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) 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“
THE ‘LONG TAIL’ IS RELEVANT

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


Source: Ferguson et al. (2014): Big data from small data: data-sharing in the ‘long tail’ of neuroscience. DOI: 10.1038/nn.3838
Consulting service, marketing and publicity (together with FIZ Karlsruhe)

Data storage and IT infrastructure

Data storage

Registration of DOI

This domain involves the core research team as they undertake the research, usually within a single institution. Access is often tightly controlled as hypotheses and analyses are developed.

Involves object selection, migration, assignment of access controls, augmentation of metadata.

May link to data objects stored in

This domain involves researchers outside the core team as they collaborate with colleagues, often across institutions. Access is more open, but not everything is shared.

This domain involves the public sphere (publication in the sense of making public). Access will usually be open to all.

RADAR SERVICE LEVELS

**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, Clarivate, Google

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

Total 297,342.00 €
### 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

**Total Fixed Costs Per Annum:** €165,342.00
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 €
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
427 universities and colleges in Germany, of which ~250 produce research data (2015/16)

- ~1000 research institutions outside of universities (2014)

Goal for the first five years:

- 30 contracts
- 1 PB (= 1 million GB) of archived/published data
## PROJECTIONS

<table>
<thead>
<tr>
<th></th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Year 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of contracts</td>
<td>3</td>
<td>8</td>
<td>13</td>
<td>20</td>
<td>30</td>
</tr>
<tr>
<td>Data volume</td>
<td>13 TB</td>
<td>51 TB</td>
<td>205 TB</td>
<td>484 TB</td>
<td>911 TB</td>
</tr>
<tr>
<td>Costs</td>
<td>167,157 €</td>
<td>169,831 €</td>
<td>172,570 €</td>
<td>175,372 €</td>
<td>178,240 €</td>
</tr>
<tr>
<td>Revenues</td>
<td>4,247 €</td>
<td>15,885 €</td>
<td>55,343 €</td>
<td>132,101 €</td>
<td>258,242 €</td>
</tr>
<tr>
<td>Marginal return</td>
<td>-162,910 €</td>
<td>-153,946 €</td>
<td>-117,227 €</td>
<td>-43,271 €</td>
<td>80,002 €</td>
</tr>
</tbody>
</table>
SCOPING YOUR SERVICE
COSTS
BUSINESS PLAN
PRICING
PRICING FOR ARCHIVAL

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

• Temporary storage of 200 GB
• Contract management and invoicing

<table>
<thead>
<tr>
<th>Cost Type</th>
<th>Cost (€)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed costs</td>
<td>0.17</td>
</tr>
<tr>
<td>Variable costs</td>
<td>0.18</td>
</tr>
<tr>
<td>Risk loading</td>
<td>0.04</td>
</tr>
<tr>
<td>VAT</td>
<td>0.07</td>
</tr>
</tbody>
</table>

per year/GB 0.46 €
Pricing for Data Publication (One-Time Payment)

- 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)

<table>
<thead>
<tr>
<th>Costs</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed costs</td>
<td>1.13 €</td>
</tr>
<tr>
<td>Variable costs</td>
<td>2.97 €</td>
</tr>
<tr>
<td>Risk loading</td>
<td>2.27 €</td>
</tr>
<tr>
<td>VAT</td>
<td>1.21 €</td>
</tr>
</tbody>
</table>

Total for 25+ years/GB: **7.58 €**
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