

# A Collaboration to Clarify the Costs of Curation – The 4C Project

Neil Grindley  
Jisc  
London  
UK  
+44 (0)203 006 6059  
n.grindley@jisc.ac.uk

## ABSTRACT

This poster will describe actions being taken in the context of the 4C Project (a Collaboration to Clarify the Costs of Curation), an EC-funded two-year coordination action that aims to promote a better understanding of the potential for undertaking digital curation activity. The approach it is taking is to focus firstly on costs but then to link that concept to related ones such as benefits, value, risk and sustainability, therefore taking a holistic economic view of digital curation. This is important as it links up with various strands of previous work, both on costs and activity models; and on benefits, sustainability and the broader economic framework for digital preservation and access. The purpose of the poster is to describe the novel framework for activity that was proposed in response to the EC FP7-ICT-2011-9 call, and to summarily describe some of the outputs and objectives in a graphical and accessible format.

## Keywords

digital curation, costs, economics, benefits, sustainability, risk, 4C, European Commission

## 1. INTRODUCTION

The European Commission's FP7 ICT 9 programme [1] invited proposals that would be "promotion schemes for the uptake of digital preservation research outcomes including outreach to new stakeholders and roadmapping activities." On the face of it, this meant coordination actions that would synthesise existing work and improve its uptake and implementation across a range of different communities. Underpinning this, however (and articulated at the briefing meeting organised by the EC to describe the aims of the call), was a sense that - despite significant investment - not enough tangible progress had been made with devising workable and competitive solutions and services in the digital preservation realm.

It was suggested that what was required was a healthy and diverse market for mature technical solutions that tackled the real long-term digital asset management problems that all types of organisations face on a daily basis. The conclusion seemed to be that if this could be orchestrated, then the consequent supply and demand would provoke the sort of activity - particularly the flow of services and solutions from SME's (small to medium commercial entities) towards institutions - that was so urgently required at macro-economic European level.

The challenge was, therefore, to design a project that usefully synthesized an area of digital preservation research; was of wide interest to a variety of stakeholders in different working domains; was capable of representing and enhancing that existing work; and

tackled the topic in such a way that it would shed useful light on the barriers to uptake and the implementation of solutions and services.

The answer alighted upon by a consortium of partners brought together under the banner of the 4C project was to focus on the costs and economics of digital preservation. The particular challenge thereafter was to devise the best way of drawing together a substantial and heterogeneous body of work; to enhance and then present it afresh to new stakeholders; and finally to make it easier for future actors in the domain to either demand or supply digital preservation solutions and services.

## 2. EARLY ASSUMPTIONS

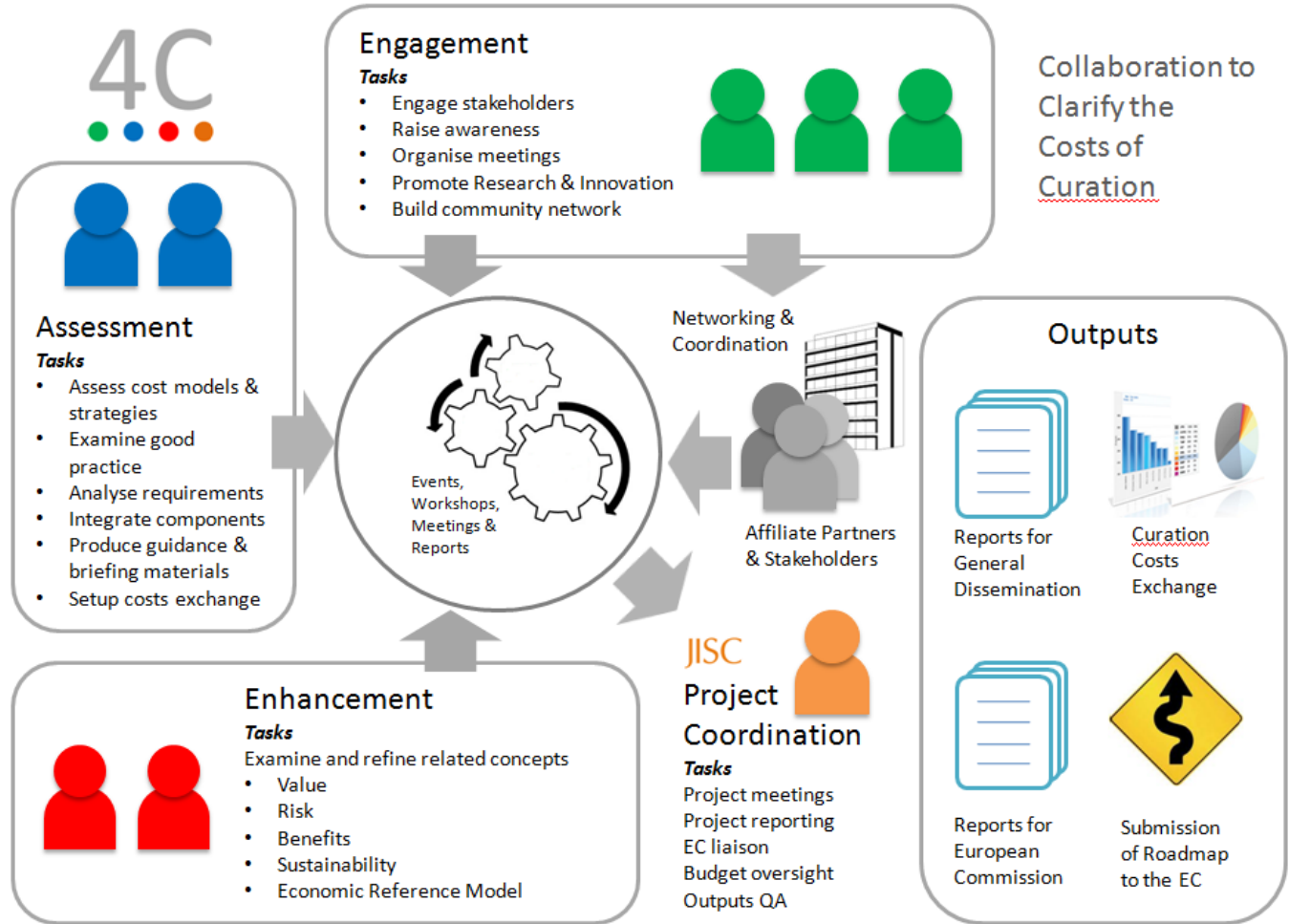
There is, as stated above, a substantial amount of work that already exists on the topic of the costs and economics of digital preservation and curation. This can be found in bibliographies [2] and in project listings [3] and includes initiatives that have: formulated cost models (e.g. LIFE project, CMDP project); suggested frameworks (e.g. Keeping Research Data Safe); formed task forces (e.g. Blue Ribbon Task Force on Sustainable Digital Preservation and Access - BRTF); written reports (e.g. APARSEN); and doubtless produced other types of output over nearly twenty years of activity. It was clear at the outset that the 4C project did not need (and should not try) to formulate a cost model to surpass and encapsulate all other cost models. Even if it were possible within the constraints of an EC-funded coordination action, it was apparent that what was required was to build on, join up and communicate this wealth of existing work rather than do new development work and risk duplicating prior effort.

What was also apparent was that the cost of digital *curation* (a term used interchangeably with *preservation* and *archiving* for the purposes of this project) was not a concept that could exist in isolation from a whole raft of other issues. Taking the broader economic view, the 4C Project classes digital curation as an investment, and whilst there are costs associated with an investment, the point is to realise a return or a benefit. Understanding what sort of ratio of cost to benefit organisations will be able to realise (and over what timescales) starts to unpack the whole complexity of digital curation and necessitates examination of other issues such as the level of risk that organisations are willing to accept and other issues such as the value that they attach to those assets. Additional factors and issues that might affect the cost of digital curation could include: trustworthiness, quality, sensitivity, confidentiality, authenticity, capability etc. These factors are referred to as *indirect economic determinants*.

There was also an assumption that the 4C project needed to build a firm foundation for future work and to set an agenda. It was, therefore, agreed that although it did not seek to build another detailed cost model, it could reasonably aspire to: assemble conceptual models; define generic specifications; and to provide a

platform for collecting information that ought to be of common interest to a broad range of organisations using a mechanism called the Curation Costs Exchange (CCEX). These form some of the deliverables of the project and will be described in more detail in the poster. Some of them are referenced in figure 1 below.

Figure 1: Conceptual Structure and Mechanism of the 4C Project



### 3. PROJECT ELEMENTS

The project structure was purposefully conceived as comprising of a minimal set of entities. The complexity of the conceptual issues and the difficulties around terminology were felt to be severe enough without introducing disparate and numerous working groups. The project was therefore formed into five work packages, the first of which was a standard component of project management. An articulation of the other four work packages usefully outlines the purpose and the internal dynamic of the project, also set out in figure 1, which graphically illustrates the project mechanism. (This image will feature as the core of the poster. Further graphical elements will surround the core and will elaborate on the individual components and the approaches taken.)

#### 3.1 Engagement

Engagement is the key activity of the 4C project and informs the entirety of the rest of the work. In line with its coordination action status, it is the principal purpose of the project and will determine whether or not the initiative makes a positive and lasting impact.

The main objective of the Engagement group is to engage with a wide range of stakeholders in memory institutions, universities, SME's, government, data intensive research, industry etc. It will identify, get involved and build partnerships with individuals, groups and institutions that are active or interested in the issue of curation costs and it will attempt to foster a better understanding of the issue amongst the community more broadly. This will be done using well-rehearsed outreach techniques but will also be facilitated by the analysis, expertise and outputs from the other two main project groups. It is this aspect of the project that provides the necessary enhancement on existing work and should enable the discussions to be threaded through with a more

sophisticated understanding of the complex conceptual relationships involved with cost issues.

### 3.2 Assessment

The main objective of the Assessment group is to establish the most effective current methods for private and public sector organisations to estimate and compare the cost of digital curation, and to identify the most beneficial paths for future development of solutions and services. This will enable stakeholders to more effectively and comprehensively assess the investment of resources that may be required to sustain their digital preservation activities; and allow comparisons of existing and future tools and models with the knowledge that a broad range of criteria: e.g. price, savings, quality, value, risks, benefits, sustainability, etc., are implicit to the comparison.

### 3.3 Enhancement

One of the key objectives of this group is to ensure that comprehensive and appropriate consideration is given to all indirect factors that might be considered economic determinants of digital curation. Whilst the Assessment group is trying to harmonise and synthesise, the purpose of the Enhancement group is to entertain complexity, consider broader conceptual issues, and to worry about how and when indirect factors should feature in the organizational planning of stakeholder organisations.

Two key factors identified at the outset by the project are:

- Levels of trustworthiness aspired to by organisations and consequent activities around certification
- The level of risk an organisation is prepared to accept

Other economic determinants will emerge as more or less of a priority in the course of the work and as part of the engagement dialogue with the community.

The Enhancement work is also driven by two other key imperatives. The first focuses on sustainability and builds on existing work [4] to develop a draft Economic Sustainability Reference Model (ESRM). This will elucidate the threats to digital assets, the timing of those threats, and help to define terminologies.

The other imperative is to move from the area of costs and into the realm of business models, which addresses the concerns raised by the European Commission at the outset (see above) and also segues into the last of the work packages which is to look forward and define an agenda for research, development and collaboration.

### 3.4 Roadmap

The purpose of this activity is to arrive at coherent and evidence-based recommendations for future action and strategy in relation to the economic aspects of digital curation. The focus will be on measures that will assist diverse types of organisations to better understand and take control of the cost of managing digital assets over varied timescales, including the provision of cost-effective solutions and services to others. This roadmap report will synthesise and exploit the valuable intelligence that emerges from the other work packages and will also ensure that the content and conclusions are complementary and non-duplicative of work being taken forward by others.

## 4. PROJECT OBJECTIVES

As should be clear from the preceding information, the project has a number of different objectives. Perhaps the most practical and immediate is to provide organisations that have a need to curate data with more effective and accurate ways of working out how much this activity will cost them. This will help them to do more effective planning and resource allocation.

This first objective is straightforward in cases where the principal business of an organisation (e.g. a national library or archive) is curation. However, this is not the case for most organisations, so the purpose of clarifying the cost of curation should also serve to bring into sharper relief the reasons why curation should or shouldn't be resourced. Therefore, the 4C work will engage not only with the costs of curation but also the benefits that it might realise. Or to put it another way - from an economic perspective - it will examine curation as an investment and be mindful that investments require returns, involve elements of risk, and connect with notions of sustainability and business planning.

The second principal objective is to synthesise, make sense, re-present - and where appropriate enhance - the previous and emerging valuable work that has been done in this area over the years. The costs, economics and sustainability of curating digital assets has been tackled from many perspectives by initiatives across the world but is still not widely understood or effectively embedded into practice. 4C will therefore undertake advocacy and promote relevant work to existing and new stakeholders.

A third important objective is to try and help address the underperforming market place for digital curation solutions and services. One practical way that 4C can assist with this is to establish more effective and accurate ways of predicting the cost of curating materials. If more accurate costs can be relied upon, then more confident designs can be produced for services and solutions (e.g. third party archiving services) that can either run at a profit in the commercial realm, or be assured of breaking even if in the not-for-profit sector.

## 5. POSTER OBJECTIVES

This poster will address a number of objectives:

- It is a publicity and dissemination opportunity for the 4C project
- It is an invitation to stakeholders to engage with the project and to identify with its aims and objectives
- It sets out a concise description of a project with complex objectives
- It tests out some assumptions and approaches which will require broad community acceptance and endorsement if the project is to be influential and have a positive impact

## 6. ACKNOWLEDGMENTS

Some of the content of this proposal has been assembled from the input of other 4C project partners. The list of organisations who are participants in the project is as follows:

1. Jisc (UK)
2. Royal Danish National Library (DK)
3. INESC-ID (PT)
4. Danish National Archives (DK)

5. German National Library (DE)
6. University of Glasgow – DCC (UK)
7. University of Essex – UKDA (UK)
8. KEEP Solutions (PT)
9. Digital Preservation Coalition (UK)
10. Secure Business Austria (AT)
11. University of Edinburgh - DCC (UK)
12. Data Archiving & Networked Services (NL)
13. National Library of Estonia (EE)

- [2] Lazorchak, B. 2012. A Digital Asset Sustainability and Preservation Cost Bibliography, Library of Congress, <http://1.usa.gov/RYMZW2> (accessed 27/04/2013)
- [3] Wheatley, P. (and other wiki contributors) (2013). Digital Preservation and Data Curation Costing and Cost Modelling, <http://wiki.opf-labs.org/display/CDP/Home> (accessed 27/04/2013)
- [4] Lavoie, B., Rusbridge, C. (2011). Introduction to Economic Sustainability Reference Model, Blog, <http://unsustainableideas.wordpress.com/2011/10/06/intro-reference-3/> (accessed 27/04/2013)

## **7. REFERENCES**

- [1] European Commission participant portal (2011), <http://bit.ly/15PNtnA> (accessed 27/04/2013)