

FDM an der Charité – Bauten und Baustellen

Evgeny Bobrov, BIH QUEST Center

27.5.2020



Aus Forschung wird Gesundheit

1.

QUEST Center

THE BIH QUEST CENTER FOR TRANSFORMING BIOMEDICAL RESEARCH

QUEST – Quality | Ethics | Open Science | Translation

Mission Statement

QUEST strives to increase the value of biomedical research at BIH and beyond.

QUEST Approaches



Open Science



Incentives



Education



Stakeholder
Engagement



Quality



Meta
Research

COMMUNITY PAGE

Improving the trustworthiness, usefulness, and ethics of biomedical research through an innovative and comprehensive institutional initiative

Daniel Strech ^{1,2}, **Tracey Weissgerber** ^{1,2}, **Ulrich Dirnagl** ^{1,2*}, on behalf of **QUEST
Group**^{1¶}

1 QUEST Center for Transforming Biomedical Research, Berlin Institute of Health (BIH), Berlin, Germany,
2 Charité - Universitätsmedizin Berlin, Berlin, Germany

2.

Forschungsdaten-Policy

Hinweis

Die zitierten Formulierungen stammen aus einem Entwurf der Forschungsdaten-Policy, und können sich somit noch ändern.

Definition Forschungsdaten

„**Forschungsdaten:** Forschungsdaten im Sinne dieser Policy sind alle Daten, die Gegenstand, Arbeitsschritt oder Ergebnis von Forschungsprozessen sind **und für die Nachvollziehbarkeit und Reproduktion von Forschungsergebnissen relevant sind.** [...] Für **nicht digitalisierbare Produkte des Forschungsprozesses** [...] können gesonderte Regelungen zur Dokumentation und Aufbewahrung gelten, die jedoch nicht Teil dieser Policy sind.“

1. Berücksichtigung von Relevanz in der Definition
2. Analoge Daten werden grundsätzlich mit adressiert
3. + Verzicht auf Begriffe „Primär-“ und „Sekundärdaten“, da nicht von allen gleich aufgefasst

Definition Forschende

„**Forschende:** Forschende sind alle aktiven Mitglieder der Charité, die Forschungsdaten sammeln und verarbeiten. Ebenso eingeschlossen sind Personen, die wie z.B. Gastwissenschaftler*innen, externe Studierende und Kollaborationspartner nicht unmittelbar der Charité angehören, die Einrichtungen aber für ihr Forschungsvorhaben nutzen.“

Andere Formulierungsbeispiele:

- **Postdam:** „alle wissenschaftlich und forschungsunterstützend tätigen Mitglieder und Angehörigen der Universität Potsdam“
- **TU:** „alle an der TU tätigen Wissenschaftlerinnen und Wissenschaftler“
- **HU:** „alle forschenden HU-Angehörigen“

→ Unterschiedliche Adressaten? Beispiel für Genauigkeit auf Kosten von Knappheit?

Stärke und Verbindlichkeit der Policy

„Falls ein Forschungsvorhaben durch Dritte gefördert wird und die zu Grunde liegende Vereinbarung Bestimmungen zum Forschungsdatenmanagement (FDM) enthält, **solte sie den Regelungen der Policy nicht grundlegend widersprechen.**“

Formulierung abgemildert → Unverbindlichkeit der Policy; auch andere Beispiele

„Für die Auffindbarkeit, Nachnutzbarkeit und Qualitätssicherung der Daten ist zudem die Verwendung fachspezifischer Daten- und Metadatenstandards, reporting guidelines und standardisierter Vokabulare...

- (1) ...zu empfehlen.
- (2) ...die Voraussetzung.
- (3) ...die Voraussetzung, und sollte von Projektbeginn an berücksichtigt werden.

Formulierung verstärkt → sich widersprechende Interessen und Wünsche zu berücksichtigen

→ Balanceakt: Policy nicht bindend, aber auch keine „Empfehlung“

(Teilweise) ungelöste Fragen

- Verantwortung der Organisation für Veröffentlichung von Forschungsdaten → streng genommen müsste jede Veröffentlichung (somit auch von Daten) durch Institution abgesegnet werden (zumindest für Angestellte), aber nicht annähernd praktikabel
- Wer entscheidet über Veröffentlichung von Forschungsdaten? PI, an Institution verbleibende Mitarbeitende, alle Autor*innen von Artikel(n),...?
- Aufbewahrung für 10 Jahre ab welchem Zeitpunkt? Mindestens in Charité-Satzung potenziell widersprüchlich
- Datenschutz:
 - Widerspruch Datenschutz und Gute Wissenschaftliche Praxis (Aufbewahrung)
 - Unklar, ob für Aufbewahrung zusätzlich zu Anonymisierung auch Einwilligung gefordert wird → derzeit Klärung
 - Wer ist bei Ausscheiden verantwortlich für Löschung?

3.

Inzentivierung für Open Data

Open Data LOM (=Leistungsorientierte Mittelvergabe)

Bisher zwei „Säulen“: Publikationen + Drittmittel

→ Nun zusätzlich Inzentivierung von Open Data im Rahmen der LOM-Vergabe

Open Data-LOM

- **Definition Open Data 2019**
 - Rohdaten zu einer Publikation
 - Würden analytische Replikation mindestens eines Teils erlauben
 - Auf Daten wurde klar hingewiesen; Daten sind zugänglich
 - Maschinenlesbares Format (keine PDF)
- **Kriterien sind weiter zu entwickeln: Interoperabilität & Nachnutzbarkeit; restricted access?**
- Detektion durch QUEST-Mitarbeitende; Nachmeldungen möglich
- Weltweit erste institutionsweite Open Data-Inzentivierung?!



Wie die offenen Daten finden?

ODDPub Algorithmus (Open Data Detection in Publications)

Text Mining-Algorithmus, Entwicklung seit August 2017 am QUEST → Github + Preprint

bioRxiv

THE PREPRINT SERVER FOR BIOLOGY

ODDPub – a Text-Mining Algorithm to Detect Data Sharing in Biomedical Publications

 Nico Riedel, Miriam Kip,  Evgeny Bobrov

doi: <https://doi.org/10.1101/2020.05.11.088021>

Erste Förderperiode 2015-2017: Ergebnisse auf Zenodo

The screenshot shows a Zenodo record for a poster titled "Einführung von Open Data als zusätzlicher Indikator für die Leistungsorientierte Mittelvergabe (LOM) Forschung an der Charité – Universitätsmedizin Berlin". The record is dated October 19, 2019, and has 371 views and 53 downloads. It is indexed in OpenAIRE. The publication date is October 19, 2019, with a DOI of 10.5281/zenodo.3511191. The keyword(s) are LOM, Inzentivierung, Incentives, Open Data, Open Research Data, Biomedical data, Biomedizinische Daten, Offene Daten, ODDPub. The communities are Open Research Data in Berlin und Brandenburg 2019. The license is Creative Commons Attribution 4.0 International. The versions section shows Version 1 from October 19, 2019. A preview of the poster is visible at the bottom, featuring the logos of the QUEST Center and the Berlin Institute of Health.

zenodo Search Upload Communities Log in Sign up

October 19, 2019 **Poster** **Open Access**

Einführung von Open Data als zusätzlicher Indikator für die Leistungsorientierte Mittelvergabe (LOM) Forschung an der Charité – Universitätsmedizin Berlin

Kip, Miriam; Bobrov, Evgeny; Riedel, Nico; Scheithauer, Heike; Gazlig, Thomas; Dirnagl, Ulrich

Seit 2009 vergeben alle medizinischen Fakultäten in Deutschland einen bestimmten Anteil des Landeszuschusses Forschung leistungsorientiert an ihre Einrichtungen (interne LOM). Die interne LOM ist ein institutionelles Incentivierungsinstrument mit dem Ziel, über eine kompetitive Vergabe nach definierten Kriterien qualitativ hochwertige Forschung zu belohnen. Die üblichen Kriterien dafür sind verausgabte Drittmittel und/oder die Anzahl der Zitationen in einem definierten Zeitraum (Journal Impact Factor (JIF) bzw. Relative Citation Ratio (RCR)). Als erste medizinische Fakultät hat die Charité 2019 das freie Verfügbarmachen von Originaldaten zu wissenschaftlichen Publikationen (Open Data) als zusätzlichen Indikator für die LOM eingeführt. Dafür wurde eine bottom-up, partizipatorische Implementierungsstrategie angewendet, die durch transparente Kommunikation und politisch-organisatorische Weiterentwicklung seit Beginn 2018 die Diskussion der Evidenz von Nutzen und Risiken von Open Data in der biomedizinischen Forschung vorantreibt.

In dieser Arbeit berichten wir über die wichtigen Schritte der Implementierung und die Herausforderungen dieses Kollaborationsprojektes zwischen dem BIH QUEST Center for Transforming Biomedical Research und dem Geschäftsbereich Forschung der Charité. Des Weiteren werden die Häufigkeit von Open Data im Publikationskorpus, der Open-Access-Status der Publikationen, die häufigsten Journals, in denen Open Data publiziert wurde, und die genutzten Repositorien beschrieben.

Indexed in **OpenAIRE**

Publication date: October 19, 2019
DOI: DOI 10.5281/zenodo.3511191
Keyword(s): LOM, Inzentivierung, Incentives, Open Data, Open Research Data, Biomedical data, Biomedizinische Daten, Offene Daten, ODDPub
Communities: Open Research Data in Berlin und Brandenburg 2019
License (for files): Creative Commons Attribution 4.0 International

Versions

Version	Date
Version 1	Oct 19, 2019

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.3511191. This DOI represents all versions.

Preview

Seite: 1 von 1 Automatischer Zoom+

QUEST CENTER Transforming Biomedical Research Berlin Institute of Health
BERLIN INSTITUTE OF HEALTH Charité & Max Delbrück Center

Einführung von Open Data als zusätzlicher Indikator für die Leistungsorientierte Mittelvergabe (LOM)-Forschung an der Charité – Universitätsmedizin Berlin

Miriam Kip, Evgeny Bobrov, Nico Riedel, Heike Scheithauer, Thomas Gazlig & Ulrich Dirnagl

4.

Fortbildungen

Open Data Workshops

Monatlich, je 3h, abwechselnd Deutsch und Englisch



Inhalte:

- Repositorien
- Metadaten
- Lizenzen
- Datenschutz
- Datensatzsuche
- U.v.m.



Seminar ‚Data sharing‘ – Rechtssicheres Teilen sensibler Daten

- Ganztägig, erstmals 1/2020; ca. 60 Personen auf Warteliste
- **3. Juni, 9-5pm als Webinar**
- Expert*innenbeiträge zu u.a.:
 - Datenschutz und Einwilligung
 - Umgang mit Gefahr von Re-Identifizierung & Tools für Anonymisierung
 - Aufgaben und Dienste einer Datentreuhandstelle
 - FAIR data & Interoperabilität
 - Infrastrukturen, Repositorien
 - Verteilte Analysen und Analysis-to-data
- Teilnahme für alle offen + Aufnahme und Online-Stellung geplant (soweit Referent*innen zustimmen)

5.

Ressourcen

FAQ Open Data in QUEST Toolbox



FAQ OPEN DATA

✓ [What is Open Data?](#)

✓ [Where is Open Data?](#)

✓ [Why is Open Data important?](#)

✓ [How might data sharing benefit me?](#)

∧ [Which data should I share?](#)

Data in this context includes all digital outputs of research. Thus, it includes raw and pre-processed (e.g. normalized or aligned) data, metadata (i.e. “data on data”), software code, protocols and any other output necessary to understand, reproduce or reanalyze data. The sharing of heavily processed data alone (e.g. summary statistics data underlying figures) is, however, not considered Open Data. Sharing of data which underlie publications is more common, but sharing of well-documented stand-alone datasets can be very valuable for (meta-)analysis by other researchers as well. The desired case would be providing raw data, pre-processed data and – if applicable - the code or steps necessary to arrive at the processed data.

✓ [Which data might not be suitable?](#)

✓ [Where can I share data?](#)

✓ [Where can I share data that do not accompany a full-length article publication?](#)



Quelle: Open Science
Foundation

FIDDLE – „Liberate“ data from the drawer

FIDDLE Options Scenarios



Transforming Biomedical Research
Berlin Institute of Health

FIDDLE - File Drawer Data Liberation Effort

Where and how to publish null / neutral results?

You would like to publish research outcomes which do not fit into the regular publication pipeline? FIDDLE provides guidance on publishing outcomes like null results or unanalyzed datasets, which all too often end up in the file drawer. Choose the options that are most important to you to find the most suitable publishing venue.

Options

I have a

- unanalyzed dataset
- stand-alone finding or results from a small study
- results from a full-scale study
- rejected manuscript

Money I have to cover publication cost (in €)

- 0
- up to 1000
- up to 2000

I want the outcome to be indexed in

- Pubmed
- Google scholar
- Google

I want my outcome to be peer-reviewed

- any
- yes
- no

I want my publication or dataset to appear immediately

- yes
- no

Choose the options that best reflects your case. None of these above option are helpful? Try to our scenario approach:

Go to scenarios

	Data Repository	Micropublication	Preprint publication	Data journals	Publishing platform	Journal open to null results
Description	Platforms that allow to upload research datasets to make them citable and reusable.	Aimed for unpublished observations, negative/neutral results and does not require a narrative story.	Platforms for yet unpublished research manuscripts that allow an immediate visibility of the results.	Journal article that focuses on presenting a dataset together with metadata and the methods used to acquire the dataset.	Articles are published without editorial filtering; peer-review happens after (immediate) publication of article.	Traditional journals that also publish null results.
Providers	Zenodo, FigShare or Dryad, use re3data to search for disciplinary repositories	ScienceMatters, BMC Research Notes	bioRxiv, ASAPBio, osf.io	Scientific Data, Data, Data In Brief, many disciplinary journals (e.g. GigaScience, Biomedical Data Journal)	F1000Research	PeerJ, PLoS One, multiple BMC journals and many other disciplinary journals
Effort	low effort	low effort	medium effort	some effort to prepare manuscript/data	some effort to prepare manuscript/data	some effort to prepare manuscript/data
Costs	free of charge	600 - 1300 €	free of charge	up to 1500€	up to 1000€	up to 1600€
Time to publication	immediate	typically 1-3 months	immediate	typically 1-4 months	immediate	typically 1-6 months
Recognition	citations of the dataset	citations of article + article in publication list (future handling of such articles is open)	citations of article + article in publication list (not universally accepted at this point)	citations of article + article in publication list	citations of article + article in publication list (not universally accepted at this point)	citations of article + article in publication list
Publishing venue can have Impact Factor	no	yes	no	yes	no	yes
Peer-review	no		post-publication review		peer-review	peer-review
Indexing: Pubmed /	no/no/yes	no/yes/yes	yes/yes/yes	yes/yes/yes	yes/yes/yes	yes/yes/yes

s-quest.bihealth.org/fiddle/

Vielen Dank!

[https://www.bihealth.org/de/quest-center/
quest-toolbox.bihealth.org](https://www.bihealth.org/de/quest-center/quest-toolbox.bihealth.org)

Contact:

DR. EVGENY BOBROV

Berlin Institute of Health (BIH)

QUEST – Quality | Ethics | Open Science | Translation

BIH Center for Transforming Biomedical Research

Open Data and Research Data Management Officer

Tel. +49 (0)30 450 543 069

evgeny.bobrov@bihealth.de

BIH Berlin Institute
of Health
Charité & MDC