



Technische
Universität
Braunschweig



http://www.data.cam.ac.uk/sites/www.data.cam.ac.uk/files/styles/leading/public/team_crop.png?itok=3GKODs4r

RADAR als eine Option für das institutionelle Forschungsdatenrepositorium der TU Braunschweig

Robert Strötgen (TU Braunschweig)
mit Unterstützung durch Angelina Kraft (TIB Hannover) & Matthias Razum (FIZ Karlsruhe)

8. DINI/nestor-Workshop "Forschungsdatenrepositorien"

RADAR – Research Data Repository

RADAR (**R**esearch **D**ata **R**epository) ist

- ein disziplinübergreifendes Forschungsdaten-repositorium
 - für die **Archivierung** und **Publikation** von Forschungsdaten
 - abgeschlossener wissenschaftlicher Studien und Projekte.
-
- September 2013 – August 2016
 - Gefördert durch die DFG



Features

Service Level

Datenarchivierung

- „dark archive“
- Flexible Haltefrist (5, 10, 15 Jahre)
- Datengeber bestimmen Zugriffsrechte (privat/geteilt/öffentlich)



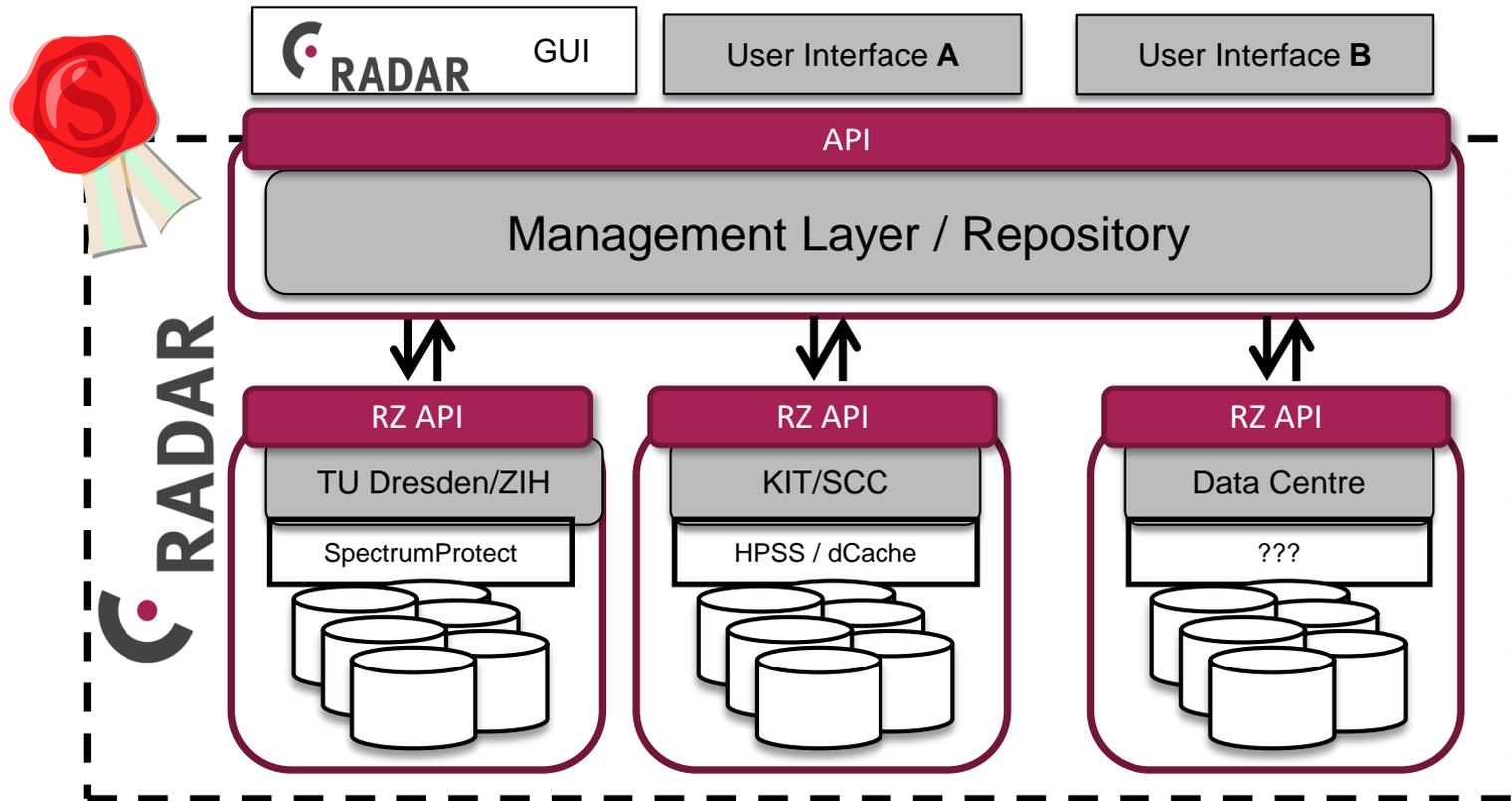
Quelle: Wikimedia Commons/Simon A. Eugster. CC 3.0 BY SA

Service Level

Datenarchivierung mit Datenpublikation

- Unbegrenzte Haltefrist (mindestens 25 Jahre)
- Optionale Embargos
- Vergabe eines DOI (Datacite)
- Schnittstelle für Peer Review
- Auswahl von Lizenzen, kein CC0 erforderlich
- Metadaten indexiert (RADAR, Datacite, OAI, Google)

Schematische Architektur

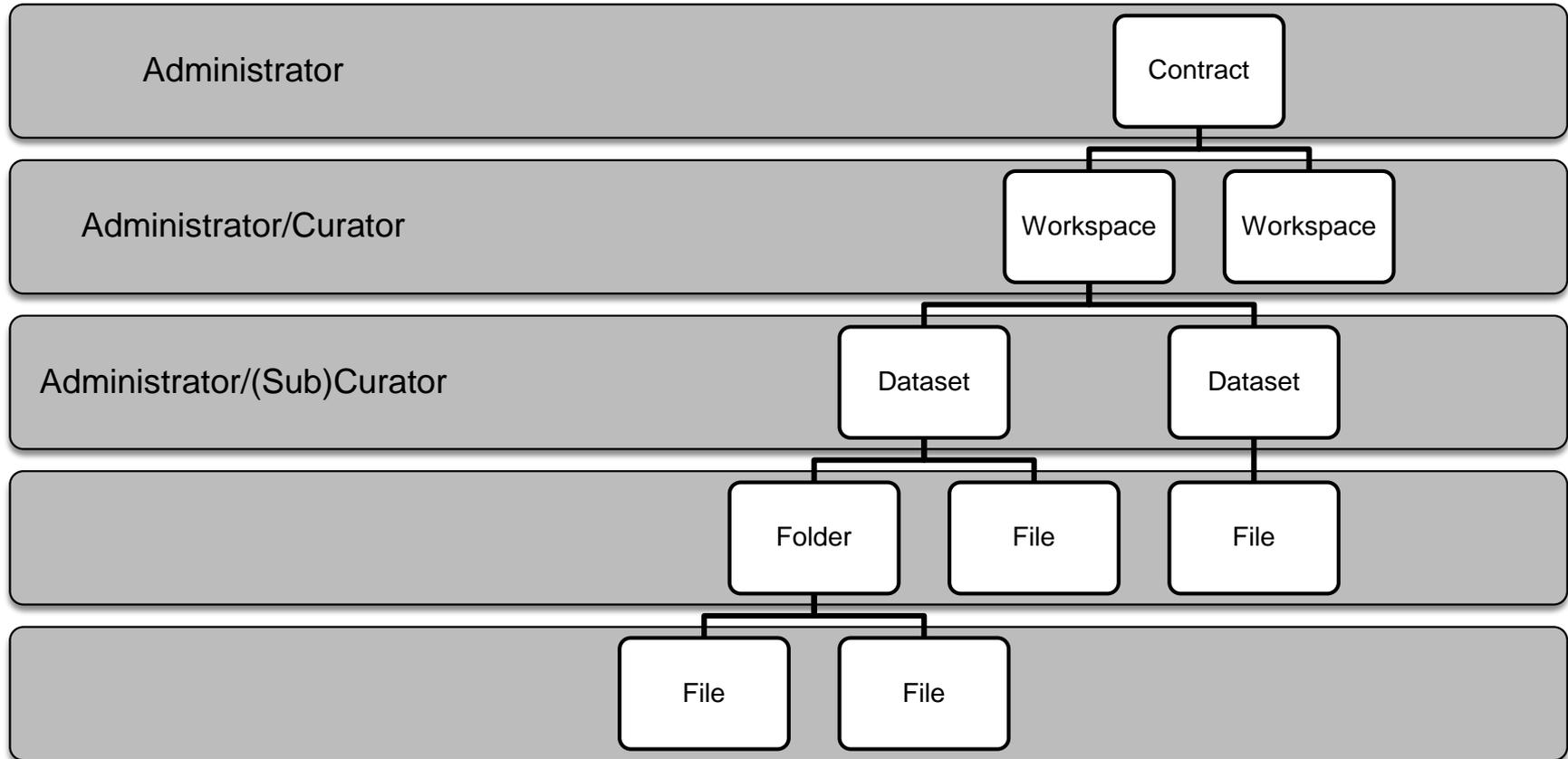


Datensicherheit

- Für archivierte/publizierte Daten:
 - Datenpakete werden mit Checksum versehen
 - Nach dem Schreiben wird Checksum direkt geprüft
 - Drei Kopien an drei Standorten (2x Karlsruhe, 1x Dresden)
 - Unterschiedliche Hard- und Software sowie Administration
 - Regelmäßige Migration auf neue Speichermedien, dabei Prüfung der Datenintegrität
- Für temporäre Daten
 - Redundantes Plattensystem (RAID-6) mit täglicher Sicherung



Hierarchische Organisation der Inhalte und Zuständigkeiten



Bedarf der Forscher/innen

-> Subsidiär zu fachlichen Repositorien

Optionen

- Betrieb des Repositoriums in der TU
- Hosting des Repositoriums durch externen Dienstleister
 - Hosting des Repositoriums durch KIT / TIB im Rahmen des RADAR-Dienstes
 - Verbundzentrale Göttingen im Rahmen des Reposis-Dienstes auf Basis von MyCoRe (in Analogie zum TU-Publikationsserver)

Betrieb des Repositoriums in der TU

- (Open Source)-Software für institutionelle Repositories existiert z.B. DSpace, Invenio, Dataverse, EPrints, OPUS, MyCoRe
- V.a. Generische Repository-Software, die um Funktionalität für Forschungsdaten erweitert wird
- Mögliche Arbeitsteilung:
 - GITZ: Bereitstellung Server, Storage, Backup
 - UB: Entwicklung, Betrieb
- Beispiel: Deposit Once an der TU Berlin

Vor- und Nachteile

- + Alle Daten innerhalb der TU
- + große Flexibilität
- + Austausch in entsprechenden Communities
- Lange Vorlaufzeit
- Hoher Personalaufwand
- Schwierigkeit bei Personalgewinnung
- Risiken bei der Anpassung (Mehraufwand, Verzögerungen)
- Risiken bei Versionssprüngen (Aufwand für lokale Anpassungen)
- Abhängigkeit von Entwicklung der Softwarebasis
- Noch keine Erfahrungen mit sehr großen Datenmengen (Skalierung?)

Repository im Rahmen des RADAR-Dienstes

- Entwickelt als generisches Forschungsdatenrepository im Rahmen eines DFG-Projekts
- Service betrieben durch FIZ Karlsruhe für wissenschaftliche Einrichtungen (nicht kommerziell)
- Beratung und Marketing durch FIZ Karlsruhe und TIB
- Angebote für Archivierung (jährliche Kosten) und Publikation (Einmalkosten mit Garantie für 25 Jahre, beinhaltet Archivierung)
- Daten liegen auf Bändern in den Hochschulrechenzentren Karlsruhe und Dresden

Kosten / Aufwände (grobe Schätzungen)

- Personalaufwand
 - Konzept, Integration etc. durch vorhandenes UB-Personal
- Sachaufwand
 - ggf. Kosten für Einrichtung und Anpassung
 - Grundgebühr (595 € oder 11.900 € - ab 50 TByte wirtschaftlicher)
 - Evtl. Betrieb einer eigenen Instanz (19.500 € jährlich)
 - Speicher für Archivierung (0,46 € bzw. 0,23 € pro GByte jährlich) und Publikation (7,56 € bzw. 6,12 € pro GByte einmalig)
- sofort nutzbar
- Als mandantenfähiges System / eigene Instanz: produktiver Start ca. 6 Monate nach Projektbeginn

Vor- und Nachteile

- + Kurze Vorlaufzeit
- + Bereits mehr spezifische Funktionen für Forschungsdaten und Mandantenfähigkeit
- + Weitere DFG-Förderung möglich
- + Integration in Infrastruktur über APIs möglich
- + Nachhaltigkeit und Exit-Strategie garantiert
- + Geringer Personalaufwand
- + Relativ niedrige Speicherkosten
- + Einmalzahlung für Forschungsdatenpublikation bei Projektfinanzierung vorteilhaft
- + Skaliert relativ gut
- RADAR-Service gerade gestartet, muss sich noch etablieren
- Integration in bestehende Infrastruktur evtl. aufwändig (z.B. Verknüpfung von Publikationen u. Forschungsdaten)
- Nötige Flexibilität evtl. nur mit eigener Instanz

Hosting bei Verbundzentrale Göttingen (VZG)

- VZG betreibt Dienstleistung „ReposiS“, die bereits für den TU-Publikationsserver genutzt wird
- Softwarebasis ist MyCoRe (Open Source, aktive Community in Deutschland)
- MyCoRe wurde um Funktionalitäten für Forschungsdaten erweitert
- Arbeitsteilung
 - Technischer Betrieb durch VZG, Anpassung und Weiterentwicklung durch Auftrag der UB
 - Redaktionelle Betreuung durch UB
- Beispiel: OpenAgrar als Repositorium von Bundesressortforschungseinrichtungen

Kosten / Aufwände (grobe Schätzungen)

- Personalaufwand
 - Konzept, Integration etc. durch vorhandenes UB-Personal
- Sachaufwand
 - Kosten für Anpassungen des schon genutzten Reposis-Dienstes
 - Erhöhung des Supportstunden-Kontingents für Reposis
 - Speicher und Backup der Verbundzentrale (350 € pro TByte jährlich)
 - Zukünftig möglicherweise noch günstigerer Speicher bei Ablage auf Bändern
- produktiver Start nach Projektbeginn: ca. 3 Monate

Vor- und Nachteile

- + Kurze Vorlaufzeit
- + Gute Integration in bestehende Infrastruktur - auch über APIs
- + Dadurch „1-Stop-Shop“ für die Forscher/innen
- + Gute Verknüpfbarkeit von Forschungsdaten und Publikationen
- + Austausch in Communities MyCore und Repositis
- + Geringer Personalaufwand
- + Relativ niedrige Speicherkosten
- + räumlich verteiltes Backup
- + 10 Jahre Garantie für dauerhaften Betrieb via SLA
- Entwicklung der FD-Funktionalität noch nicht „vollständig“
- (noch) kein räumlich verteiltes Backup
- Abhängigkeit von Repositis-Dienst und MyCore-Community (Entwicklungsdynamik, Nachhaltigkeit?)
- Noch keine Erfahrungen mit sehr großen Datenmengen (Skalierung?)

Ungelöste Fragen bei allen Optionen

- Umgang mit personenbezogenen Daten
- Umgang mit hochsensiblen Daten
- Echte Langzeitarchivierung der Forschungsdaten (immer nur *bitstream preservation*)
- Fachspezifische Funktionen (Schemata, Vokabulare) – generische Angebote sind hier den fachlichen unterlegen

Fazit

- Betrieb des institutionellen Forschungsdatenrepositoriums für die TU Braunschweig eine ernsthafte Option
- RADAR ist dabei eine Option neben weiteren
- Entscheidung steht im 1. Quartal 2018 an (auf der Grundlage einer detaillierten Entscheidungsmatrix)



Technische
Universität
Braunschweig



http://www.data.cam.ac.uk/sites/www.data.cam.ac.uk/files/styles/leading/public/team_crop.png?itok=3GKODs4r

Vielen Dank für Ihre Aufmerksamkeit !

forschungsdaten@tu-braunschweig.de