

Collaboration to Clarify the Cost of Curation



Report of Workshop #2 at IDCC 2014 on 24th February 2014 at Omni Hotel, San Francisco, CA

Project funded by the European Commission within the Seventh Framework Programme		
Dissemination Level		
PU	Public	✓
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Version History

Version	Date	Changed pages / reason	Modified by
0.01	06 March 2014	First draft	KH
0.02	10 March 2014	Review and comments	JD
0.03	11 March 2014	Adaption of comments	KH
1.00	07 Jan 2015	Finalised version	KH

Attendees

Neil Grindley, Jisc
Rachel Bruce, Jisc
Kevin Ashley, DCC
Joy Davidson, DCC
José Borbinha, IST- University Lisbon
Katarina Haage, DNB

Lizzy Rolando
Stephen Abrams
Kyle Rimkus
Bronwen Sprout
Su Nee Goh
Emily Lin
Brian Westra
Angeletta Leggio
Driek Heesakkers
Alex Ball
Rita van Duinen
Nathan Tallman
Carla Graebner
Daureen Nesdill
Laura Vilela R. Rezende
Edward McCain
Wendy Kozlowski
Daniele Balduzzi
Limor Peer
Melanie Davis
Viv Hutchison
David S.H. Rosenthal
Melissa Cragin
Helen Tibbo
Beth Yakel

Agenda

Section 1 (09:00-10:30)

1. Introduction to the 4C project and the costs of curation (presentation)
2. What is your organisation interested in? (Presentation and Q&A using the 4C Indirect Economic Determinants and the more broadly defined benefits of curation)
3. How do different organisations count the cost of curation? (Exemplars & participants invited to briefly share experiences)

Section 2 (11:00-12:30)

4. Introduction to the 4C draft Cost Concept Model (Presentation)
5. How would you break down the cost of curation? (Exercise in small groups supported by 4C team member)

Section 3 (13:30-15:00)

6. The CCEx and sharing costs (Presentation and Q&A)
7. From costs to business models via risk (Presentation and Q&A)

Section 4 (15:15-16:30)

8. Sustaining solutions and services using the ESRM (Exercise)
9. Recap, summing up and feedback

Minutes

Note: Because all presentation slides are available on the [4C website](#), this report focuses on the discussions following or in between the presentations or rather the results of the exercises.

The workshop “Costing curation: Are we on the right track?” was a full day workshop held on February 24th 2014 in the scope of 2014’s IDCC conference in San Francisco. It was visited by 25 participants. After a brief but comprehensive introduction to the 4C project, its purposes, approaches and goals by all six present 4C attendees the workshop went straight off to the heart of the matter.

For the purposes of feedback, these notes capture the outputs of points 2, 3, 5, 6, 8.

(2)What is your organisation interested in? (Presentation and Q&A using the 4C Indirect Economic Determinants and the more broadly defined benefits of curation)

Having been introduced to the Indirect Economic Determinants the workshop participants were asked to “rank” the following 15 terms with regard to their importance for their organisation’s backgrounds. The majority of participants came from universities and academic libraries respectively memory institutions with data repositories. Other organisations like publisher, international development organisation, provider of preservation software and services and federal government were represented individually. The results of the ranking are as follows:

1. University sector:
 - Authenticity – high (5), medium (2), low (2)
 - Benefit** – high (8), medium (1), low (-)
 - Efficiency – high (2), medium (6), low (1)
 - Impact – high (5), medium (2), low (2)
 - Innovation – high (-), medium (7), low (2)
 - Interoperability – high (4), medium (2), low (2)
 - Quality – high (4), medium (4), low (1)
 - Reputation – high (4), medium (5), low (-)
 - Risk** – high (6), medium (3), low (-)
 - Sensitivity – high (4), medium (3), low (2)
 - Skills – high (2), medium (4), low (1)
 - Sustainability – high (5), medium (3), low (1)
 - Transparency – high (2), medium (4), low (3)
 - Trustworthiness** – high (6), medium (2), low (1)
 - Value – high (5), medium (4), low (-)

The results show that benefit, risk and trustworthiness are the three terms with the highest importance, closely followed by authenticity, value, impact and sustainability. Added terms from university sector representatives: compliance, research quality assessments.

2. Library sector:
 - Authenticity – high (8), medium (3), low (-)
 - Benefit – high (4), medium (6), low (-)
 - Efficiency – high (4), medium (6), low (1)
 - Impact – high (3), medium (7), low (1)
 - Innovation – high (2), medium (4), low (5)
 - Interoperability – high (7), medium (3), low (1)
 - Quality – high (8), medium (3), low (-)

Reputation – high (10), medium (1), low (-)
 Risk – high (5), medium (2), low (4)
 Sensitivity – high (5), medium (4), low (1)
 Skills – high (3), medium (6), low (2)
 Sustainability – high (8), medium (2), low (1)
 Transparency – high (5), medium (4), low (2)
Trustworthiness – high (9), medium (2), low (-)
Value – high (10), medium (1), low (-)

The results show that value, reputation and trustworthiness are the three terms with the highest importance, closely followed by authenticity, quality and sustainability. Added terms from the library sector: versioning (2), compliance (2), scalability, ease of use, willingness of faculty, capacity (data set size).

3. Other organisations:
 - Authenticity – high (2), medium (3), low (-)
 - Benefit – high (1), medium (3), low (1)
 - Efficiency** – high (4), medium (1), low (-)
 - Impact – high (3), medium (1), low (1)
 - Innovation – high (2), medium (1), low (1)
 - Interoperability – high (3), medium (1), low (1)
 - Quality – high (2), medium (3), low (-)
 - Reputation – high (2), medium (3), low (-)
 - Risk** – high (4), medium (1), low (-)
 - Sensitivity – high (2), medium (1), low (2)
 - Skills – high (1), medium (3), low (1)
 - Sustainability – high (2), medium (3), low (-)
 - Transparency – high (3), medium (2), low (-)
 - Trustworthiness – high (3), medium (2), low (-)
 - Value** – high (4), medium (1), low (-)

The results show that value, risk and efficiency are the three terms with the highest importance, closely followed by trustworthiness, transparency, impact and interoperability. Added terms from other organisations: usability, scalability, compliance, versioning.

The discussion subsequent to this exercise revealed that the 15 terms are not static and their importance might change over the hands of time. Certain terms need to be viewed from different angles regarding the background and kind of organization that is ranking them. Another difficulty in building a digital repository also seems to be the different skills in staff and activity as well as in the organisational and management level. Some people would like to take more assets in their digital repository but simply do not have the monetary and other resources for more storage spaces and positions.

(3) How do different organisations count the cost of curation? (Exemplars & participants invited to briefly share experiences)

The presentation showed a few examples of how differently organisations can or do count the cost of curation, these were based on examples from Advisory Board members and project partners.

In the subsequent discussion it became clear that not only calculating but also comparing the costs in digital long term preservation can be seen as the “Holy Grail” that needs to be achieved/ found. Questions like “Is 4C gathering information about costs in curation in terms of finding a solution?” and about the difference between price and cost arose. Another participant mentioned to take a look at records management and how the large amount of digital material is managed there. One participant stated that they know how much digital curation costs but what seems impossible is to break down the costs to the process itself and the typical kind of content and also that the real challenge lies in the different skills of people. The costs of curation for one year or more is easy to calculate; however, the mission of long-term preservation is to store data forever. It was also mentioned that digitization projects are easy to calculate but beyond this it is much more difficult to do the costing; once a project is finished the danger can occur of losing the incentives to do on-going curation activity. It seems also sensible to calculate the costs in advance and not during a project although this seems more difficult to approach/achieve in an early stage of preservation planning. Some participants also wanted the ability to compare costs against particular scenarios (i.e., type of collection, type of organization, tier/level of service, doing something vs. doing nothing). Some participants wanted to be able to compare themselves with like organisations while others wanted to be able to compare themselves with a range of organization types.

(5) How would you break down the cost of curation? (Exercise in small groups supported by 4C team member)

After the introduction to the 4C draft Cost Concept Model (CCM) (point 4 on agenda) the question about the difference between activity, quality and categories in the mind-map to the CCM arose; this needs to be clarified with the composer of the CCM (Hervé L’Hours). Other comments on the CCM were to maybe shorten the time scale and if so, how and how usefully short; to maybe monetize some of the benefits. Subsequent to the presentation the participants were asked to break down the cost of curation for their organisation with the help of the “journey” exercise that has been introduced to them beforehand (see slides 32-46 in [presentation](#)). The results of this exercise were presented in plenum and showed different approaches to manage and handle costing curation.

(6) The CCEx and sharing costs (Presentation and Q&A)

The Curation Costs Exchange was presented theoretically and the mock ups were shown to the participants. Having been introduced to the purpose and functionality of the CCEx the participants were asked to answer the following questions:

As a ... - Please indicate your professional area (researchers, administrator, librarian, etc.)

I would expect to find ... - What information would you expect to see here?

I’d like to ... - What functionality would you like to see in CCEx?

So I can ... - What could the information you’d aim to get out of CCEx help you to do?

I’d be prepared to share... - What data would you be willing to share with others via CCEx? Under what conditions (anonymity?)

The results of these questions are captured in feedback spreadsheet:

<https://drive.google.com/?tab=wo&authuser=0#folders/0B2LvYlqQS0VLTWptZ2gOU1pLMFU>

(8) Sustaining solutions and services using the ESRM (Exercise)

Having been introduced to the Economic Sustainability Reference Model (ESRM) the participants were asked to fill in the ESRM Appendix. This exercise showed the variation or rather dependency of the background of the digital repository or archive; for example, digital assets from research sector are highly different to digital assets from other sectors. The answers also highly depend on how much into detail you want to or can go.