

# Forschungsdaten – Was kann eine Bibliothek tun?

Dr. Dominik Schmitz  
RWTH Aachen Hochschulbibliothek

06.03.2013

# Was sind Forschungsdaten?

„Unter Forschungsdaten sind [...] digitale und elektronisch speicherbare, Daten zu verstehen, die im Zuge eines wissenschaftlichen Vorhabens z.B. durch Quellenforschungen, Experimente, Messungen, Erhebungen oder Befragungen entstehen.“ (DFG, 2010)



... jeder Output einer Forschungstätigkeit, der keine Publikation ist.  
(Rombouts & Prinčič, 2010)

# Forschungsdaten ≠ Forschungsinformationen

- Forschungs**primär**daten als Arbeitsmittel/-gegenstand der Forschenden
  - Beobachtungsdaten z.B. Sensoren
  - Experimentelle Daten
  - Simulationsdaten
  - Abgeleitete oder kombinierte Daten
  - Referenzdaten
- **Nicht** „Kennzahlen“ zu Forschungsaktivitäten (= Forschungsinformationen)
  - Kerndatensatz Forschung
  - EuroCRIS (European Current Research Information System)
  - CERIF (Common European Research Information Format)

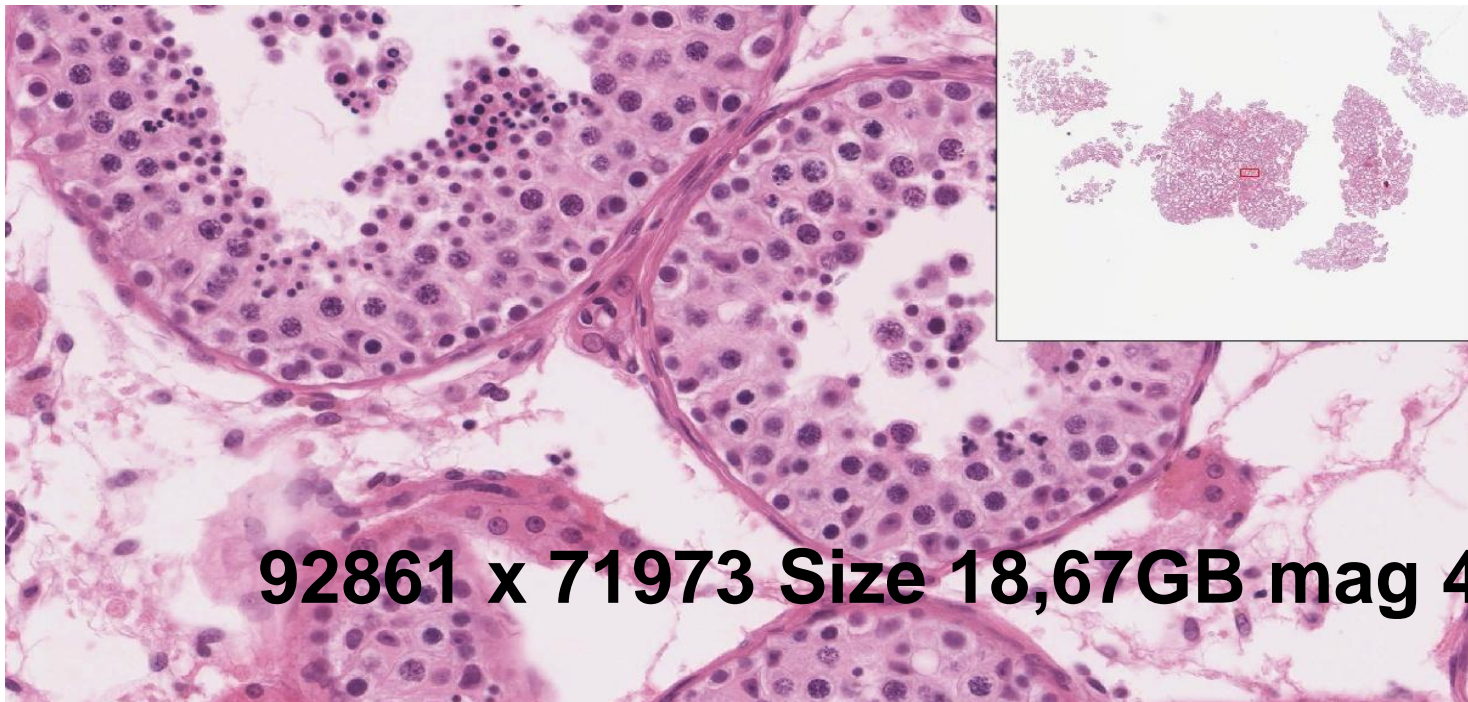
- Histologische Mikroskopaufnahmen: Wg. Datenvolumen nur ein Verweis auf das eigentliche Artefakt

→ hohe Auflösung (ca. 25.000 Pixel × 25.000 Pixel)

→ Größe der Einzelbilder (ca. 8-10 GB pro Bild)



DFG ProjektRepository



**92861 x 71973 Size 18,67GB mag 40X**

**Warum sollte man sich  
für Forschungsdaten interessieren?**

## Digitale Forschungsinfrastruktur und Bibliotheken

### Relevante Aspekte für Bibliotheken

- Erstberatung an Universitäten und bei Forschungsinstitutionen
- Beratung für die Einrichtung und den Betrieb von Virtuellen I
- Beschreibung und Erschließung von Informationsressourcen
- Interoperabilität
- Professionelles Informationsmanagement
- Archivfunktion/Repositorien
- Elektronisches Publizieren
- Digitale Langzeitarchivierung
- IT-Werkzeuge

## Bibliothekartag 2012

DARIAH-DE 101. Bibliothekartag, Hamburg

## Rolle der Bibliothek beim Thema Forschungsdaten

- Bibliothek als *ein* Akteur im Zusammenspiel mit *anderen*.
- Bibliothek als *ein* zentraler Anlaufpunkt („Verteiler“) zum Thema.
- Klassische Kompetenzen nutzen:  
Publikation und offener Zugang  
Beratung  
Information



## Jahrestagung 2012:

Forschung – Information –  
Infrastruktur: Bausteine  
für Open Science  
insbes. Vortrag Prof. Thomas,  
Geoforschungszentrum Potsdam

## AG Elektronisches Publizieren / Forschungsdaten (Pampel, Kindling)

## RESEARCH DATA ALLIANCE

Researchers around the world  
sharing and using research data  
without barriers

<http://rd-alliance.org>

Launching Event: 18.-20.03.2013, Göteborg

Candidate Working Groups on

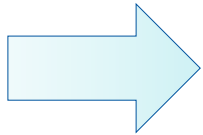
Harmonization and Use of PID Information Types, UPC Code for Data, Data Type Registries, Metadata, Pub/Data Citation/Linking, Data Foundation and Terminology, Practical Policy, Legal Interoperability, Defining Urban Data Exchange for Science, The Engagement Group, Marine Data Harmonization, Repository Audit and Certification, Preservation e-Infrastructure, Contextual Metadata

- DFG, Leitfaden für Antragsteller (2010)
  - „[...] legen Sie bitte dar, welche Maßnahmen ergriffen wurden bzw. während der Laufzeit des Projektes getroffen werden, um die Daten nachhaltig zu sichern und ggf. für eine erneute Nutzung bereit zu stellen [...]“
- Wissenschaftsrat 2012: Empfehlungen zur wissenschaftlichen Informationsinfrastruktur
  - Richtlinien für Datenmanagement, Konventionen für Metadaten (als fester Teil Projektmanagement, Personal!)
  - **Anerkennung als eigenständige Forschungsleistung**
  - Langfristiger Zugang, Kriterien für Archivierungsentscheidung
- Weitere Empfehlungen/Vorgaben
  - Allianz der Wissenschaftsorganisationen, National Science Foundation, OECD, EU, Zeitschriften: Nature, PloS



## Forschungsdaten...

- sind Quelle und zentrales Objekt wissenschaftlicher Forschung,
- werden nur eingeschränkt (nach-)genutzt oder zugänglich gemacht – auch schon innerhalb eines Lehrstuhls/Instituts/...,
- und häufig nicht langfristig archiviert.



## **Beratungs-/Unterstützungsbedarf**

Datenmanagement, Qualitätskontrolle,  
Nachnutzung, Archivieren und Teilen

**Was gibt es zu tun?**

Daten-  
management

Recherche

Repository

Daten  
<->  
Publikation

## Forschungsdaten

Beratung

Metadaten

Daten-  
publikation

Archivierung

...

## About

[...]

The goal of re3data.org is to create a global registry of research data repositories. The registry will cover research data repositories from different academic disciplines.

re3data.org will present repositories for the permanent storage and access of data sets to researchers, funding bodies, publishers and scholarly institutions. In the course of this mission re3data.org aims to promote a culture of sharing, increased access and better visibility of research data. In the first phase of the project the following tasks are prioritized:

- the conception and construction of a web-based registry of research data repositories,
- the definition of selection criteria of research data repositories,
- the formulation of a metadata schema to describe research data repositories.

According to the schedule the registry will go live in autumn 2012.

re3data.org is funded by the [German Research Foundation DFG](#) in the period 2012 to 2014.

# PANGAEA®

Data Publisher for Earth & Environmental Science



**All** Water Sediment Ice Atmosphere

[Help](#) [Advanced Search](#) [Preferences](#) [more...](#)

[About](#) – [Submit Data](#) – [Projects](#) – [Software](#) – [WDC-MARE](#) – [Contact](#)

This work is licensed under a [Creative Commons License](#)

“The information system PANGAEA is operated as an Open Access library aimed at archiving, publishing and distributing georeferenced data from earth system research. The system guarantees long-term availability of its content through a commitment of the operating institutions.”

Not logged in ([log in](#) or [sign up](#))

& Environmental Science

Always quote citation when using data!

[Show Map](#) [Google Earth](#) [RIS](#) [BrisTex](#)

Table 1) Thermoluminescence datings of several sites in Schleswig-Holstein, north 4/PANGAEA.785372, ks, Leszek; Piotrowski, Jan A; gen; Fedorowicz, Staniskaw; Butrym, oluminescence indications of the Middle n) glaciation in northwest Germany. doi:10.2312/meyniana.1995.47.69

ce (TL) datings of glacial deposits from in northwest Germany (Schleswig-Holstein) dle Weichselian Glaciation in this region. imply an ice sheet at about 70-40 ka B.P. in this part of Europe has not been The reliability of the TL dates was cross- ologies; the dating was performed in two



independent laboratories. These dates and their interpretations, however, must still be confirmed both geologically and paleontologically on suitable sites where glacial sediments are sandwiched between Eemian deposits and interstadial deposits older than the Late Weichselian.

**Coverage:** *Median Latitude:* 54.486367 \* *Median Longitude:* 9.753867 \* *South-bound Latitude:* 54.154600 \* *West-bound Longitude:* 9.308400 \* *North-bound Latitude:* 54.807300 \* *East-bound Longitude:* 10.128900

**Event(s):** **Brueggerholz** \* *Latitude:* 54.184900 \* *Longitude:* 10.089900 \* *Location:* Schleswig-Holstein, Germany

**Daen\_Woehld** \* *Latitude:* 54.483800 \* *Longitude:* 10.128900 \* *Location:* Schleswig-Holstein, Germany

**Ellund** \* *Latitude:* 54.795800 \* *Longitude:* 9.308400 \* *Location:* Schleswig-Holstein, Germany

**Parameter(s):**

#	Name	Short Name	Unit	Principal Investigator	Method	Comment
1	Event label	Event				Metadata
2	LATITUDE	Latitude				Geocode
3	LONGITUDE	Longitude				Geocode
4	Area/locality	Area		Marks, Leszek		
5	Sediment type	Sediment		Marks, Leszek		
6	Sample ID	ID		Marks, Leszek		UG = Gdansk Lab., Lub = Lullin Lab.
7	Dose rate	D s	mGy/a	Marks, Leszek		d alpha
8	Dose rate	D s	mGy/a	Marks, Leszek		d beta
9	Dose rate	D s	mGy/a	Marks, Leszek		d gamma

### In this section

- [Briefing Papers](#)
- [How-to Guides](#)
- [Curation Reference Manual](#)
- [Curation Lifecycle Model](#)
- [Policy and legal](#)
- [Data Management Plans](#)
- [Tools](#)
- [Case studies](#)
- [Repository audit and assessment](#)

### Standards

[Disciplinary Metadata](#)

DIFFUSE

- [Publications and presentations](#)
- [Roles](#)
- [Curation journals](#)
- [Informatics research](#)
- [External resources](#)

## Disciplinary Metadata

While data curators, and increasingly researchers, know that good metadata is key for research data access and re-use, figuring out precisely what metadata to capture and how to capture it is a complex task. Fortunately, many academic disciplines have supported initiatives to formalise the metadata specifications the community deems to be required for data re-use. This page provides links to information about these disciplinary metadata standards, including profiles, tools to implement the standards, and use cases of data repositories currently implementing them.

For those disciplines that have not yet settled on a metadata standard, and for those repositories that work with data across disciplines, the General Research Data section links to information about broader metadata standards that have been adapted to suit the needs of research data.

### Search by Discipline



Biology



Earth Science



General Research Data



Physical Science



Social Science & Humanities

# Was kann (m)eine Bibliothek tun?

# RWTH Aachen: DFG-Projekt „ProjektRepository“

- Niederschwellige, webbasierte Kooperationsinfrastruktur
- Pandisziplinär: Bibliothek, Rechenzentrum, drei Institute

Pathologie



Wasserbau/  
-wirtschaft

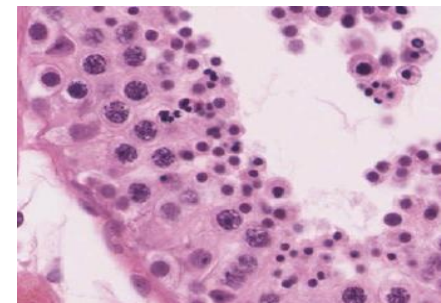
- Rolle der Bibliothek Stadtbaugeschichte
  - Fachspezifische Individualanalyse Daten
  - Identifizierung relevanter Metadatenschemata, Thesauri, Klassifikationen, Normdaten
  - Integration in Pilot-Forschungsumgebung (MS SharePoint)
  - Beratung zu Persistent Identifier, Repositories, ...



# Beispiel: Pathologie

## Eigenes Metadatenschema

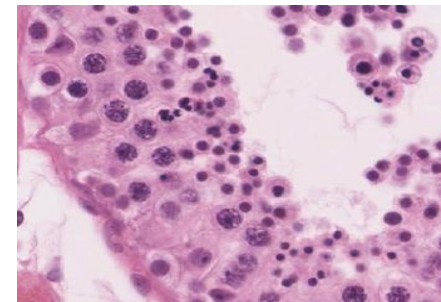
- Aktuell umfasst das Metadatenschema 26 Felder
- Dublin Core für pandisziplinäre Datenfelder
  - Krankheitsgruppe (ICD-10)
  - Fach
  - Topographiegruppe
  - Beobachtung Organsystem
  - Diagnose
  - Lokalisation /  
Tumorlokalisierung des  
Primärtumors (ICD-O-3)
  - Tumormorphologie (ICD-O-3)
  - BI-RADS
  - B-Klassifikation
  - Seite
  - Alter
  - Geschlecht
  - Färbung
  - Immunhisto
  - Verschlagwortung nach  
MESH (DC.Subject)
  - Vergrößerung der Histologie
  - 10 weitere organisatorische  
und techn. Metadatenfelder



# Beispiel: Pathologie

## Eingesetzte Thesauri

- ICD-O: duale Klassifikation
- ICD-10: Klassifikationssystem: Konkordanz zu GKD
- MeSH (Medical Subject Headings)-Codes zur Anbindung an die zentrale medizinische Publikationsdatenbank PubMed
- BI-RADS Klassifikation: Mammographiebefundung
- B-Klassifikation: Klassifikation von Mammastanzbiopsien



Firefox

Demonstrator - Datenblattansicht

Websiteaktionen

Durchsuche

Speichern Abbrechen

Einfügen

Ausschneiden

Kopieren

Element löschen

Datei anfügen

Rechtschreibung

Comit ausführen

Zwischenablage

Aktionen

Rechtschreibung

ProjectRepository

Homepage

ProjectRepository

Bilder

Makro

Histo

VA KM 1

Bibliotheken

Websiteseiten

Freigegebene Dokumente

Filme Angios

Folien

Listen

Kalender

Aufgaben

Demonstrator

Diskussionen

Teamdiskussion

Papierkorb

Gesamter Websiteinhalt

E-N

1029

URL (Pathologie)

Geben Sie die Webadresse ein: (Klicken Sie hier zum Testen)

XX

Geben Sie die Beschreibung ein:

Tumormorphologie

8500/2 Nichtinfiltrierendes intraduktales Karzinom o.n.A.; k

BI-RADS

B-Klassifikation

Alter

Seite

Färbung

Immunhisto

Vergrößerung

Title

Vorschläge

9301/0 Kalzifizierende odontogene Zyste (Kalzifizierende odontogene Zyste)

9340/0 Kalzifizierender epithelialer odontogener Tumor (Kalzifizierender epithelialer odontogener Tumor)

9131/0 Kapilläres Hämangiom (Kapilläres Hämangiom) [Tumormorphologie (ICD-O)]

9171/0 Kapilläres Lymphangiom (Kapilläres Lymphangiom) [Tumormorphologie (ICD-O)]

9140/3 Kaposi-Sarkom (Kaposi-Sarkom) [Tumormorphologie (ICD-O)]

8934/3 Karzinofibrom (Karzinofibrom) [Tumormorphologie (ICD-O)]

8240/1 Karzinoidtumor fraglicher Dignität (Karzinoidtumor fraglicher Dignität)

8240/3 Karzinoidtumor o.n.A. (Karzinoidtumor o.n.A.) [Tumormorphologie (ICD-O)]

8390/3 Karzinom der Hautanhangsgebilde (Karzinom der Hautanhangsgebilde)

8941/3 Karzinom in pleomorphem Adenom (Karzinom in pleomorphem Adenom)

8035/3 Karzinom mit osteoklastenähnlichen Riesenzellen (Karzinom mit osteoklastenähnlichen Riesenzellen)

8035/6 Karzinom mit osteoklastenähnlichen Riesenzellen, Metastasen (Karzinom mit osteoklastenähnlichen Riesenzellen, Metastasen)

8035/9 Karzinom mit osteoklastenähnlichen Riesenzellen, unbestimmtes Organ (Karzinom mit osteoklastenähnlichen Riesenzellen, unbestimmtes Organ)

8010/3 Karzinom o.n.A. (Karzinom o.n.A.) [Tumormorphologie (ICD-O)]

8010/9 Karzinomatose (Karzinomatose) [Tumormorphologie (ICD-O)]

8589/3 Karzinome mit thymusähnlichen Anteilen (Karzinome mit thymusähnlichen Anteilen)

8010/6 Karzinom-Metastase o.n.A. (Karzinom-Metastase o.n.A.) [Tumormorphologie (ICD-O)]

8980/3 Karzinosarkom o.n.A. (Karzinosarkom o.n.A.) [Tumormorphologie (ICD-O)]

9121/0 Kavernöses Hämangiom (Kavernöses Hämangiom) [Tumormorphologie (ICD-O)]

9172/0 Kavernöses Lymphangiom (Kavernöses Lymphangiom) [Tumormorphologie (ICD-O)]

Tumormorphologie

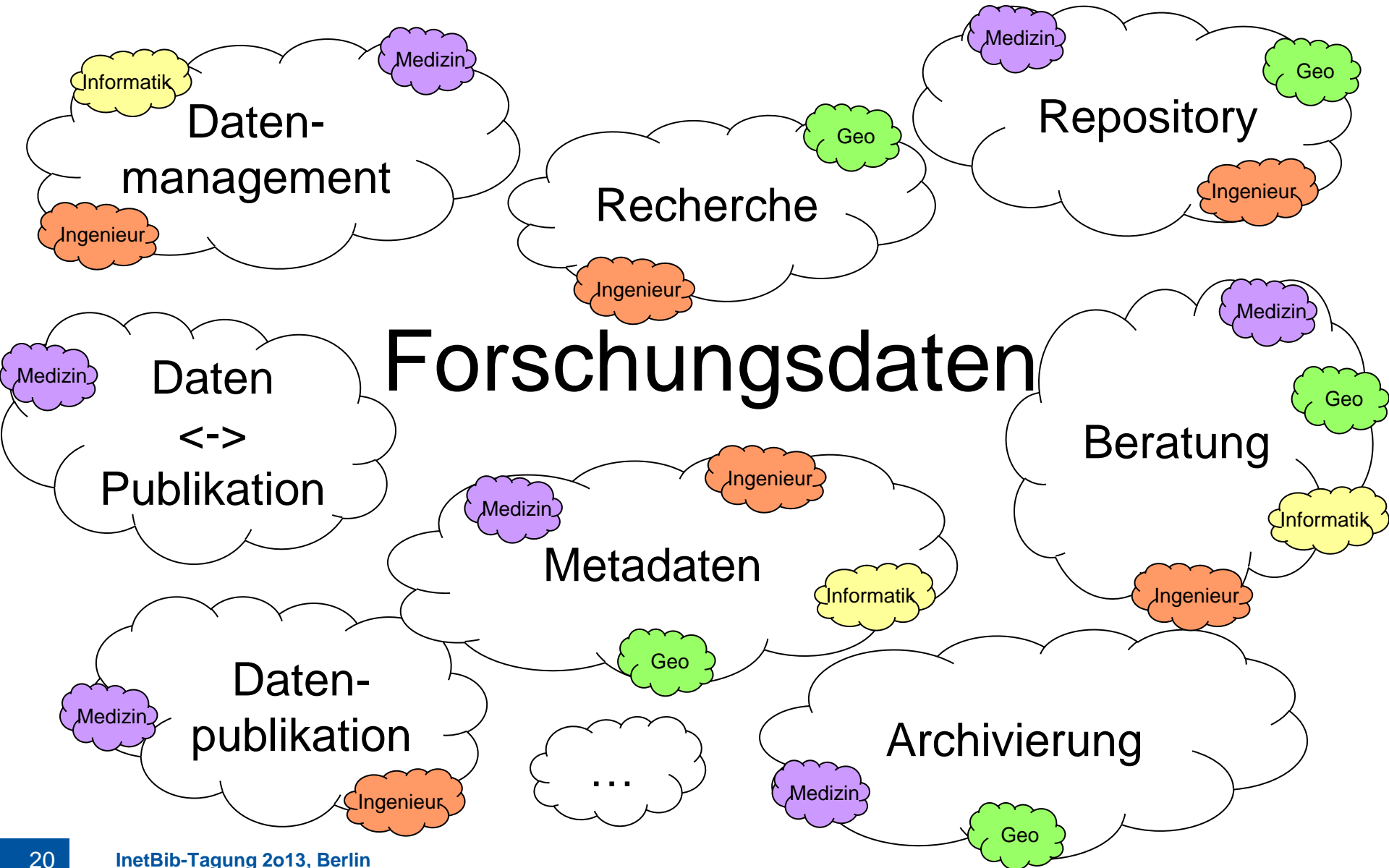
8500/2 Nichtinfiltrierendes intraduktales Karzinom o.n.A.;

BI-RADS

BI-RADS 4b: mittelgradige Wahrscheinlichkeit

gewebsnekrose, Zyste

# Fachspezifisch!



- Individueller Beratungsbedarf
- Fachspezifische Strukturen vor Ort, eingebunden in (inter-)nationale Netzwerke
  - Rechenzentrum: Keine Fachreferenten/-innen
  - Fachgesellschaften: Nicht vor Ort!
- Aufgaben
  - Informieren, ggf. Problembewusstsein wecken
  - Konzeptionierung und Anwendung von Metadatenschemata, Sicherung der Interoperabilität und Standardisierung
  - Repositorien, Datenarchive und das Management von Informationsressourcen
  - Veröffentlichung, Zitierbarkeit der Daten
  - Sicherung der digitalen Langzeitverfügbarkeit

- Individueller Beratungsbedarf
- Fachspezifische Strukturen vor Ort,

## Aber

Sehr breites Wissensspektrum erforderlich!

→ „Resource Sharing“

Einrichtungsübergreifende Kooperationen erforderlich,  
z.B. in *Deutsche Initiative für Netzwerkinformation (DINI)*,  
– Bibliotheken – Rechenzentren – Fachgesellschaften –

- U.a. DINI AG Kompetenzzentrum interoperable Metadaten (KIM),  
Unter-AG zu Metadaten für *Forschungsdaten* im Aufbau  
25.-26.3., Mannheim: Workshop KIM, Session Forschungsdaten

## ■ Kontakt

- Forschungsdatenteam, RWTH Aachen Hochschulbibliothek
- [forschungsdaten@bth.rwth-aachen.de](mailto:forschungsdaten@bth.rwth-aachen.de)

## ■ DINI AG KIM, Unter-AG „Forschungsdaten“

- <https://wiki.d-nb.de/display/DINIAGKIM/Forschungsdaten>
- Mailingliste

## ■ Termine

- 18.-20.03.2013, Göteborg: Launching Event  
Research Data Alliance (<http://rd-alliance.org>)
- 25.-26.03.2013, Mannheim: DINI-Workshop KIM  
<https://wiki.d-nb.de/display/DINIAGKIM/KIM+WS+2013>