Collaboration to Clarify the Costs of Curation

Jisc Digital Festival, 11-12 March 2014

The Economics of Digital Curation: Crunching the Numbers, Comparing the Costs

Neil Grindley – Jisc & 4C Project Coordinator
Project Summary

The Collaboration to Clarify the Costs of Curation (4C) project will help organisations across Europe (and beyond) to more effectively invest in digital curation and preservation.

Vision

The 4C vision is to create a better understanding of digital curation costs through collaboration.

Mission

Our mission is to provide useful, useable resources which support the process of cost management in digital curation.
**Assessment**

*Tasks*
- Assess cost models & strategies
- Examine good practice
- Analyse requirements
- Integrate components
- Produce guidance & briefing materials
- Setup costs exchange

**Engagement**

*Tasks*
- Engage stakeholders
- Raise awareness
- Organise meetings
- Promote Research & Innovation
- Build community network

**Enhancement**

*Tasks*
- Examine and refine related concepts
  - Value
  - Risk
  - Benefits
  - Sustainability
  - Economic Reference Model

**Networking & Coordination**

**Outputs**

- Reports for European Commission
- Curation Costs Exchange
- Submission of Roadmap to the EC

**Events, Workshops, Meetings & Reports**

- Affiliate Partners & Stakeholders

**Jisc**

**Project Coordination**

*Tasks*
- Project meetings
- Project reporting
- EC liaison
- Budget oversight
- Outputs QA
Collaboration to Clarify the Costs of Curation

Partners

**Engagement**
- Jisc
- Deutsche Nationalbibliothek
- nestor
- Digital Preservation Coalition
- KEEP SOLUTIONS

**Assessment**
- Statens Arkiver
- THE ROYAL LIBRARY
- DANS
- Jisc
- sba-research.org

**Enhancement**
- National Library of Estonia
- University of Essex
- UK-Data Archive
- inesc id tisboa
Why should we concern ourselves about the cost of curation? (What are the stakeholders saying ...)

- Understanding the cost of preservation may mean we can offer realistic and cost effective curation services to others.

- Understanding costs can support strategic planning.

- Understanding costs can support tactical decision-making.

- Understanding costs can provide evidence of cost-effectiveness and value.

- Clarifying and publishing the cost of digital curation can be used to enhance our organisation’s credibility. But this must be done along with the context of how the costs were calculated.

- Understanding economic drivers can help to strategically align an organisation.
There is a sizeable canon of research into cost modelling for digital curation. This research has tended to emphasize the cost and complexity of digital curation and preservation, but the research is in many ways preliminary and there has been little uptake of the tools and methods that have been developed—for example, tools to manage and estimate costs have not been integrated into other digital curation processes or tools. The question is why? That’s where the 4C project comes in.

The 4C project’s mission is to provide useful, useable resources which support the process of cost management in digital curation. Major outputs from the project will include:
Collaboration to Clarify the Costs of Curation

What we don’t want to do ...!

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Acronym</th>
<th>Owner</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Test bed Cost Model for Digital Preservation</td>
<td>T-CMDP</td>
<td>National Archives of the Netherlands</td>
</tr>
<tr>
<td>2</td>
<td>NASA Cost Estimation Tool</td>
<td>NASA-CET</td>
<td>National Aeronautics &amp; Space Administration</td>
</tr>
<tr>
<td>3</td>
<td>LIFE$^3$ Costing Model</td>
<td>LIFE3</td>
<td>University College London and The British Library</td>
</tr>
<tr>
<td>4</td>
<td>Keeping Research Data Safe</td>
<td>KRDS</td>
<td>Charles Beagrie Limited</td>
</tr>
<tr>
<td>5</td>
<td>Cost Model for Digital Archiving</td>
<td>CMDA</td>
<td>Data Archiving and Networked Services (DANS)</td>
</tr>
<tr>
<td>6</td>
<td>Cost Model for Digital Preservation</td>
<td>CMDP</td>
<td>Danish National Archives and The Royal Library, DK</td>
</tr>
<tr>
<td>7</td>
<td>DP4lib Cost Model</td>
<td>DP4lib</td>
<td>German National Library</td>
</tr>
<tr>
<td>8</td>
<td>PrestoPRIME Cost Model for Digital Storage</td>
<td>PP-CMDS</td>
<td>The PrestoPRIME project</td>
</tr>
<tr>
<td>9</td>
<td>Total Cost of Preservation</td>
<td>CDL-TCP</td>
<td>California Digital Library</td>
</tr>
<tr>
<td>10</td>
<td>Economic Model of Long-Term Storage</td>
<td>EMLTS</td>
<td>Rosenthal, D.</td>
</tr>
</tbody>
</table>

Come up with another Cost Model ...!
Only 15% of people in the 4C stakeholder consultation indicated that they had tried to use a cost model.

What can be done to ensure that models are developed in line with users’ needs?

15 drivers for development are listed in the Needs & Gaps report along with 11 recommendations.

Draft good practice proposals for cost model developers

1. Use a standardised definition of digital curation

2. Limit the purpose of the model and define clearly the expected users of the model and its scaling capacity

3. Start out and continue to prioritise simplicity; be explicit about limitations on accuracy

4. Limit the time scope

5. Use simple formulae

6. Implement the model in a simple and widespread tool
We are working on a Cost Concept Model ...

And a Gateway Specification...
So the Cost Concept Model and Gateway Specification should help in various ways ...

- It will be a one-stop shop for thinking about the components of cost models
- It will enable comparisons of various existing cost models
- It will help people to design their own modelling approaches
- It will help to build consensus around definitions and terminology
- It will be an accommodating structure to showcase and raise awareness of future cost modelling work
- It will feed into the 4C Roadmap work which will recommend future activity in relation to further clarifying the economics of curation
<table>
<thead>
<tr>
<th>authenticity</th>
<th>reputation</th>
</tr>
</thead>
<tbody>
<tr>
<td>benefit</td>
<td>risk</td>
</tr>
<tr>
<td>efficiency</td>
<td>sensitivity</td>
</tr>
<tr>
<td>impact</td>
<td>skills</td>
</tr>
<tr>
<td>innovation</td>
<td>sustainability</td>
</tr>
<tr>
<td>interoperability</td>
<td>transparency</td>
</tr>
<tr>
<td>quality</td>
<td>trustworthiness</td>
</tr>
<tr>
<td></td>
<td>value</td>
</tr>
</tbody>
</table>
Other Concepts and models we are developing ...

A Taxonomy of Indirect Economic Determinants

authenticity  reputation
benefit  risk
efficiency  sensitivity
impact  skills
innovation  sustainability
interoperability  transparency
quality  trustworthiness
value
Collaboration to Clarify the Costs of Curation

There will be various 4C outputs including reports and a Roadmap but let’s focus on two in particular ...

The Economic Sustainability Reference Model (ESRM)

The Curation Costs Exchange (CCEx)
We need to think carefully about what exactly is the problem that we are trying to solve ...

If they want to, organisations can work out how much it costs them to manage their digital assets
### Collaboration to Clarify the Costs of Curation

**4C Data Gathering Exercise**  
**Organisation A**

<table>
<thead>
<tr>
<th>Curation Categories</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-Ingest</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Ingest</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Preservation Planning</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Data Management</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Archival Storage</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Access</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Administration</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

**Accounting Principles**

- **Labour Direct**: 0.2, 0.0, 0.7, 0.6, 1.3
- **Labour Indirect**: 0.1, 0.0, 0.7, 0.6
- **Capital Direct**: 1.3
- **Capital Indirect**: 0.0, 0.0, 0.3
- **Note**: €5,000, €5,000, €1,000, €1,000

**Time Period**  
2012

**Total Cost**  
€252,000
Collaboration to Clarify the Costs of Curation

4C Data Gathering Exercise
Organisation B

Curation Categories
- Ingest
- Curation
- Access

Cost Categories
- Hardware
- Software
- Employment
- Accommodation
- External Services
- Transfer

Accounting Principles
- Direct
- Costs of Service
- Absorbed Indirect Costs of Service
- Unabsorbed Indirect Costs of Service

Time Period
2012

Total Cost
€645,683.26
Collaboration to Clarify the Costs of Curation

4C Data Gathering Exercise
Organisation C

Curation Categories
- Ingest
- Data Management
- Archival Storage
- Preservation Planning
- Access
- Administration
- Common Services

Accounting Principles
- Labour Costs
- Capital Costs
- Offset By Revenue

Size of Collection
393 TB

Time Period
2012

Total Cost
€15,800,000
Collaboration to Clarify the Costs of Curation

4C Data Gathering Exercise
Organisation D

Curation Categories
- Ingest
- Archival Storage
- Metadata Management
- Access
- Administration

Cost Categories
- Product/Service
- Hardware

Time Period

Total Cost
€349,665
4C Data Gathering Exercise
Organisation E

Curation Categories
Digital Archiving

Cost Categories
- Hardware
- Software Maintenance
- Software Development
- Staff/Other

Time Period 2007
Total Cost 123,000

Time Period 2012
Total Cost 348,500
Collaboration to Clarify the Costs of Curation

4C Data Gathering Exercise
Organisation F

Accounting Principles
Development & Improvement
Operation

Staff
Development, Technical Support
Training, Communications, Public Relations

Expenses
Software Design
Software Licenses
Support
External Development
Hardware Purchase
Hardware Operating costs
Graphic Design

Curation Categories
Long term Digital Preservation

Size of Collection
2 TB

Assumption
€8k per TB per year for Storage Costs

Time Period
2007-2012

Total Cost
€205,000
Collaboration to Clarify the Costs of Curation

4C Data Gathering Exercise
Organisation G

Curation Categories
Digital Archiving

Content Management
Data Development
Hardware
Production
Systems Development
Delivery
User Support
Overhead
Operations Management

Time Period
2012-13

Total Cost
€ 3,130,110
Collaboration to Clarify the Costs of Curation

So ... What exactly are the problems we need to tackle?

• The random numbers problem - How can we meaningfully compare the numbers that we end up with? \textit{[cost data]}

• Activity based costing versus financial accounting methods

• Describing what the organisation does \textit{[cost metadata]}

• Describing the amount and type of data that is being looked after \textit{[cost metadata]}

• Sensitivity around data – Many organisations are not particularly happy to broadcast what it costs them to manage their data. How can we effectively anonymise the sharing of data?

• Complexity - The detail builds up very quickly across different organisations and it doesn’t map together easily

• And we somehow have to make sure that the benefits are presented alongside the costs
Collaboration to Clarify the Costs of Curation

What the CCEx can do for you

- Input your cost data
- Compare your expenditure with others
- Know how others are spending their budget
- Browse through cost models, descriptions, and comparisons
- Know what cost models you should use
- Know about other's risks and benefits
- Connect with others with the same motivations
- Stay informed about all news and events
- Know about suppliers and their services
**My curation costs**

Information about the costs of curation on your organization.

### Ingest sub-activities

<table>
<thead>
<tr>
<th>Name</th>
<th>Cost</th>
<th>Breakdown</th>
<th>OAIS entity</th>
<th>OAIS function</th>
<th>Raw data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Validation</td>
<td>14100</td>
<td>0 sub-activities</td>
<td>ingest</td>
<td>validation</td>
<td>✓</td>
</tr>
<tr>
<td>Incorporation</td>
<td>2200</td>
<td>0 sub-activities</td>
<td>ingest</td>
<td>incorporation</td>
<td>✓</td>
</tr>
</tbody>
</table>

*Sum 16300 €*
Collaboration to Clarify the Costs of Curation

Compare your expenditure with others

http://cc-ex.org/comparison

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Ingest

- Fixity check
- Format validation

Average
Yours
The Economic Sustainability Reference Model proposes that a sustainability strategy requires consideration of four categories of issues:

- The Economic Lifecycle
- Sustainability Conditions
- Key Entities
- Economic Uncertainties
The activity of digital curation is assumed to be the central active component and the engine that will ensure the sustainability of digital assets.
Investment into curation will in turn facilitate use (or the potential for use)
And use (or the potential for use) will realise value, thereby delivering a return on the investment.
This could play out in a linear fashion with assets being created, curated and then deleted according to a retention schedule. But in the context of sustainability, it is more likely to be a cyclical process.
There will be a gap in the cycle where a technical or business issue introduces a threat to the continued viability of the assets. This becomes a decision point ... Do we allocate more resources to tackling the problem?
Sustainability Conditions

Five Sustainability Conditions are set out to maximise the prospects for sustaining assets.
The assets must be understood (or perceived) to have tangible or intangible value
Relevant stakeholders must be sufficiently motivated to support curation
Where resources are scarce then discretion must be used to prioritise curation of the most valuable assets
The organisation should have an appropriate mandate; a supportive governance structure; and be optimally configured to sustain the assets.
Collaboration to Clarify the Costs of Curation

There must be a sufficient flow of ongoing resources (including financial and human capital) to achieve long-term goals.
Key Entities

Three Key Entities are set out which are found in all digital curation contexts. Sustainability requires the nature of these entities to be understood.
Key Entities

Three Key Entities are set out which are found in all digital curation contexts. Sustainability requires the nature of these entities to be understood.

**ASSETS**

Every type of digital asset exhibits various attributes or properties that to a greater or lesser extent may affect the how they are curated.

**STAKEHOLDERS**

The stakeholder ecosystem for digital assets can be complex and the supply side and demand side should be understood in relation to who is undertaking the curation for the benefit of whom.

**PROCESSES**

The processes involved must be capable of (and optimised for) efficiently enhancing the value of the assets.
The inclusion of Economic Uncertainties is an acknowledgement that even the best sustainability strategy cannot accurately predict the future and that some expectation or mitigation of uncertainty (both threats and opportunities) should be built into the strategy where possible.