

Defining a Roadmap for Economically Efficient Digital Curation – a 4C Project Workshop

Neil Grindley

JISC

Brettenham House (South Entrance)

5 Lancaster Place

London WC2E 7EN

n.grindley@jisc.ac.uk

Katarina Haage

Deutsche Nationalbibliothek

Information Technology

Adickesallee 1 D-60322

Frankfurt am Main

k.haage@dnb.de

Paul Stokes

JISC

One Castlepark Tower Hill

Bristol BS2 0JA

p.stokes@jisc.ac.uk

ABSTRACT

The 4C Project is tasked with delivering a Roadmap report and it is this drive towards ‘economic efficiency’ in relation to digital curation that will be central to the agenda that it sets out. This workshop is an important opportunity to connect with stakeholders and get input for a critical deliverable of the project. But it is also an opportunity for participants to learn more about the economics of digital curation and to critically assess the efficiency and sustainability of their own services and solutions.

General Terms

Communities, strategic environment, digital preservation marketplace, theory of digital preservation.

Keywords

Economics, policy, strategy.

1. INTRODUCTION

The 4C Project (a Collaboration to Clarify the Costs of Curation) is a European Commission funded two year coordination action which has been funded to provide useful, useable resources that provide better support to identify and quantify the cost of digital curation. From the outset, however, the project has taken the view that costs cannot be dealt with in isolation from a number of other related concepts (e.g. benefits, risk, quality, sustainability) and this holistic view might more accurately be described as an economic perspective on digital curation.

Borrowing the language of economics and mapping it onto digital curation needs to be done selectively and carefully. Digital assets do not have the same attributes as other kinds of (financial) assets and equally, it may not be possible to define when digital assets become (economic) liabilities in any objectively quantifiable way. However, there is still terminology from the field of economics that may help to define what the digital curation community might aspire to over the next few years and the starting point for this workshop is the concept of ‘economic efficiency’- which might be defined as the optimised situation where it is no longer possible to add quantity or value given a finite availability of resources.

The 4C Project is tasked with delivering a Roadmap report and it is this drive towards ‘economic efficiency’ in relation to digital curation that will be central to the agenda that it sets out. The consultation, stakeholder engagement, analysis and modelling work that have been done allow some principles to be proposed and some assertions to be made that will form the backbone of the report. The purpose of a Roadmap – particularly where it seeks to

set out an action agenda for a range of stakeholders across various communities – is to make politically astute observations and to arrive at plausible conclusions. This is only possible via early interaction with stakeholders and by achieving some level of community validation before publication and this is the purpose of the workshop. One of the guiding principles of the 4C Project is to create a better understanding of the economics of digital curation through collaboration; and also to be an ‘open and social’ project and to listen to the needs of the community. iPRES 2014 occurs at roughly the three quarter point of the two year project and provides a timely opportunity to check and refine the draft Roadmap.

Early ideas and discussions about the structure and content of the Roadmap have indicated that it will need to address various questions.

- What vision should we advocate and what principles should we espouse to bring about economically efficient digital curation?
- What current economic inefficiencies do we need to eliminate?
- What or who is the most influential mechanism to bring that about and where will that influence most be felt?
- What is the policy, business and regulatory framework for digital curation and how is it likely to change?
- Over what timescales should we advocate action?
- How can we most economically sustain and exploit existing work? (including the 4C Project outputs)
- How are the economic requirements of stakeholders changing?
- Is it possible and economically desirable to try and align digital curation practice (including standards and terminology)?
- How can we most effectively invest in digital curation at the institutional, national and international level?

This workshop is an important opportunity to connect with stakeholders and get input for a critical deliverable of the project. But it is also an opportunity for participants to learn more about the economics of digital curation and to critically assess the efficiency and sustainability of their own services and solutions.

2. Intended content

Half day workshop

09.00 – Introduction to the aims of the session, the purpose of the 4C Roadmap and a perspective on the economics of digital curation (presentation and Q&A)

09.30 – Presentation of the 4C project outputs

10.00 – Breakout groups to discuss (and then briefly feedback on) digital curation economic needs & gaps

10.30 – Break

10.45 – Presentation of the draft Roadmap for Economically Efficient Digital Curation

11.15 – Breakout groups to discuss the Roadmap

12.00 – Feedback from the groups

12.45 – Summing up

13.00 – Lunch

Open to public

This workshop will be paid for by the 4C Project and be open to all participants interested in the economics of digital curation.

Requirements for its organisation

All that is required is a room, a screen and a data projector. The ideal audience would be a mixture of those with opinions and information to offer the 4C project and those who would take information back to their organisation and prepare the way for effective dissemination towards the end of the project. This is both an input and an output opportunity. A workshop of anywhere between 10-20 people would be a useful size and a great opportunity to have a detailed conversation with an interested audience.

Local capability

See above

Speakers

The lead for the workshop will be the 4C Project coordinators Neil Grindley and Paul Stokes. Other 4C project partners will contribute and if possible, affiliate stakeholder organisations will also present.

Intended Audience

Practitioners, Managers and Funders – this has applicability at all levels and should be of practical, tactical and strategic interest.

3. WORKSHOP OUTCOMES

This workshop was the first opportunity to get face-to-face feedback from the community on the draft 4C project roadmap. 'Investing in Curation: a shared path to sustainability' states six messages and sets out a number of actions that various stakeholder groups should act upon to realise a suggested shared vision that could be realised by the year 2020.

The draft Roadmap is available at: <http://4cproject.eu/d5-1-draft-roadmap>

The vision is as follows:

In five years time (2020) it will be easier to design or procure more cost effective and efficient digital curation services because

the costs, benefits and the business cases for doing so will be more widely understood across the curation lifecycle and by all relevant stakeholders. Cost modelling will be part of the curation planning and management activities of all digital repositories.

The workshop was divided into two main sections. Firstly participants were asked to consider the main challenges they and their institutions faced with curating digital assets (particularly in relation to economic issues). Secondly, they were asked to think about the draft 4C Roadmap messages and to consider how relevant they were to their own local context and to what extent they were plausible and sensible as an agenda for action and change.

The first discussion (challenges) surfaced the following issues:

- The scale and type of issues that will need to be faced is difficult to predict but international collaboration and knowledge exchange will mitigate the impact of that uncertainty
- There are important stakeholders (e.g. certain areas of government and publishing) who don't yet feel that curation planning is their problem or who don't yet understand that 'digital is not technology'. Or to put it another way, they haven't yet understood that digital assets are a business issue and not an IT problem.
- We need better models to understand the cost of collaboration; and to understand the scale and costs of the R&D that may be needed
- There are ownership issues that cause problems around the openness (or not) of data; about how to define the costs of distributed costs centres; who actually owns digital collections; and monolithic IT budgets that can't be broken down into departmental figures.
- Human & managerial issues (rather than technical) require additional focus and resource
- Joined up infrastructure is expensive but is a requirement
- Sustainability is a big challenge and this has to be tackled by robust business and use cases; through automation rather than manual curation processes; and by making the activities (and the assets) more visible and apparent to the organisation
- Selection is happening but techniques need to evolve to cope with appraisal at scale
- The current software solutions are inadequate so demand and requirements need to be better articulated and tools need to be more carefully specified
- There is a lot of inertia and inflexible legacy working practices within organisations that slow down ingest; limit file format choices; hinder policy development and changes to working practices
- Finding properly qualified staff and the right kind of curation expertise is hard

The second discussion prompted the following thoughts in response to the Roadmap:

- The focus of the Roadmap is very much on the 'asset' nested within an 'organisation'; structures may change

over time and an alternative or additional focus might be on people and skills and emerging technologies

- Many organisations (especially libraries and archives) are still very wrapped up in dealing with analogue collections and the transition to digital and the curation challenges associated with this
 - Predicting 5 years into the future is a long or a short time depending on organisational context; the predictions for 2025 in the Roadmap are already being tackled in practice now
 - Message 1 ('Make choices and select') was one of the more problematic statements. Selection may be incompatible with 'big data' techniques and may also be in conflict with the mission of some libraries; but it may also be stating the obvious or rehearsing accepted practice in environments where digital curation is established
 - Message 2 ('Demand efficient systems') skews activity towards procurement rather than in-house development and assumes that there is already an effective marketplace and market analysis that can be drawn upon
 - Message 3 ('Build scalable infrastructure') was an uncontroversial message
 - Message 4 ('Sustainability') should extend beyond thinking about organisations and assets and should also include software and applications and embedding sustainability into up-front funding arrangements
 - Message 5 ('Make funding dependent on lifecycle costing') should be clearer about what the funding will actually support and be wary of inhibiting activity entirely
 - Message 6 ('Be transparent and share') should reference the power of open source and other 'open' concepts and emphasise the potential to improve quality
- There are general issues with definitions throughout the Roadmap, for example it may not be clear to everyone what is meant by, 'lifecycle', 'value' and 'efficient' in the context they are used
 - There are important contextual organisational differences that need to be acknowledged, particularly in cases where assets are generated internally or acquired from external sources; and where activity is community-led or where it is commercially-driven
 - Curation and preservation thinking needs to happen at the content (assets) level but also at the application (systems) level and at the platform (environment) level and this has economic implications
 - The issue of standards alignment and the convergence of practice is complicated and it is not clear whether it is an opportunity or a problem and how the economics work out in terms of community practice and functional markets
 - The roadmap needs to be clear about the ownership problem (see 'challenges' above) and who should be taking responsibility and in what context
 - There is much that can be learnt and taken from business and big data industries; public sector organisations should be more open to these ideas to introduce more economic practices
 - The Roadmap could set out more of a research agenda and provide an innovation platform for students and early-career researchers