# Forschungsdatenmanagment in den (Göttinger) Geowissenschaften

Wo stehen wir und wo wollen wir hin?



– Ein Praxisbericht –



Dr. Bianca Wagner, Geowissenschaftliches Zentrum der Universität Göttingen

6. Workshop der DINI-nestor-AG Forschungsdaten "Kooperationstreffen Forschungsdaten" 13./14.10.2016, Göttingen

# Inhalt

Forschungsfeld Geowissenschaften Forschungsdatenmanagement am GZG

Forschungsdatenmanagement in der Community

Herausforderungen

### Dank

Bernd Leiss Klaus Wemmer Alexander Gehler Puneet Kishor Volker Thiel Gernot Arp Tobias Licha Jens Dierkes Wiebke Warner Geo-Gender-Chancenfond

# Forschungs- (und Spannungs-) feld Geowissenschaften

# Geowissenschaften

Wie war es einmal?

Wie ist es jetzt?

Wie wird es sein?

Erde - Vergangenheit - Erde

- Gegenwart -

Erde

- Zukunft

Was passiert jetzt?



Was passierte?



Was wird passieren?

# Geowissenschaftliches Zentrum Göttingen

Geoinformatik

Pharmazie |

Biologie

Vermessung Struktur-Geophysik geologie Ingenieurwesen Mineralogie Angewandte Material-Geologie Geographie wissenschaften Kristallographie Sedimentologie Physik Klimatologie Isotopengeologie Geobiologie Mikrobiologie Astrophysik Geochemie

Umweltwissenschaften

# Forschungsobjekte









# Forschungsmethoden

"Objekt-" Untersuchung

Experiment

Modellierung

Direkt oder indirekt?

Geometrie: 1D, 2D oder 3D?

Abbildend oder nicht abbildend?

Einmalig oder mehrmals?

Verortbar oder nicht verortbar?

# Externe Geodaten

Landesämter





Weltraumagenturen

# "Stake holders"

Landesämter

Universitäten

Bundesbehörden Geowissenschaftliche Daten

Museen

Industrie

Forschungsinstitute

Amateure

# Forschungsdatenmanagement am GZG

# Datenmanagement

Daten speichern

Daten auswerten

Daten austauschen

Daten bereitstellen

Daten sichern

Daten archivieren

Externe Festplatte Interner Cloud Server

Web-Portal

# Repository

• Thematisches Repository ("Pangäa")

• Repository eines Verlags

 Repository eines anderen Fachbereichs ("Genebank



# Projekte & Aktivitäten

- Sammlungsdatenbanken (Objekte)
- Thematische Datenbanken (Messdaten)
- DFG-Rundgespräch
- Verbundprojekt

#### J. F. Blumenbach

Blumenbach als Wissenschaftler Wissenschaftshistorischer Kontext Anthropologie Internationale Dimension Chronologie Briefregesten

Blumenbachs Werke

Vorlesungen

Projektbeschreibung

Leitung und Mitarbeiter

Publikationen

Tagungen

Links

Kontakt

Interner Bereich

#### Einleitung

In seinem Nachruf formulierte der Göttinger Mediziner und Historiker Karl Friedrich Heinrich Marx (1796-1877) die folgende Einschätzung von Blumenbachs Bedeutung:

An sein Gedächtniss, an seinen Namen knüpfen sich seit mehr als einem halben Jahrhunderte die bedeutungsvollsten Ereignisse dieser Universität, ja der Entwicklungsgang einer der grössten und wichtigsten Wissenschafts-Zweige ist mit dem, was er unternommen, geleistet, gefördert hat, auf das Innigste verschlungen.

Aus der Reihe derjenigen, die mit ihm denselben Weg betreten, gestrebt, geforscht hatten, stand er zuletzt wie eine einsame Säule, wie eine Pyramide der Vorzeit da, zum erweckenden Beispiel uns Jüngeren, wie die Natur zuweilen einer hohen inneren Seelenkraft auch durch Festigkeit und Dauer der äusseren Form das Siegel ihrer Vollendung aufdrücke.



Aus: Zum Andenken an Johann Friedrich Blumenbach: eine Gedächtniss-Rede gehalten in der Sitzung der Königlichen Societät der Wissenschaften den 8. Februar 1840. Göttingen 1840.

Impressum Akademie

#### Über naniweb

naniweb ist eine Erfassungsdatenbank für die Bestände naturwissenschaftlicher Objektsammlungen und ermöglicht das Archivieren, Verwalten und Recherchieren von digitalen Bildund Multimedia-Daten sowie die entsprechende Eingabe von objektbezogenen musealen und fachwissenschaftlichen Metadaten. Da über 200 verschiedene Dateiformate unterstützt werden, kann es zur Erfassung und Verwaltung verschiedenartigster Bestände genutzt werden. Digitale und digitalisierte Aufnahmen von organischen oder mineralischen Fundstücken und Präparaten, wissenschaftliche Bilddokumentation, komplette Präsentationen oder Videos sind nur einige der Möglichkeiten, wie naniweb zur Inventarisierung und Verwaltung des akademischen Erbes eingesetzt werden kann.

naniweb ist komplett web- und browserbasiert und beruht ausschließlich auf Open-Source-Komponenten. Die Systemvoraussetzungen für die Nutzer beschränken sich auf einen Internetbrowser und eine Internetverbindung. Die intuitiv verständliche Handhabung ermöglicht auch für wenig Computer-affine Nutzer den leichten Einstieg.

Etwa 20 verschiedene Datenfelder sind mit (inter-)nationalen Normdaten hinterlegt. Aus der Gemeinsamen Normdatei (GND) der Deutschen Nationalbibliothek werden alle Personen- und Institutionsfelder, Objektkategorien sowie die Schlagwörter gespeist. Internationale Ortsangaben werden aus der freien Ortsdatenbank geonames befüllt, nationale aus der Datenbank der Geographischen Namen 1: 250 000 des Bundesamts für Kartographie und Geodasie. Literaturangaben zum Objekt können mit dem entsprechenden Datensatz im Gemeinsamen Verbundkatalog (GVK) verknüpft werden.

Mit naniweb wird den teilnehmenden Einrichtungen außerdem die Möglichkeit eröffnet, Objekte in das Portal der Universitätssammlungen Göttingen auszuspielen und dort online zu publizieren. Die Daten werden automatisch validiert und durch eine entsprechend angepasste Schnittstelle überführt. Die Weitergabe valider Daten in verschiedenen Formaten nach internationalen Standards erlaubt auch die Übergabe an Portale wie die Deutsche Digitale Bibliothek (DDB) oder Europeana.

#### Ansprechpartner:

Frank Dührkohp
Verbundzentrale des GBV (VZG)
37073 Göttingen
Tel.: (49)551 3910405
E-Mail: duehrkohp@gbv.de

Abteilung Geobiologie, Prof. Reitner., Dr. Gehler

naniweb



Über naniweb

Impressum

https://naniweb.gbv.de

# Geochemische Datenbank

THE CENTRAL ANDES GEOCHEMICAL GR

- Panama Geochemical Database
- Central Andes Geochemical
   Database
- Central Andes time scale map
  Database
- Andes Ignimbrite Database

Abteilung Geochemie, Prof. Wörner

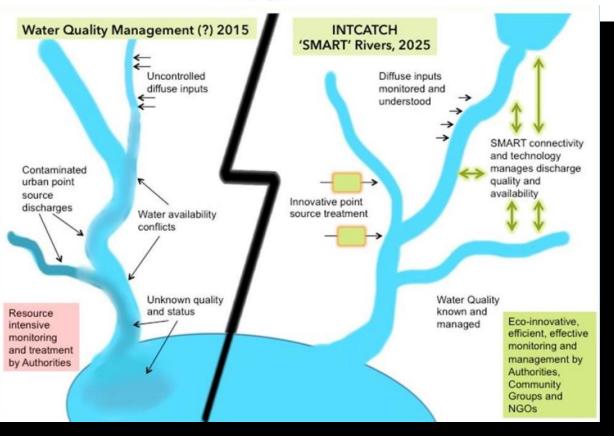
# DFG-Rundgespräch

- Februar 2016
- "hard rock geochemistry"
- EarthChem und assoziierte Datenbanken (z.B. GEOROC 2020)
- Nationales Archiv für Forschungsproben
- Internationale Einbindung
- Weitere nationale Vernetzung

Abteilung Geochemie, Prof. Wörner

# INTCATCH 2020

,SMART' Rivers
Big data analysis
Citizen science



Abteilung Angewandte Geologie, Dr. Licha

# Forschungsdatenmanagement in der Community

# World Wild Web

- "klassische Datenportale" (NIBIS, NASA, ...)
- Spezielle Suchdienste (Geo Leo)
- Thematische Datenbanken (GEOROC, PBD, ...)
- Thematische Datenportale (SAFARI, Macrostrat)
- Internationale Initiativen (EarthCube, IEDA, EGDI)
- Projekte der geologischen Dienste (One Geology)
- Data Publisher (Pangäa)
- Uni-/Industrie-Kooperationen (GeoClass, SAFARI)

Our services are generally open for archiving, publishing, and re-usage of data. The World Data Center PANGAEA is member of the ICSU World Data System.



ALL TOPICS

Search for measurement type, author name, project, taxa,...

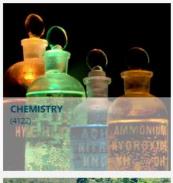
~



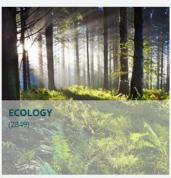






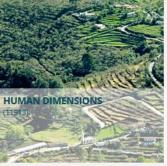






















A P





#### Trusted Data Services for Global Science



#### **Data Portal**

Share ®

The ICSU-WDS Data Portal is a prototype allowing retrieval of datasets from participating WDS Members. As a first step, it provides access to currently available metadata catalogues using international and community agreed standards.

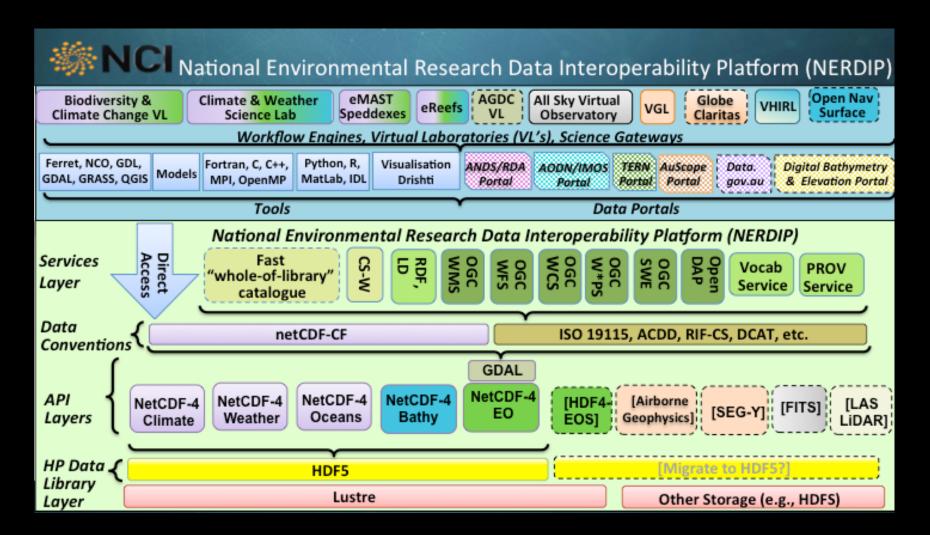
The prototype was started with initial candidates willing to participate in setting up the portal framework. It is currently being updated to include more participating WDS Members and will be relaunched.

Participating WDS Members need to implement current standards in the field of Spatial Data Infrastructures (SDI) and can generally use the resulting interoperability to network also with other communities and portals such as the Ocean Data Portal (IODE), the Global Change Master Directory (GCMD), or the Global Earth Observations System of Systems (GEOSS).

The prototype is hosted by PANGAEA (WDS Regular Member).

Find Data		
Query (syntax help):		
Anywhe	ere in data description:	
Dataset	citation:	
Principa	al investigator:	
Publish	er/Source:	<b>~</b>

## Nationale Plattform



# Nationale Plattform

Wyborn & Evans (2015):

"Integrating 'Big' Geoscience Data into the Petascale National Environmental Research Interoperability Platform (NERDIP): successes and unforseen challenges"

- "heterogeneous data"
- "scarcity of tools and people skills"
- "lingering impact of legacy practices and standards"

"People and tools need to be brought to the data"

# AGU Fall Meeting 2016

#### IN12A

Big Data Analytics I

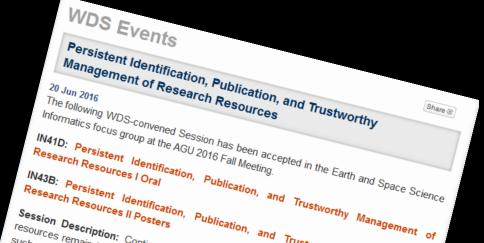
#### **IN13E:**

Exploring Value in Our Research Products: Approaches to Assessment of Research Products Such as Data Sets, Metadata, and Research Software within a Policy Framework that Supports Scalable, Reusable, Sustainable, and Open-Access Outcomes for Research and Societal Benefit

#### IN41D:

Persistent Identification, Publication, and Trustworthy Management of Research

Resources I



# Crowd-sourcing/ Citizen Science



# Crowd-sourcing/ Citizen Science

# New app turns smartphones into worldwide seismic network

By Robert Sanders, Media relations

FEBRUARY 12, 2016













C Berkeley scientists today are releasing a free Android app that taps a smartphone's ability to record ground shaking from an earthquake, with the goal of creating a worldwide seismic detection network that could eventually warn users of impending jolts from nearby quakes.



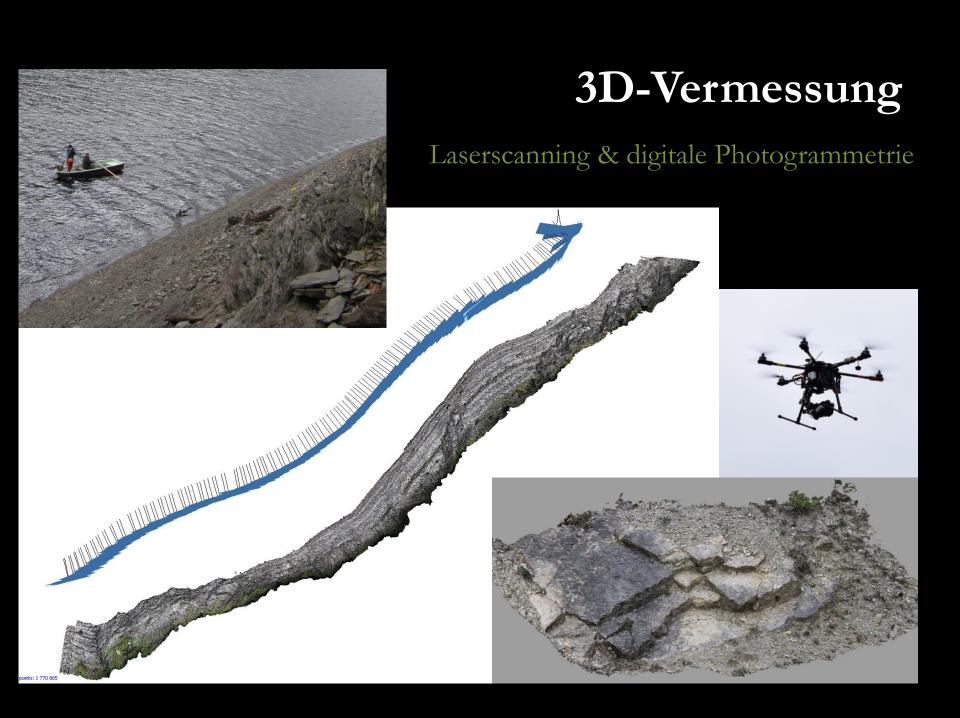
# Data Mining

# GeoDeepDive



# GeoDeepDive. Bringing dark data to light.

Many geological sample-based data are hard-earned and often unique. Their primary mode of dissemenation has traditionally been via publication, which makes aggregating data synthetic analyses prohibitively time consuming. We are building a new type of digital library to support machine reading and learning approaches to geological data extraction.



# Sicht aus der Praxis I

- viele heterogene Weblösungen
- monothematische Weblösungen
- fehlende Weblösungen

• Initiativen auf verschiedenen Ebenen

- Zunahme der Recherchezeit
- Datamanagement als zusätzliche Arbeitsaufgabe

# Sicht aus der Praxis II

- Zurückhaltung gegenüber "open source"-Gedanken
- heterogene Datenformate & Softwarepakete
- umfangreiche Metadaten (Aufbereitung, ...)
- Auftragsdaten
- verschiedenste Projektideen

# "Lokale" Herausforderungen

- Aufarbeitung historischer bzw. analoger Daten
- Weiterführung älterer Datenbanken
- Weiterführung von eigener Software
- Bewertung und Pflege von Projektdatenportalen
- Konsistente Pflege der eigenen "Forschungsmetadaten"
- Unklarheiten zu Urheberrechten

Umdenken
Austausch
Vernetzung
Kooperation

Information
Beratung
Schulung
Finanzierung

## **Fazit**

#### PID

meta-analyses
text- and data-mining
cross-disciplinary
multiple source input
Spatiotemporal Big Dat
Data-intensive Science

Research Interoperability Data Platform analysis ready data cubes, data arrays

Ein Stein ist immer noch ein Stein –

und ein Buch ist immer noch ein Buch.



Suche

Thematische Suche

e-Zeitschriften

Geo-Guide

GEO-LEO e-docs

#### Suche nach Informationsquellen zum System Erde und Weltall

z.B. Venustransit oder Vulkan\*

SUCHEN

erweiterte Suche

keine Suchabfrage

GEO-LEO ist ein zentrales Portal für die Recherche nach Informationsquellen zum System Erde und Weltall, das mit GEO-LEOe-docs auch eine Publikationsplattform für elektronische Volltexte bietet.

GEO-LEO ist ein Gemeinschaftsprodukt der Universitätsbibliothek der TU Bergakademie Freiberg (UBF) und der Niedersächsischen Staats- und Universitätsbibliothek Göttingen (SUB), die mit dieser Metasuchmaschine ihre fachspezifischen Literaturbestände und digitalen Medien sowie die anderer fachrelevanten Anbieter für die Forschung in Deutschland erschließen. Neben der einfachen und erweiterten Suche gibt es auch thematische Sucheinstiege. Von den einzelnen Treffern aus haben Sie direkten Zugriff auf digitale Versionen oder auf die Bestelloption via Subito oder die Online-Fernleihe der Bibliotheksverbünde.

#### OneGeology Portal

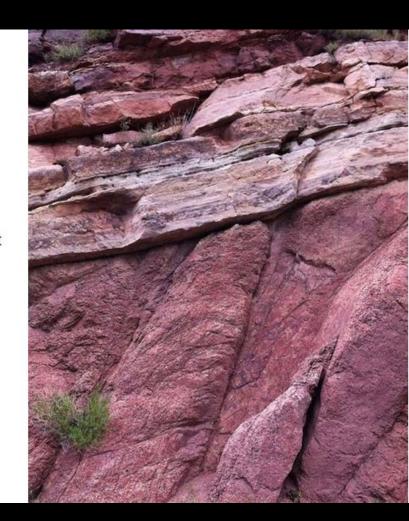


# Macrostrat

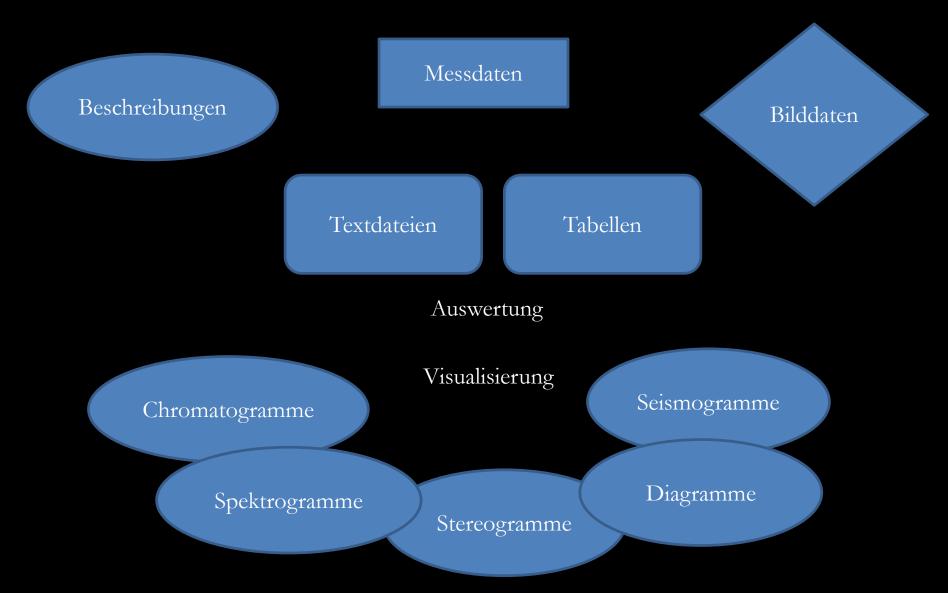
# Macrostrat. Database for geological data.

Macrostrat was originally built to test hypotheses for the correlation between the sedimentary rock and fossil records. It is spatially explicit and chronostratigraphically focused and serves as a scaffolding for organizing and analyzing geological data of all types.

Learn more »



# Typische Datentypen





### WP1. Management, Coordination and Reporting to the EC (Leader: UBRUN)

- T1.1 Management and Reporting to the EC
- T.1.2 Communication among partners
- T.1.3 Managing the legal and ethical obligations
- T.1.4 Risk management

# WP2. Innovative approach towards water monitoring (Leader: ESAL)

- T.2.1 Monitoring strategies and tools
- T2.2 Social dimensions
- T2.3 Innovative monitoring strategies
- T2.4 Set-up of DSS

# WP3. Innovative sensors and technologies, their integration and validation (Leader: BOKU)

- T3.1 Requirements for monitoring
- T3.2 Heavy metal sensors
- T3.3 Validation of sensors
- T3.4 Validation of integrated system
- T3.5 Biosensors
- T3.6 Mobile genomics laboratory

# WP4. Autonomous and radio controlled boats for water quality monitoring (Leader: UNIVR)

- T4.1 Optimization of the boats
- T4.2 Controlling the boats
- T4.3 Validation of the boats

# WP5. Innovative stormwater treatment and management (Leader: UVIVR)

- T5.1 CSO treatment system
- T5.2 Validation of the CSO system
- T5.3 Treatment systems in the UK
- T5.4 Stormwater management plans

# WP6. Development and validation of water quality model and DSS (Leader: TCH)

- T6.1 Hydrodynamic water quality model
- T6.2 Integrate water quality model in DSS
- T6.3 Adaptation of DSS for UK

# WP7. Data management systems and visualization techniques (Leader: ALG)

- T.7.1 BlueGate customization
- T.7.2WAIS Cloud System
- T.7.3 Connection to DSS
- T.7.4 Web and mobile applications

### IntCatch