

## Collaboration to Clarify the Cost of Curation



### 4C Workshop at iPRES 2013 in Lisbon, Portugal

## Report

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

**Version History**

<b>Version</b>	<b>Date</b>	<b>Changed pages / reason</b>	<b>Modified by</b>
0.01	Sep 06 2013	First draft	SS
	Mar 06 2014	Edited	SS
	Mar 07 2014	Edited	KH

## Attendees

### 4C:

1. William Kilbride, DPC
2. Kevin Ashley, DCC
3. Luis Faria, KEEPS
4. Sarah Norris, DPC
5. Paul Stokes, Jisc
6. Neil Grindley, Jisc
7. Alex Thirifays, DNA
8. Sabine Schrimpf, DNB
9. Katarina Haage, DNB
10. Diogo Proença, INESC-ID

### Participants:

11. Vincent Joguín, Eupalia Sas
12. Kirnn Kaur, British Library (APARSEN)
13. Marcin Ostasz, Barcelona Supercomputing Centre
14. Jette Junge, State and University Library Denmark
15. Catherine Jones, Science + Technology Facilities Council
16. Yvonne Fries, ZBW Kiel
17. Paul Wheatley, University of Leeds
18. Pauline Sinclair, Tessella
19. Anna Henry, TATE
20. Angela Dappert, DPC
21. Sharon McMeekin, DPC
22. Nancy Deromedi, University of Michigan
23. Christina Bankhardt, AbbVie
24. Elisha Pavlaskova, Charles University in Prague
25. Jamie Shiers, CERN
26. Mette von Essen, National Archives of the Netherlands
27. Nancy McGovern, MIT Libraries
28. Andrew Wilson, (no affiliation)
29. Maurizio Lunghi, Fondazione Rinscimento Digitale
30. Angela Holzer, German Research Foundation

## Agenda

- 09:00-09:05 [Welcome - Katarina Haage, DNB](#)
- 09:05-09:20 [4C Workshop - Introduction \(iPRES 2013\)](#) - Neil Grindley, Jisc
- 09:20-09:50 [Presentation of APARSEN results on analysis and testing of cost models](#) - Kirnn Kaur, British Library
- 09:50-10:20 [4C Workshop - Web Consultation Results \(iPRES 2013\)](#) - Ulla Bogvad Kejser, KBDK / Alex Thirifays, DNA
- 10:20-10:45 Q & A and [Thoughts on the Curation Costs Exchange](#)
- 10:45-11:15 Coffee Break
- 11:15-11:30 [4C Workshop - KE Funding Presentation \(iPRES 2013\)](#) - Angela Holzer, DFG
- 11:30-11:45 Outcomes from the recent DCC Research Data Management Forum - Kevin Ashley, UEDIN-DCC
- 11:45-12:00 [4C Workshop - ESRM Presentation \(iPRES 2013\)](#) - Neil Grindley, Jisc
- 12:00-12:15 [Short presentation of CERN cost data](#) - Jamie Shiers, CERN
- 12:15-12:50 Open discussion - William Kilbride, DPC
- 12:50-13:00 Wrap up

All slides can be downloaded from the 4C website: <http://4cproject.eu/community-resources/focus-groups/ipres-workshop>

## Workshop Report

*Note: Because all presentation slides are available on the 4C website, this report focuses on the discussions following or in between the presentations.*

William Kilbride invited the workshop participants to share what comes to their mind when they think about the issues of cost of curation. The participants said that they thought about disasters (cost of data loss), cost of business systems, cost of outsourcing vs. doing curation in house.

After Kirnn Kaur's APARSEN presentation, the discussion focused on the fact that most cost models were not designed for reuse in the first place. As the only exception, the LIFE cost model was developed explicitly for the wider community. Altogether, it was concluded that there is a need for simplicity in the models. Also an abstraction of the existing models was proposed. The "Cost Concept Model" that will be developed in the 4C project takes this line and might foster progress in this area.

Following Alex Thirifays' presentation of the initial results of the 4C project, several participants expressed wishes for topics that 4C should take into consideration: It would be interesting to look at and identify the "cost of inaction" in digital curation that may lead to data loss. Also benefits should be investigated (in fact, 4C has a dedicated task on benefits). One participant said he wanted to see some evidence for the money save that results from small amounts of DP activities. He believes that this would be the most powerful benefit to allow for early preservation investments. Another participant stressed the importance to engage with industry and learn from their general cost modeling experiences.

It was emphasized how interesting it was that APARSEN and 4C came to the same conclusions in their cost model analysis. Both projects should use this congruency to strengthen their points.

It was debated in how far the standardization of workflows can simplify the costing of digital curation activities. One participant shared her experience from her work in the area of educating DP practitioners: Most important is "what" should be preserved, "how" should it be preserved and "how much" should be preserved. From this perspective, it is rather unlikely that standard workflows (to map cost parameters against) make cost modeling easier. It is probably unrealistic to assume that standard workflows can be derived because things \*are\* different in reality.

William Kilbride sent the workshop participants to the Coffee break with four proposals on what the "Curation Cost Exchange" platform that the 4C project will be developing might become. After the break, the different options were discussed:

1. Recommender service: The system tells you what to think about, which models to look at when you want to implement a cost model  
→ The participants thought that none of the existing cost models is good enough that the recommender service should truly \*recommend\* it. The recommender service could, however, direct people to existing work on which they can build their own cost modeling activities: "This is where and how you can start your own cost modeling exercise."
2. The global oracle: The system gives you an answer how much money you need to sustain your DP strategy on the basis the information that you put in  
→ The participants doubted that the existing cost models are developed enough to support this option. For the oracle, one participant warned, 4C would have to create a new cost model and that is the last thing the project should be doing.  
As a variation of the oracle option, one participant proposes that the tool guides the interested person through a set of questions and helps to build one's own algorithm based on these. (Questions like: What do you need to preserve? How do you preserve it? How much needs to be preserved? Who does it, with what percentage of his/her time?)

3. The business case generator: The systems provides you with links to case studies based on some benefits assumptions that you put in
  - One participant calls attention to the fact that the cost model generator is technologically *\*not\** an exchange. It is different from the other three options.
4. The cost model generator: The system gives you the tools to create your own cost model
  - None of the participants picked up this option.

After the presentations of the afternoon session, Neil Grindley concluded the workshop. He invited the participants to contribute to the development of the 4C project resources by providing input, discussion initial results, or, if possible, sharing cost information to improve the Cost Concept Model and ultimately the Curation Cost Exchange.