Collaboration to Clarify the Cost of Curation

Report of Focus Group Meeting on May 13ᵗʰ 2014 at ARCHIVING Berlin

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	19 May 2014	First draft	KH
1.00	23 May 2014	Ratified version	KH

Attendees

4C:

Alex Thirifays, DNA

Hervé L'Hours, UK Data Archive

Katarina Haage, DNB

Ulla Bogvad Kejser, KBDK

Participants:

Dr. Harry Enke, aip Potsdam

Elena Simukovic, HU Berlin

Joost van der Nat, NCDD

Karen Colbron, Presto4U

Lindio Ligios, Presto4U

Marcel Ras, NCDD

Marcin Ostasz, Barcelona Supercomputing Centre

Martin Iordanidis, hbz Köln

Reiner Mauer, gesis

Agenda

Time	Topic	Responsible	Description
13.30-14.00	Introduction round Overview of the 4C Project	Katarina Haage, German National Library	
14.00-15.30	General introduction to CCM and GRS	Hervé L'Hours, UK Data Archive	A general introduction to the challenges of the Cost Concept Model/Gateway Requirements Specification and how we have chosen to structure our approach
	A 'specification' exercise	Hervé L'Hours	Exercise to consider your requirements from a cost modelling perspective
15.30-15.45	Break		
15.45-16.00	Excursus – Netherlands Coalition for Digital Preservation (NCDD)	Joost van der Nat and Marcel Ras, NCDD	Introduction to the NCDD and the present project on digital preservation and the relation to the 4C project
16.00-17.00	General introduction to the Curation Costs Exchange (CCEx)	Alex Thirifays, Danish National Archives	15 minutes introduction to the concept, the purpose, the hopes, the sustainability and the what's-in-it-for-you
	In-depth discussion about the submission of costs	Alex Thirifays	Discussions on the usefulness of sharing and comparing costs; exercises regarding the submission template; feedback and exchange of views and experiences
17.00-17.30	Round-up and possible future collaborations		

Minutes

After a welcome and brief overview of the 4C project by KH the participants introduced themselves and formulated their expectations of the meeting. Goals for the day were: Share knowledge across projects – learn and bring, work to find out what it costs and get a cost "template".

HLH presented the CCM (core concepts) and GRS. What it is – and what it isn't.

Prior to the event attendees were provided with an outline of the deliverable structure and the workshop questions which would be presented

At the event the structure and content of the deliverable were described in more detail. 9 attendees and guest speakers were then split into two groups, mediated by members of the 4C project team. Though forms were provided for annotating responses to questions (see some extracts below) the general tone was informal as the project team took advantage of an opportunity to test concepts and definitions in the real world. Responses are anonymised.

Attendees were primarily those with a role similar to repository managers or data managers within a project scenario so potential implementers of curation cost models rather than theorists. This presented an opportunity to work on the themes of stakeholder context, organisational profile and attitude to benefits. Resources were addressed through a question relating to categorisation of labour forces and activities were addressed via questions about the approach to the structure and quality of activities and the structures of collections; these latter themes are represented under separate sections of the core cost concepts but have a clear relationship to the 'facts' about a system which form the organisational profile.

These themes together are critical to the structure of the framework as their clear communication between implementers and theorists is critical. Potential implementers of curation cost models must have a certain level of understanding and maturity around the issues to select the appropriate model and implement it with an appropriate degree of success. Theorists must understand these issue to correctly model the underlying curation systems relevant to their method. These subjects are essential for designing use cases or scenarios which are a common means of communication between the developers of a methodology and their uses.

Overall the structure of the deliverable was well received and provided an effective means of standarising communications around this complex area. As expected there are some artefacts from current cost models (e.g. common descriptions of collection profiles and a standard approach to grouping types of labour) which were not well established within the attendee's organisations. One area for improvement in the deliverable is a clearer distinction between stakeholder context (stakeholder identification and management, perhaps closest to Designated Community monitoring in OAIS terms) and Organisation Context which will be more clearly defined as relevant business intelligence derived from Stakeholder interaction in future versions of the deliverable.

Original question text below is quoted and in bold. Each bullet collects key parts of the responses from a participant.

1. Stakeholder Context

It became clear during discussions that the defining factor for the separation of stakeholder context and organisational profile levels of the systems was the artefacts they produce. Stakeholder engagement produces contact lists, questionnaires, minutes and other records to manage the communication process. The outcomes of that process are then operationalised into actionable parts of the organisation profile

through the creation of targets, mission statements, service level agreements, reporting procedures, mandates, contracts, licences, risk registers etc.

"Do you identify relevant stakeholders for your organisation If so who are they…"

- Funders, legislators, standards developers, staff, users

- Some organisations are very complicated amalgams of projects and services, this increases the complexity of stakeholder identification and interaction.

- Funders, user communities, sites which harvest our content

- Customers, software providers, depositors, financial supporters (ministry)

- Creators, content providers, users, professionals, industry. Decision makers are differentiated from influencers.

- Creators, users, decision makers, influencers, depositors, professionals.

- Domain memory institutions, libraries and archives

- Customers, libraries, researchers, financial supporters

- Definition of use cases

"…and how do you manage and communicate with them?"

- Reporting

- We diagram them in terms of their roles and influence

- They support in developing mandates and mission statements

- 1 to 1 contact

2. Organisational Context

"What aspects of your organisation (your 'organisational context') do you consider critical to cost model issues?"

- Skills, budget, collection profiles, data standards, legislative environment

- Staff costs, software licencing, storage (network operations.

- Sustainability ("forever"), zero cost to users, trustworthiness (TDR status), and openness of data (minimal access restrictions).

- Our Mission and Vision must be met by delivering our primary business processes.

- Policies, processes, legal framework.

- Business plans, use cases and business cases

- User needs (e.g. tiered storage and use on demand)

3. Incentives/Benefits

"Do you address Indirect Economic Determinants, incentives, risks and benefits explicitly within the organisation?

- Not at this stage, only risk is talked about.

- Risk assessment

- Through development plans and strategies, preservation policy

- Efficiency is considers an operational issue, not a cost issue. Benefits are only addressed at a high level.

- Delivering our tasks and responsibilities within budget and within our legal obligations is the focus. Risks are dealt with if we can't do that.

- Budgets are provided and we deal with this by making choices.

- As a project we directly address risks and opportunities around all activities and review them annually

- Funder driven focus

- Long-term goals are damaging for science

"If so does this follow a formal approach (ISO standard, risk analysis etc.)?"

- Data Seal of Approval. Looking at the ISO16363 for trust and ISO27001 for information security.,

- ISO16363, DRAMBORA, DIN31644

- Audit and Certification through the Data Seal of Approval is planned.

4. Service/Activities

"How do you structure repository activities…"

- We're currently in a project phase so via a project management system

- Based on OAIS and organisational structure

- To be decided, we're still designing the system

- Organisational structure and divisions

- We sit at the middle and don't know what each site does in detail.

"and manage their quality?"

- Validation of standardised XML but otherwise by trusting the opinion of expert teams who document their actions.

5. Asset Adjustments

"How do you structure your collections? Are format, complexity and quantity the critical criteria?"

- A data collection is made up of data, metadata (structured documentation) and documentation (text files and descriptions which aren't that structured. Formats are approved because of value to users and preservation efficiency. No specific complexity consideration. Quantity is monitored at Archival Storage level but could be better monitored through quantity estimates pre-Ingest.

- This is done by the user, we just take what they give us

- PDF/A is the required format with quantity being an issue for both transmission and dissemination.

- Structured by type and format, evaluated through significance 2.0 http://arts.gov.au/sites/default/files/resources-publications/significance-2.0/pdfs/significance-2.0.pdf

6. Labour

"Do you categorise your labour force? In terms of Qualifications, experience, training received, job description etc."

- No, the requirements are defined on a per role basis. It would not be useful to band staff costs by this method.

- Not really

- By speciality (developers vs maintenance), job description, hierarchy and salary scales.

After a short coffee break Marcel Ras from the NCDD gave a guest presentation and introduced the participants to the structure of the organisation followed by an overview on NCDD's current project on cost modelling and the goals by his colleague Joost van der Nat.

Subsequently the participants were introduced to the CCEx . Beforehand the meeting information material on the CCEx and the submission template was sent to the attendees to prepare the session. After a brief introduction to the CCEx by AT the group started a discussion and gave feedback on what they had received for preparation. In the following comments, questions and remarks on the CCEx are listed:

- It took some time in the beginning to understand how to start enter information in the submission template due to the fact that an example in the cost chart was missing.
- A cost submission template is needed: If organisations were able to fill in the submission template effortlessly, they would not need it at all.
- Terminology – be very clear about the definitions
- The products that are developed are not always only curation related and are used for many departments in an organisation, which means that activities costs cannot always be attributed to digital curation only (complicates the submission of costs)
- Create more filters/refine the 'Profile'tab
- Narrow the scope of stakeholders or develop various templates for types of stakeholders/institutions.
- Focus on defining clearly the pre-ingest (Production) process and make clear on which assets you put the focus.
- It's not possible to operate with FTE's only

The attendees were then asked to think about the benefits of the submission template or rather why they would submit data:

- Average calculation
- Benchmarking; look at the change from year 1 to year 2 – have we done better than the year before? Have our plans, strategies, choices had the intended impact?
- Self-assessment
- For budgeting, planning, predictions
- Important information for funders
- Important information for sending bills
- Development of new services and to inform business models
- Communication
- Collect and compare figures and facts

- Raise awareness
- Submission of cost data requires clear outlining of issues of anonymity and confidentiality
- Doing the exercise spurs discussions within the organisation – the exercise itself is beneficial.
- Mentality change: Maximise the culture of sharing – it instigates faster development, enhances trust raises awareness and leads to increases in efficiency

In the end of this meeting further engagement between 4C and the NCDD and Presto4U was agreed on.

AT mentioned that the Curation Costs Exchange Platform may encounter some sustainability issues beyond the lifetime of the project, but that the Digital Preservation Coalition and Nestor both have showed an interest in keeping it alive. NCDD stepped in at this point and gracefully offered to engage in sustainability negotiations with the 4C-project as well.

As an outlook the attendees mentioned the idea to calculate the benefits of curation and the benefits of collaboration. KH informed that all presentations will be online and circulated soon after the meeting.

Action Summary

Item	Description	Who	Owner	Due by
1	Upload and circulate all presentations on the 4C website [done]	KH		May 19th 2014
2	Write report [ongoing]	KH, all	KH	June 2nd 2014