FIZ Karlsruhe

Leibniz-Institut für Informationsinfrastruktur

WHAT DOES THE PRACTICE OF OPEN RESEARCH DATA COST?
Developing a Financing Model for RADAR

Matthias Razum
Open Science Days, Berlin - October 16, 2017

SCOPING YOUR SERVICE
COSTS
BUSINESS PLAN
PRICING

RADAR – RESEARCH DATA REPOSITORY

RADAR (**R**esea**r**ch **Da**ta **R**epository) is a

- cross-disciplinary research data repository

- for the archival and publication of research data

- of completed scientific studies and projects.

Funded by DFG from September 2013 until August 2016

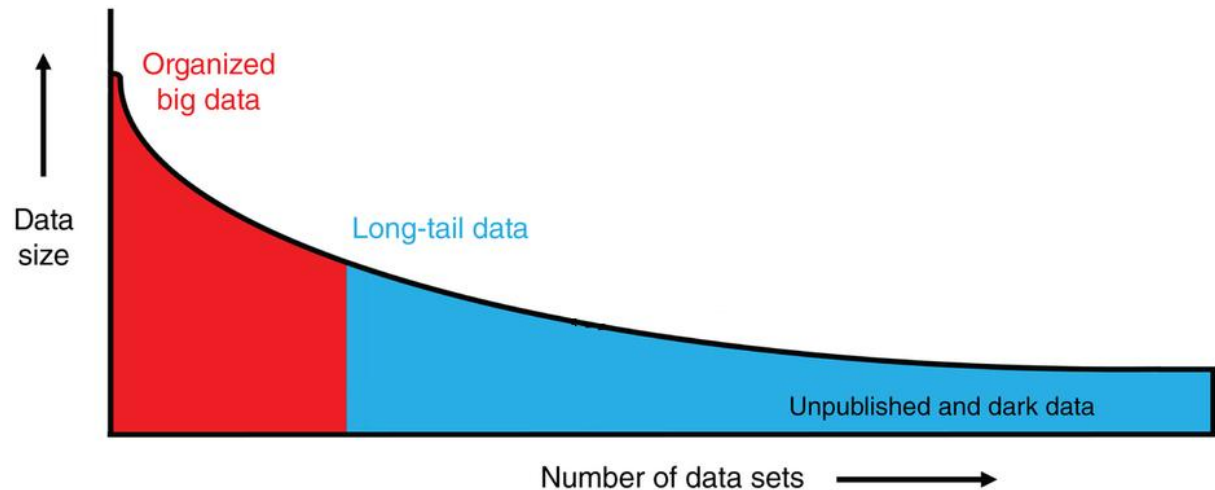No competition with established disciplinary data repositories

- Emphasis on „Long tail"

What does the practice of Open Research Data cost?   Open Science Days 2017, Berlin

"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

FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

COOPERATION PARTNERS

TIB LEIBNIZ INFORMATION CENTRE FOR SCIENCE AND TECHNOLOGY UNIVERSITY LIBRARY — Consulting service, marketing and publicity (together with FIZ Karlsruhe)

SCC — Data storage and IT infrastructure

ZIH Zentrum für Informationsdienste und Hochleistungsrechnen — Data storage

DataCite FIND, ACCESS, AND REUSE DATA — Registration of DOI

What does the practice of Open Research Data cost?   Open Science Days 2017, Berlin

FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur
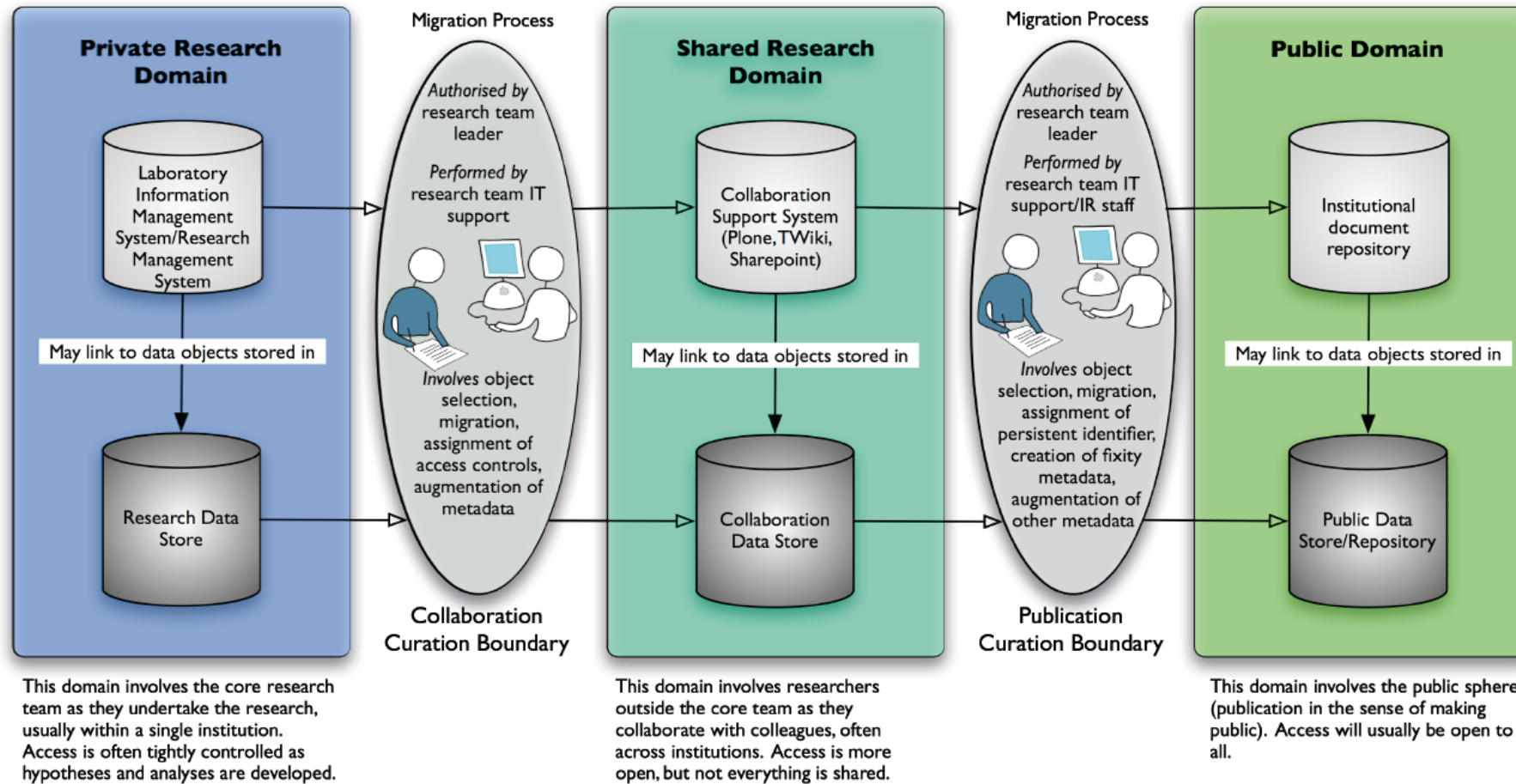
Data Archival

- „dark archive"

- Flexible retention period (5, 10, 15 years)

- Data providers control access rights (private/shared/public)

Data Archival with Publication

- Unlimited retention period (25+ years)

- Optional embargos

- DataCite DOI

- Interface for peer review

- Choice of licenses, no CC0 required

- Metadata indexed, e.g. by RADAR, DataCite, Clariavate, Google

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

FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

DATA SECURITY

- For each dataset, a checksum is calculated upon ingest

- The checksum is re-calculated after writing to tape ('read after write')

- Three copies at three locations (2x Karlsruhe, 1x Dresden)

- Different hardware, software and administration

- Routine migration of data to new storage media, including fixity check
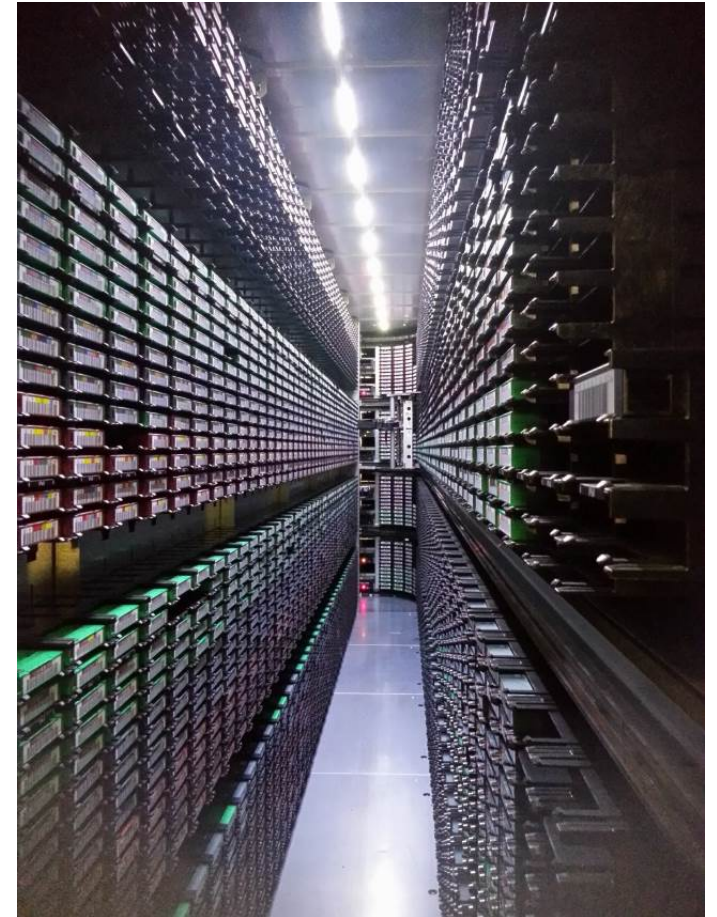
Quelle: SCC, Karlsruher Institute for Technology (KIT)

What does the practice of Open Research Data cost?   Open Science Days 2017, Berlin

FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

LEGAL ISSUE

- For offers with costs, liability for cardinal obligations cannot be limited (when using standardized contracts)

- Liability risks include loss or corruption of archived data

- The monetary value of research data is hard to calculate, therefore little indication of the amount of damage

- Nearly impossible to find an insurance willing to cover such a risk

→ Organizational and technical solution to minimize risks

Quelle: Wikimedia Commons/User Walknboston, CC-BY-SA 2.0

FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

Within the scope

- Data of completed research projects

- Bitstream preservation

- Three copies, two data centers

- 25+ year retention phase for published data

- 'Long-tail' data

Outside of the scope

- Management of active data

- Data curation

- Functional long-term preservation

- Regular fixity checks

- Insurance

- No 'big data' disciplines

Quelle: Pixabay/User moritz320, CC0

FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

SCOPING YOUR SERVICE
COSTS
BUSINESS PLAN
PRICING

FIXED COSTS PER ANNUM

Personnel costs (24 PM EG 13)

- Program Manager (12 PM)   132,000.00 €
- System administration (3 PM)   33,000.00 €
- Software maintenance and development (9 PM)   99,000.00 €

Material costs

- Temporary storage   1,694.00 €
- Virtual machines   10,648.00 €
- Licenses and maintenance contracts   1,000.00 €
- Travel expenses   10,000.00 €
- Information materials   5,000.00 €
- Other   5,000.00 €
                                                                        ——————————
                                                                                  297,342.00 €

What does the practice of Open Research Data cost?   Open Science Days 2017, Berlin

FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

FIXED COSTS PER ANNUM

Personnel costs (24 PM EG 13)

- Program Manager (12 PM)   132,000.00 €
  - System administration (3 PM)   33,000.00 €
  - Software maintenance and development (9 PM) 99,000.00 €

Material costs

- Temporary storage   1,694.00 €
- Virtual machines   10,648.00 €
- Licenses and maintenance contracts   1,000.00 €
- Travel expenses   10,000.00 €
- Information materials   5,000.00 €
- Other   5,000.00 €

                165,342.00 €

What does the practice of Open Research Data cost? Open Science Days 2017, Berlin

VARIABLE COSTS

Archive storage (Tape) – per annum and GB	0,18 €
Temporary storage – per annum and GB	0,42 €
Additional virtual machine – per annum (currently none)	2.662,00 €

FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

SCOPING YOUR SERVICE

COSTS

BUSINESS PLAN

PRICING

TARGET GROUPS

- Universities and research institutions

  - Institutional offering for research data management

  - Integration with existing institutional portals

- Researchers

  - Publication and archiving of research data from projects

- GLAM (cultural heritage)

  - Long-term preservation of digitized materials (master scans)

  - Access copies for online access

Quelle: Wikipedia Commons, Arthurrh. CC 3.0 BY SA

FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

POTENTIAL MARKET

427 universities and colleges in Germany, of which ~250 produce research data (2015/16)

- ~1000 research institutions outside of universities (2014)

Source: destatis

Goal for the first five years:

- 30 contracts

- 1 PB (= 1 million GB) of archived/published data

What does the practice of Open Research Data cost?   Open Science Days 2017, Berlin

FIZ Karlsruhe

Leibniz-Institut für Informationsinfrastruktur

PROJECTIONS

	Year 1	Year 2	Year 3	Year 4	Year 5
Number of contracts	3	8	13	20	30
Data volume	13 TB	51 TB	205 TB	484 TB	911 TB
Costs	167,157 €	169,831 €	172,570 €	175,372 €	178,240 €
Revenues	4,247 €	15,885 €	55,343 €	132,101 €	258,242 €
Marginal return	**- 162,910 €**	**- 153,946 €**	**- 117,227 €**	**- 43,271 €**	**80,002 €**

FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

SCOPING YOUR SERVICE
COSTS
BUSINESS PLAN
PRICING

Fixed costs	0,17 €
Variable costs	0,18 €
Risk loading	0,04 €
VAT	0,07 €
per year/GB	**0,46 €**

Additional yearly service charge of 500.00 € (net price) per contract includes

- Temporary storage of 200 GB

- Contract management and invoicing

FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

PRICING FOR DATA PUBLICATION (ONE-TIME PAYMENT)

Fixed costs	1.13 €
Variable costs	2.97 €
Risk loading	2.27 €
VAT	1.21 €
25+ years/GB	**7.58 €**

- Personnel costs are included for the first five years only

- Increased risk loading of 200%

- **34,7 % discount** (7.58 € instead of 25*0.46 € = 11.50 € per year/GB)

FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

FURTHER EVOLVEMENT OF BUSINESS PLAN

- Discount for data volumes > 50 TB

- Offer for consortia

- Open source with maintenance contracts

- Local storage of one copy/all copies

- One-time payment for archived data

- Annual fee for published data

What does the practice of Open Research Data cost?   Open Science Days 2017, Berlin

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

FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

Mitglied der
Leibniz
Leibniz-Gemeinschaft