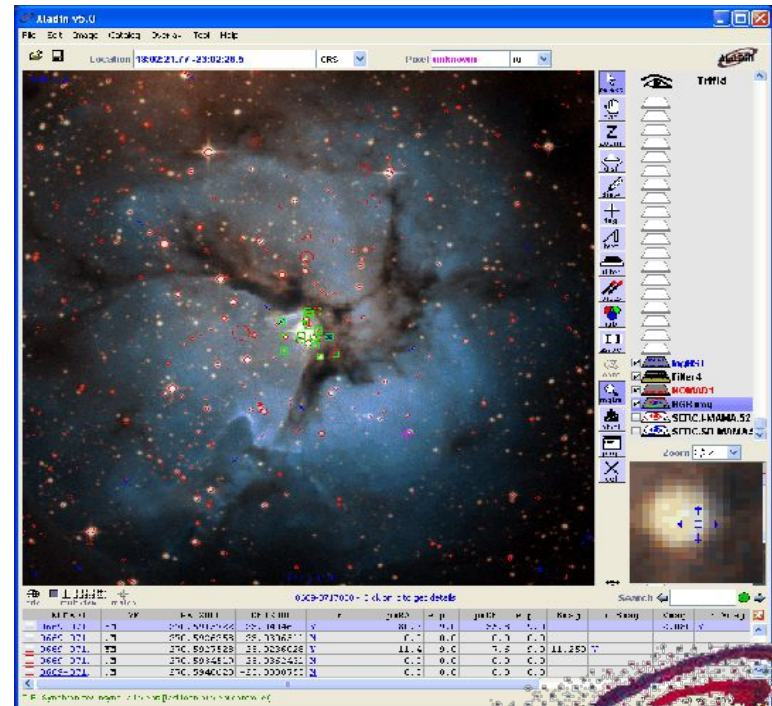
1. Empirie
Beschreibung natürlicher Phänomene
2. Theorie
Entwurf von Modellen
3. Simulation
Berechnung komplexer Zusammenhänge
4. Datengetriebene Wissenschaft / eScience



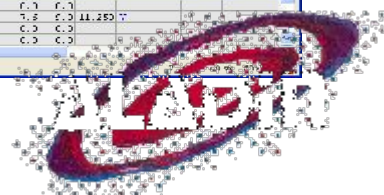
Beispiel: Virtuelles Observatorium



Beispiel: Virtuelles Observatorium



<http://aladin.u-strasbg.fr/aladin.gml>



Webtool unter:
<http://aladin.u-strasbg.fr/java/nph-aladin.pl>

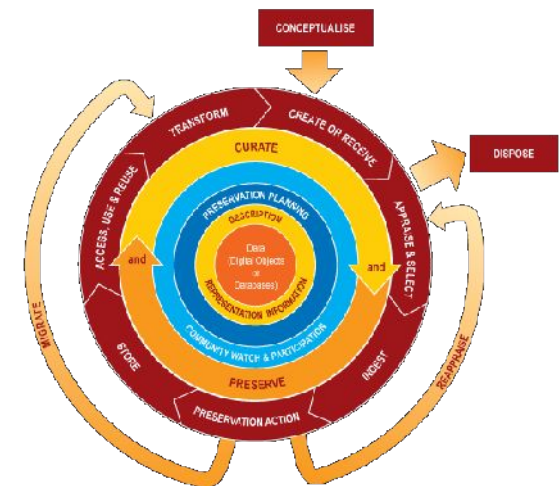


Erforderlich: Datenmanagement Datenkuratierung / Data Curation



Erforderlich: Datenmanagement Datenkuratierung / Data Curation

- Dokumentieren der Daten (Metadaten)
- Archivieren (dauernde Aufbewahrung, Referenzierbarkeit, Sichern der Nachweiskette)
- Zugänglich machen (Auffindbarkeit, Interpretierbarkeit)
- Idealerweise: Publikation der Forschungsdaten



<http://www.dcc.ac.uk/resources/curation-lifecycle-model>



Forschungsdomäne Klimawandel

- Projekt Wibaklidama:
Wissensbasiertes Klimadatenmanagement



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

- Empirische Untersuchung des Datenmanagements in der Klimaplattform
- Ziel: Forschungsdatenmanagement und -austausch unterstützen



Daten in der Klimaplattform



Daten in der Klimaplattform

- Klimaforschungsrelevante Daten sind:
 - Heterogen
 - Großvolumig



Daten in der Klimaplattform

- Klimaforschungsrelevante Daten sind:
 - Heterogen
 - Großvolumig
- Handhabung der Forschungsdaten:
 - Repositorien für Messinstrumente
 - Disziplinspezifische Repositorien
 - Projektbezogene Publikationen
 - Interne Ablage



Beispiel: Metadaten zu einem Datensatz

Quelle: Information Systems and Data Center, GFZ Potsdam

(1)



Beispiel: Metadaten zu einem Datensatz

(1)

Quelle: Information Systems and Data Center, GFZ Potsdam

Entry_ID: IGS-NRT-NAV+ban2_2009_224_00_00
Entry_Title: GPS data for ISDC NRT, GNSS ephemerides

Group: Data_Set_Citation
Dataset_Creator: GFZ Potsdam
Dataset_Title: ISDC NRT, GNSS ephemerides, RINEX 2.1, Unix compressed
Dataset_Release_Date: 2009-10-30 11:34:22
Dataset_Release_Place: Potsdam, Germany
Dataset_Publisher: GFZ IGS-ISDC
Version: 1.000
Data_Presentation_Form: profile
Online_Resource: <http://www.gfz-potsdam.de>
End_Group

Group: Personnel
Role: Investigator
First_Name: Markus
Last_Name: Rothacher
Email: reigber@gfz-potsdam.de
Phone: +49-331288-1100
FAX: +49-331288-1111
Group: Address
Prof. Dr. M. Rothacher
GFZ Potsdam, Dep. 1
Telegrafenberg
14473 Potsdam
GERMANY
End_Group
End_Group

Group: Personnel
Role: Technical Contact
First_Name: Markus
Last_Name: Ramatschi
Email: maram@gfz-potsdam.de
Phone: +49-331288-1746
FAX: +49-331288-1111
Group: Address
Dr. Markus Ramatschi
GFZ Potsdam
Telegrafenberg
14473 Potsdam
GERMANY
End_Group
End_Group

Group: Personnel
Role: DIF Author
First_Name: Markus
Last_Name: Ramatschi
Email: maram@gfz-potsdam.de
Phone: +49-331288-1746
FAX: +49-331288-1111
Group: Address
Dr. Markus Ramatschi
GFZ Potsdam
Telegrafenberg
14473 Potsdam
GERMANY
End_Group
End_Group



(2)

Originating_Center: GFZ Potsdam

Group: Data_Center

Data_Center_Name: ISDC > NRT Information System Data Center

Data_Center_URL: http://isdc.gfz-potsdam.de/gnss

Group: Personnel

Role: Data Center Contact

First_Name: Bernd

Last_Name: Ritschel

Email: rit@gfz-potsdam.de

Phone: +49-331288-1685

FAX: +49-331288-1703

Group: Address

Bernd Ritschel

GFZ Potsdam

DRZ

Telgrafenberg A3

14473 Potsdam

GERMANY

End_Group

End_Group

End_Group

Discipline: EARTH SCIENCE

Group: Parameters

Category: EARTH SCIENCE

Topic: SOLID EARTH

Term: GEOPHYSICAL FIELDS

Variable: ISDC NRT GPS data

End_Group



Kontaktdaten

Group: Parameters

Category: IGS-ISDC

Topic: processing facility

Term: GFZ

Variable: varchar(80)

End_Group

Group: Parameters

Category: IGS-ISDC

Topic: igs-filename

Term: ban2224a.09n.Z

Variable: varchar(20)

End_Group

Group: Parameters

Category: IGS-ISDC

Topic: isdc-filename

Term: IGS-NRT-NAV+ban2_2009_224_00_00.Z

Variable: varchar(80)

End_Group

Group: Parameters

Category: IGS-ISDC

Topic: generation date

Term: 2009-10-30 11:34:22

Variable: datetime

End_Group



(3)

Group: Parameters
 Category: IGS-ISDC
 Topic: mission
 Term: GSTB
 Variable: varchar(20)
 End_Group

Group: Parameters
 Category: IGS-ISDC
 Topic: data access
 Term: restricted
 Variable: varchar(80)
 End_Group

Group: Parameters
 Category: IGS-ISDC
 Topic: revision
 Term: 1.00
 Variable: numeric(4,3)
 End_Group

Group: Parameters
 Category: IGS-ISDC
 Topic: start date
 Term: 2009-08-12 00:00:00
 Variable: datetime
 End_Group

Group: Parameters
 Category: IGS-ISDC
 Topic: stop date
 Term: 2009-08-12 00:00:00
 Variable: datetime
 End_Group

Group: Parameters
 Category: IGS-ISDC
 Topic: station id
 Term: ban2
 Variable: varchar(80)
 End_Group

Group: Parameters
 Category: IGS-ISDC
 Topic: location
 Term: Asia > India > Bangalore (GFZ)
 Variable: varchar(80)
 End_Group

Group: Parameters
 Category: IGS-ISDC
 Topic: longitude
 Term: 77.51
 Variable: numeric(10,7)
 End_Group

Group: Parameters
 Category: IGS-ISDC
 Topic: latitude
 Term: 12.95
 Variable: numeric(10,7)
 End_Group



(4)

Group: Parameters
 Category: IGS-ISDC
 Topic: altitude
 Term: 917.
 Variable: numeric(10,3)
 End_Group

Group: Parameters
 Category: IGS-ISDC
 Topic: observation file reference
 Term: IGS-NRT-OBS+ban2_2009_224_00_00.hat.Z
 Variable: varchar(80)
 End_Group

Group: Parameters
 Category: IGS-ISDC
 Topic: navigation file reference
 Term: IGS-NRT-NAV+ban2_2009_224_00_00.Z
 Variable: varchar(80)
 End_Group

Group: Parameters
 Category: IGS-ISDC
 Topic: quality check file reference
 Term: IGS-NRT-QCK+ban2_2009_224_00_00.Z
 Variable: varchar(80)
 End_Group

Group: Parameters
 Category: IGS-ISDC
 Topic: observables
 Term: GNSS ephemerides
 Variable: varchar(80)
 End_Group

Keyword: GPS ground data

Source_Name: GPS > Global Positioning System Satellites
 Sensor_Name: GPS-Receiver > AOA BENCHMARK ACT

Project: IGS-NRT > Near RealTime GPS data

Group: Quality
 TEQC Quality file available: IGS-NRT-QCK+ban2_2009_224_00_00.Z
 End_Group

Group: Distribution
 Distribution_Size: 1096
 Distribution_Format: RINEX 2.1, Unix compressed
 End_Group

Group: Summary
 This data set comprises GPS ground data of a sample rate of 30 sec,
 generated by decoding the original measurement data. The quality check was performed with TEQC.
 End_Group



Data Practices in der Klimaplattform



Teilen von Daten: Bedarf



Teilen von Daten: Bedarf

„Wir erzeugen eigentlich gar nicht selber Daten, sondern nehmen die Daten Anderer. (...) Wir produzieren als Output in erster Linie Paper bzw. Modell-Outputs.“



Teilen von Daten: Bedarf

„Wir erzeugen eigentlich gar nicht selber Daten, sondern nehmen die Daten Anderer. (...) Wir produzieren als Output in erster Linie Paper bzw. Modell-Outputs.“

Am besten decken den Bedarf "Geeichte und korrigierte, vergleichbare Messwerte (Saubere Rohdaten ohne Fehler und prozessierte Daten).„



Teilen von Daten: Bedarf

„Wir erzeugen eigentlich gar nicht selber Daten, sondern nehmen die Daten Anderer. (...) Wir produzieren als Output in erster Linie Paper bzw. Modell-Outputs.“

Am besten decken den Bedarf "Geeichte und korrigierte, vergleichbare Messwerte (Saubere Rohdaten ohne Fehler und prozessierte Daten).„

Wunsch nach Zugriff auf Daten- und
Expertenverzeichnisse



Weitergabe von Daten (1/2)



Weitergabe von Daten (1/2)

"Metadaten werden nicht automatisch generiert, sondern das muss händisch gemacht werden. Und genau das ist das Problem. Es gibt (...) einige kleinere Informationssysteme, 'der große Schlag' ist uns aber noch nicht gelungen."



Weitergabe von Daten (2/2)

Die "Gefahr der Missinterpretation (...) hat dazu geführt, dass wir sehr restriktiv geworden sind im Umgang mit Open Access bei Daten (eben mit Ausnahme von Daten, die „publiziert“ wurden, da müssen wir ja).„



Weitergabe von Daten (2/2)

Die "Gefahr der Missinterpretation (...) hat dazu geführt, dass wir sehr restriktiv geworden sind im Umgang mit Open Access bei Daten (eben mit Ausnahme von Daten, die „publiziert“ wurden, da müssen wir ja).„

"Ich würde allerdings die Daten nicht jedem so einfach geben, sie nicht frei ins Netz stellen, dafür braucht man Zugangsregelungen. Neben dem rechtlichen Aspekt ist hierbei auch bedeutsam die Rückkoppelung mit anderen, Datenaustausch ist für mich an Personen gekoppelt."



Nutzung fremder Daten



Nutzung fremder Daten

„Ich selber bin vorsichtig mit Daten, die ich nicht kenne. So muß dann auch eine zentrale Datenbank sicherstellen, dass die Daten verlässlich sind. (...) Außerdem brauche ich den Kontakt. Es ergeben sich immer Fragen.“



Teilen von Daten praktizieren

"Der Zugriff auf Daten muß im Dialog erfolgen"



Teilen von Daten praktizieren

"Der Zugriff auf Daten muß im Dialog erfolgen"

- Bedarf ist vorhanden, aber:
- große Vorbehalte bei Datenerzeugern und -nutzern



Teilen von Daten praktizieren

"Der Zugriff auf Daten muß im Dialog erfolgen"

- Bedarf ist vorhanden, aber:
- große Vorbehalte bei Datenerzeugern und -nutzern
- Vermittlung des Vertrauensverhältnisses zwischen Datenerzeuger und -nutzer durch Qualitätsgeprüfte wissenschaftliche Publikationen von Forschungsdaten

Zitate aus Interviews mit Datenmanagern, geführt von Silke Grossmann, Juni – September 2009 und einer Online-Befragung zum Informationsbedarf in der Klimaforschung.



Forschungsdaten in der „Small Science“



Forschungsdaten in der

„Small Science“

- überschaubare Projekte
- Daten befinden sich am Rechnerarbeitsplatz
- Datenmanagement nur innerhalb der Arbeitsgruppe
- ungenügende Erfassung von Metadaten
- Verbleib der Daten nach Projektende ungewiss
- keine Policies
- Zugang über persönliche Ansprache



Forschungsdaten in der

„Small Science“

- überschaubare Projekte
- Daten befinden sich am Rechnerarbeitsplatz
- Datenmanagement nur innerhalb der Arbeitsgruppe
- ungenügende Erfassung von Metadaten
- Verbleib der Daten nach Projektende ungewiss
- keine Policies
- Zugang über persönliche Ansprache

„Big Science“

- Betrieb von „Data Centern“ / disziplinspezifischen Datenrepositorien
- Einsatz internationaler Standards und Formate
- Langzeitarchivierung ist gesichert
- Datenzugriff ist geregelt, häufig open access



Einrichtungstypen

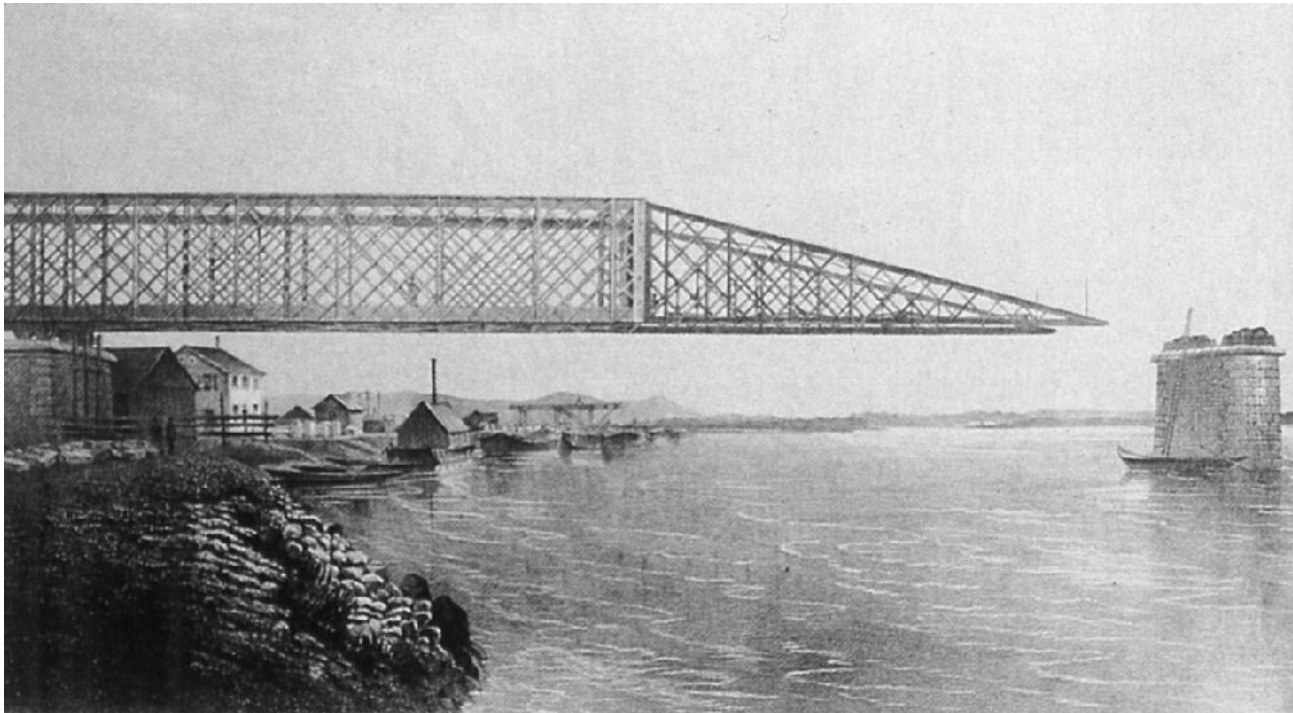


Einrichtungstypen

- **Im Umgang mit Forschungsdaten gibt es mindestens drei Einrichtungstypen:**
 - reine Forschungseinrichtungen
 - Hochschulen
 - Behörden



Vier Brücken zur Forschungsdateninfrastruktur



Vier Brücken zur Forschungsdateninfrastruktur

- **Zwischen den Disziplinen**
 - interdisz. Klimaforschung verwendet Daten anderer Disziplinen
 - übergreifende Metadatenstandards - Übergänge zwischen den Disziplinen
- **Von Big Science zu Small Science**
 - große Systeme müssen auch die Daten der kleinen Projekte aufnehmen
- **Zwischen den Institutionen**
 - Einrichtungen müssen typenübergreifend kooperieren
 - Abstimmung zu Policies und Standards
- **Zwischen technischen Systemen**
 - automatische Metadatenerfassung
 - Aufbereitung für Archivierung
 - direkter Zugriff mit Analyse- und Visualisierungstools



Praktische Ansätze zum Ausbau der Dateninfrastruktur



Policies



Policies

- Zielen auf Publikationskultur / wiss. Anerkennung
- Schaffen Rahmenbedingung für Einsatz von Standards und Kooperationen
- Unterstützen angemessene technische und personelle Ausstattung der Datenkuratierung

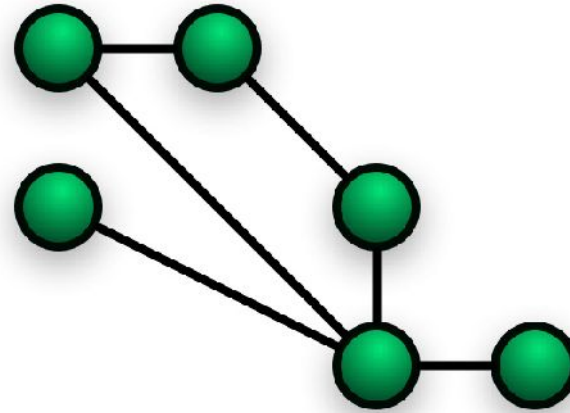


Semantic Web - Linked Data



Semantic Web - Linked Data

- Vernetzung muss koordiniert aber dezentral erfolgen
- "selbstgenerierende Verknüpfung" erforderlich



Fazit

- **Es gibt tragfähige technische und organisatorische Lösungsansätze zur Ausweitung des Forschungsdatenmanagements**
- **Die „datengetriebene Wissenschaft“ ist aus Perspektive des Forschungsdatenmanagements eine *Infrastrukturaufgabe***



Vielen Dank!

Kontakt:

Lars Müller

Fachhochschule Potsdam
Fachbereich Informationswissenschaften
Friedrich-Ebert-Str. 4
14467 Potsdam
Telefon: +49 (331) 580 1533

lars.mueller@fh-potsdam.de

<http://wibaklidama.fh-potsdam.de/>

