

Forschungsdaten

Bachelor Informationsmanagement
Modul Digitale Bibliothek (SS 2014)

Dr. Jakob Voß


2014-05-26

Forschungsdaten

Forschungsdaten sind Daten, die im Zuge wissenschaftlicher Vorhaben z.B. durch Digitalisierung, Quellenforschungen, Experimente, Messungen, Erhebungen oder Befragungen entstehen

Allianz der deutschen Wissenschaftsorganisationen¹

¹http:

[//www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/](http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/) 

Research Data Alliance (RDA)

- ▶ Gegründet 2013 von Wissenschaftsorganisationen aus EU, USA und Australien
- ▶ Schaffung von einheitlichen Standards über Fach- und ländergrenzen

<https://rd-alliance.org/>

Warum Forschungsdaten?

- ▶ Rasanter Anstieg von datenbasierter Forschung (“Data Science”): Messwerte, Erhebungen, Analysen, Bilder, Karten...
- ▶ Rasanter Anstieg des Umfangs der Daten
- ▶ Daten sind meist nicht mehr in Forschungspublikationen direkt enthalten (z.B. als einfache Tabelle) sondern nur noch als Diagramme und Interpretationen

Beispiele

`http://dx.doi.org/10.1594/PANGAEA.125848`

`https://www.youtube.com/watch?v=hrHM3bUym3g`

Publikation von Forschungsdaten

- ▶ Umfang inzwischen kein Problem mehr
- ▶ Bislang nicht so akzeptiert und honoriert wie Fachartikel
- ▶ Unterschiedliche Standards je nach Fachdisziplin
- ▶ Förderung von Forschungsdaten-Infrastrukturen

ij Welche weiteren Hemmnisse mag es geben, Forschungsdaten zu veröffentlichen ?!

Nachnutzung von Forschungsdaten

- ▶ Erklärtes Ziel
- ▶ Verschiedene Auffassungen von Nachnutzung

Interne Nachnutzung Datenautoren sowie ausgewählte Kollegen
und Mitarbeiter

Externe Nachnutzung Fachöffentlichkeit

Was sind Forschungsdaten für Bibliotheken?

Die drei Vorstellungen von Daten

Daten als Fakten : das wird aus den Forschungsdaten erhofft Daten

als Beobachtungen : das sind Forschungsdaten als Messungen,

Berechnungen, Aufzeichnungen... wenn sie erzeugt werden Daten

als Dokumente : das sind Forschungsdaten für digitale Bibliotheken

Erschließung von Forschungsdaten: Was?

- ▶ Metadaten zur Findbarkeit (*discovery metadata*)
 - ▶ Allgemeine Beschreibung
 - ▶ Möglichst einheitlich (TODO: FORMATE?)
- ▶ Metadaten zur Nachnutzung (*re-use metadata*)
 - ▶ Fachliche Beschreibung und Tiefenerschließung
 - ▶ Fachspezifisch sehr unterschiedlich

ij Wo liegen die Grenzen zwischen beschreibenden Metadaten, Tiefenerschließung und den Daten selber ?!

Grenzfall: Nano- und Atomare Publikationen

- ▶ einzelne Aussagen
- ▶ einzelne Beispiele
- ▶ einzelne Textbausteine
- ▶ einzelne Annotationen oder Verknüpfungen
- ▶ ...

Nicht nur im wissenschaftlichen Bereich. Vgl Simon Appleby:
Dropping the A-Bomb on publishing. 14.2.2014

Fachspezifische Standards

- ▶ durch Fachwissenschaftler festgelegt
- ▶ möglichst
 - ▶ einheitlich
 - ▶ gut dokumentiert
 - ▶ kontrolliert

Beispiel: <http://www.biosharing.org>

Erschließung von Forschungsdaten: Wie?

- ▶ Sammlung und Erschließung in Forschungsdaten-Repositories
- ▶ Verbreitung der Metadaten mittels üblicher Verfahren wie OAI-PMH
- ▶ Identifizierung der Forschungsdatensätze mit persistenten Identifiern (üblicherweise DOI)

Forschungsdaten-Repositories und Verzeichnisse

- ▶ Verzeichnisse
 - ▶ DataCite
 - ▶ R3data (2012-2015)
 - ▶ DataBib (2012-2015)
- ▶ Repositories
 - ▶ Fachspezifische Entwicklungen
 - ▶ CKAN (viele verschiedene Installationen)

Beispiele

- ▶ Wissenschaftliche Repositories
 - ▶ Forschungsdaten im engeren Sinne (Wissenschaft, frei & unfrei)
 - ▶ siehe z.B. <http://www.biosharing.org/biodbcore/>
- ▶ Freie Daten / Open Data
 - ▶ Aus dem öffentlichen Sektor
 - ▶ data.gov
 - ▶ data.gov.uk
 - ▶ datahub.io
 - ▶ Community-Projekte
 - ▶ Wikidata
 - ▶ MusicBrainz
 - ▶ IMDb
 - ▶ ...

- ▶ Gegründet 2009 von mehreren führenden Forschungsbibliotheken
- ▶ Derzeit 22 Mitglieder (British Library, TU Delft, TIB Hannover...) und weitere angeschlossene Einrichtungen
- ▶ Fast 300 angeschlossene Datenlieferanten (*data centers*)
- ▶ Über 3,2 Millionen DOIs
- ▶ Koordination durch die TIB Hannover

<http://search.datacite.org>

<http://oai.datacite.org>

DataCite Metadatenschema

`http://schema.datacite.org`

Notwendige Felder

- ▶ Identifier (**DOI** & ggf. weitere)
- ▶ **Autoren** (ggf. mit Identifier)
- ▶ Titel (**Original**, Untertitel, Übersetzung)
- ▶ **Herausgeber**
- ▶ Datum (**Jahr** u.A. Veröffentlichung, Gültigkeit, Erstellung...)

DataCite Metadatenschema

Empfohlene Felder

- ▶ Beitragende (ggf. mit Identifier und konkreten Rollen)
- ▶ Thema (ggf. mit Identifier)
- ▶ Publikationstyp (ggf. nach CASRAI²)
- ▶ Verknüpfungen zu anderen Datensätzen (Teile, Nachfolger, Varianten...)
- ▶ Beschreibung(en)
- ▶ Geokoordinaten

Weitere Felder

- ▶ Mime-Type
- ▶ Umfang
- ▶ Version
- ▶ Lizenz

²<http://dictionary.casrai.org/research-personnel-profile/contributions/outputs/publications>

Übung (in Gruppen)

Wieviel Prozent der Einträge in DataCite sind Texte oder haben keinen Publikationstyp?

Zur Beantwortung unserer Forschungsfrage

- ▶ Welche Daten werden benötigt?
- ▶ Wo sind die Daten vorhanden?
- ▶ Wie kommen wir an die Daten?
- ▶ Wie werten wir die Daten aus?

Zur Beantwortung unserer Forschungsfrage

- ▶ Welche Daten werden benötigt?
 - ▶ Anzahl der Einträge in DataCite
(Gesamt, *T*ext, *O*hne Publikationstyp)
- ▶ Wo sind die Daten vorhanden?
 - ▶ In den Metadaten bei DataCite
- ▶ Wie kommen wir an die Daten?
 - ▶ OAI oder Suche bei DataCite
- ▶ Wie werten wir die Daten aus?
 - ▶ $(T + O)/G$

Bibliotheken und Forschungsdaten

- ▶ Bibliotheken sind Infrastruktureinrichtungen
- ▶ Sammeln, Erschließen, Zugänglich machen
- ▶ Erfassung von Forschungsdaten als Dokumente

Beispiel: ETH Zürich

Fachstelle Digitaler Datenerhalt an der ETH-Bibliothek

<http://www.library.ethz.ch/Digitaler-Datenerhalt>

- ▶ Dauer- oder zeitweise Aufbewahrung von Daten
- ▶ Bei Bedarf Zugriffskontrolle
- ▶ Veröffentlichung mit DOI
- ▶ Planung von Massnahmen zur Migration in lesbare Formate
- ▶ Hilfe und Beratung bei der Erschließung von Daten
- ▶ Hilfe und Beratung bei Anforderungen von Förderorganisationen

Alternative: internationale, fachspezifische Datenrepositories

Weitere Informationen zu Forschungsdaten

- ▶ Wiki zu Forschungsdaten
<http://http://www.forschungsdaten.org/>
- ▶ Aktuelle Publikationen zu Forschungsdaten
- ▶ Selbstdarstellung von Projekten zu Forschungsdaten