

New Perspectives on Economic Modeling for Digital Curation

Neil Grindley
Jisc
Brettenham House
5, Lancaster Place, London
+44 (0)203 006 6059
n.grindley@jisc.ac.uk

Ulla Bøgvad Kejser
The Royal Library, Denmark
Postbox 2149
1016 København K
45 91324747
ubk@kb.dk

Hervé L'Hours
UK Data Archive
University of Essex, Wivenhoe Park
Colchester CO4 3SQ
+44 (0)1206 822669
herve@essex.ac.uk

ABSTRACT

Society is increasingly dependent on the availability of digital information assets however the resources that are available for managing the assets over time (curating) are limited. As such, it is increasingly vital that organizations are able to judge the effectiveness of their investments into curation activities. For those responsible for digital curation, it is an ongoing challenge to ensure that the assets remain valuable in a sustainable manner. Digital curation and preservation practices are still evolving and they are not well aligned across different organizations and different sectors. The lack of clear definitions and standardization makes it difficult to compare the costs and benefits of multiple curation processes, which again impedes identification of good practice. This paper introduces a new perspective on modeling the economics of curation. It describes a framework of interrelated models that represent different aspects of the economic lifecycle based around curation. The framework includes a sustainability model, a cost and benefit model, a business model, and a cost model. The framework provides a common vocabulary and clarifies the roles and responsibilities of managers with a demand for curation of digital assets and suppliers of curation services and solutions. Further, the framework reflects the context in which managers operate and how this context influences their decision-making. This should enable managers to think through different scenarios around the economics of curation and to analyze the impact of different decisions to support strategic planning. The framework is intended to serve as a basis for developing tools to help managers analyze the costs and benefits associated with curation. The models are being developed and refined as part of the EU project 4C "Collaboration to Clarify the Cost of Curation", which is bringing together and bridging existing knowledge, models and tools to create a better understanding of the economics of curation.

General Terms

Strategic environment, digital preservation marketplace, theory of digital preservation.

Keywords

Economics, models, curation, preservation, strategy, decision-making, costs, benefits, risks, sustainability.

1. INTRODUCTION

It is difficult for organizations responsible for managing and curating digital assets to know whether they are managing those assets cost-effectively. Irrespective of the sort of data they are managing (e.g. business records, research data, cultural heritage collections, personal archives, etc.), all organizations investing in curating digital assets will expect these assets to realize some form of value over short, medium or longer timescales.

The language used to describe the management of assets over time to release value should reflect commonly used economic principles and it is through this lens that the 4C project (a Collaboration to Clarify the Costs of Curation) examined the management of digital assets and developed our framework. The framework looks at the costs of curation activities; what benefits these activities bring to stakeholders (and society as a whole); and how knowledge about these costs and benefits can help stakeholders develop sustainable digital curation strategies. More specifically though, recognizes that the management of digital assets, the realization of value, and the ability to sustain those assets for as long as needed (to realize some value) all rely on an organizations ability to make sound investments into digital curation. Or to put it another way, digital curation is the pivot around which strategic and economic planning turns and it requires a sustainable flow of resources to support it.

To ensure timely resourcing, organizations that undertake digital curation need to understand the economic lifecycle that they operate within, the costs that are incurred, and the benefits that their assets may realize. This understanding must encompass their own business processes as well as the incentives that drive funders and other stakeholders. Suppliers of asset management systems and services need to have detailed knowledge of what activities are involved, how much they cost and what the cost drivers are. They also need to understand how the systems and services generate value for their customers.

Stakeholders from the demand and supply side depend on the availability of sound financial information for accounting and budgeting. As well as knowing the factual costs, for example, records of the capital and labor costs required to develop and operate a specific system, they must also have contextual information. Context includes underlying assumptions about what is being priced, for example, the quality of the service as well as an indication of the benefits – and thus the value – that such investments represent. This financial information allows financial transactions to be recorded and analyzed for internal management purposes and may also provide greater evidence and transparency for meeting external legal requirements. It can also provide a basis for the evaluation of possible solutions and thus support budgeting

iPres 2014 conference proceedings will be made available under a Creative Commons license.

With the exception of any logos, emblems, trademarks or other nominated third-party images/text, this work is available for re-use under a Creative Commons Attribution 3.0 unported license. Authorship of this work must be attributed. View a [copy of this licence](#).

and decision-making. This need for reliable and comparable financial information is exacerbated by the general growth in the amount and complexity of digital information assets that require management. This in turn puts curation budgets under pressure.

Models and tools have been developed to help organizations operate in the economic landscape and to assess the costs and benefits of digital curation. At first, interest was on assessing the costs of curation, but soon the importance of understanding the associated benefits, and stakeholder incentives for funding digital curation was also recognized by the community. This was not least owing to the extensive work of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access [4]. The 4C's Economic Sustainability Reference Model (ESRM) springs from this work and aims to assist the development of sustainable strategies for digital curation [9]. Tools have also been developed to support the definition and measurement of benefits of curating research data [2,3]. An overview of models and bibliographies can be found at the Open Planets Foundation website¹, in a blog post on the Signal [13] and in a deliverable report by the 4C project [8]. A more detailed description and evaluation of current cost and benefit models can be found in another 4C report [10].

Today's trends are towards developing a unified theory of how to model the costs and benefits of digital curation in a way that will facilitate comparison of alternative scenarios and selection of good practices to ultimately gain efficiencies in digital curation [15]. Despite all the effort being put into investigations of the economics of digital curation, there is still a need to improve the map of the economic digital curation landscape and to provide practical tools that help stakeholders navigate and better understand how curation investments become more sustainable.

2. ECONOMIC MODELS

The Economic Sustainability Reference Model (ESRM) is intended to be used as a strategic tool to support planning and provoke discussion and is primarily aimed at executive and managerial staff with responsibility for managing organizational budgets rather than operational level staff undertaking curation activities [9]. The ESRM provides a foundation for the development of sustainability strategies for digital curation by organizing the problem space; providing a common reference point of concepts and vocabulary; and introducing a layer of abstraction that hides the complexities and idiosyncrasies of individual implementations and contexts, while at the same time embodying sufficient detail to support substantive discussions of shared issues.

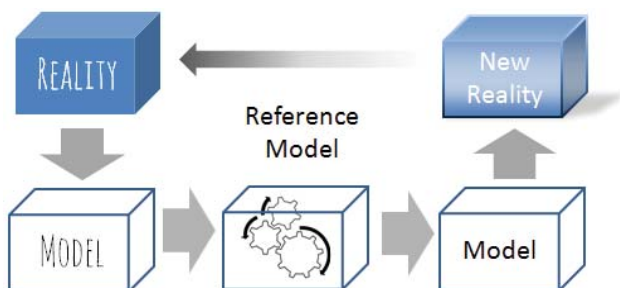


Figure 1. The purpose of a reference model.

¹ Open Planets Foundation, Digital Preservation and Data Curation Costing and Cost Modelling, <http://wiki.opf-labs.org/display/CDP/Home>

The intention of this reference model as represented in Figure 1, is to provide people with a method of comparing current practice with an abstracted and exemplary view of alternatives; and then to provide them with an approach to advocating for change.

In relation to the modeling of digital curation, the ESRM nests within the category of economic models and is a planning resource that does not require any technical knowledge of digital curation tools and techniques. Figure 2 shows a graphical depiction of the relation between a costs model, a benefits model and an overarching economic sustainability model.

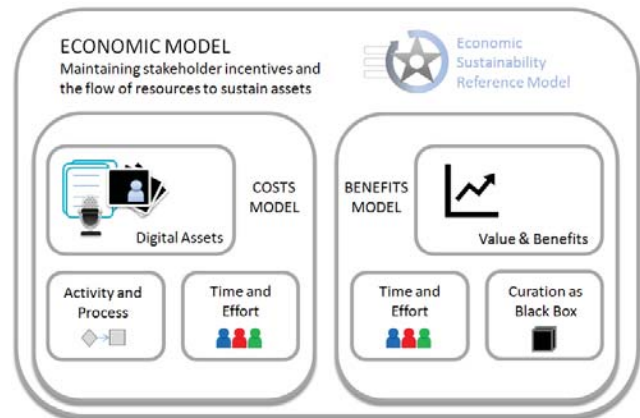


Figure 2. The nesting of costs and benefits modeling activities within the overarching framework of an economic model.

The aim of the nested model is to highlight that tackling the economics of digital curation requires a number of different perspectives and is comprised of a series of disparate tasks that occur across the curation lifecycle. Each of these tasks will be more or less achievable at different points in time depending on the organizational objectives, what resources are available to carry out curation tasks, and what information is available to help assess the potential impact of undertaking these specific tasks. At the most general economic modeling level, the motivation is to provide an understanding of why and how overall curation processes are likely to be economically affordable. This can be summarized as understanding the incentives to curate; and understanding how a flow of sufficient resources can be maintained to support these processes over time.

The ESRM maps out the key elements of the problem space planners face when designing a sustainability strategy for the digital curation processes they apply. It focuses on the general concept of a sustainability strategy, breaks it down into its key components, and draws planners' attention to the properties of those components most relevant for economic sustainability. The ESRM breaks down into four primary components:

- The Economic Lifecycle;
- Sustainability Conditions – value, incentives, selection, organization and resources;
- Key Entities - digital assets, curation processes and stakeholders (and stakeholder ecosystem);
- Uncertainties (Risks).

2.1 The Economic Lifecycle

Digital curation processes are assumed to be the central active component that require investment and are the mechanism that will ensure the sustainability of digital assets. Investment into

curation will in turn facilitate use (or the potential for use) of digital assets and will realize value, thereby delivering a return on the investment. This could play out in a linear fashion with assets being created, curated, used and then deleted according to a retention schedule; but in the context of sustainability, it is more likely to be a cyclical process with decision points occurring from time to time when some disruption is experienced. There will be a gap in the cycle when some kind of issue (e.g. financial, technical, business, reputational) introduces an uncertainty and this will provoke a decision point, as depicted in Figure 3. The decision might be articulated as, "are we willing to change the nature of our investment to respond to the issue(s) in order to ensure the sustainability of our assets?" The decision point would more usually be prompted by a threat rather than an opportunity but it is feasible that both scenarios could be substantially disruptive in different ways.

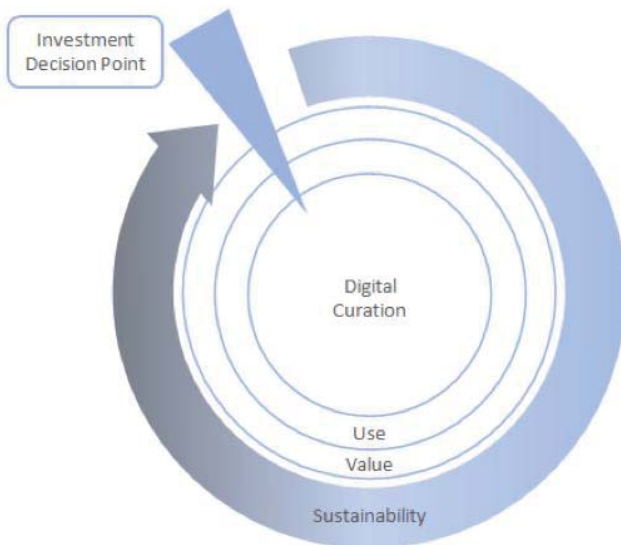


Figure 3. The ESRM Economic Lifecycle.

2.2 Sustainability Conditions

Five *Sustainability Conditions* set out issues that must be tackled to maximize the prospects for sustaining assets:

- Value – the assets must be perceived to have tangible or intangible value to relevant stakeholders;
- Incentives – relevant stakeholders must be sufficiently motivated to support and fund curation;
- Selection – where resources are scarce then discretion must be used to prioritize curation of the most valuable assets;
- Organization – the organization responsible for the curation of the assets should have an appropriate mandate; a supportive governance structure; and be optimally configured to sustain the assets;
- Resources – there must be a sufficient and ongoing flow of resources (including capital and labor) to achieve curation objectives.

2.3 Key Entities

Three *Key Entities* are proposed which are found in all digital curation contexts. Sustainability requires the nature of these entities to be understood:

- Assets – every type of digital asset exhibits various attributes or properties that to a greater or lesser extent may affect how they are curated;
- Stakeholders – the stakeholder ecosystem for digital assets can be complex and the supply side and demand side should be understood in relation to who is undertaking the curation and for the benefit of whom;
- Processes – they must be capable of (and optimized for) efficiently maintaining and possibly enhancing the value of the assets.

2.4 Economic Uncertainties (Risks)

The inclusion of *Economic Uncertainties (Risks)* is an acknowledgement that even the best sustainability strategy cannot accurately predict the future and that some expectation or mitigation of uncertainties should be built into the strategy (Figure 4).

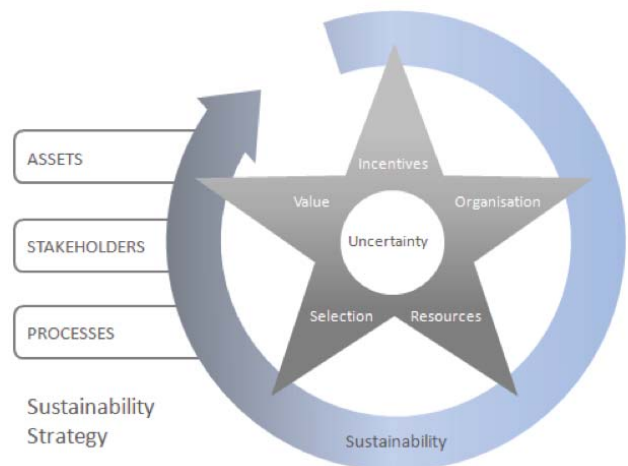


Figure 4. The ESRM components support the creation of a sustainability strategy for curation.

There is an enormous body of work on risk management and these methodologies should be employed, including the concept of negative and positive risks. Building flexibility into planning will allow the possibility of taking advantage of any opportunities that may present themselves (e.g. a cheaper service option becomes available from a different supplier; or a plan is mooted to massively upscale operations). It should also cope when a threat arises (e.g. a natural disaster substantially reduces world stocks of hard disks, or one of the major sponsors of activity unexpectedly withdraws support).

Examining the ESRM with its focus on sustainability is a useful approach to understanding the economic level of modeling, which encompasses the costs, benefits, and risks levels discussed below.

3. FRAMEWORK OF MODELS

The 4C project is developing a framework of models, terms and concepts to discuss and clarify economic decisions about digital curation and to provide common reference points. The framework is centered on the concept of the Curation Service, offered by a Provider to a Consumer (concepts are written with capital letters). The Provider and Consumer are decision-makers. Around this simple structure we then model different aspects of the economic lifecycle to explain the factors and mechanisms that impact on decision-making. The framework is shown in Figure 5.

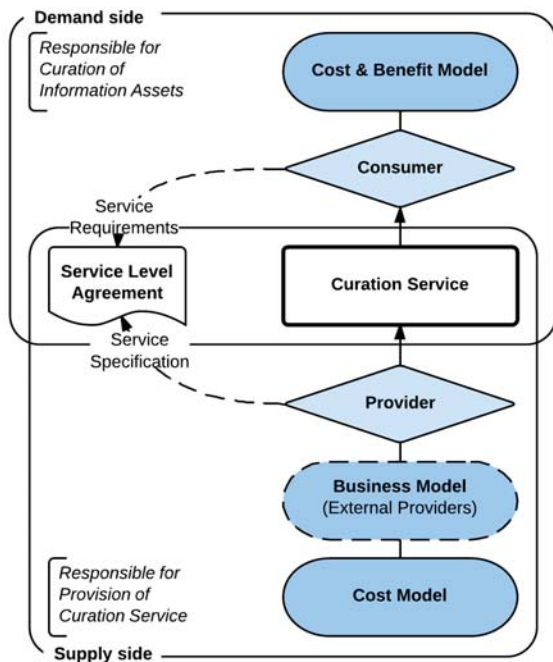


Figure 5. The 4C framework of economic models representing the demand and supply side of curation services.

The distinction between the two roles – representing the demand and supply side – is useful because the roles have different responsibilities reflecting different incentives for curation and different needs for tools. Even when services are provided in-house and the role of the Consumer and the Provider both reside within the same organization (or even with the same stakeholder group) it is useful to keep this distinction in mind when analyzing decision-making processes.

3.1 Curation Service

The Curation Service represents a value proposition; it incurs costs and should deliver benefit. It may cover the whole digital curation lifecycle or it may signify selected parts of the lifecycle, such as an ingest service or a storage solution. When it is provided in-house the Consumer can usually specify the requirements for the quality of the service – the Service Level – directly. When it comes to services that are outsourced, it may in some cases be possible for the Consumer to specify the required Service Level, while in other cases it may only be possible to select one or more predefined services.

The Curation Service can be defined in an agreement between the Provider and the Consumer, also known as the Service Level Agreement (Figure 5). Such agreements may be legally binding or have a more informal or ad hoc character, which is often the case with internal agreements, for example between two departments in an organization.

3.2 Consumer

The Consumer is responsible for the curation of information assets and must ensure that the applied Curation Service meets the organization's requirements in a sustainable way. To facilitate decision-making and strategic planning they typically use tools for costs and benefits analysis and risk management. In the framework, the demand side of the economic lifecycle is modeled by the Cost & Benefit Model.

Consumers, such as memory institutions, are of course also likely to use business models although not to address curation specifically. The value they propose to their users (and what needs to be addressed in their business case) is the services that curation enable, such as the ability to search for information assets across multiple collections. And the Cost & Benefit Model is intended to capture such benefits. Likewise, Consumers only need to know the overall costs and specifications of the quality levels of the services in order to balance cost and benefit. They see curation as a black box and do not normally need models to provide detailed cost information.

3.3 Provider

The Provider is responsible for delivering the Curation Service as agreed. The Curation Service can be supplied in-house or by outsourcing or in combination. External Providers need to generate sound business cases for services they offer, and ensure they provide return on investments (profit). Therefore, they need an exhaustive understanding of the costs associated with the services, and the cost drivers, as well as the value that the proposition brings to potential Consumers (customers). To facilitate these analyses they need business models and detailed cost models (see section 5). If the curation service is provided in-house, there may not be a need to develop a business case for curation, because the service may not be expected to realize a profit (this is indicated by the dotted line in Figure 5). In this case, Providers only need detailed cost models. The Business Model and the Cost Model represent the supply side.

Providers are also likely to use cost and benefit, and risk analysis tools, but not to optimize the curation of assets per se. Rather, these analyses are used to optimize their services, and for external Providers also their business cases and, as such, captured by the Cost Model and the Business Model.

4. COST & BENEFIT MODEL

In this section we describe the components of a conceptual Cost & Benefit Model for curation and explain how it can be used to analyze decision-making processes from the perspective of the Consumer. The model is depicted in Figure 6.

4.1 Objectives & Strategies

The Objectives & Strategies concept describes an organization's goals in terms of curation of the digital assets for which the organization, represented by the Consumer, is responsible, and outlines how it will reach these goals. The Consumer defines the Service Requirements for the Curation Service based on the Objectives & Strategies, and evaluates the Cost & Benefit of the service against these.

4.2 Organizational Context

The Objectives & Strategies are defined by the Organizational Context. Thus, Consumers make decisions in the light of the nature of the organizations and the information assets they hold, as well as stakeholders and the interests that they represent. Thus, they have to navigate a complex landscape consisting of a range of conditions where different influencers are likely to have different – and potentially conflicting – agendas. All of these intertwined internal and external conditions influence the decision-making process. To clarify the conditions we divide the Organizational Context into three key aspects:

- Organization (Mission, People, Systems)
- Information Assets (Quantity, Quality)
- Stakeholders (Internal, External)

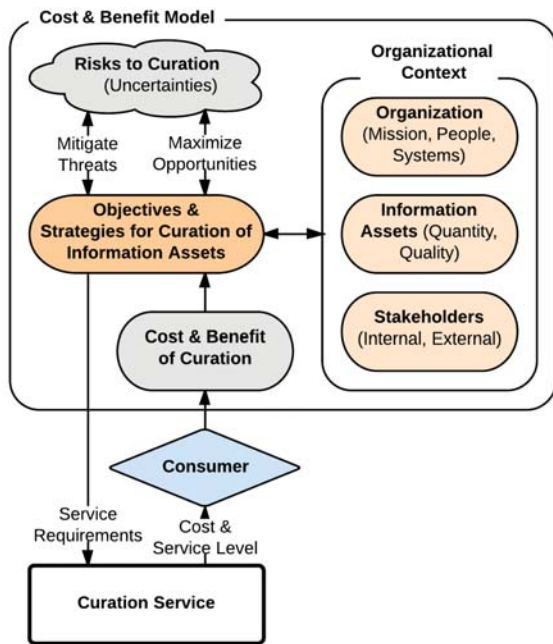


Figure 6. The Cost & Benefit Model for Curation represents the Consumer perspective.

4.3 Risks

The Objectives & Strategies are also influenced by Risks to Curation. This concept represents the effect of uncertainty on curation objectives. It encompasses both negative risks (threats) and positive risks (opportunities). The risks must be articulated and managed through curation strategies to minimize threats and maximize opportunities as illustrated in Figure 6. There are costs and benefits associated with mitigating or maximizing risks. The ability of a Curation Service to enhance positive risks is obviously a benefit, but so is the ability to mitigate negative risks. Again the value of the benefit will depend on the organization that the Consumer represents. If for example, an investment results in mitigation of a negative risk, this only represents value proportionally with the Consumer's incentive to reduce this risk.

4.4 Cost and Benefit

There are Cost and Benefit associated with meeting an organization's curation objectives, materialized as the Curation Service. As described above, an organization's objectives and strategies are likely to change over time influenced by its context and any risks that may be encountered. The changes further impact the requirements for services and, eventually, the Cost and Benefit of curation.

4.4.1 Costs

The costs of a Curation Service depend on which activities are included in the service and on the quality of the activities undertaken – the Service Level. Once all the involved activities have been identified and qualified, and resources attached to them, it is in principle possible to calculate the cost of the specified Curation Service. The core cost concepts needed to model these relations are described in section 5.

4.4.2 Benefits

In contrast the benefits – the advantages – of a Curation Service can only be identified and evaluated from a specific Consumer perspective. For example, if the proposed service consists of a

system designed to minimize loss of data by providing multiple replicas, the perceived benefits of this service will depend on the Consumer's willingness to accept the risk of losing data. This subjective nature of benefits is illustrated in Figure 6 where the Cost & Service Level represents the information associated with the delivered service. Through the Consumer the Cost & Service Level is transformed to Cost & Benefit.

4.4.2.1 Valuation of benefits

In formal cost and benefit analysis the value of the benefits of the curation service are summed up and then the costs of providing the service are subtracted to ideally reveal the net value of the service to a given Consumer. Some benefits have a market price and it is therefore relatively easy to measure their value. Examples include the benefits of a music service that offers streaming of songs based on user fees or licenses, or the benefits of cost savings gained by investments in more efficient curation services. These benefits are also called financial or economic benefits. However, if there is no conventional market on which a benefit can be traded, no market price can be applied. It is for example difficult to assess the benefits of Europeana.eu, which aggregates European memory institutions' cultural heritage assets to make them more easily accessible to the general public, or benefits in the form of good will returned to an organization from investments in better trustworthiness of a repository. Even though, such non-financial or non-economic benefits do not have a direct market price, they still represent real value to stakeholders. Economists measure the value of benefits that do not have a market price by so-called non-market valuation techniques such as revealed preferences, which analyze past behaviors, and stated preferences (also known as contingent valuation), which asks hypothetical questions, for example about willingness to pay for a predefined change in the quality a service.

4.4.2.2 Identification of Benefits

To justify costs it is important for organizations (Consumers) to elicit and describe what the benefits of curation are, who they will benefit, how valuable they are to stakeholders, and possibly also indicate how likely it is that the benefits will realize value, and when this value will be realized. The Cost & Benefit Model provides a structure that can be used as a starting point for the identification of benefits. Thus, extending the concepts to actual instances and describing an organization's Objectives & Strategies, Stakeholders, Risks, and so on, should make it more clear to the Consumer what the benefits are.

5. BUSINESS MODEL AND COST MODEL

In this section we describe the Conceptual Cost Model (CCM) for curation and show how it relates to the Business Model. The models are depicted in Figure 7. The Business Model is not described in detail in this paper since it is still in its development phase and has not yet been fully conceptualized. Further information about the conceptual cost modeling can be found in a deliverable report by the 4C project [14].

The intention of the CCM is to provide a common foundation on which tools for assessment of curation costs can be built and to enable the specific costs of curation services and solutions to become more comparable. A concept is an abstract idea generalized from specific instances, and building on a common foundation, should enable the tools to provide comparable cost calculations at some level. The closer a tool gets to representing specific curation scenarios the more accurate the calculations are likely to be. However, the closer to specific scenarios, the less comparable the resulting cost calculations will be.

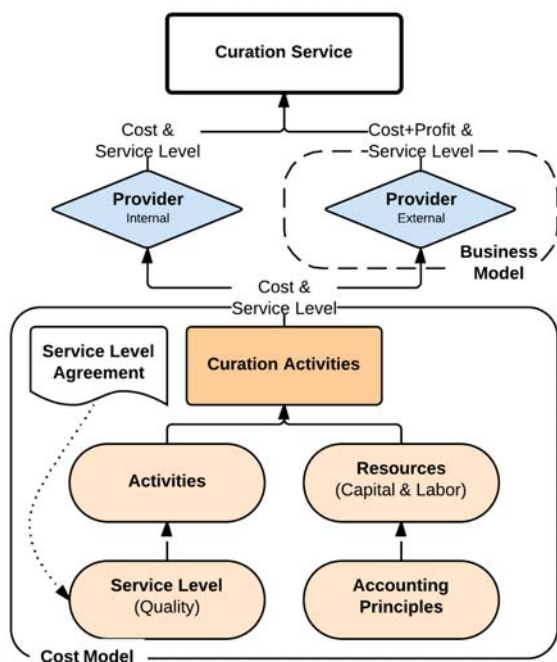


Figure 7. The Business Model and Conceptual Cost Model (CCM) represents the Provider perspective.

A cost model for curation in this context is defined as a representation that describes how Resources – direct capital and labor costs, as well as indirect costs (overheads) – required for accomplishing digital curation activities relate to costs. Cost models can further be characterized by their cost structure – the way they define and breakdown Activities and Resources, and by the way they define and handle the variables that influence the costs.

It is important for any organization providing a Curation Service to understand the distribution of costs, and what the most important curation costs are because these costs need special attention and careful management. Service Providers have to understand the factors that drive the costs up or down, such as the quantity and quality of the information assets and the length of time that the information assets will need to be curated – short or longer-term. Thus, there are many dependencies that the Provider must be aware of, for example, the costs of any systems and staff skills that are critical for delivering the service. They also need to consider how costs are likely to develop in the future, including considerations of possible financial adjustments caused by inflation or deflation. Costing digital curation is not a trivial task for a number of reasons, not least because we do not have a common understanding of the component Curation Activities [12].

5.1 Curation Activities

The costs of a Curation Service depend on the Curation Activities required to accomplish the service and on the Service Level (quality) of the activities. If the service is supplied by an external business Provider profit is normally added to the cost of delivering the Curation Service (Figure 7). Thus, the output of the CCM is a specification of the Service Level and the corresponding Cost, while the output of the Business Model, among other things, is a specification of the Service Level and the Cost including any profit.

There are many interrelated activities involved in curation and these can be implemented in many different ways and they can be set up to meet different quality requirements. This complexity makes it hard to specify the Curation Activities in a precise and clear-cut way, and it makes it difficult to delimit the costs from other business costs. Thus, there are no standardized ways of breaking down and accounting for the cost of Curation Activities. On top of this, the activities depend on constantly evolving technologies, which in turn leads to repeated changes in systems and procedures, and thus also in the costs.

5.1.1 Activities

There are numerous ways to define and breakdown activities. From the curation cost perspective we simply define an activity as a measurable amount of work performed by systems and/or people to produce a result. In order to achieve a measurement of an activity we need to break it down to a level at which we can specify the required resources, and thus get an estimate of the costs of performing the activity. The required level of granularity is also related to the required level of accuracy of the estimate.

The 4C project has used the OAIS standard [5] for a trustworthy repository as the basis for defining curation activities. The standard includes a functional model that describes a conceptual repository and three roles that interact with the repository, namely Manager, Producer and Consumer. The functional entity model divides digital preservation activities into seven functional entities: Ingest, Data Management, Archival Storage, Access, Preservation Planning, Administration, and Common Services, and these entities are further broken down in individually described functions. The PAIMAS standard [6] is an adjunct to OAIS, which provides more detailed specification of the activities around the transfer of information assets from the Producer to the repository.

Given our aim to design a generic framework to support the full breadth of possible future research and development in cost and benefit methods, we have concluded that the OAIS model, which is a well-established international standard in the field of digital preservation, provides the best starting point for breaking down Curation Activities. In fact the OAIS functional model has also been applied as a basis for the description of activities in most of the current cost models [11]. However, there are a series of challenges with applying the OAIS functional model directly to curation cost modeling.

First of all, the OAIS functional descriptions are intentionally described at an abstract and implementation neutral level. It is intended as a 'reference model'. However, costs can only be assessed against actual processes and systems. Both off the shelf services and solutions developed for specific purposes may cover multiple OAIS entities/functions or only parts of them. In these cases some mapping between the Curation Activities and OAIS entities/functions is required. Such mapping is difficult and it is further complicated by the fact that, due to the complexity of the involved activities, some of the OAIS terms are not easily understood or self-explanatory.

Second, the OAIS standard only addresses long-term digital preservation within the 'archival phase', whereas 4C aims to take a broader approach to curation such as expressed by the DCC Lifecycle viewpoint², which incorporates conceptualization, data creation/capture and the use and reuse of information assets.

² DCC Curation Lifecycle Model, <http://www.dcc.ac.uk/digital-curation/digital-curation-faqs/dcc-curation-lifecycle-model>

Further, it also applies to organizations and projects with a remit limited to short and medium term storage. Thus, curation covers the full lifecycle of information assets, and these activities may be expressed by the three OAIS roles covering production, use, and management activities in addition to the repository activities.

In conclusion we have decided to use the OAIS standard to populate the activity model in the framework as far as possible, but we also acknowledge there may be a need to bend the standard in some ways to make it more applicable to costing. Any such amendments would need to be justified by a particular curation cost model developer. The proposed framework extends to support the full curation lifecycle and divides activities in levels, starting from the high-level roles, functional entities and functions, which are used by the OAIS standard and, if required, allowing for further breakdown of OAIS functions into measurable activities.

The activity breakdown structure includes the following entities:

- Production: including for example conceptualization, creation of information assets, capture, and digitization
- Pre-ingest: including for example appraisal, selection, and preparation for ingest
- Ingest: ingest of information assets
- Storage: short and long-term storage and maintenance of information assets
- Data Management: management of descriptive and administrative data
- Access: provision of access to information assets
- Lifecycle Planning: planning, research and development of curation activities
- Administration: administration of repository systems, standards and policies
- Common services: including services necessary to support a repository such as inter-process communication, name services, temporary storage allocation, exception handling, security, and directory services
- Use: use and re-use of information assets, including for example interfaces for crowdsourcing
- Management: including for example the provision of overall budgets and policies, and any certification related activities

5.1.1.1 Service Level

The Service Level defines the quality of the Activities. It is usually specified in a Service Level Agreement (Figure 5 and 7). The lack of a clear way of defining and measuring Service Levels represents an important challenge in cost and benefit modeling because of the close relationship between the Service Level of the Curation Activities and the Cost, as well as between the Service Level of the activities and the Benefits perceived by the Consumer. If for example we consider the activity to 'store information assets' the Service Level of the activity may among other things specify that three copies of the assets are stored. All other things being equal, the Cost of this activity will be proportional to the number of copies specified. Likewise, the number of copies will normally be proportional with the level of information integrity because the more copies the lower risk of data loss. However, it will be inversely proportional to the level of confidentiality because the more copies that exist, the higher the risk of compromising access. Therefore, the same Service Level (quality) of the activity may have different value to different

Consumers, depending on the Service Requirements in relation to costs, integrity, and confidentiality.

The Service Level may be evaluated through quantitative (e.g. pass/fail, minimum score, certification level) or qualitative measures (such as descriptions of the quality). Thus, the Service Level can be a defined quality criteria for an activity; a more complex and formal agreement between two or more units; or a higher level of service 'quality' formalized through a certification process, for example through ISO 9000³ or ISO 27000⁴. There are also more or less standardized ways to certify the quality of repositories for long-term preservation and access. For example, ISO 16363 [7], Data Seal of Approval (DSA)⁵, Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC)⁶, Information and Documentation - Criteria for Trustworthy Digital Archives (DIN 31644)⁷. These audit and certification instruments can help to establish quality measurements.

5.1.2 Resources

Activities are performed both by systems and people. Thus, to complete an activity a certain amount of resources are required, and for accounting purposes these are often divided into Capital and Labor costs. Resources are what must be expended to deliver activities.

Capital Costs include, for example, building space (server space, office space, and so on), equipment (servers, network, and the like), energy (for systems, cooling, et cetera) and materials (storage media, and so on). Depreciation (for tangible assets) and amortization (for intangible assets) are mechanisms for distributing capital costs over the estimated useful lifetime of an asset to indicate how much of an asset's value has been used. For example, the time in which a server becomes obsolete may be five years. With a 5-year time period the cost of using this resource will be its acquisition cost, whereas with a 1-year period the cost would be the depreciated acquisition cost.

Labor costs consists of salaries and any benefits paid to staff for a period of time or for a certain job. Salaries are normally differentiated by job functions (developer, metadata officer, etc.) and possibly also by skill level, seniority and/or performance. The labor costs required to complete an activity can be expressed as a monetary value – the cost of salaries multiplied by time expended on the activity – but they may also be expressed simply in time – as the time it takes to complete the activity for a certain job function. The advantage of measuring labor costs in time is that it makes the figures more comparable across organizations and countries, where there may be significant differences in salaries. If needed the time measure can be translated into monetary values for a specific scenario. If for example the cost of running a system

³ ISO 9000 Quality Management, http://www.iso.org/iso/home/standards/management-standards/iso_9000.htm

⁴ ISO 27000 Information Security Management Systems, http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?csnumber=63411

⁵ Data Seal of Approval (DSA), <http://datasealofapproval.org/en/>

⁶ Center for Research Libraries (CRL), <http://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/trac>

⁷ Nestor Seal for Trustworthy Digital Archives, http://www.langzeitarchivierung.de/Subsites/nestor/EN/nestor-Siegel/siegel_node.htmltml

takes a developer 20 hours per week, this figure can be multiplied with salaries applicable to the job functions in different countries. Along this line, the unit Full Time Equivalent (FTE) is used to make workloads comparable. FTE expresses the workload as the ratio of the total number of working hours during a certain period by the number of full-time working hours in that period. 1 FTE is equivalent to that of a person working full time for a year.

Capital and labor costs can also be divided in direct and indirect costs. Direct costs are those directly used for performing digital curation activities, such as costs of acquisition of storage media or the costs of staff employed to add metadata. Indirect costs, also called residual costs or overheads, are those incurred by the usage of shared resources, such as general management and administration or common facilities and systems, where it has not been feasible to allocate the cost to specific activities.

Variable costs fluctuate depending on the amount of activities being undertaken and are differentiated from fixed costs, which do not depend on the amount. For example, the cost of materials used to complete an activity is a variable cost, as opposed to salaries and rents, which are fixed regardless of the amount of activities. Thus, variable costs are normally equal to direct costs and fixed costs to indirect costs. However, given enough scale and time, no cost is really fixed.

Costs can also be divided in one-time costs, periodic (term) costs or recurring costs, depending on the time period. The term capital or investment cost is often used to denote a one-time cost incurred upon the acquisition of equipment such as a storage system. The term periodic cost is used to indicate that the cost will be incurred at irregular intervals. Recurring costs, also known as running costs or operating costs, include costs relating to the consumption of media, energy and labor.

Other important time related aspects of costs include inflation (general price increases), individual price changes that are related to specific resources – such as storage media, energy, office space, computer scientist wages – and interest, which reflect economic growth and cost of capital. Even though the cost of resources has in general been increasing, the cost of both capital and labor per unit of digital information assets has, due to technological innovation, been decreasing over the past decades (although at very different rates). Therefore, in order to calculate the present value of estimated future costs different discount rates are preferable. The present value is needed in order to compare different cost scenarios over time.

Costs can be divided by accounting periods to capture past cost (ex post) and/or future costs (ex ante). Records of past cost are used in accounting whereas estimations of future costs over certain time periods (such as months, quarters, and years) are used for budgeting.

5.1.2.1 Accounting Principles

Accounting can be defined as a set of concepts and techniques that are used to measure and report financial information about an economic unit [16]. In order to make financial reports understandable and comparable between organizations, the reports need to follow generally accepted accounting principles (GAAP) defined by national and international standardization bodies. The International Financial Reporting Standard (IFRS) Foundation is an independent, not-for-profit private sector organization working in the public interest to develop and promote the use of a single set of globally accepted, international financial reporting standards through its standard-setting body the International

Accounting Standards Board (IASB)⁸. Thus, the Accounting principles, delivered as national or international standards should govern standard accounting practices.

Just as it can be difficult to segregate the costs, which are incurred when carrying out Curation Activities it can be difficult to segregate costs that are incurred within Resources. The Transparent Approach to Costing (TRAC)⁹, which is applied in Higher Education in the UK, has been suggested as a concrete tool for recording resource cost data in relation to research data [1].

6. DISCUSSION

The approach taken has been to accept that the models have different purposes (communication, simplification, common understanding of basic relationships, complex expression of curation concepts in a specific context) and that where there are overlaps, either in purpose or terminology, perfect interaction and synchronization between them will not always be apparent. But the important factor is to understand that no particular approach or view of a system exists in isolation and that, where possible, models should be designed and expressed within the context of the higher level and more granular surrounding models. The ESRM and the framework help to clarify and signpost these relationships.

The establishment of the framework with its distinction between those with a demand for curation of assets and those that supply curation services has enabled us to clarify roles and responsibilities at the conceptual level, namely that of the Consumer and the Provider. The distinction may seem rigid and indeed in real life roles are often less clearly defined, but it has proved useful for identifying the kind of models and tools that are required to support decision-making related to the economics of digital curation. Further, it has been useful for clarifying the relationships between the different models (Cost & Benefit, Business Model, Cost model), as well as to define the kind of financial information the models deliver.

On the demand side we found that to ensure that the information assets remain sustainable Consumers basically need tools for analyzing the cost and benefits of Curation Services. This includes the ability to assess the cost and benefit of alternative services and of managing risks. As a first step to facilitate such analyses the Cost and Benefit Model defines and describes – at a conceptual level – the dynamics of the determinants that influence the costs and benefits of curation including risks. The model is still under development, but we have shown how it may already help identify potential benefits of curation.

On the supply side we found that Providers need tools that will help them assess how the costs vary with the quality of the service being applied. To this end it became clear that it is also necessary to distinguish between internal and external Providers. The reason is that the latter need business models in addition to cost models, to generate profitable business cases.

An ongoing challenge is the tension between the need for very specific local application of terms and concepts and the need to have common terms and classifications if models and their outputs are to be more generally understood and ideally comparable. These tensions between generally applicable and

⁸ International Financial Reporting Standards (IFRS) Foundation, <http://www.ifrs.org/The-organisation/Pages/IFRS-Foundation-and-the-IASB.aspx>

⁹ TRAC, <http://www.jcpsg.ac.uk/guidance>

understood concepts and the need for local specifications apply throughout complex systems of all types. There is not yet any authority to yield a 'big stick' when encouraging the use of standardized terms and classifications, And only by researching, defining and presenting likely 'controlled vocabularies' and promotion of the benefits of their re-use will we see the slow agreement and use of common definitions.

The framework we have described here is conceptual. There are more advantages of describing the models at a conceptual level. First of all, it provides a common framework for defining the cost and benefit of a curation service unambiguously which is a prerequisite for making cost and benefit comparable across different scenarios. At the conceptual level the model should in principle be able to encompass all use cases and in this sense it may serve as a guide for developers of cost and benefit models.

Also the concept models supports the clarification of central economic terms and encourages a common language around costs and benefits, and in this way it also supports communication and exchange of knowledge. The lack of a universally accepted terminology and clarification of cost and benefit concepts has previously been shown to be an important obstacle for reaching consensus on how to model these [11]. The 4C project is developing a Curation Costs Exchange platform (CCEx) where cost data and information about the cost data can be shared¹⁰. A key aim for CCEx is to employ standard use of terms and classifications.

Given the complexity of assessing costs and benefits and the entailed complexity of any tool aiming to simulate this complexity, it is unlikely that any single tool will be able to handle all scenarios. However, it may be realistic that tool developers can use the concept model as a basis to ensure that the resulting assessments are comparable, and then develop tools on top of the model for different groups of similar stakeholders (profiles). It should be possible for developers of cost and benefit tools to interpret and populate the concepts according to the context they need to address whilst maintaining references to more generic elements. This should make it possible to provide financial information that maps onto comparable entities, which in turn may mean that profiles for specific types of organizations working in similar environments can be developed.

Tackling complexity by the application of detailed models is likely to come with increased costs of collecting the required cost data and information, and these costs must be justified by a correspondingly greater utility of the results. So it is important for users to define the purpose of the modeling in order to understand their requirements in terms of the degree of granularity and accuracy that they will expect the model to deliver. The process to define activities is in general beneficial to any organization since it will improve their understanding of the activities and workflows and allow for possible optimizations.

We have decided to base the generic CCM on the functional model defined in the OAIIS standard. Even though OAIIS is a reference standard and does not define the entire digital curation lifecycle it is still the most detailed and widely used standard that relates to the field of digital curation. However, in order to encompass curation scenarios other than those for long-term trustworthy preservation, there is a need to relax some of the requirements, for example, to encompass scenarios where

information assets only need to be retained for the short or medium term.

Extensions of the OAIIS model to cover the full lifecycle are critical to the remit of 4C and curation costing in general, as are exceptions which support those with responsibility for storing information assets over the short and medium term (e.g. encompassing storage as well as full archival storage) but until the OAIIS has been specifically researched and found appropriate for cost-assignment, or a commonly accepted alternate approach has been developed, these core functional entities should remain our common benchmark and deviations from that benchmark should be documented and justified when applied to a particular curation cost methodology. These may be primarily for practical reasons such as dividing the more esoteric costs of planning, management and administration into more direct cost centers such as production, ingest, storage and access.

Similarly maintaining a clear link between terminology and the OAIIS benchmark and those used in a particular approach will support the ongoing comparison of approaches. This will help to drive adoption of a common approach by defining how the model and specification should be updated over time to take account of changes in the broader environment.

7. CONCLUSIONS

In this paper we investigate the usefulness of new approaches to modeling the economic landscape of curation and have set out a nested Economic Sustainability Reference Model, which indicates some hierarchy of scope. An economic level of modeling is the broadest and most encapsulating activity and subsumes not only all of the other approaches referenced in this paper but also has a relationship with business models. This has not been touched upon in detail here but is, in fact, being addressed by ongoing work on the 4C project. Sustainability planning is proposed as a form of economic modeling and one that can largely stand in to represent how to think about digital curation from an economic perspective. 'Largely' rather than 'wholly' to acknowledge the gap left by business planning and the related analyses and assertions that would form part of that process.

The next nested layer focuses on costs and benefits modeling considered as a dual concept and providing a framework for sensibly informing decisions that may need to be taken in relation to adopting or rejecting curation services.

Also we have described a framework of conceptual models, including a Cost & Benefit Model, a Business Model and a Cost Model focusing on the roles and responsibilities of the Consumer and Provider of Curation Services and shown how it can help clarify decision-making processes. More specifically it has clarified the relation between the models and their outputs. In addition, it has highlighted that while the costs of curation can in principle be assessed objectively once you have identified the activities involved and the resources required to complete them, the value of benefits of curation can only be assessed in relation to a specific stakeholder.

The work set out in this paper leads to some conclusions about future work and much of this follows from the points made above (see section 6 - Discussion).

* This is a complex area and there is further work to do to adequately join up existing models and to define new ones that will help to make sense and provide a more coherent perspective on the economics of digital curation;

¹⁰ Curation Costs Exchange (CCEx):
<http://www.curationexchange.org>

* Related to that complexity, a lot more work needs to be done to standardize terminology and all types of modeling (economic, costs, benefits and business) need further validation from diverse groups of stakeholders;

* The OAI is an imperfect foundation for breaking down activity-based costing approaches but it is the only real practical and widely accepted standard that can currently be referenced.

Looking specifically at two of the diagrammatic representations in this paper (Figure 3 and Figure 6) another conclusion that presents itself is the importance of the decision-making moment as a fundamental design feature of economic modeling.

It is also clear, in terms of the work that the 4C project has done, that the models and other resources are beginning to usefully join up concepts and link the whole area together but there is a great deal more work that can now more clearly be set out. This can usefully be described and addressed by the 4C Roadmap [17], which will be the final output of the project and will synthesize all of the learning and conclusions into an action agenda for the wider community.

8. ACKNOWLEDGMENTS

We thank 4C Project colleagues who have done extensive and collaborative work on many aspects of this paper.

The 4C project is funded by the European Commission as part of the FP7 ICT Cooperation programme, contract 600471.

9. REFERENCES

- [1] Beagrie, N., Chruszcz, J., and Lavoie, B. 2008. *Keeping Research Data Safe: A Cost Model and Guidance for UK Universities*. JISC, p.13, <http://www.jisc.ac.uk/publications/reports/2008/keepingresearchdatasafe.aspx>
- [2] Beagrie, N. 2011. *User guide for keeping research data safe. Assessing costs/benefits of research data management, preservation and re-use*, Version 2.0., Charles Beagrie Limited, p.24-31, http://www.beagrie.com/static/resource/KeepingResearchDataSafe_UserGuide_v2.pdf
- [3] Beagrie, N. and Houghton, J. W. 2014. *The Value and Impact of Data Sharing and Curation: A synthesis of three recent studies of UK research data centres*, Jisc., p. 8-11 http://repository.jisc.ac.uk/5568/1/iDF308_-_Digital_Infrastructure_Directions_Report%2C_Jan14_v1-04.pdf
- [4] Blue Ribbon Task Force on Sustainable Digital Preservation and Access. 2010. *Sustainable Economics for a Digital Planet: Ensuring Long-Term Access to Digital Information*, Final report, http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf
- [5] CCSDS (Consultative Committee for Space Data Systems). 2012. *Reference Model for an Open Archival Information System (OAIS)*, CCSDS 650.0-M-2, Magenta Book, (ISO14721:2012), <http://public.ccsds.org/publications/archive/650x0m2.pdf>
- [6] CCSDS (Consultative Committee for Space Data Systems). 2004. *Producer-Archive Interface Methodology Abstract Standard (PAIMAS)*, CCSDS Magenta Book, (ISO 20652:2006, reviewed and confirmed in 2009) <http://public.ccsds.org/publications/archive/651x0m1.pdf>
- [7] CCSDS (Consultative Committee for Space Data Systems). 2011. *Audit and Certification of Trustworthy Digital Repositories*, CCSDS 652.0-M-1, Magenta Book, (ISO 16363:2012), <http://public.ccsds.org/publications/archive/652x0m1.pdf>
- [8] Ferreira, M. and Farier, L. 2013. *Baseline Study of Stakeholder & Stakeholder Initiatives*, Deliverable report D2.1, Collaboration to Clarify the Costs of Curation (4C), <http://www.4cproject.eu/d2-1-stakeholders>
- [9] Grindley, N. and Lavoie, B. 2013. *Draft Economic Sustainability Reference Model*, Milestone report, Collaboration to Clarify the Costs of Curation (4C), <http://www.4cproject.eu/ms9-draft-esrm>
- [10] Kejser, U.B. et al. 2014. *Evaluation of Cost Models and Needs & Gaps Analysis*. Deliverable report D3.1, Collaboration to Clarify the Costs of Curation (4C), <http://www.4cproject.eu/d3-1>
- [11] Kejser, U.B., et al. 2014. State of the Art of Costs and Benefit Models for Digital Curation. In *Proceedings from IS&T Archiving Conference*, Archiving 2014 Final Program and Proceedings, ISSN 2161-8798, Online ISSN: 2168-3204, 144-149, <http://ist.publisher.ingentaconnect.com/content/ist/ac>
- [12] Kejser, U.B, Nielsen, A. B., Thirifays, A., 2012. Modelling the Costs of Preserving Digital Assets, In *Proceedings of UNESCO Memory of the World Conference "The Memory of the World in the Digital Age: Digitization and Preservation"*, Eds. L. Duranti, E. Schaffer, 529-539, http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/mow/VC_Kejser_et_al_27_B_1350.pdf
- [13] Lazorchak, B. 2014. A Digital Asset Sustainability and Preservation Cost Bibliography, *The Signal – Digital Preservation*. Blog post. <http://blogs.loc.gov/digitalpreservation/2014/01/a-national-agenda-bibliography-for-digital-asset-sustainability-and-preservation-cost-modeling/>
- [14] L'Hours, H. et al. 2014. *Cost Concept Model and Gateway Specification*, Deliverable report D3.2, Collaboration to Clarify the Costs of Curation (4C), <http://www.4cproject.eu/d3-2-ccm-grs>
- [15] Lungu et al. 2012. Economic Alignment. In *Aligning National Approaches to Digital Preservation*, Ed. N. McGovern, 195-268, <http://educopia.org/publications/ANADP>
- [16] Walther, L. M. 2013. Welcome to the World of Accounting. In *Financial Accounting*. <http://www.principlesofaccounting.com/chapter1/chapter1.htm>
- [17] 4C Project, 2014. *Investing in Curation: A Shared Path to Sustainability*. 4C Draft Roadmap. <http://4cproject.eu/d5-1-draft-roadmap>