



FIZ Karlsruhe

Leibniz-Institut für Informationsinfrastruktur

RADAR

Archivierung und Publikation von Forschungsdaten

Matthias Razum

DINI Jahrestagung 2017

Göttingen, 05. Oktober 2017

RADAR IN A NUTSHELL
VOM PROJEKT ZUM PRODUKT
AUSBLICK



RADAR – RESEARCH DATA REPOSITORY

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

- disziplinübergreifendes Forschungsdatenrepositorium
- für die Archivierung und Publikation von Forschungsdaten
- abgeschlossener wissenschaftlicher Studien und Projekte.

Gefördert durch die DFG von September 2013 – August 2016

Keine Konkurrenz zu etablierten disziplinären Datenzentren

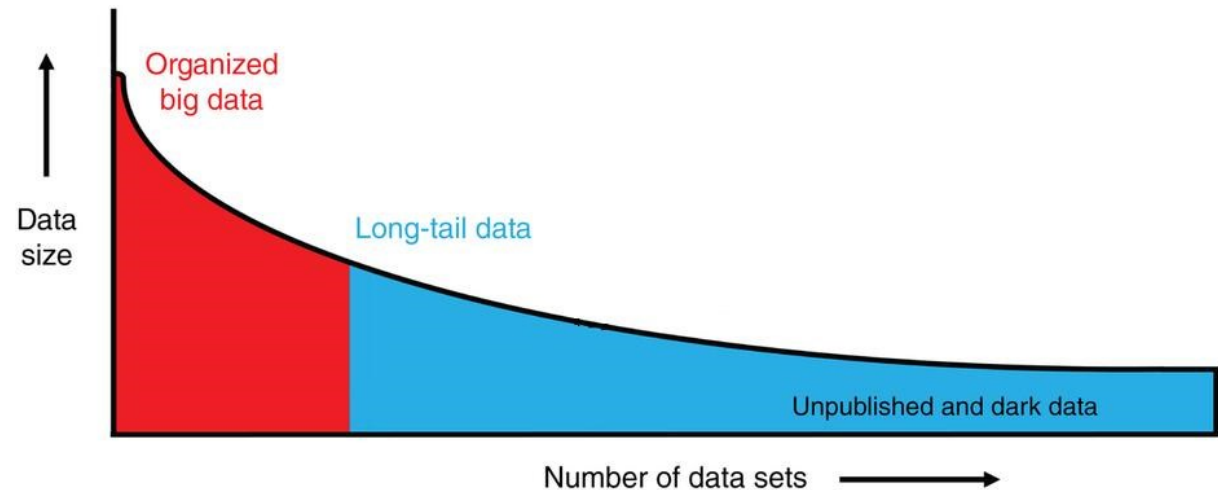
- Schwerpunkt „Long tail“



DER ‚LONG TAIL‘ IST WICHTIG

“The majority of datasets produced through research are part of the ‘Long Tail of Research Data’”

Source: Humphrey C (2014): OpenAIRE-COAR Conference, Athens



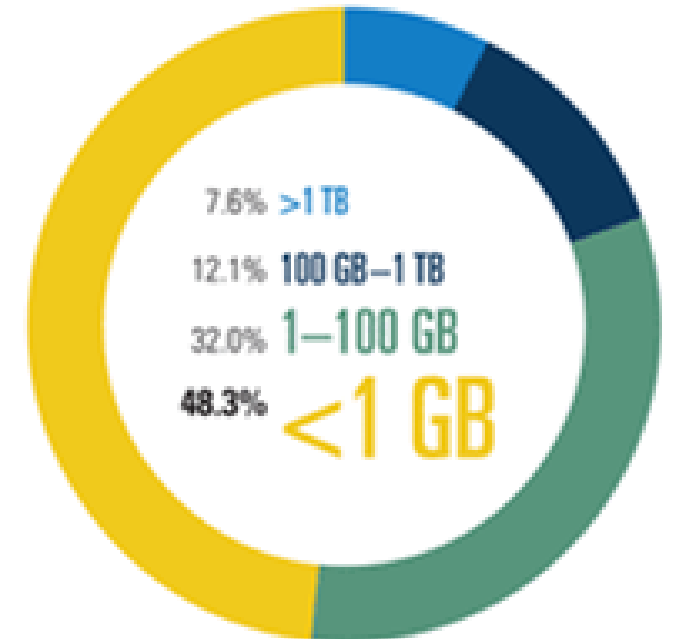
Source: Ferguson et al. (2014): Big data from small data: data-sharing in the 'long tail' of neuroscience. DOI: 10.1038/nn.3838

FORSCHUNGSDATEN SIND NICHT IMMER „BIG DATA“

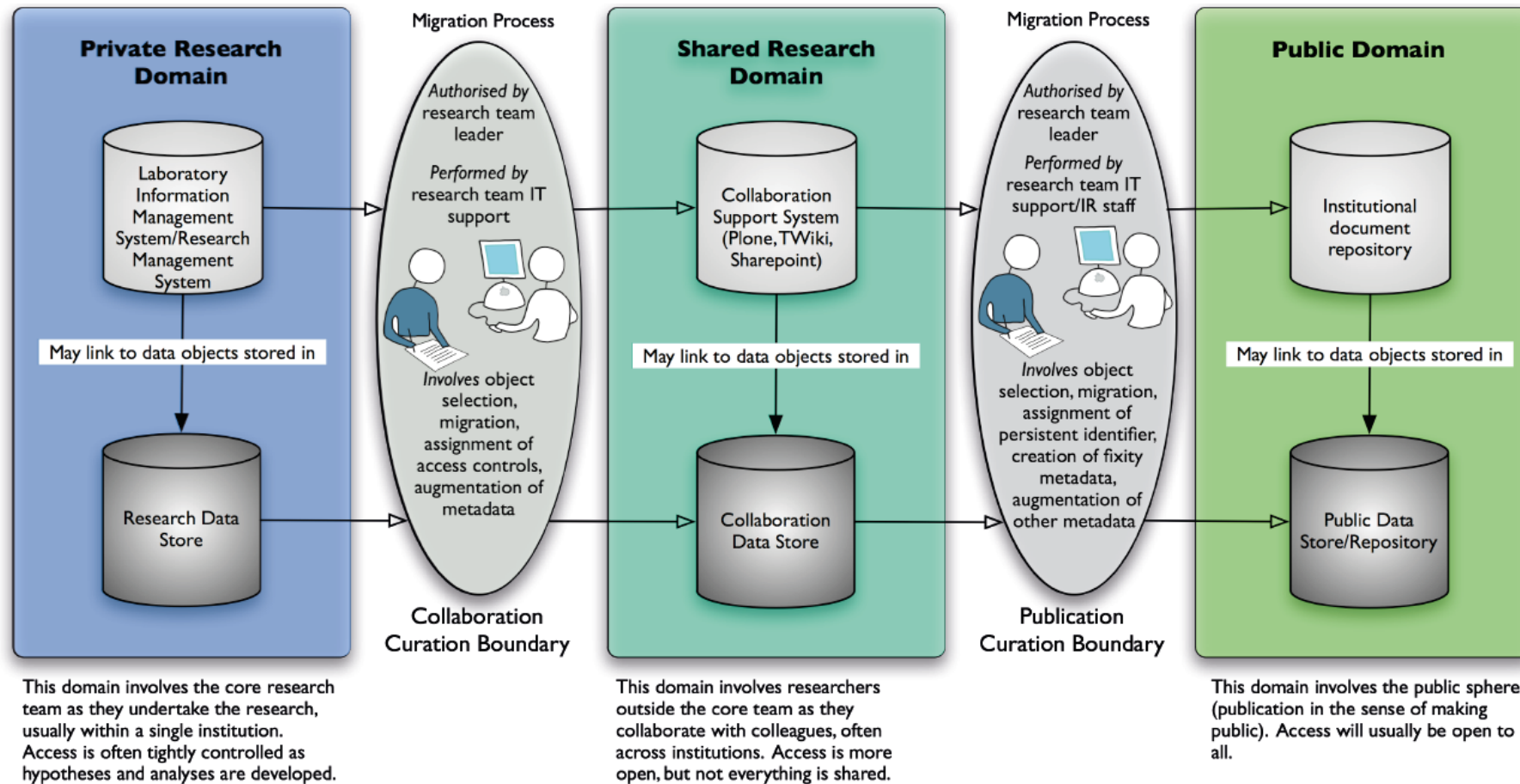
Science Survey 2011:

- 48 % of respondents were working with datasets that were <1GB in size
- 50 % stored data exclusively! in labs

Source: Science (2011): 331(6018), p. 692-693. DOI: 10.1126/science.331.6018.692



LEBENSZYKLUS UND FOKUS



Version 1.4, <http://andrew.treloar.net/>, 07Dec07

SERVICE LEVEL VON RADAR

Datenarchivierung

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

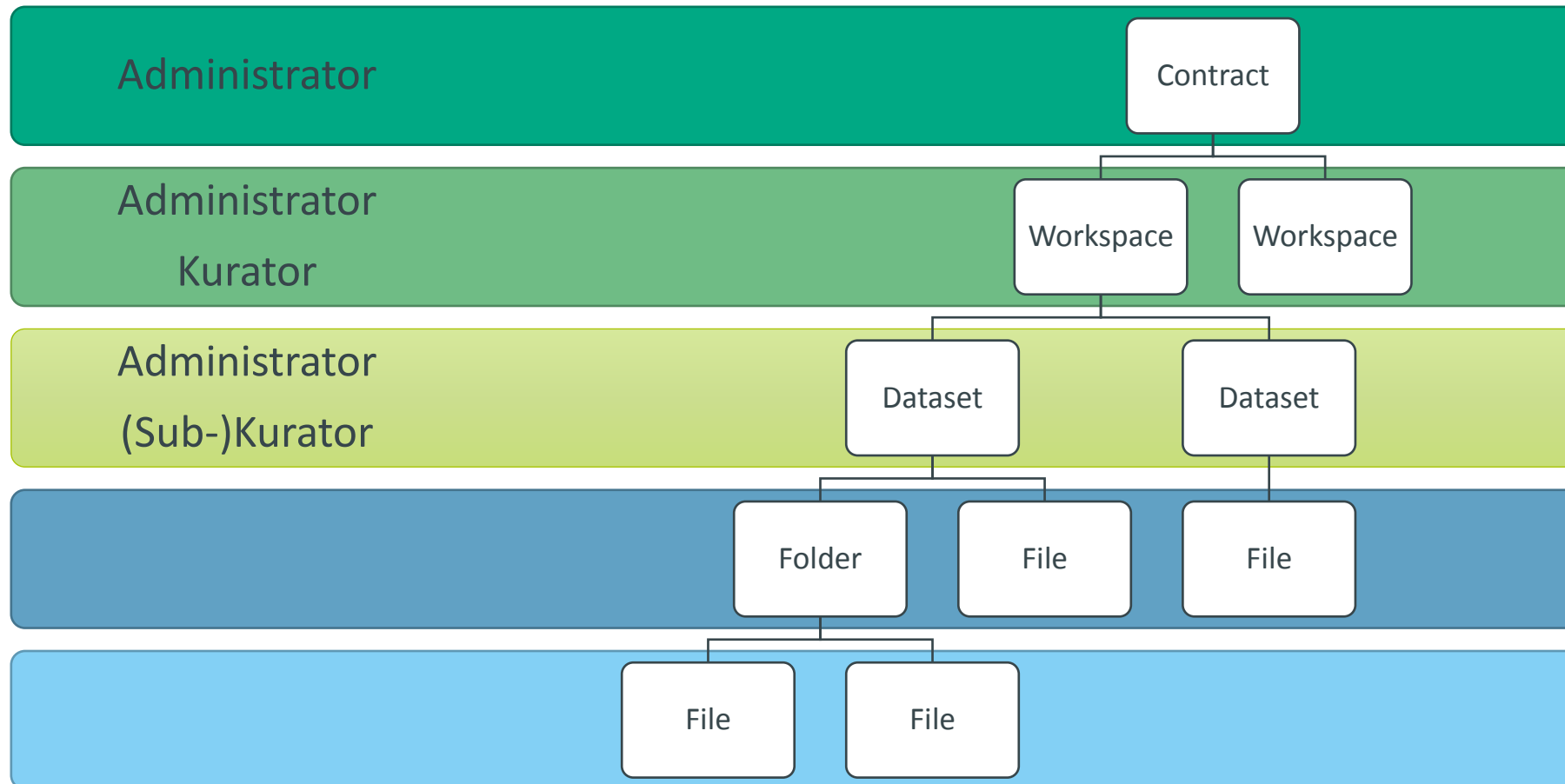


Quelle: Wikipedia Commons, Simon A. Eugster. CC 3.0 BY SA

Datenarchivierung mit Publikation

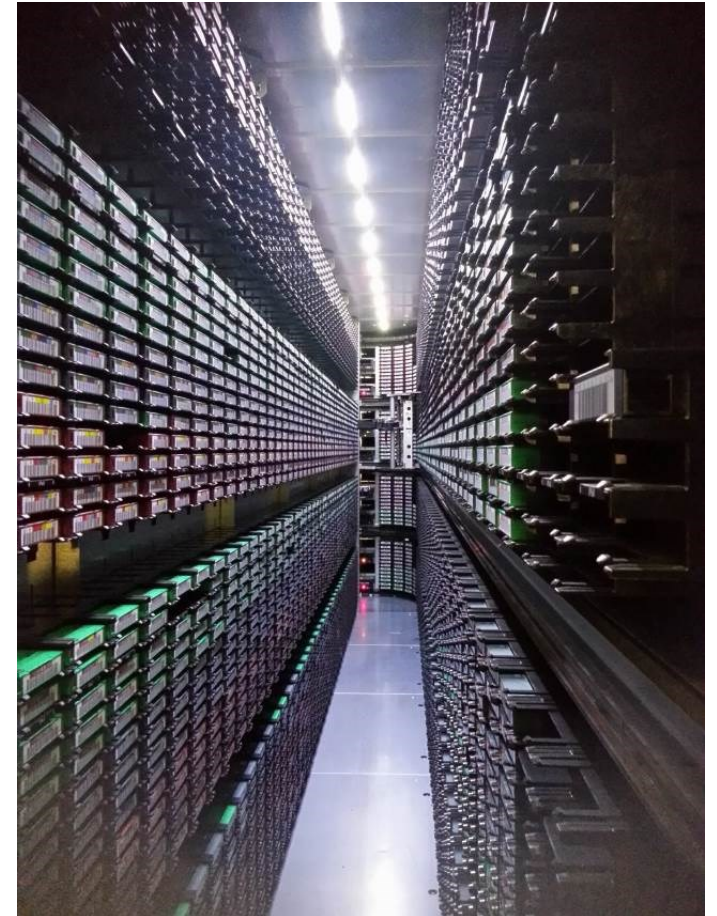
- Unbegrenzte Haltefrist (mindestens 25 Jahre)
 - Optionale Embargos
 - Vergabe eines DOI (DataCite)
 - Schnittstelle für Peer Review
- Auswahl von Lizenzen, kein CC0 erforderlich
- Metadaten indexiert (RADAR, DataCite, Google)

ROLLEN UND HIERARCHISCHE ORGANISATION DER INHALTE



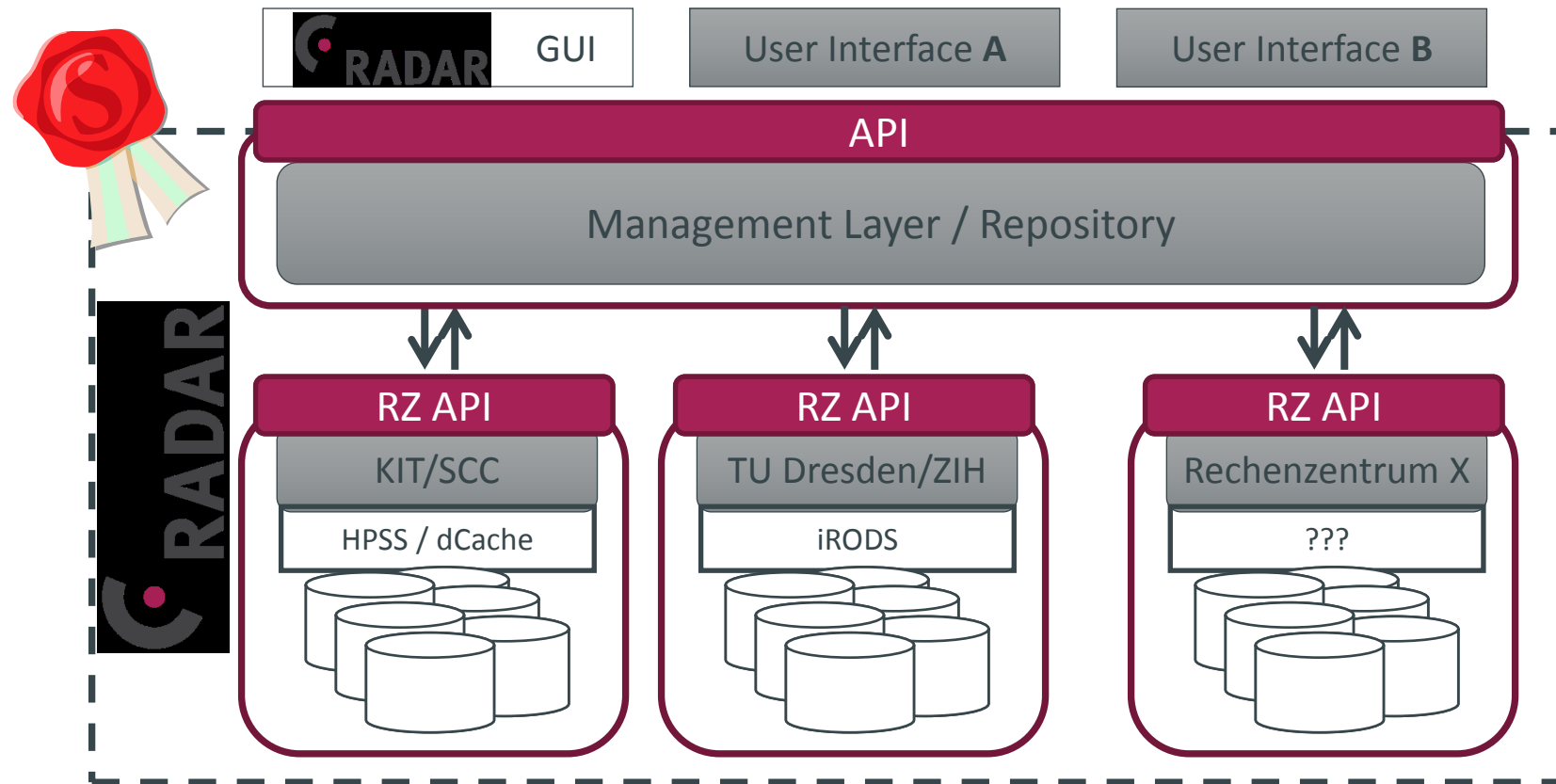
DATENSICHERHEIT

- Für archivierte/publierte Daten:
 - Datenpakete werden mit einer 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



Quelle: SCC, Karlsruher Institute for Technology (KIT)

SCHEMATISCHE ARCHITEKTUR



RADAR IN A NUTSHELL
VOM PROJEKT ZUM PRODUKT
AUSBLICK



KOOPERATIONSPARTNER



Beratung, Marketing und Öffentlichkeitsarbeit



Datenarchivierung und IT-Infrastruktur

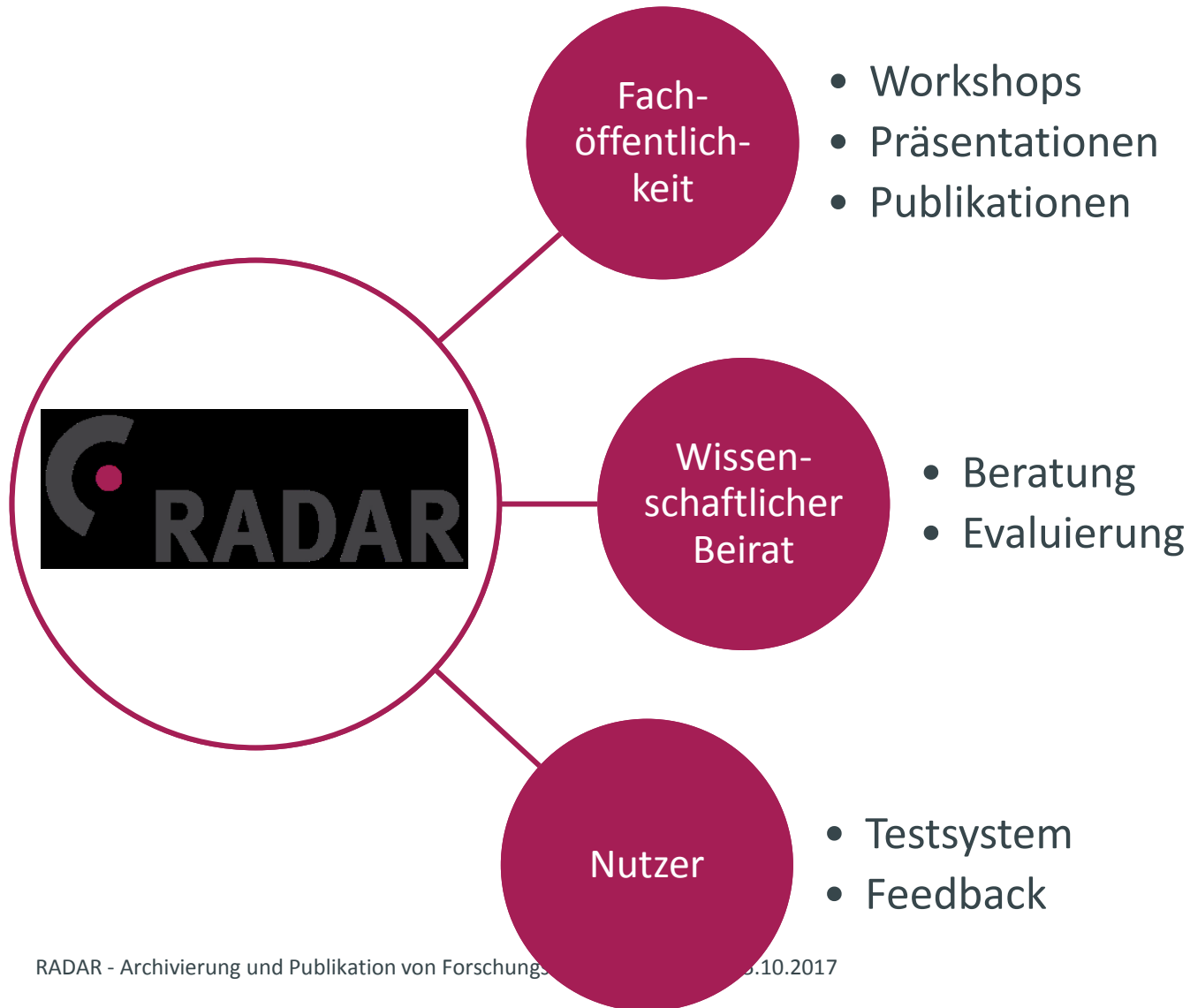


Datenarchivierung



Vergabe von DOI

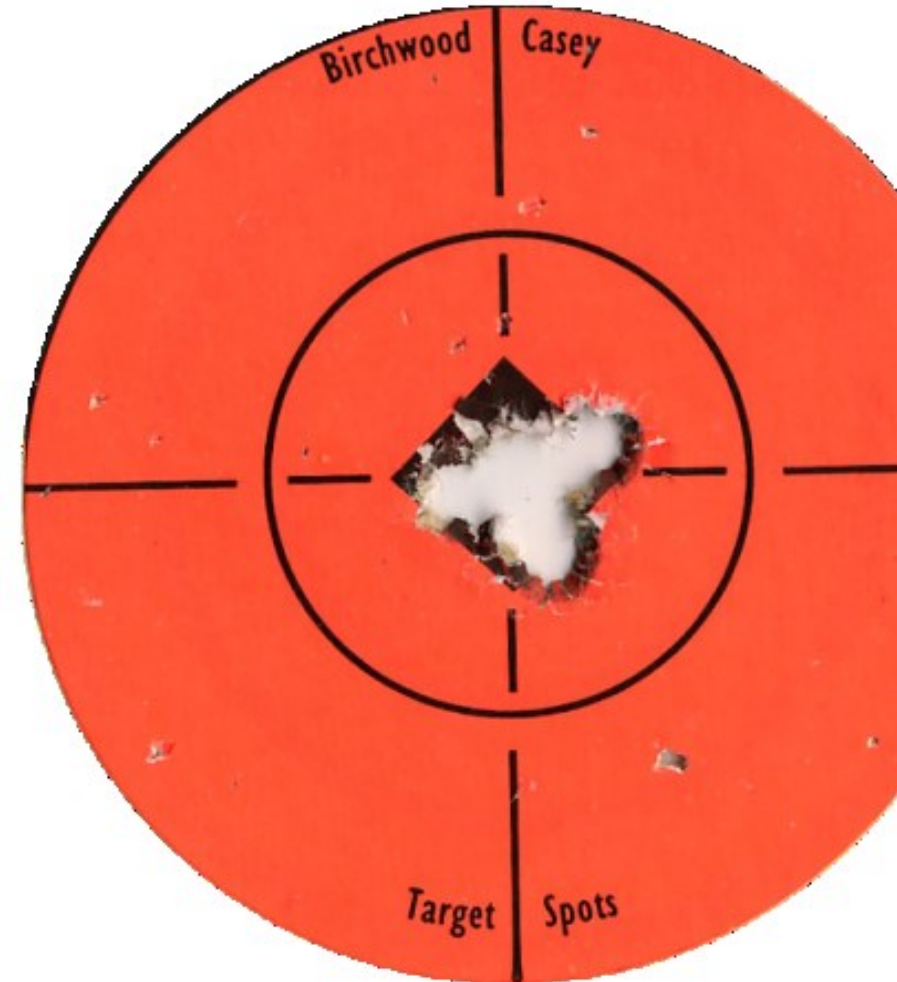
AUS DER COMMUNITY – FÜR DIE COMMUNITY



„Aufbau muss bedarfsgetrieben erfolgen“

ZIELGRUPPEN

- Hochschulen und außerhochschulische Forschungseinrichtungen
 - Institutionelles Angebot zum Forschungsdatenmanagement
 - Integration mit existierenden institutionellen Portalen
- Forscher
 - Publikation und Archivierung von Forschungsdaten aus Projekten
- Kultureinrichtungen
 - Langzeitarchivierung von Digitalisaten (Master)
 - Access-Kopien für Online-Zugriff



Quelle: Wikipedia Commons, Arthurrh. CC 3.0 BY SA

MARKT

427 Hochschulen in Deutschland, davon geschätzt ca. 250 forschende Hochschulen (2015/16)

- 955 außeruniversitäre Forschungseinrichtungen (2014)

Quelle: destatis

Ziel für die ersten 5 Jahre:

- 30 Verträge
- Gesamtvolumen der archivierten/publizierten Daten: 1 PB (= 1 Mio. G)



Quelle: Wikimedia Commons/Ildar Sagdejev, CC-BY-SA 3.0

PREIS FÜR „ARCHIVIERUNG“

Anteil Fixkosten	0,17 €
Variable Kosten	0,18 €
Risikozuschlag	0,04 €
Mehrwertsteuer	0,07 €

pro Jahr/GB (netto) **0,46 €**

Hinzu kommt jährliche Grundgebühr von 500,00 € (netto) pro Vertrag

- temporärer Speicher (200 GB)
- Vertragsmanagement

PREIS FÜR „ARCHIVIERUNG MIT PUBLIKATION“ (EINMALZAHLUNG)

Anteil Fixkosten (5 Jahre)	1,13 €
Variable Kosten	2,97 €
Risikozuschlag (200%)	2,27 €
Mehrwertsteuer	1,21 €

25+ Jahre/GB **7,58 €**

- Berücksichtigung der Personalkosten nur in den ersten 5 Jahren
- Risikozuschlag bei 200%
- **34,7 % Rabatt** bei Speicherung von 1 GB über 25 Jahre: 7,58 € statt $25 \cdot 0,46 \text{ €} = 11,50 \text{ €}$

UND DIE ANDEREN?

- Freie Lizenzwahl (CC0 nur für Met
- Daten unterliegen deutscher Juris
- Neben Datenpublikation auch reir
- 25 Jahre garantierte Haltefrist für
- Schnittstellen zur Einbindung in ei
- Rollenkonzept mit delegierter Adr
- Datensparsamkeit



RADAR IN A NUTSHELL
VOM PROJEKT ZUM PRODUKT
AUSBLICK



WAS KOMMT ALS NÄCHSTES?

- Weitere Nutzer gewinnen
 - Geschäftsplan sieht 3 Nutzer im ersten Jahr vor (bisher vier Verträge)
 - Hochschulen als erste Nutzergruppe
 - Leibniz-Institute folgen als zweite Kundengruppe
- Weiterentwicklung des Geschäftsmodells
 - Konsortien
 - Inhouse-/Hybrid-Nutzung
- Zertifizierung (Core Trust Seal)



Quelle: Wikimedia Commons/User:Tasma3197, CC-BY-SA 3.0

REPRISE: LONG TAIL

„The amount of data needed for research has grown tenfold within the last four years”

Mikko Hakala, Aalto Univer



FUNKTIONALE ERWEITERUNGEN DER SOFTWARE

- Effiziente Unterstützung großer Datenvolumina
 - Alternative Datenübertragungsmethoden (GridFTP)
 - Projektion
 - „Streaming“
- Versionierung
- Verbesserte Unterstützung fachspezifischer Metadaten
- EduGAIN (bisher DFN-AAI)
- Lösung für besonders schutzbedürftige Daten (personenbezogen, Auftragsforschung)
- Erweiterte Administrationsmöglichkeiten



FRAGEN?

matthias.razum@fiz-karlsruhe.de
Twitter: @RADAR_Service



Except where otherwise noted,
content on this site is licensed under
a Creative Commons Attribution 4.0
International License.

