Cost and Value analysis of digital data archiving

ANNA PALAIIOLOGK
Detailed and meaningful **cost information** allows:

- more accurate planning
- better forecasting and control
- more accountability and transparency
- prioritise/control the level of ambition - realistic strategy (e.g. collection levels and preservation aims, quality-quantity balance, etc)
Challenges + Terminology

- funding does not grow in line with information growth
- curation vs. storing of the data
- acquisition and ingest
- guidelines vs. regulation on preferred formats
- legal requirements and grant terms
- access - most variable area of costs
DANS case study

- Data Archiving & Networked Services (DANS) is an institute of the Dutch Royal Academy of Arts and Sciences (KNAW)
- an independent digital archive
- collection: 14,000 datasets (1.5 TB) available to public and 10 datasets (20 TB) not available to the public
- 51 employees
- work processes based on Open Archival Information System (OAIS) - ISO 14721:2003
- mixed budget of approximately 3.8 million euro/year
- costs measured in Euros per dataset
- next slide depicts processes and vision of DANS
IMPROVE RESEARCH DATA INFRASTRUCTURE IN SOCIAL SCIENCES & HUMANITIES

MISSION

INNOVATION & GROWTH

Big institutional collectors make their data available for free through DANS

Leading partner in standards of infrastructure (provide innovative solutions)

Increased use & re-use of data stored in EASY

Complete coverage of fields we are active in

CUSTOMERS

Users, clients and partners satisfaction

Increase awareness amongst research community, students and partners

Increased demand for consultancy

Increased number of datasets available to end user

INTERNAL PROCESSES

Effective administrative management

Compliance to international standards

Efficiency of archiving process

INNOVATION & GROWTH

Dataset access aligned to national and international law

All datasets available from one portal

SUPPORTERS

Sustainable sources of revenue

Growth in the number of supporters
Budget distribution

Staff is the major resource pool in digital archiving, up to 65–70% of total expenses.

<table>
<thead>
<tr>
<th>Data Acquisition</th>
<th>Office</th>
<th>IT services and equipment</th>
<th>Staff</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>14,2%</td>
<td>14,3%</td>
<td>7%</td>
<td>64,5%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Staff needs to do tasks bringing the most value. Rest needs to be automated.
% of money spent on each activity
Workload allocation per discipline

- Ingest
- Archival Storage
- Data Management
- Access
- Preservation
- Archival Administration

Categories:
- History
- Social Sciences
- Archaeology
Key findings

In long term data archiving:

1. Up-front costs of acquisition and ingest of data (70-90% of total) dominate the long-term costs of storage and preservation.

2. Up-front costs dominated by staff time rather than hardware or other technology costs.

3. Long-term costs scale weakly, if at all, with the size of an archive. Preserving 10 TB is not that much more expensive than preserving 10 GB.
ABC methodology

Resources | Costs | Cost drivers
---|---|---
Activities | Activity cost drivers
Cost Objects
ABC data collection “How-To”

- Dedicating a person to be responsible for collecting the cost information
- Do not overwhelm staff with information
- Do not expect all staff to be “on the same page” from the beginning
- Run a trial for a day or a week
- Ask staff to report separately on activities outside the Model
- Allow for a general comments field
- Leave, sickness or absence should be specified separately
Value + Economic Impact Analysis

Methods being applied to:

- report published
- in progress
- in progress
Benefits data collection

Desk-research sources:
- Organisation and infrastructure evaluation reports
- Documentation on data usage and users
- Internal (management) reports
- Annual and mid-term reports

Interviews with:
- Organisation management and staff
- Policy makers and practitioners
- Government institutions
- Non-academic and private sector representatives

Online-survey addressed to:
- Depositors and users
**Economic measures of value**

- **Investment value**: annual operational funding & the costs that depositors face in preparing data for deposit and in making that deposit

- **Use value**: average user access costs \( \times \) number of users

- **Contingent value**: the amount users are "willing to pay" or "willing to accept" in return for giving up access

- **Efficiency gain**: user estimates of time saved by using the Data Service resources

- **Return on investment**: estimated return with time (30yrs)
### Investment & Use Value

<table>
<thead>
<tr>
<th>Description</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Investment Value</strong></td>
<td>Amount spent on producing the good or service</td>
</tr>
<tr>
<td><strong>Use Value</strong></td>
<td>Amount spent by users to obtain the good or service</td>
</tr>
</tbody>
</table>

### Contingent Value (Stated)

- **Willingness to Pay**: Maximum amount user would be willing to pay
- **Consumer Surplus**: Total willingness to pay minus the cost of obtaining
- **Net Economic Value**: Consumer surplus minus the cost of supplying

### Efficiency Impact (Estimates)

- **Willingness to Accept**: Minimum amount user would be willing to accept to forego good or service
- **Survey User Community**
  - Estimated value of efficiency gains due to using service
- **Wider User Community**
  - Estimated value of efficiency gains due to using service
- **Wider Research Community**
  - Estimated increase in return on investment in data creation arising from the additional use facilitated by service

### Return on Investment (Scenarios)

- **Increased Return on Investment in Data Creation**
  - Estimated increase in return on investment in data creation arising from the additional use facilitated by service

### Wider Impacts (Not Measured)

- **Society**
  - ?

---

Survey User Community (registered users)
Investment & Use Value (Direct)

- Investment Value: £23m per annum
- Use Value: £24m per annum

Contingent Value (Stated)

- Survey User Community (active registered users excluding school and under-graduate students)

Willingness to Pay
- £25m per annum

Consumer Surplus
- £21m per annum

Net Economic Value
- £18m per annum
  (More than 5 times ESDS operational budget)

Efficiency Impact (Estimates)

- Willingness to Accept: £81m – £111m per annum

Survey User Community Efficiency Gain
- £68m - £112m per annum

Wider User Community Efficiency Gain
- £100m plus per annum

Return on Investment (Scenarios)

- Wider User Community
- Wider Research Community

Increased Return on Investment in Data Creation over 30 years
- £58m - £233m (NPV)
  (2.5-fold to 10-fold RoI)

Wider Impacts (Not Measured)

- Society

?
Next steps

**Costs**
- Refine cost drivers
- Allocate the other-than-staff costs to activities
- Experiment with other cost objects
- Develop the “matrix of dataset complexity”
- Apply economic adjustments
- Test reliability and accuracy
- Develop/Customise software to make ABC easy to use

**Value/Benefits**
- Develop the benefits framework further
- Collect more diverse/detailed data
- Verify results
References


