
Wohin mit meinen Daten?

Niederschwellige Archivierung in den Sozial- und Wirtschaftswissenschaften

Mathis Fräßdorf

Berlin, WGL, 15.09.2014

Agenda

- 1 Warum überhaupt Bereitstellung von Forschungsdaten?
- 2 SowiDataNet in Kürze
- 3 Überlegungen aus Sicht eines Repositoriums
- 4 Überlegungen aus Sicht einer Datengebenden
- 5 Überlegungen aus der Sicht einer Nachnutzerin
- 6 Herausforderungen

Allianz der Deutschen Wissenschaftsorganisationen (2010):

„[...] unterstützt die Allianz die langfristige Sicherung und den grundsätzlich offenen Zugang zu Daten aus öffentlich geförderter Forschung.“

DFG (2012):

„[Die] Maßnahmen zielen darauf ab, den möglichst offenen Zugang zu wissenschaftlich relevanter Information, zu Forschungsdaten sowie zu Arbeits- und Kommunikationsplattformen zu optimieren.“

Wissenschaftsrat (2012):

„Forschungsdaten sollten zum Zweck der Qualitätssicherung und Nachnutzung in geeigneten Forschungsdaten- und Datenservicezentren veröffentlicht werden.“

Horizon 2020, EU (2013)

“Regarding the digital research data generated in the action (‘data’), the beneficiaries must:

deposit [the data] in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate — free of charge for any user — [...], including associated metadata, needed to validate the results presented in scientific publications as soon as possible[...].”

Böckler Stiftung (2014):

„Wenn aus Projektmitteln systematisch Daten erhoben werden, die für die Sekundärauswertung geeignet sind, legen Sie bitte dar, welche Maßnahmen ergriffen werden sollen, um die Daten nachhaltig zu sichern und für erneute Nutzung bereit zu stellen.“

Auf Seiten der Drittmittelgeber:

- Nachhaltige Sicherung der Daten
 - Förderung der Transparenz
 - Effiziente Nutzung von (Steuer-) Geldern durch
 - Nachnutzung
 - Kontrolle
 - Qualitätssicherung
- => Über kurz oder lang wird eine Bereitstellung von Forschungsdaten empirisch forschende Wissenschaftlerinnen und Wissenschaftler aller Disziplinen betreffen.

Gemeinschaftsprojekt von 4 WGL-Instituten:

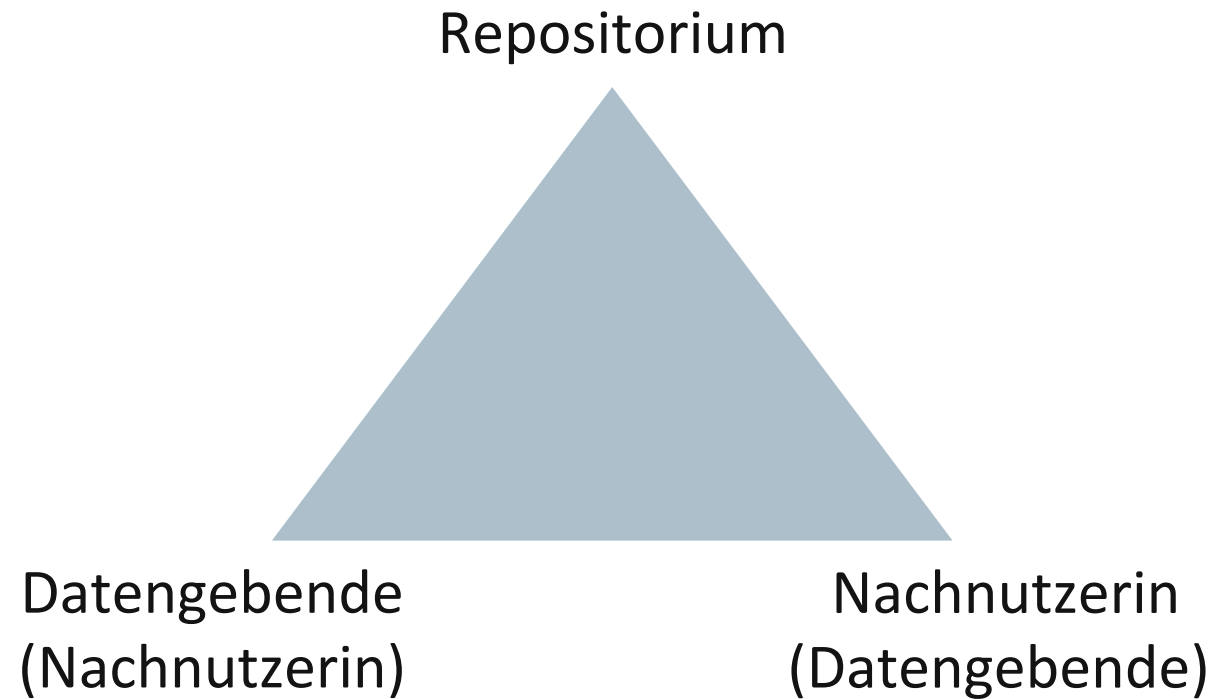


Zielsetzung: Eine digitale Infrastruktur bereitstellen, die die eigenständige Archivierung und Distribution quantitativer Forschungsdaten in den Sozial- und Wirtschaftswissenschaften erlaubt.

Gefördert im WGL-Wettbewerbsverfahren von 2014-2017 (36 Monate).

Arbeitspakete

1. Anforderungsanalyse
2. Organisationsmodell
3. Metadaten
4. Softwareentwicklung
5. Beispielinhalte
6. Transfer



Ein Repository hat Interesse an

1. Qualität
2. Dokumentation
3. Relevanz
4. Kontinuität

1. Qualität: Gute Daten...

- ...wurden nachvollziehbar erhoben/generiert
 - Sowi: z.B. Stichprobenplan
- ...sind gut dokumentiert
 - Sowi: z.B. Beobachtungseinheit, Datengrundlage, Methode, Umfang
- ...benötigen keine Rückkopplungsprozesse
- ...können durch das Repository zur Verfügung gestellt werden

2. Dokumentation durch Metadaten

- Metadaten erleichtern die Arbeit eines Repositoriums
 - Standardisierte Felder zur Einordnung der Daten
 - Ermöglicht Suchen und Gesucht-Werden (Harvesting)
 - Vergleichbarkeit wird ermöglicht
 - Steigerung der Qualität
- Sollten daher möglichst umfangreich sein
- Sollten von der Person eingegeben werden, die die Daten am Besten kennt -> Datengebende
- Metadatenstandards sind fachspezifisch
 - Sowi: Data Documentation Initiative (DDI)
 - Ähnliche in Bibliothekswesen, Geodaten, Internetinhalten

3. Relevanz: Häufigkeit der Datennutzung

- Daten, die nie heruntergeladen werden, verbrauchen nur Platz
- Ein Repository hat seine Berechtigung auch über die Nachnutzung, nicht nur über die Archivfunktion
- Im Idealfall gibt es in einem Fachgebiet nicht zu viele Repositorien, sondern nur eins

4. Kontinuität

- Einrichtung eines Repositoriums ist aufwendig
- Wissensverlust sollte vermieden werden

Wer sind potentielle Datengeber für Repositorien?

- Eher: Forscher mit kleineren Projekten, z.B. Doktoranden, Forscher mit kleineren Drittmittelprojekten, Forscher ohne Anbindung an größere Infrastrukturen
- Weniger: Forscher in Datenzentren, Forscher mit starker institutioneller Einbindung

Eine Datenproduzentin hat in erster Linie Interesse daran,

- die Pflicht der Archivierung der Forschungsdaten zu erfüllen
- möglichst wenig Zeit mit der Archivierung zu verbringen -> niederschwellig
- auch etwas von der Archivierung zu haben (außer der Pflichterfüllung)

Nutzen der Archivierung für die Datenproduzentin

- (Pflichterfüllung gegenüber Geldgebern)
- Sichtbarkeit erhöhen:
 - Wissenschaftliche Reputation durch Publikationen
 - Spannungsfeld: Datenproduktion vs. Publikationsoutput
 - Durch Zitierbarkeit von Forschungsdaten wird die Leistung honoriert
 - Wie? Persistente Identifikatoren, z.B. DOI
 - Sowi: da|ra als Vergabestelle von DOIs in Deutschland
 - (Text-) Publikationen mit Daten verlinken
- Transparenz herstellen
 - Erstellen von Datensätzen nachvollziehbar machen
 - Missbrauch der Daten vorbeugen
 - Versionierungen zeigen Veränderungen

Metadatenbereitstellung

- Kann für Datengeberin aufwendig sein
- Allerdings klare Vorteile, wenn die Metadaten ausführlich sind
 - Bessere Suchbarkeit
 - Erhöhte Wahrscheinlichkeit der Nachnutzung
 - > erhöhte Wahrscheinlichkeit der Zitation
 - > erhöhtes wissenschaftliches Standing

Daten anderen zur Verfügung zu stellen ist keine Selbstverständlichkeit.

Tenopir et al. (2011)

„Sind Ihre Daten verfügbar?“

- Knapp 1/3 gibt keine Antwort
- Von denen die antworten, geben 46% an, ihre Daten seien nicht verfügbar.

Fecher, Friesike & Hebing (2014) analysieren 98 Forschungsarbeiten aus verschiedenen Disziplinen zur Frage *„Welche Faktoren beeinflussen Data-Sharing in der Wissenschaft?“*.

Ergebnisse (stark vereinfacht):

- Soziodemographie (Alter, Nationalität, Karrierepunkt, Umfeld (Disziplin))
- Kontrollmöglichkeiten, Rechte
- Aufwand (Zeit, Geld)
- Nutzen (Zitationen, Austausch, Qualität)

SowiDataNet: Anforderungsanalyse

Vorbehalte gegen Data Sharing

- Anspruch auf Erstnutzung der eigenen Daten -> wann wird geteilt?
- Fehlerhafte Reanalysen könnten die eigene Arbeit diskreditieren
- Fragen des Datenschutzes (dürfen Befragungsdaten einfach geteilt werden?)
- Lizenzfragen
- Findet eine Nachnutzung überhaupt statt?
- Geld- und Personalmangel

Anforderungen der Nachnutzenden

- Schnelle Auffindbarkeit
- Gute Suchfunktion
- Gute Dokumentation zur Relevanzprüfung
- Möglichst wenige (nur eine) Anlaufstellen
- Mindeststandards und Qualitätsprüfungen
- Keine großen Formalitäten zur Nutzung
- Direkter Download der Daten
- Replikation

Repositorium

- Gute Daten
- Gute Metadaten
- häufige Nutzung
- Existenzsicherung

Datengebende (Nachnutzerin)

- Pflichterfüllung
- Geringer Aufwand
- Zitationen ermöglichen
- Transparenz erzeugen

Nachnutzerin (Datengebende)

- Auffindbarkeit
- Suchfunktion
- Dokumentation zur Relevanzprüfung
- Mindeststandards
- Wenige Anlaufstellen
- Keine Formalitäten
- Direkter Download
- Replikation

Repositorium

Gute Daten
Gute Metadaten
häufige Nutzung
Existenzsicherung

Datengebende (Nachnutzerin)

Pflichterfüllung
Geringer Aufwand
Zitationen ermöglichen
Transparenz erzeugen

Nachnutzerin (Datengebende)

Auffindbarkeit
Suchfunktion
Dokumentation zur Relevanzprüfung
Mindeststandards
Wenige Anlaufstellen
Keine Formalitäten
Direkter Download
Replikation

Ziele bei SowiDataNet

SowiDataNet wird eine Infrastruktur für Forschungsdaten in den Sozial- und Wirtschaftswissenschaften

- Zentrale Recherchemöglichkeit
- Metadatenerfassung in mehreren Stufen
- Direkte Datenregistrierung mit DOIs über da|ra
- Instituten den Aufbau einer Infrastruktur abnehmen
- Einzelforschern ohne institutionelle Einbindung helfen
- Kommentarmöglichkeiten durch Nachnutzende
- Langzeitarchivierung ermöglichen
- Beispiele und Vorschläge zum Datenmanagement erarbeiten

Immer: Orientiert am konkreten Bedarf der Scientific Community

Wer kann entscheiden, welche Daten eingestellt werden können?

Ist die Datenproduzentin immer eindeutig identifizierbar?

Wie lassen sich qualitative Daten archivieren?

Welche Finanzierungsmodelle sind vorstellbar?

Wie fachübergreifend kann ein Repository sein?



Vielen Dank für Ihre Aufmerksamkeit!

Vielen Dank für Ihre Aufmerksamkeit.



**DIW Berlin — Deutsches Institut
für Wirtschaftsforschung e.V.**

Mohrenstraße 58, 10117 Berlin
www.diw.de

Redaktion

Mathis Fräßdorf
Abteilung Forschungsinfrastruktur
mfraessdorf@diw.de
