



DepositOnce

Das Repository der TU Berlin für Forschungsdaten und Publikationen

Pascal Becker | Universitätsbibliothek | 8. DINI-/nestor-Workshop „Forschungsdatenrepositorien“





Projekt „Forschungsdaten“

- Gemeinsames Konzept von UB, Rechenzentrum und Forschungsabteilung (03/2012)
„Organisatorisch-technisches Konzept für eine Forschungsdaten-Infrastruktur in der TU Berlin“
 - www.szf.tu-berlin.de/menue/info-materialien_publicationen
- TUB-internes Projekt zum Aufbau einer universitätsweiten Forschungsdaten-Infrastruktur
 - Projektlaufzeit 09/2012 – 08/2014
 - Getragen durch: Universitätsbibliothek (Projektleitung), IT-Service-Center tubIT, Forschungsabteilung
 - Unterstützung durch Kanzlerin und Vizepräsident für Forschung
 - UB stellt zur Initiierung des Projekts eine Stelle Wiss. MA (Informatiker) zur Verfügung
 - Projektdurchführung in enger Zusammenarbeit mit den Wissenschaftler/innen (**Pilotpartner**)
 - Zustimmung der Datenschutzbeauftragten und des Personalrats zur Einführung des Repositoriums
- **Speicherung + Verifizierung + Langzeitverfügbarkeit + Nachnutzung von Forschungsdaten**

Folie: M. Kuberek

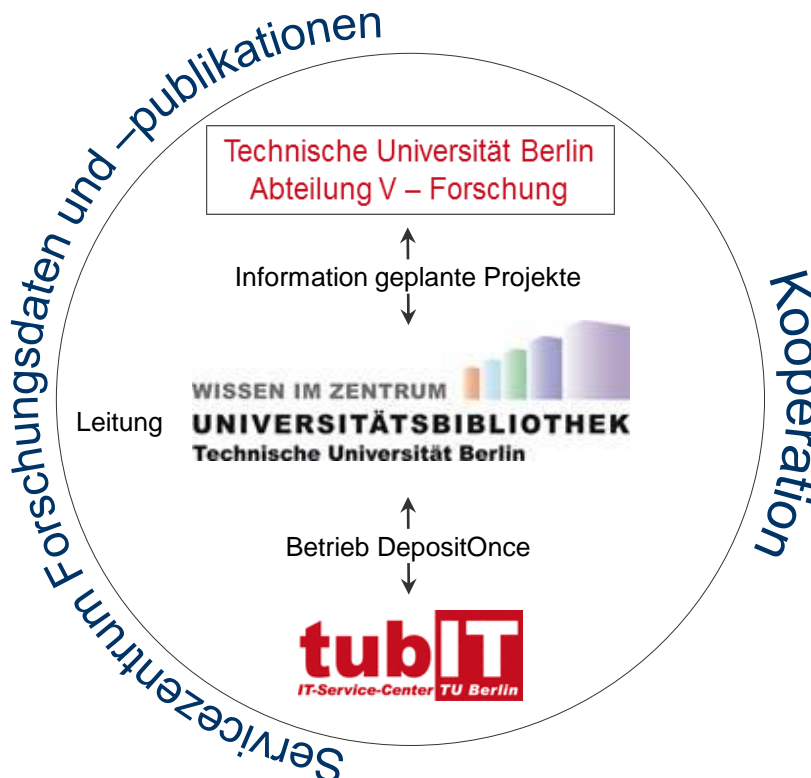


Forschungsdaten-Infrastruktur an der TU Berlin

Virtuelle Organisationseinheit SZF



Services für Forschende

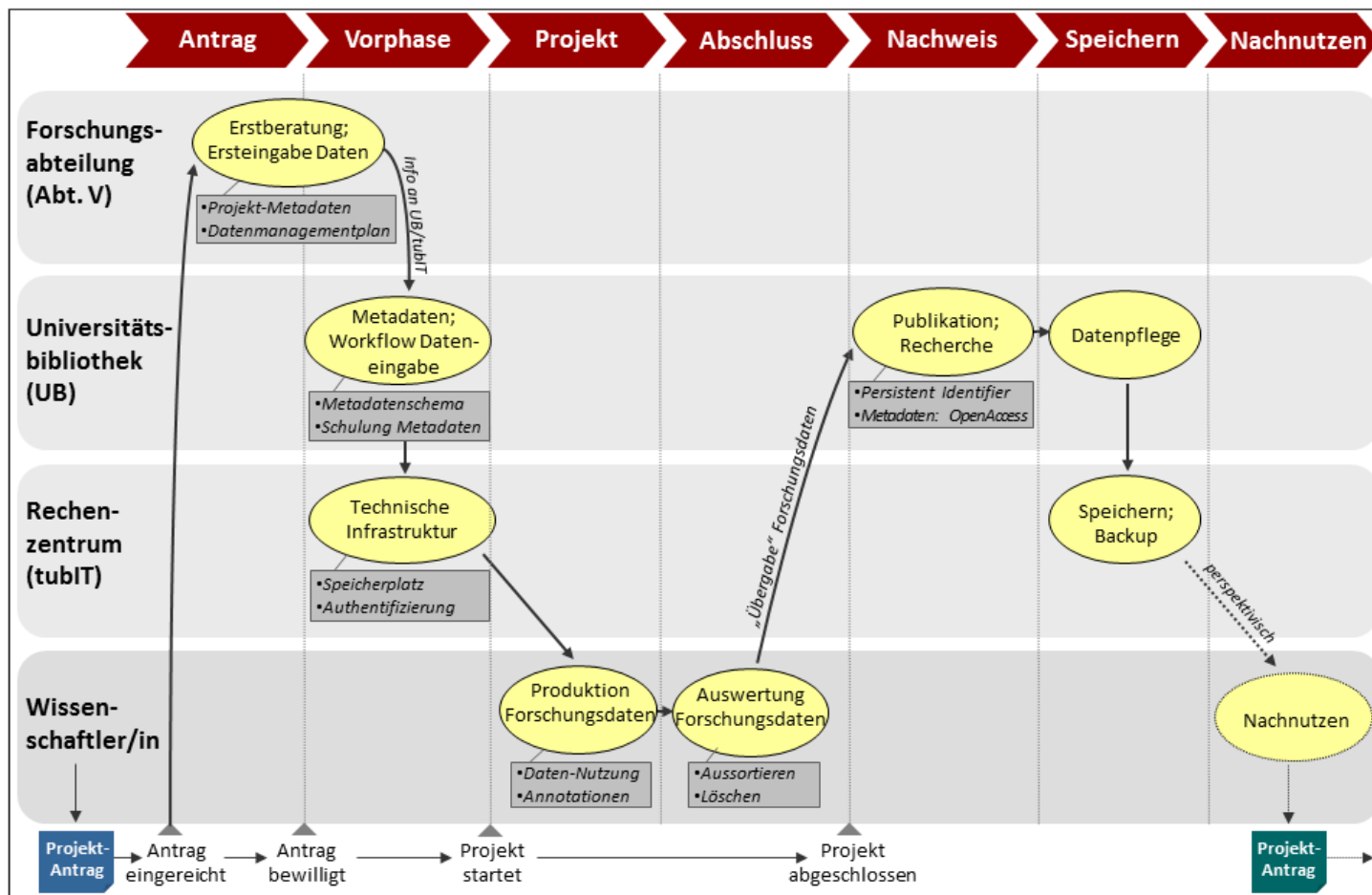


- **SZF-Website** • Informationsplattform im Internet • szf.tu-berlin.de
- **DepositOnce** • Repositorium für Forschungsdaten und Publikationen • depositonce.tu-berlin.de
- **TUB-DMP** • Web-Tool zur Erstellung von Datenmanagementplänen • dmp.tu-berlin.de
- **Helpdesk** • SZF-Team • team@szf.tu-berlin.de

Folie: M. Kuberek

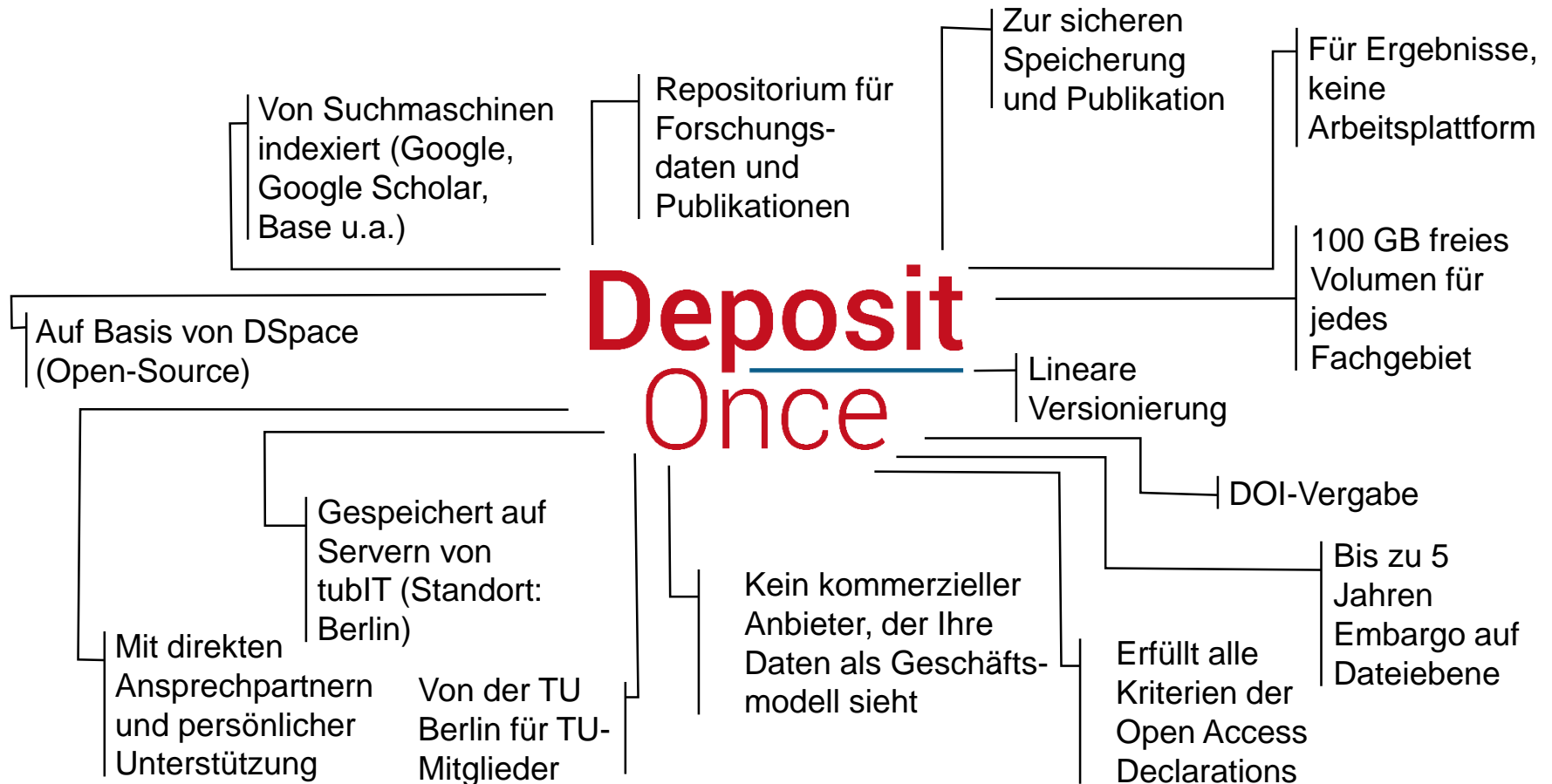


Forschungsdaten-Lifecycle





Argumente für die Nutzung von DepositOnce



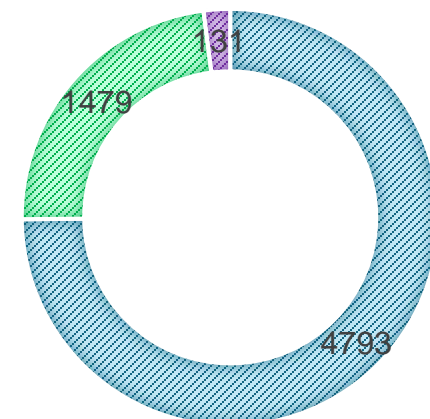


DepositOnce

- Repository für Forschungsdaten und Publikationen
- Institutionelles Repository für Mitglieder der TU Berlin
- Auf Basis von DSpace (inzwischen 6.2x)
- Code-Contribution an DSpace
 - DOI-Vergabe
 - HTML5-Upload
 - Zahlreiche Bugfixes und Verbesserungen (z.B. im Bereich Versioning)
- Eigene spezifische Ergänzungen
(<https://github.com/tuub/DSpace>)

INHALTE

- Dissertationen
- Weitere Publikationen
- Forschungsdaten





Forschungsdaten

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

*... und/oder Daten, die zur Reproduktion eines Ergebnisses notwendig sind, wie Skripte, Zeitschriebe, Berechnungen, usw.
(Definitionserweiterung im Projekt „Forschungsdaten“ 2013)*

Folie: M. Kuberek



Forschungsdaten und Publikationen

- Enger Zusammenhang zwischen Publikationen und den Daten auf denen sie basieren
 - Forschungsdaten von Interesse um detaillierteres Wissen zu erlangen
 - Forschungsdaten von Interesse für die Nachnutzung
 - Publikationen zu Forschungsdaten um zu erfahren, woraufhin bestimmte Daten bereits untersucht wurden und wie andere mit den Daten umgegangen sind
 - Publikationen über die Entstehung einzelner Forschungsdatensätze von Interesse
- Wir bezeichnen Forschungsdaten und Publikationen zusammenfassend als „Forschungsergebnisse“



Verknüpfung in DSpace

- Verknüpfung von Forschungsdaten und Publikationen in DepositOnce über DOIs in entsprechenden Metadatenfeldern
- Derzeit manuell befüllt
- In Vorbereitung: „Curation Task“ um automatisch Rücklinks zu setzen
- Idee: „Search“-Button für die Selfsubmission
- Traum: Automatische Zitationsextraktion
 - Nicht alles was Forschungsdaten zitiert basiert auf ihnen
 - Aber jede Publikationen, die einen Forschungsdatensatz erwähnt, ist einen Hinweis wert



Verknüpfung in DSpace

<http://dx.doi.org/10.14279/depositonce-1>

<http://dx.doi.org/10.14279/depositonce-31>

Is Supplemented <http://dx.doi.org/10.14279/depositonce-31>
By:

Is Supplement <http://dx.doi.org/10.14279/depositonce-1>
To:



Die große Masse

- Forschungsdaten sind sehr heterogen und sehr disziplinspezifisch
 - Das gilt für Format
 - Das gilt für Dateigrößen
 - Das gilt für die Anzahl von Dateien
- Größe eines Datensatzes weniger wichtiger als Größe der größten Datei
- In einigen wenigen Disziplinen sehr große Dateien (hunderte GB, Terabytes, ...)
 - Oft nur mit spezifischen Geräten/spezifischer Software auszuwerten
 - Eine mehrere Terabyte große Datei wird nicht nebenbei auf den persönlichen Arbeitscomputer geladen
 - Sehr große Dateien sollten über spezifische Protokolle übertragen und nicht über „normales“ https
- Der Großteil der Disziplinen hat keine so großen Dateien




Umsetzung in DSpace

- DSpace kann mit allem Umgehen, was sich gut von einem Servlet Container über https übertragen lässt
 - Die Größe der größten Datei ist wichtig, nicht die Gesamtmenge
 - Größte Einzeldatei in DepositOnce (Stand 11/2017): ca. 14 GB (<http://doi.org/cgng>)
 - JSPUI: HTML5-Upload überträgt Dateien in einzelnen kleinen Stücken
 - Gesucht: ein gutes Konzept, um das Laden vieler Dateien auf einmal zu vereinfachen
 - Datensatz mit den meisten Dateien in DepositOnce (Stand 11/2017): 414 Dateien (<http://doi.org/cgnh>)
- Für den absoluten Großteil der Forschungsdaten ausreichend



Umsetzung in DSpace

<http://dx.doi.org/10.14279/depositonce-87.2>

Files in This Item:				
File	Description	Size	Format	
pictures.zip	Pictures of the measurement setup	40,97 MB	ZIP Archive	View/Open
RIRs.zip	Data: single-channel room impulse responses	142,78 MB	ZIP Archive	View/Open
BRIRs.zip	Data: binaural room impulse responses	13,79 GB	ZIP Archive	View/Open
AdditionalInformation.pdf		1,49 MB	Adobe PDF	 View/Open



Repositorien für Forschungsdaten – Repositorien für Publikationen

- Volltext-Indizes helfen bei Forschungsdaten nicht -> die Bedeutung guter deskriptiver Metadaten steigt wieder
- Werden Forschungsdaten aufgrund von Ordner-Strukturen in Archiven bereitgestellt, sollte die Dokumentation zusätzlich außerhalb der Archive bereitgestellt werden
 - Suchmaschinen-Kompatibilität
 - Nutzer/innen können sich einfach einen Eindruck verschaffen, ob die Forschungsdaten von Interesse sind
- DOI-Vergabe motiviert zur Nutzung von Repositorien
 - Zeitschriften verlangen zunehmend supplemental material und akzeptieren dafür nur DOIs



Versionierung

- Versionierung insbesondere (aber nicht nur) für Forschungsdaten sehr hilfreich

Notice This is not the latest version of this item. The latest version can be found at: <http://dx.doi.org/10.14279/depositonce-87.2>

Version History

Version	Item	Date	Summary
2	doi:10.14279/depositonce-87.2 *	2015-04-16 18:32:36.035	Update of AdditionalInformation.pdf
1	doi:10.14279/depositonce-87	2015-03-13 14:56:05.0	

* Selected version



Repositorien für Forschungsdaten – Repositorien für Publikationen

- Repositorien speichern Dateien und Metadaten, unabhängig vom Inhalt der Dateien
- Visualisierung und Einbindung in die Weboberfläche für Publikationen (PDFs) einfacher als für die Formatvielfalt von Forschungsdaten (speziell in disziplinübergreifenden Repositorien)
- Volltextindizes helfen beim Finden von Publikationen, können für die Dokumentation von Forschungsdaten hilfreich sein, helfen für die eigentlich Daten jedoch nicht
- Forschungsdaten können andere Größenordnungen haben als Publikationen, meistens ist das aber kein Problem
- Forschungsdatensätze umfassen oft mehrere Dateien



Forschungsdaten und DSpace – ein Fazit

- Für uns waren wichtig: DOI-Vergabe und HTML5-Upload
 - Beides von uns umgesetzt und in DSpace enthalten
- Für die meisten Daten sehr gut geeignet, Beschränkungen eher durch Browser und HTTPS, als durch DSpace
- Herausforderungen liegen weniger im technischen, mehr im organisatorischen Bereich und in der Betreuung der Nutzer/innen



Technische Universität Berlin
Universitätsbibliothek
Pascal-Nicolas Becker
p.becker@tu-berlin.de

Servicezentrum Forschungsdaten und -publikationen

<http://www.szf.tu-berlin.de>

DepositOnce

<https://depositonce.tu-berlin.de>

<https://github.com/tuub/DSPACE>