Public report of the 1st 4C Advisory Board Meeting

The first Advisory Board Meeting was held on 11 June 2013 at the Jisc Office in London, UK. In attendance were:

1. Neil Grindley, Jisc
2. Paul Stokes, Jisc
3. Rachel Bruce, Jisc
4. David Rosenthal, LOCKSS
5. Matthew Addis, Arkivum
6. Ron Dekker, Netherlands Organisation for Scientific Research
7. Sabine Schrimpf, Deutsche Nationalbibliothek
8. Raivo Ruusalepp, National Library of Estonia
9. Sean Barker, BAE Systems
10. Hildelies Balk, Koninklijke Bibliotheek
11. Alex Thirifays, Danish National Archive
12. Brian Lavoie, OCLC

Some highlights of the meeting:

Context, background, introduction to project
Provided by Neil Grindley:
- The project is a co-ordination action, intended to clarify and synthesise current information to make it more usable
- In the curation domain complex models are rife
- The chosen term for the process is "curation"
- Curation is an investment
- The EU is keen to use projects such as 4C to drive economic recovery through the engagement of SMEs

Reaction from Board members about high-level aims
Particular discussion points included:
- The Curation Costs Exchange (CCEx)
- The role of SME in relation to larger companies
- The sustainability of the project
- The governance of the project
- The Advisory board agreed to maintain an informal flow of information in the intervening period and to allow some information regarding the board and its remit to be published on the project web site.

Summary of Work Packages and their Objectives
WP1 - The question of the openness and published status of documents relating to the project was discussed. At present most documents/outputs in development are not open, but when they are complete they will be. It was suggested that the 4C glossary is a document that should be open even during development.
WP2 - Board members were invited to engage with the on-line consultation and provide feedback on the stakeholder analysis. The board noted that commercial stakeholder should include more than just SMEs. Other stake holders were identified including: Media companies; Pharma; Oil and gas; Scientific archivist groups (who meet on a regular basis); "stake holders we don't know yet"— potentially the biggest group. Other related projects mentioned include BBC, Presto, APARSEN, TAPE (a Finish project).

WP3 – The Work Package covers three areas: 1) Finding out what (cost model data) is out there; 2) Establishing what stake holders want; 3) Establishing the gaps and discrepancies between the models and the wants. The appropriate number of models (or frameworks) needed to cover "all" use cases was discussed, and the number and types of parameters for those models and the expectation regarding the use of model results. There was also discussion regarding the increasing adoption of "models (complex digital objects and the proprietary software required to access them) as the unit of curation (as opposed to "documents") and how that might affect any curation models. Once again it was noted that the different use and ascribed meaning of the same word—"model"—in alternative domains could lead to confusion.

WP4 - The Work Package sets out to fill some gaps in the existing cost models. It was noted that, although the CCEx could be a rich source of information for WP4, it would not be available in time. The board discussed the issues surrounding the reasons for curation, specifically the differences between keeping materials because of their intrinsic value, and keeping materials for compliance reasons. Comparisons were made between research data kept for its future research value and pharmaceutical data kept for compliance (although it was acknowledged that in most cases data will be kept for both reasons even if the relative weighting is different). Further comparisons were made between the standards in the US and in Europe. The Board recommended developing a map of the reasons for data curation.

WP5 - Discussion of Work Package 5 was passed over in favour of increasing the time devoted to discussion of the project challenges.

Discussion of the main challenges the project faces

- Terminology—how do we most clearly explain to stakeholders what we want from them? - There is a potential tension between the use of the glossary internally (and in externally facing communications) to ensure that we put across a consistent, internally understood message and the life of the glossary beyond the lifetime of the project. What may be useful and pragmatic within the 4C arena may not be suitable for or accepted by a wider audience.

- Models and specifications—how can we ensure that they are useful and understandable? – Neil Grindley introduced the Economic Sustainability Reference Model (ESRM), currently in its 0.6 iteration. The glossary describes a model as "a simplified version of reality". Neil Grindley's hypothesis is that a reference model has a process that enables the user to change things for the better. The board discussed possible uses for the ESRM (and compared it to the OAIS model). It was postulated that a reference model should contain the building bricks to allow the creation of new models, or as a tool for comparing differences in models, but the ESRM as described is a tool to help the user to think about their needs and requirements in a different way (and is particularly focused on the preservation of digital objects). A public draft of the model is to be made available on the web site for comment.

- Gathering Costs Data—who will give it to us and how should we analyse it? - It was proposed that the 4C consortium members should lead by example with all partners pledging to come up with 2 sets of costs each (giving a starting point of 26 sets). The board agreed that 4C should be smart in the data gathering process, using secondary indicators such as effort, using statistical methods, making the metrics gathered as easy for the supplier to source as possible and embracing "ball park" figures. Relative costs were deemed to be useful, especially in situations where costs and prices are being obfuscated by other supplier/customer arrangements.

- The Curation Costs Exchange—what is it and how do we make it useful? - A paper prototype with a group of "friendly" participants was deemed to be a good idea in order to get some unbiased opinions of needs (what they want from the CCEx) and concerns (what they fear about the CCEx). The people who had indicated their willingness to be contacted from the initial consultation exercise might form just such a group. Questions were raised about quality control, specifically would there be checks on who put in data and checks on data validity.
- Another potential use case could be those who wish to offer services using the CCEx to see where such services could be best targeted.