Collaboration to Clarify the Cost of Curation

D3.2- Cost Concept Model and Gateway Specification

Deliverable Lead:	UK Data Archive (UESSEX)
Related Work package:	WP3—Assessment
Author(s):	Hervé L'Hours (UK Data Archive, UESSEX) Ulla Bøgvad Kejser (KBDK) Kathrine Hougaard Edsen Johansen (DNA) Alex Thirifays (DNA) David Wang (SBA) Stephan Strodl (SBA) Kevin Ashley (UEDIN-DCC) Joy Davidson (HATII-DCC) Patrick McCann (HATII-DCC) Jaan Krupp (NLE) Neil Grindley (JISC)
Dissemination level:	PU
Submission date:	04 August 2014
Project Acronym:	4C
Website:	http://4cproject.eu
Call:	FP7-ICT-2011-9
Project Number	600471
Instrument:	Coordination action (CA)—ERA-NET
Start date of Project:	01 Feb 2013
Duration:	24 months

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	2014-01-14	First draft	HL'H
0.02	2014-02-18	Restructure	HL'H
0.03	2014-02-26	Re-review version	HL'H
0.04	2014-03-24	Stable pre-review version	HL'H
0.05	2014-05	Interim release pre-delivery of additional content from partners	HL'H
0.06	2014-06-20	Stable release	HL'H
0.07	2014-06-27	Pre-Edinburgh Meeting Release	HL'H
0.08	2014-07-11	Stable release to project team	HL'H
0.08	2014-07-22	Integration of additional text and diagrams	HL'H
0.10	2014-07-23	Handover for final project review	HL'H
0.11	2014-07-29	Internal reviewer comments. Final restructuring. Minor version history removed for brevity.	KGA/JD
1.00	2014-08-04	Final edit. Release version	PLSS

Acknowledgements

This report has been developed within the project “Collaboration to Clarify the Cost of Curation” (4cproject.eu). The project is an ERA-NET co-funded by the 7th Framework Programme of the European Commission.

The 4C participants are:

Participant organisation name	Short Name	Country
Jisc	JISC	UK
Det Kongelige Bibliotek, Nationalbibliotek Og Kobenhavns Universitetsbibliotek	KBDK	DK
Instituto de Engenharia de Sistemas e Computadores, Investigacao e Desenvolvimento em Lisboa	INESC-ID	PT
Statens Arkiver	DNA	DK
Deutsche Nationalbibliothek	DNB	DE
University of Glasgow	HATII-DCC	UK
UK Data Archive (University of Essex)	UESSEX	UK
Keep Solutions LDA	KEEPS	PT
Digital Preservation Coalition Limited by Guarantee	DPC	UK
Verein Zur Forderung Der It-Sicherheit In Osterreich	SBA	AT
The University of Edinburgh	UEDIN-DCC	UK
Koninklijke Nederlandse Akademie van Wetenschappen -Knaw	KNAW-DANS	NL
Eesti Rahvusraamatukogu	NLE	EE

Disclaimer: The information in this document is subject to change without notice. Company or product names mentioned in this document may be trademarks or registered trademarks of their respective companies.

D3.2- Cost Concept Model and Gateway Specification by 4cproject.eu is licensed under a Creative Commons Attribution-ShareAlike 3.0 Unported License.

This document reflects only the authors’ view. The European Community is not liable for any use that may be made of the information contained herein.

Author(s):	Hervé L'Hours (UK Data Archive, UESSEX) Ulla Bøgvad Kejser (KBDK) Kathrine Hougaard Edsen Johansen (DNA) Alex Thirifays (DNA) David Wang (SBA) Stephan Strodl (SBA) Kevin Ashley (UEDIN-DCC) Joy Davidson (HATII-DCC) Patrick McCann (HATII-DCC) Jaan Krupp (NLE) Neil Grindley (JISC)

Table of Contents

Figures

Executive summary

This document introduces a Framework supporting the implementation of a cost concept model against which current and future cost models for curating digital assets can be benchmarked. The value built into this cost concept model leverages the comprehensive engagement by the 4C project with various user communities and builds upon our understanding of the requirements, drivers, obstacles and objectives that various stakeholder groups have relating to digital curation. Ultimately, this concept model should provide a critical input to the development and refinement of cost models as well as helping to ensure that the curation and preservation solutions and services that will inevitably arise from the commercial sector as 'supply' respond to a much better understood 'demand' for cost-effective and relevant tools. To meet acknowledged gaps in current provision, a nested model of curation which addresses both costs and benefits is provided. The goal of this task was not to create a single, functionally implementable cost modelling application; but rather to design a model based on common concepts and to develop a generic gateway specification that can be used by future model developers, service and solution providers, and by researchers in follow-up research and development projects.

The Framework includes:

- *A Cost Concept Model*—which defines the *core concepts* that should be included in curation costs models;
- An *Implementation Guide*—for the cost concept model that provides guidance and proposes questions that should be considered when developing new cost models and refining existing cost models;
- A *Gateway Specification Template*—which provides standard metadata for each of the core cost concepts and is intended for use by future model developers, model users, and service and solution providers to promote interoperability;
- *A Nested Model for Digital Curation*—that visualises the core concepts, demonstrates how they interact and places them into context visually by linking them to *A Cost and Benefit Model for Curation;*

This Framework provides guidance for data collection and associated calculations in an operational context but will also provide a critical foundation for more strategic thinking around curation such as the Economic Sustainability Reference Model (ESRM).

Where appropriate, definitions of terms are provided, recommendations are made, and examples from existing models are used to illustrate the principles of the framework.

1 Introduction

Definition: Digital Curation

Digital curation involves maintaining, preserving and adding value to digital content throughout its entire lifecycle. The active management of digital material reduces threats to its long-term value and mitigates the risk of digital obsolescence. As well as reducing duplication of effort in digital object creation, curation enhances the long-term value of existing content by making it available for further use in a wide variety of contexts.[1]

1.1 Purpose

Despite reaching a level of maturity in terms of academic study the implementation of cost models for digital curation and preservation is still fragmented. Some standardisation of approach is critical to embedding curation cost collection and analysis into operational curation processes. A range of cost models exist but few have been validated in real world scenarios other that in which the models were created. The difficulty experienced in trying to apply existing cost models beyond the projects and organisations which developed them has meant that we have yet to reach a critical mass of adoption.

For those wishing to use cost models in practice, there remains an acknowledged gap in standard descriptions and supporting documentation which ensures there is a high barrier to entry. Potential users are faced with a significant time investment in order to understand the variety of scopes and approaches before being presented with the challenges of applying a cost model. For the wider curation community the disparity between approaches and the lack of standardisation presents a challenge when attempting to compare the outputs of a cost models.

The practice, tools, technology and market for digital preservation and curation solutions and services are all changing rapidly. These changes can and do have significant effects on costs and as such models need to be robust enough to withstand changes which take place during the life of a digital object and the life of a collection of objects. Manual processes become automated; risk tolerances change; service components are outsourced or federated; creators of material may switch allegiances to other curation archives as new actors emerge with new pricing or benefit structures.

The increased involvement of the commercial sector adds more actors to the curation systems. The need for service agreements between parties in these often 'distributed' systems—both in terms of organisational entity and geography—increases the demand for clear cost and price data and, critically a need to justify prices in relation to costs or costs with reference to risks and benefits.

One result of this increased complexity is a greater tendency to characterise groups of curation activities as 'services'. Whether or not a service is presented as a commercial proposition this 'packaging' of activities and an agreed 'service level' to denote a value proposition can play a useful role in standardising approaches to curation cost capture and analysis.

Embedding the collection of cost data directly into curation workflows supports real time curation cost reporting and timely managerial decisions, but academic research into the economics of curation remains a critical input as curation practices evolve.

[1] As defined by the Digital Curation Centre—http://www.dcc.ac.uk/digital-curation/what-digital-curation

Costs alone are not enough. Without a thorough understanding of the organisational context, it is impossible to assess the true value and effectiveness of any investment. Earlier research by the 4C project has shown that approaches which address curation costs without addressing the associated risks and benefits are of limited value—both locally and to the wider community. Indeed, a key finding of the 4C project has been the need for a shared understanding of how both costs and benefits influence the economic models of digital curation.

1.2 Issues

The lack of a standardised approach to defining and structuring curation activities as well as the lack of common accounting principles and practices represents a major challenge in making financial information comparable, which is essential for selecting the most cost effective and efficient curation services.

This gap in current provision calls for a common vocabulary to describe the costs and benefits of curation as well as a conceptual cost and benefit model. As these 'concepts' are abstract ideas generalised from specific instances, the concept model should be able to represent all curation scenarios and reflect any type of system or organisation holding any type of information assets and for any retention period.

The high level of abstraction necessary to ensure the cost concept model and gateway specification cover the wide range of possible cost models means that it is not possible to specify curation scenarios in detail; the purpose is to provide common reference concepts from which cost model developers can drill down to actual implementations. Depending on how similar organisations are this approach should enable them to—more or less directly—compare their financial information.

1.3 Sustainability & Roadmap

The 4 C report "D3.1—Evaluation of Cost Models and Needs & Gaps Analysis" provided a 'snapshot' in time of curation cost models. This snapshot could be periodically repeated using the benchmark of the cost concept model and gateway specification. If the Framework proposed here is widely adopted such periodic review will be vastly simpler. The cost concept model and gateway specification presented here are intended to be used and reviewed periodically to maintain alignment with the state of the art in curation cost modelling.

2 Getting started with the 4C Framework

This Framework is primarily aimed at developers of cost models but will also be of value to those developing curation solutions and services as it provides a common set of reference points. Building on the Framework, cost model developers can extend the core cost concepts and apply them to specific curation systems spanning the digital curation lifecycle. When using the Framework, we recommend that clearly defined use cases are produced and strongly encourage the provision of sufficient documentation to enable future users to evaluate, select and implement appropriate cost models and services within their own organisations.

2.1 Components

The Framework includes the cost concepts model, an implementation guide, the gateway specification template, and the nested model for digital curation.

- Cost concepts model—cost model users will benefit from the provision of clearer documentation when selecting and applying a model and the overall landscape of cost models will become clearer. Core cost concepts are intended to evolve over time alongside developments in curation cost and benefit modelling.
- Implementation guide—rather than a manual for developing a specific model, the framework offers guidance for a standard approach to be adopted by cost model developers.
- Gateway specification—as the required outputs may include comparisons between options or between cost figures from several systems, issues surrounding accuracy and comparability must be sufficiently explained in the specification.
- Nested model for digital curation—the nested model provides an implementation example of the core cost concepts which will help users to visualise how the core concepts interact in an organisational context.

For the purposes of this Framework, a service is a group of activities 'bundled' together with varying degrees of formality. Application of the Framework alongside a service-oriented approach to curation will support the standardisation and comparability of curation costs and foster an understanding of the benefits. In this way the Framework fosters a mature approach to costs.

The 4C Framework assumes that addressing costs without considering the associated risks and benefits are of limited value to the community and a Cost and Benefit (C&B) model for digital curation is presented as the foundation of developing a cost model. It is intended that the terms and concepts from this C&B model should be used when completing the specification template and in other documentation associated with the cost model.

Developers and users must also be aware of the significance of organisational maturity when selecting a method. Not all organisations will have sufficient infrastructure to deliver the inputs at the necessary granularity to support a model. Equally, they may not have the means to act upon and manage the changes implied by the outputs of a cost model.

2.2 Limitations of the framework

Potential use cases for cost models are extremely varied, ranging from calculations of capital and labour costs per terabyte (TB) of data to estimating the effect of a technological, process or market change on

future operational costs. No framework can hope to define a detailed approach to this vast range of options. For this reason the 4C Framework provides an abstraction of core concepts to be better able to represent a broad spectrum of curation scenarios and reflect the diverse needs of organisations holding different types of information assets for a range of retention periods.

3 Core Cost Concepts

Essentially, identifying the costs of curation relies upon an organisation's ability to identify their curation-related ***activities*** and to measure these against the ***resources*** required to undertake them. However, while an accurate understanding of curation costs is essential, in many ways it is more important for an organisation to be able to weigh the costs of curation activities against a potential return on their investment—either through derived benefit or through mitigation of risk. In this respect, the ***organisation context*** is also a core concept.

Stakeholders include any individual, organisation or body which has a demand for or an interest in asset curation. Within this stakeholder ecosystem one or more organisations will usually deliver services which cover the whole or a defined portion of the digital object lifecycle. Bearing these in mind, we propose that there are three core concepts that must be understood in relation to curation costing. These are:

1. Curation services/activities
2. Resources
3. Organisation context

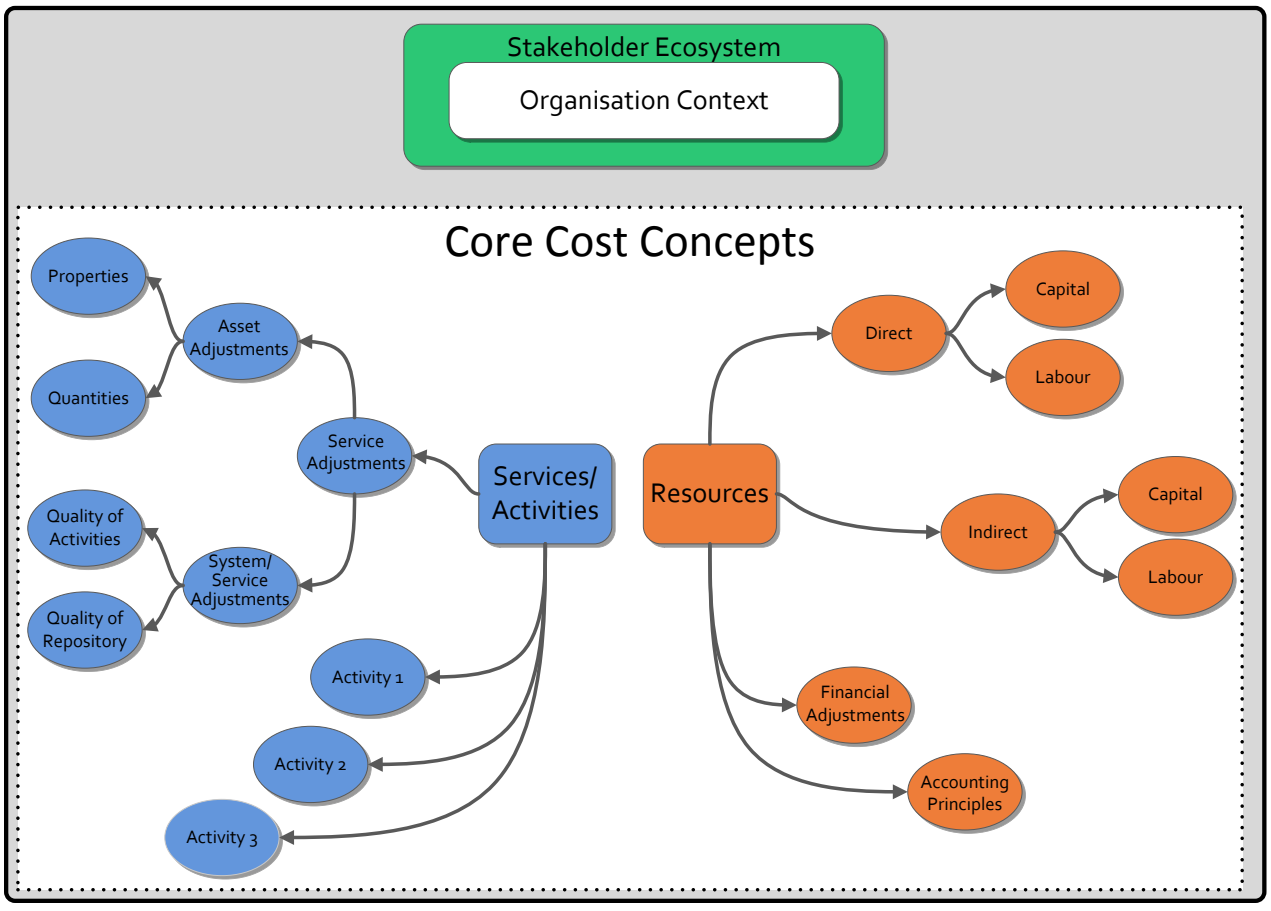

Figure 1—Core Cost Concepts and the Stakeholder Ecosystem

Curation activities are regulated by service adjustments and resources (direct/indirect capital and labour) and reflect the investment required to undertake curation activities. Both are regulated by the accounting principles applied within the organisational context.

The term ‘core cost concept’ has several implications that must be clarified:

- The core concepts are as granular as possible before the details of ***Organisation Context*** become relevant. Beyond the core concepts particular approaches to stakeholder management, local and national accounting principles and the particular information assets maintained by or services offered by an organisation start to play a role in the details of designing a cost model.
- The term ’core’ indicates that most cost models will either need to include these concepts or document why a given core concept is not relevant to their implementation.
- Clear, commonly applied definitions have been identified as a key driver for the communication of curation costing issues and therefore in the adoption of models and methods (4C D3.1 2014). Core concepts are defined simply and in a methodologically-neutral way. Future cost model developers are urged to use these definitions in their work and to contribute to advancing and improving these definitions.

The following sections will introduce the core concepts more fully and describe a range of issues that should be considered in relation to each of the core concepts.

3.1 Service/Activity Costing

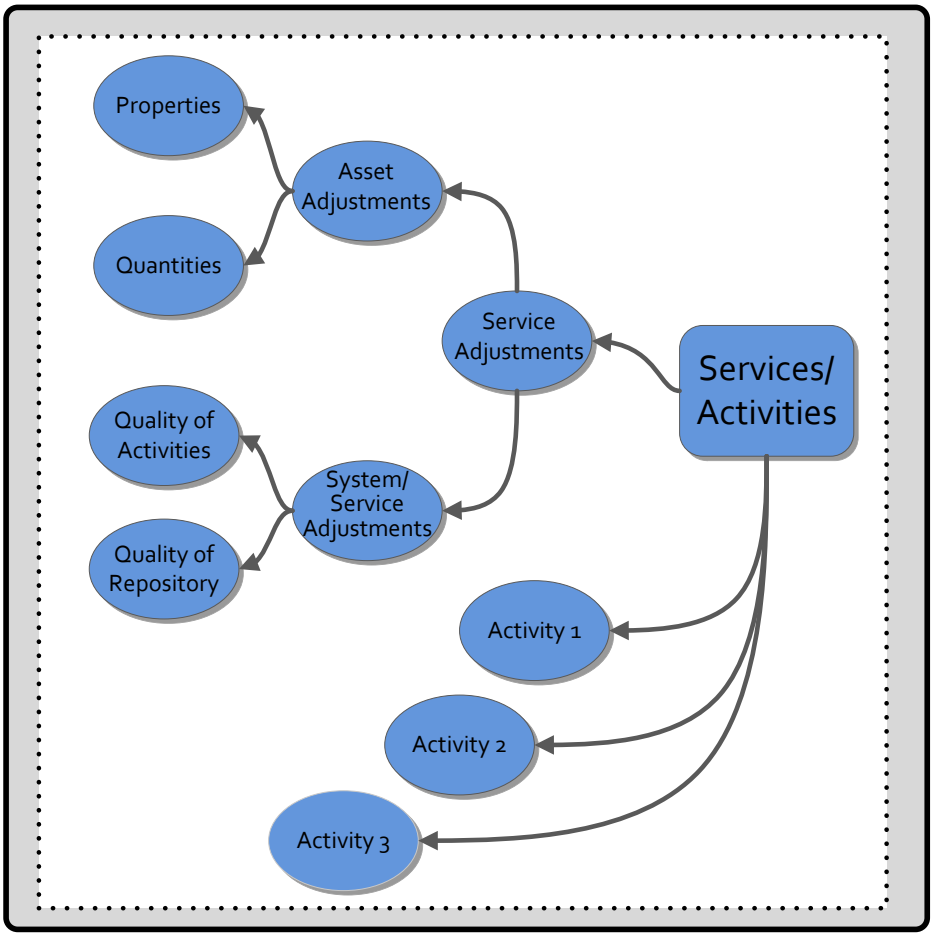

Figure —Services/Activities

Definition: Activity

Measureable amount of work performed by systems and/or people to produce a result

For the purposes of this Framework a service is a group of activities 'bundled' together with varying degrees of formality. For an illustration of this within an organisational context, see the section describing *A Nested Model for Digital Curation.*

While useful calculations can be made using only labour costs and capital expenditure information, many models have activities as a critical component. The costing of activities must take into account service adjustments including quality criteria and the properties and quantities of the digital assets being curated.

Recommendation:

Consider whether your model addresses activities and how these will be broken down.

3.1.1 Structuring Activities

Models can vary in which areas of the curation lifecycle area covered and the level of detail they apply to parts of the lifecycle.

The definition of digital curation and underlying digital curation activities are not universally accepted and understood, nor sufficiently detailed. There are challenges in extracting a subset of activities defined as 'curation' from the overall business processes of an organisation even if the organisational remit is primarily curation.

Any breakdown of costs by activity represents an effort to users, especially if such breakdown is not part of standard accounting practices; in addition there may be issues of accuracy as the breakdown becomes more granular.

Most models use the OAIS reference model[2] (see section 3.1.3) as the starting point for considering activity structure, often starting with Archival Storage and Ingest.

Many current cost models only calculate the cost of curation from the point of ingest into an archive and the OAIS model focusses on the 'Archival' phase of the lifecycle, but curation challenges start well before assets are deposited in a managed repository with a long term digital preservation remit. Activities around the 'Production' phase of the lifecycle generally receive less attention.

Increasingly, non-traditional stakeholders must take on curation and preservation activities for their digital assets. For instance, with increasing mandates in the UK to retain research data, research intensive Universities must now make informed decisions at the grant application stage about both the in-project research data management costs and the ongoing costs to preserve access to the data for as long as required—in most cases a period of at least ten years. This implies a need to identify directly incurred costs during the active phase of the research project ('Production') as well as assessing the cost of retaining the data over the longer term.

Model Example: CMDA includes the DANS-Activity-based Reference Model (DANS–ABRM) which describes activities taking place in a trusted repository.

[2] OAIS—Open Archival Information System

Model Example: CMDP provides very detailed activities around archival storage.
Model Example: CDL-TCP model provides some defined activities, however these are not presented as a checklist, but as the so-called 'Intervention' sheet that allows the user to enter activities themselves.
Model Example: Perhaps one of the most defined activity models is NASA-CET. With the NASA-CET activity model, users must map planned activities to the CET Data Service Producer (DSP) reference model.
Model Example: In the case of CDL-TCP some OAIS terms have been renamed to facilitate understanding by non-specialists.
Model Example: Tools like LIFE3 encourage users to modify the model structure to reflect local circumstances.
Model Example: With KRDS, users are encouraged to adapt the language used in the model and benefits spreadsheets to reflect local strategies and objectives. Steps 4, 5 and 6 in their user guide deal with adapting model for local use.

3.1.2 Custom activity breakdown structure

Users want to be able to adapt cost models to reflect their particular ***Organisation Context***. Internal organisational costs may already be accounted for at departmental level or activities may be documented as procedures for each section. This will increase the temptation to select a custom activity breakdown using existing organisational entities as the higher level grouping for activities.

This approach has the same limitations in terms of comparability as other custom activity breakdowns, but also has other risks. Organisations are dynamic entities and departmental structures can change even without a significant change of underlying activities as sections split and merge. It's to mitigate this risk that best practices in records management identify a clear need for an activity-based approach; the same applies to curation costs. Even for organisations with no need for external comparisons and no interest in extending the body of community knowledge on curation costs, the use of activities rather than an organisational entity approach is more stable over time.

Recommendation:

Consider what level of customisation is really necessary. Any customisation represents a trade-off regarding the ability to output data that is comparable across organisations or use cases.

3.1.3 Standard activity breakdown and the OAIS model

As with the breakdown of costs by resource (see *Resources*), there is limited consensus within current models in how to breakdown activities. This has an inevitable impact on the exchange and comparison of financial information and the ability to compare outcomes across users and across models. However, if we want to compare costs between organisations or across different services—to learn from each other's practices and to identify the most efficient ways of handling digital curation—we need to define and break down costs in a more transparent and uniform way. Currently the OAIS model remains the most common reference point.

Many existing models use the OAIS standard and its functional entities: Ingest, Data Management, Archival Storage, Preservation Planning, Access, Administration, and Common Services as a point of reference for describing activities that incur costs. Beyond these repository entities, digital curation

includes pre-repository (production, pre-ingest) and post-repository (use and reuse) activities, as well as general management (see *Lifecycle Descriptions*).

OAIS provides a common framework against which archives can be more meaningfully compared and contrasted, but as a reference model it provides no concrete instructions for the implementation of a real-world system. By definition a reference model should be *"an abstract framework for understanding significant relationships among the entities of some environment… … "as such it is not directly tied to any standards, technologies or other concrete implementation details, but it does seek to provide common semantics that can be used unambiguously across and between different implementations."* (Schumann and Recker, 2012). The OAIS functional model *"aids OAIS designers of future systems and provides a more precise set of terms and concepts for discussion of current systems"*(Magenta Book, 2012, p. 44).

The formal descriptions of the OAIS processes and functions and examples of real world systems provide the insight that problem solving, planning or decision-making processes are often less formalised. If a single OAIS function is performed by multiple teams or departments a distinct mapping is difficult.

Nevertheless the OAIS functional model provides a well-known baseline for evaluating and developing archival systems because it supports communication between different teams and departments by providing a common vocabulary. Another benefit provided by mapping existing processes to the model is the ability to evaluate established workflows and thus to spot gaps in compliance.

Recommendation:

It is strongly recommended that activity structures are the based on the OAIS model with amendments or extensions clearly documented and justified. See Amending or Extending the OAIS model.

3.1.4 Service Adjustments

Service adjustments are made in relation to the assets, or the digital curation system and/or service.

3.1.4.1 System/Service Adjustment

Digital curation system and/or service-related adjustments include the quality, reliability or resilience of systems and services at a high level. For example:

- The likelihood that an object is lost due to media or systems failure
- The likelihood that a collection is lost due to disaster
- The variety of formats in which the access service can deliver an object
- The quality—at a per-process level—of an error handling procedure

The Quality of Activities

"Cost models support different ways to specify the quality of activities. Six current cost models reviewed in Task 3.1 support structured specification of the quality of activities (e.g., a table with pre-defined elements). Four of the models allowed for the free text descriptions of the quality of activities. The NASA-CET model assumes that principal investigators will assess their confidence in the information provided back by the tool". (From the 4C project deliverable 3.1)

For many organisations, it is critical to maintain the quality of information assets. However, how quality is measured varies across organisations. The 'quality' of curation activities undertaken—whether evaluated as formal metrics (pass, fail, minimum score) or via more subjective 'expert opinion' (for example "this abstract will be understandable by our designated community of users")—is critical to maintaining the

quality of information assets. Consistency in what is meant by quality and how it is measured will be important for enabling comparison.

Recommendation:

Activity costing structures should be extended to support evaluation of the quality of the outcomes of those activities. Identical activities which do not share quality parameters are not directly comparable.

Model Example: KRDS leaves it to the user to determine how to relate costs and quality e.g. indicating the quality of digitisation.
Model Example: LIFE3 covers quality levels associated with digitisation procedures and volume as well as the QA of metadata and policies.
Model Example: CMDP covers the quality of record repairs as well as comparing the costs of different levels of archive storage.
Model Example: CMDA leaves it to the user to determine how to structure quality.

Quality of Repository

Not all quality assessments are made per-activity. The overall quality of a repository may be integrated into curation cost calculations. Some level of standard evaluation or formal certification may help to establish comparable procedures and quality measurements. Quality measurement may be undertaken internally or through external review and assessment such as an audit/certification process. Controls may address processes (ISO9000). Alternatively, measurement could focus on features such as the level of 'Information Security' applied (ISO27000), or the organisation may choose to formalise the trust the community chooses to display in them through attaining Trusted Digital Repository (TDR) status. The adoption of such standards may enable more valid comparison across different repositories and systems.

Model Example: The KRDS cost model addresses standardisation issues during the phase of evolving preservation functions and file formats (First Mover Innovation). In this phase organisations may need to develop tools, standards and best practices as first innovator.
Model Example: The CMDA defines as a prerequisite for the cost model that an organisation using the CMDA model has the philosophy of a trusted digital repository. It does not include compliance with a specific standard or certification but it assumes compliance to arbitrary standards. The model further states that all costs are related to the quality of the repository.

3.1.5 Asset Adjustments

Asset related adjustments include *quantity* of the assets, expressed as numbers of items and/or by data *volume*, and the *properties* of the assets, in terms of their type and complexity.

3.1.5.1 Properties of Assets

Only the properties of assets which are designated as important to the curation process and which have an impact on cost are designated as 'asset' adjustments. For example, if the property 'retain original embedded metadata in photos' is a cost-free outcome of a format transformation it is not an adjustment.

Most cost methods which address asset properties will attempt some sort of conceptual simplification due to the vast number of properties which may impact the curation activities undertaken.

Simple Data Formats

- A few levels and dimensions, for example 2D data (documents, images, sound, video)
- A simple relationship between the stored object and the accessed object (an image, a document)

Complex Formats

- Multiple levels and dimensions, for example chemical or meteorological models
- Frequent instances where access requests require parts of many separate stored objects
- Objects where there is no canonical rendering and thus many options for access (databases, process models, engineering models)

Model Example: LIFE3 model covers five default file types: web sites, e-journals, research outputs (theses), sound recordings, and 'other'.
Model Example: CMDP provides a list of formats to choose from including text, email, spreadsheets and databases.
Model Example: PP-CMDS addresses the storage of audio-visual assets.
Model Example: In theory KRDS can be applied to any data types.
Model Example: DP4Lib model allows costs related to any type of asset to be mapped to sub-services.
Model Example: CMDA leaves it to the user to determine what complexity means.
Model Example: File types cannot be specified in PP-DMDS and CDL-TCP.
Model Example: LIFE3 can capture quite detailed information for sound recordings but only allows small databases (up to 10MB) for research data.
Model Example: DP4Lib can handle any type of asset.
Model Example: PP-CMDS deals with audio and video assets
Model Example: CDL-TCP does not allow the user to specify data formats.
Model Example: CMDP covers databases, images and audio formats but does not include complex formats in the pre-defined menus, but they could be added.

3.1.5.2 Quantity of Assets

Estimating the annual increase in the number or volume of assets may support forward planning. An organisation may have a minimum number of assets to break even on a process or may have a maximum number of assets as a top capacity.

Model Example: The CMDA model allows the organisation to specify the number of privacy protected files.
Model Example: DP4lib and CDL-TCP account for the volume rather than the number of assets.
Model Example: DP4lib does not support specification of increasing asset sizes because results are calculated and output for a single year.
Model Example: EMLTS focuses only on storage costs and does not refer to the number of assets.
Model Example: T-CMDP breaks down costs by amount of assets and calculates the batch costs by dividing staff costs across a number of items. Costs incurred on a per-year basis or over a timespan of multiple years are supported. Assets are specified in terms of the existing and new number and size of batches per year.

Model Example: LIFE3 allows the organisations to specify increases in number and volume. It has a 'refine creation' tab in which project variables are identified (quality and volume); in the 'refine bit-stream preservation' tab, costs are broken down by storage requirements in megabytes.
Model Example: The CDL-TCP model does account for the volume of assets. Most routine preservation actions performed on content, such as characterization, fixity/integrity check, normalization, and so on are supposed to be automated. As a consequence these costs are seen as independent from the number of assets they involve. Calculations using the model allow specification of time spans of several years but do not account for increases of the volume of assets.

Maximum/Minimum amount of assets

The collection size may be defined to help define the scale at which activities must be undertaken.

Model Example: LIFE3 offers suggested volumes, which can be altered by the user. Defaults are set at fewer than 100,000 items for “low volume” and more than 1,000,000 for “high volume”.
Model Example: CMDP similarly supplies default values that can be changed, but the storage costs are estimated based on systems with a capacity between 1-500 TB.
Model Example: The lowest volume specified in CDL-TCP is “up to 100GB” with the largest (100TB) being an arbitrary limit to simplify the tool rather than a limit of the underlying model.
Model Example: KRDS makes no assumptions on the number of items.
Model Example: PP-CMDS does not provide an upper limit.

Upload/download capacity of repository system

The capacity of a system to accept deposits or provide access to content may be limited by quantity or frequency both of which can impact the requirements for infrastructure, hardware and software and therefore impact costs.

Higher access (upload or download) rates could reveal bottlenecks in the system’s infrastructure if demand exceeds available bandwidth.

Model Example: NASA-CET allows defining of activity sets and includes information about “expected number of users” and “estimated average number of requests per user, per year”.
Model Example: LIFE3 allows the user to refine access in a separate worksheet of the excel spreadsheet.
Model Example: KRDS includes the access frequency as a cost driver.
Model Example: PP-CMDS tool supports specification of the number of files which are accessed per month. Optionally an adaptive selection of storage systems responsible for the access can be activated to simulate load balancing in order to increase access rates and to mitigate overloading of system resources.

3.2 Resources

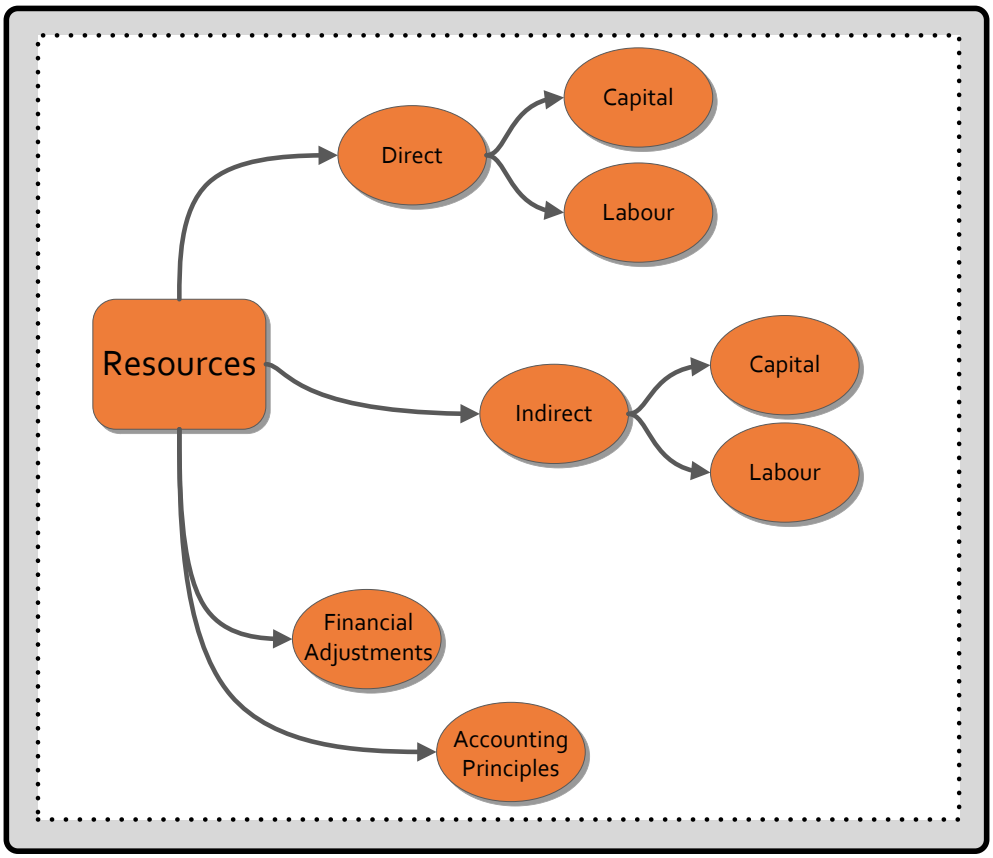

Figure 2—Core Concepts: Resources

Delivering services and activities requires the expenditure of resources in terms of capital and labour, whether as direct or indirect costs. The analysis of expenditure of resources must take into account financial adjustments (see *Financial Adjustments*) which may form part of the applied accounting practices of an organisation or may be applied to deliver some calculation within a cost method. All accounting practices are guided by national and international accounting principles.

Recommendation:

Communicate clearly to the user which direct or indirect capital and labour costs are required to apply the model. Identify the financial adjustments in place. Resource costing will be highly dependent on the individual use cases. As with activities the varied approaches to the breakdown of resource costs impact our ability to compare outputs.

Definition: Resource cost

Cost associated with a particular type of resource, capital or labour.

Organisations may seek to account for running costs (maintenance and operation) or may require curation cost support for investment (capital) costs.

Model Example: most methods do allow users to account for capital (investment), maintenance and operating costs, but the PP-CMDS provides the total cost based on experience from storage providers and a few large-scale institutions.

3.2.1 Direct cost

Definition: Direct cost

Costs associated with resources used for performing digital curation activities (for example costs of acquisition of storage media, costs of adding metadata), where the amount of resources spent can be directly measured. Also known as variable costs.

Definition: Variable cost

Costs, which vary directly with the amount of production. Often the same as direct costs.

3.2.2 Indirect cost

Definition: Indirect cost

Costs incurred by the usage of shared resources, such as general management and administration or common facilities and systems, where it has not been possible to distribute the cost on specific activities. Also known as residual cost or overhead.

Definition: Fixed cost

Costs, which do not vary with the amount of production. Often the same as indirect costs.

Indirect costs—also called residual costs or overheads—can sometimes be added to direct costs as a percentage of direct cost. In this case indirect costs are not directly equal to fixed cost. In general, given enough scale and time, no cost is really fixed.

It is common for methods which address indirect costs to include items such as office space for staff and equipment.

3.2.3 Capital

Definition: Capital cost

Cost incurred once, by acquisition (building space, equipment, materials) or by investments. Also known as investment cost or one-time cost.

Definition: Investment cost

See Capital cost.

Definition: One-time cost

See Capital cost.

Capital cost can be further differentiated by type—examples include building space (server space, office space, and so on), equipment (servers, network, and the like), energy (for systems, cooling, etc.) and materials (storage media, etc.). Capital costs are often abbreviated as ***capex*** (capital expenditure), and operating costs as ***opex*** (operating expenditure).

3.2.4 Labour

Labour costs can be differentiated by knowledge or skills (unskilled, skilled, 1st degree, Master's degree, Doctorate, and so on) and/or job functions (developer, metadata officer, et cetera) or by banding of salary scales.

Definition: Labour cost

Cost of wages paid to workers.

Direct labour costs are those costs paid to employees plus any benefit and payroll taxes paid by the employer. Indirect labour costs represent those paid for support labour.

A requirement for full economic costing (FEC) is often a driver for including labour costs in curation cost models.

Definition: Full- time equivalent (FTE)

A unit that indicates the workload of a worker by expressing the ratio of the total number of paid hours during a period by the number of working hours in that period. Also known as annual work unit (AWU). Used to make workloads comparable.

3.2.5 Accounting Principles

Accounting principles regulate the calculations of resources required to complete activities. These principles follow national and international standards. Organisations' accounting practices will apply these principles but the structure of accounting may not align closely with the needs of cost models. This implies that there may be an additional set of analyses necessary to bring accounting data to a point where it can be used as data for curation costing.

3.2.6 Financial Adjustments

Financial adjustments include inflation (or deflation), depreciation, and interest (or discount rates).

Most costing methods allow for depreciation costs and some allow for discounting. 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.

Estimations of costs are often based on analogy, in other words on experience from similar activities and projections of historic cost data, for example those derived from accounts.

Model Example: The KRDS guidance explains each type of financial adjustment. Users agree which adjustments will be applied and implement these alongside their data in a spreadsheet.
Model Example: T-CMDP assumes a depreciation of 33% per year for hardware and software.
Model Example: PP-CMDS allows the user to define how costs per unit change over time.
Model Example: LIFE3 includes a cost deflator variable that can be specified in the 'model variables' tab.

3.2.6.1 Depreciation/Amortization

Costs can be expressed as the depreciation (physical or through obsolescence) of assets over time.

Definition: Amortisation

A mechanisms for distributing capital costs over the estimated useful lifetime of an intangible asset to indicate how much of the asset's value has been used.

Definition: Depreciation

A mechanisms for distributing capital costs over the estimated useful lifetime of a tangible asset to indicate how much of the asset's value has been used.

In general, depreciation (for tangible assets) and amortization (for intangible assets) are mechanisms for distributing capital costs over the estimated useful lifecycle 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 (one measure of the lifetime of a server) may be five years. With a 5-year time period the cost of using this resource may simply be its acquisition cost, whereas with a 1-year period the cost would be the depreciated acquisition cost (whether linear, exponential or other).

3.2.6.2 Inflation, interest rates, discount rates

Fluctuations of the value of money over time include inflation (general price increases), individual price fluctuations that are related to specific resources—such as storage media, energy, office space, computer scientist wages—and interest, which reflects economic growth and cost of capital.

Even though the cost of resources have in general been increasing, the cost of both capital and labour 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.

A more detailed exploration of the relationships between the core concepts can be found in the section *A Nested Model for Digital Curation*. The approach to extending the core cost concepts in response to the required use cases and the particular organisation context (driven by interaction with the stakeholder ecosystem) including a benefit component, is described in the following sections.

4 Implementation Guide

This section supports the use of the Cost Concepts Model to:

- develop new cost models (potentially including a benefit component)
- refine existing models
- complete the gateway specification template
- create supporting documentation

The main content is directed at cost model developers while definitions and descriptions of core cost concepts are directed at those with less familiarity with the relevant topics. The more basic content is intended for re-use in cost model documentation.

Whether a model intends to address a small number of discrete activities in a local environment or to develop a broadly scoped model applicable to many curation systems, this guidance will be of relevance. However, the framework cannot offer a globally applicable instruction set for creating a particular cost model. Key areas of development are covered and questions are posed which the cost model developer should address if their approach is to meet their goals and gain acceptance and adoption.

4.1 Standardisation

This framework is intended to support a more standardised approach to developing cost models.

The curation community has yet to reach consensus on how to breakdown the cost-bearing activities surrounding digital curation and remains in the critical early stages of integrating benefits into such cost calculations.

The application of standards and the delivery of new proposed approaches and structures in a way that permits standardisation will be invaluable to cost model developers and users alike. The benefits of standardisation go beyond the individual curation cost model and provide for easier analysis and, potentially, greater comparability between the outputs of different models.

Models which take care to document the standards they apply, or clear reasons for variance from a standards-based approach, will support both users and the community as a whole.

Standardisation may not imply the use of identical terms in every setting but if alternate terms are used to improve communication and understanding by users who are non-curation specialists these should be clearly mapped to the standard definitions.

4.2 Use Cases

Use Cases—or some other method of presenting the curation scenario supported by the model, especially those that provide a clear, high level description of what the model does—act as an initial communication bridge between the cost model developer and cost model user.

There are numerous approaches to developing scenarios and use cases. Whichever approach is selected, it should be clearly defined and applied at the same level of detail to all use cases. For instance:

- “I want to build/use a model and associated tool which reflects a snapshot of my own organisation and make specific one off calculations”.

- "I want to build/use a model and associated tool which reflects a snapshot of my own organisation and make specific calculations which can be repeated over time. I will then compare the results over time".
- "I want to build/use a model which supports comparison between two or more known organisations, systems or services".
- "I want to build/use a model which lets me compare my results to the wider community either as a snapshot or over time".

From such high level expressions of intent the use cases can be supplemented by relevant details from the ***organisation context*** and the approach to ***resources*** and ***activities***.

Attempts to develop approaches which cover a wide range of use cases rapidly become complex to develop, validate and apply. While it may be tempting to develop a model that touches on all aspects of digital curation, it's recommended that you start with a clear, limited purpose (with clear inputs and outputs) that is conducive to detailed validation and can be supported by a usable tool. Consider one goal as a starting point, such as short term prediction or estimating historical or present costs.

It is important to clearly limit the inputs in terms of stakeholders, organisation size, activities and range and quantity of assets to be covered.

Digital curation, like most complex tasks, is subject to high degrees of uncertainty. A simple approach reduces the risk of inaccurate outcomes. The high levels of variation across curation organisations means that attempts to demand very granular input or offer very granular output will create a complex model with limited validity when results are compared over time or between organisations.

Starting with a subset of activities with fewer variables and known quantities, such as 'Archival Storage', will provide a good baseline for further explorations in more complex areas of curation such as Ingest or Preservation Planning. This approach has been adopted by the 4C project for the curation costs exchange (CCEx).

Limiting the complexity, the time scope, and applying simple formulae will all support the implementation of the model in a concise, user friendly tool.

4.3 Calculations

4.3.1 Integrating Benefits

Benefits are typically divided into financial benefits—benefits that can be expressed in monetary values such as value generated from user fees or licenses—and in non-financial benefits including an organisation's increased trustworthiness (reputation) or reduced business risks.

Whilst the costs of curation essentially depends on the quantity and the required quality of the information assets—which, in principle, can be assessed objectively for a particular scenario—the benefits of the scenario depend on the perspective of the service consumer—and as such the identification and assessment of benefits is subjective, and this should be reflected in the way that cost and benefit models are designed.

The integration of benefits into costing processes is essential for comparing alternate solutions, strategic planning and risk management as well as in the more operational analysis of efficiency gains.

Considerations:

- Will your model integrate benefits alongside cost calculations? Few current models have done so but a clear demand for approaches which include costs and benefits have been identified.

> **Model Example:** KRDS Benefits Framework Tool identifies benefits and the KRDS Value-chain and Benefits Impact Tool help identify potential measures or illustrations of the value and impact of those benefits. These two tools are meant to be used in conjunction with the KRDS activity based cost model.

The KRDS Benefit Framework Tool describes benefits along three dimensions, each further divided in two categories. The first dimension "What are the outcomes?" is divided in direct benefits ("positive impacts obtained from investing in a data curation activity") and indirect benefits ("negative impacts avoided by investing in a data curation activity"). The second dimension "When are the benefits received?" is divided into near-term and long-term benefits and the third "Who benefits?" into internal and external benefits[3]. The KRDS Value-Chain and Benefit Impact Analysis Tool support qualitative measurements of the identified benefits.

> **Model Example:** CMDA includes a balanced scorecard approach to ensure that the mission of an organisation and existing strategies are translated into strategic objectives that can be measured operationally.

The Cost Model for Digital Archiving (Palaiologk, A. et al 2012) has used the balanced scorecard (BSC) approach, which is a strategic management tool that helps organisations clarify their vision and strategy and turn it into actions, to address and categorise benefits.

4.3.2 Formulae

If your approach to gathering cost data has been sufficiently generalised and simplified to produce more generally accurate outputs then the model should require relatively simple formulas and these should be formally documented.

Simplicity in this context does not imply easily solvable or very linear formulae, but rather that the use of associated data in look up tables of model-specific conditions (start, end, transitions etc.) should be avoided.

Considerations:

- Does your model apply algebraic formulae? If so, are these clearly documented to enable users to understand the calculations being made?
- Are users able to adjust the algebraic formulae?

> **Model Example:** T-CMDP uses Excel formulas
> **Model Example:** LIFE3 model guidance describes the formulas used.

[3] From A Guide to the KRDS Benefit Framework, v. 3, 2011, http://www.beagrie.com/static/resource/KRDS_BenefitsFramework_Guidev3_July%202011.pdf

Model Example: NASA-CET uses regression techniques to develop the coefficients for a set of seven trial relationships of FTE to workload parameter for each of the selected workload parameters.
Model Example: The DP4Lib cost model provides a formula to calculate amortisation rates in the documentation.
Model Example: CDL-TCP clearly defines formulas for the "Pay-as-you-go" (annual billing cycle) and the "Paid-up" (one-time payment) price models within the documentation.

4.3.3 Modularity

Modular systems tend to made up of separately developed sub-systems, each devoted to particular aspects of the system as a whole. Theoretically such subsystems can be independently developed as long as the input/output parameters are clearly defined. This building block approach allows complexity to be added (or removed) without having to re-develop the system as a whole every time.

Considerations:

- Is your model modular? If so, what are the minimum modules necessary to practically implement the model and yield meaningful results?

Model Example: NASA-CET allows the user to select which elements to include.
Model Example: LIFE3 allows the user to modify elements within the tool quite easily.
Model Example: DP4Lib allows for elements of cost groups and sub-activities to be extended and customised.
Model Example: CMDP is currently missing three modules and yet produces results.
Model Example: KRDS leaves it to the user to describe what to include or remove when they implement their institutional spreadsheet.
Model Example: CMDA is modular in theory.
Model Example: The CDL-TCP model is somewhat modular as it consists of 11 high-level cost categories, which are defined in separate worksheets support changes to each cost category.

4.3.4 Time Parameters

Costs can be divided by time such as one-time costs, periodic (term) costs or recurring costs. he term capital or investment cost is often used to denote a one-time cost incurred on the acquisition of equipment such as a storage system. The term periodic cost is used to indicate that the cost will incur at intervals. Recurring costs also known as running costs or operating costs include costs of the consumption of media, energy and labour[4].

Definition: Periodic cost
Cost that are repeated and incur at intervals (for example some licenses). Also known as term cost.

[4] Capital costs are often abbreviated as capex (capital expenditure), and operating costs as opex (operating expenditure)

Definition: Operating cost

See Recurring cost.

Definition: Recurring cost

Ongoing cost (such as from consumption of media, energy and labour). Also known as running cost or operating cost.

Definition: Running cost

See Recurring cost.

Users need to know the cost of curating and sustaining access to digital assets but even for curation organisations with a long term digital preservation remit uncertainty increases with time. Past 4C research (D3.1) has identified that budgeting and the ability to account for running costs are the greater drivers for adoption, perhaps reflecting the fact that organisations find it easier to seek one-time funding than funding directed at ongoing running costs.

Considerations:

- Will your approach focus on budgeting or accounting?
- Will your approach focus on running costs or investment costs?
- Will your model support calculations over time such as annual increases in the number or volume of assets?
- What units of time will your model support for activities?
- If your model offers calculation over the longer term does it provide transparent statements about the likely impact on accuracy (see section 4.5)?

Model Example: The LIFE3 model allows users to choose to assess costs by year, day/hour/minute rates to calculate costs for shorter activities.
Many tools are designed as forward planning tools; those which do not offer pre-defined data and formulas can handle current, past or future costing.
Model Example: LIFE3 allows users to record procurement costs as part of ingest.
Predicting future costs over the midterm and longer can be difficult with pre-defined data and formulas, unless models incorporate changes like annual pay increases or staffing numbers.
Model Example: T-CMDP allows you to include a 'repeat after N years' calculation but the underlying assumption that activities and roles will remain identical reduce the value of such periodic repeats.
Model Example: LIFE3 indicates ten years from the point of ingest.
Model Example: EMLTS estimates how storage costs might vary over a 100 year period and aims to reflect the effects of changing technology over long periods of time.

4.4 Making Comparisons

4.4.1 Comparisons within and between systems

Comparisons are also relevant over time and between curation systems. Comparing the costs and benefits of different scenarios to support decision-making and funding requests is a driver for managers, who form the largest potential user group for cost models.

Considerations:

- What comparisons does your model support when applied to a single organisation or system?
- What further comparisons are possible (or conversely what comparisons are not practical) between two organisations implementing your model?

4.4.2 Comparison across Models and Tools

Current models don't easily align and interoperate which doesn't reflect the demand from different stakeholders within an organisation to use different models to meet different cost calculation scenarios. Ideally users want the ability to 'round trip' data through multiple models/tools across the organisation(s).

Considerations:

- Is your model intended to function within a wider landscape of models and tools?

> **Model Example**: if a researcher uses NASA-CET or KRDS to calculate in-project, directly incurred costs for a specific project, it is not easy to feed the tailored source data and/or results into another cost model such as LIFE3 or CMDP at the institution-level to help calculate indirect costs associated with longer-term archiving.

4.5 Detail, Accuracy and Validation

Any data input errors based on imprecise definitions will impact the immediate accuracy of outputs but could also impact periodic internal comparisons or comparisons between organisations. Such imprecise outputs could be critical to organisations, for example those considering in-house vs. outsource solutions for curation.

Considerations:

- Are you clear on the level of accuracy of your model will offer?
- Can it cope with increases in asset quantities which could impact infrastructure, hardware and software costs?
- Can it cope with increased service usage which could reveal bottlenecks in infrastructure or bandwidth?
- If your model offers calculation over the longer term does it provide transparent statements about the likely impact on accuracy (see section 4.5)?
- Is your approach to defining curation activities transparent or is it subject to misinterpretation by users struggling to extract and analyse curation costs independent of the costs of wider business activities?
- Are limitations on accuracy, including those implied by scalability, documented?

Model Examples: NASA-CET, LIFE3, KRDS &PP-CMDS are the only models which include access frequency.

As well as an influence on accuracy the detail and flexibility in your model will also have a direct impact on the learning curve and ease of input of any tool you develop (see section 4.7.1)

4.5.1 Detail

Determining the level of detail your model should reflect can be tricky. Stephen Abrams (CDL-TCP model owner) states,

> *"We purposefully did not attempt to model costs at a finer degree of granularity, such as would be required to break things down at the sub-OAIS entity level. We believe, perhaps somewhat paradoxically, that past a certain level of modelling granularity the accuracy in estimating costs actually decreases as the granularity increases. (In essence, we feel that it is easier to make an accurate estimate of time in terms of days rather than hours, weeks rather than days, etc.)*
>
> *We have tried very hard to ensure that the TCP does not give the impression of greater accuracy than may be justified given the many assumptions and intuitive estimates that go into it. Also, we found in many cases that it was difficult to map our local practices into the OAIS sub-functions in an obvious and unambiguous manner."*

4.5.2 Flexibility

Offering a very flexible solution which accepts a wider range of user selected parameters will improve the range of applications of your model but will also introduce more opportunities for errors in data input.

Considerations:

- How can you offer flexibility in your model without sacrificing the ability to make realistic comparisons of costs from different sources?

4.5.3 Model Validation

Trust from end users is critical to model adoption. Applying common definitions and using recognised standards and controlled vocabularies for settings are excellent starting points for any cost model as are frank and transparent statements about the likely accuracy. If possible, demonstrate that your model has been rigorously tested on a range of validated data from appropriate sources.

4.6 Understanding the Stakeholder Ecosystem

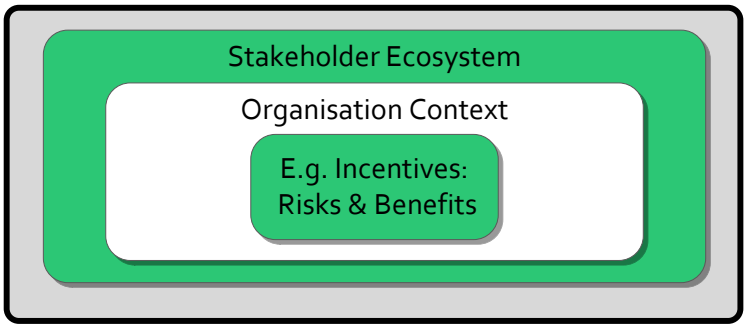

Figure 3—The Stakeholder Ecosystem and Organisation Context

An organisation's stakeholders may be defined *as "the individuals and constituencies that contribute, either voluntarily or involuntarily, to its wealth-creating capacity and activities, and that are therefore its potential beneficiaries and/or risk bearers."* (Post, 2002) Unlike some other definitions in the field, this recognises a *"mutual relationship between stakeholder and corporation"* (Du Plessis et al., 2005). For curation we recognise a wider definition of stakeholders which goes beyond 'wealth creating' to 'value creating'

The concept of stakeholders is widely used in organisational business theory and practice to denote anyone who can effect or be affected by the actions undertaken by an organisation or system. In curation cost calculations stakeholders influence policy, strategy and actions via institutional role or unit such as a managing board or a funding body. Alternatively they may be represented indirectly, for example by examining purchase records or arranging questionnaires to identify the needs of customers.

An understanding of the stakeholder ecosystem will help model developers identify a wider audience for the work (see *Understanding the Stakeholder Ecosystem*).

For model users it is important to undertake stakeholder identification and management to integrate these wider viewpoints into their practices. Furthermore the outcomes of cost modelling may indicate the need for change. Effective management of change also requires stakeholder interactions. The ability to undertake such stakeholder management is a function of organisational maturity (see *Maturity*).

At the stakeholder level priorities are conceived as 'indirect economic determinants' (IED) which must be 'operationalised' into clear strategies and procedures if they are to be integrated into cost and benefit models.

Considerations:

- Does your model need to address the wider stakeholders explicitly or does it only need to gather specific items of information regarding the organisation context?

4.6.1 Organisation Context

It is important to communicate which variables about users' organisations are relevant to the model. These may include the size of the organisation in terms of staff or budget, whether the organisation is wholly or partly tasked with curation, whether long-term digital preservation is in scope, the legal and policy framework (which might impact the selection and quality criteria for activities) and the people, roles and skills in place.

A cost model may address several parts of the digital curation lifecycle which could encompass a number of organisations. Those adopting and implementing cost models may outsource some portion of their activities to third parties or may wish to include costing for lifecycle phases which they do not directly control (for example, archives may not control data production and consumption). Several inter-organisation relationships—whether informal, contractually controlled or via agreed service levels—may need to be understood.

Organisation Type

One may reasonably expect curation costs from similar organisational 'types' to be more comparable than costs from more contrasting organisations. Clarify which types of organisation are in scope and define those types. If possible use a pre-existing classification.

Lifecycle Scope

Ideally adopt a standard lifecycle model whose structure and terminology is meaningful to your users. A high-level lifecycle description provides a clear basis for communication before addressing the detail of service/activity costing (see section 3.1). An overview of curation lifecycle models is provided in *Lifecycle Descriptions.*

Considerations:

- What parts of the digital object lifecycle are supported by the model?

Collection Profile

The collection profiles refer to the range of variables surrounding the digital assets undergoing curation activities including any mandatory retention period. See section 3.1.5

Considerations:

- Does your model address particular types or quantities of assets?

Benefits (Incentives and Risk)

The integration of risks and benefits into costing processes is essential for comparing alternate solutions, strategic planning and risk management as well as in the more operational analysis of efficiency gains.

Considerations:

- Will your model incorporate a benefits component? If so which information do you require in the organisational profile to enable the integration of the 'value' assigned to various benefits and the organisational appetite for various risks?

The organisation context topics above are considered critical but are not exhaustive, others may include those below.

Legal/Policy

Legal, funder, and other mandates and policy constraints may require that particular activities are undertaken or benefits prioritised.

Roles and Skills

Models for curation costs may be designed for and consumed by various types of user with different focuses including curation specialists and non-specialists. As such, there may be a great variety of roles and skills that need to be taken into account when developing and using cost models. See section 3.2.4.

Considerations:

- Does your model address these aspects of labour costs?

4.7 Audience, Users and Documentation

Your audience is not limited to those using the model. An understanding of your audience should also be applied when creating supporting materials.

Consider the various roles that make up your audience from decision makers selecting a model to adopt or adapt, to those entering data into the model and those making organisational changes based on the outputs of the model. The potential audiences may be limited by the lifecycle coverage or by restricting the model to a subset of issues that only apply to certain types of organisation such as. a particular subset of 'Activities' or a strong focus on storage costs.

Based on an understanding of your whole audience a documentation and communication strategy might address a wider audience to ensure adoption, but target more narrowly for the users of the model or tool.

Considerations:

- Does your documentation support the range of users likely to be called upon during implementation?

Cost models and tools which deliver a high level of accuracy may be need to be quite complex and supporting documentation may need to be equally detailed. In some existing models the detailed view, either in separate documents or embedded within the tool functionality, is the only material available.

Good quality, comprehensive documentation from high-level communications to detailed user guides can be critical to driving adoption of a cost model. Documentation facilitates the initial selection and increases the usability of models and tools if well-written and readily available. Rather than delivering a challenging quantity of mixed documentation, consider developing layers of documentation appropriate to different target audiences. The following may be helpful:

- Provide brief fact sheets for funders
- The gateway specification as a standard overview
- User guides and supporting research papers for managers considering an approach to curation costing
- A 'Quick Start' guide to let adopters start getting a feel for the tool
- Detailed Manuals or Help files for those fully committed to implementing a model

Considerations:

- Define your terms and concepts early in the material and make them easy to refer back to.
- If terminology is changed for particular target audiences provide mappings to the more standard terms and definitions.
- Embed detailed documentation within any tools.
- Maintain documentation online so it remains readily accessible.

Model Example: Research papers about T-CMDP have been published and there is guidance built into the spreadsheet.
Model Example: NASA-CET includes a 95-page user guide and 28-page technical guide.
Model Example: LIFE3 model tools are accompanied by extensive guidance.
Model Example: KRDS has a 46-page user guide including a brief "How-to guide" and there are also a number of fact sheets and flyers to introduce various elements of the framework.
Model Example: Resources relating to CMDA are limited to research papers.
Model Example: There are research papers related to CMDP and some guidance within the tool.
Model Example: DP4Lib documentation, guides and manuals are limited to the German language.
Model Example: CDL-TCP is supported by a 26-page paper which is clearly presented and clarifies the model to users.
Model Example: A paper and a series of blog posts serve to give an overview of the EMLTS, but the model and detailed user instructions for it are unavailable.

4.7.1 User Expertise and Learning Curve

In addition to the issues of tool usability the learning curve for the user will vary depending on the complexity of the model and the clarity with which it is documented. Beyond this, the time needed to understand a cost model depends a great deal upon the user's familiarity with digital curation, the range of ***activities*** involved***,*** their awareness of the ***resource*** available, and their familiarity with their ***organisational context***.

Considerations:

- Does your model require input from different roles within an organisation? Models for curation costs may be designed for and consumed by various types of user with different focuses including curation specialists and non-specialists.
- Can you clearly identify the skills necessary to understand and implement the model?
- Can your model provide results without some input from a repository manager or curation specialist role?
- Consider that non-specialists such as general financial/account managers, department directors and chief executive officers are often likely to have accounting and budgeting responsibility for digital curation actions.

Model Example: T-CMDP assumes the existence of a dedicated Archive and Preservation facility with a skilled team, familiar with digital preservation and with IT and record keeping issues.
Model Example: NASA-CET is intended for use by principal investigators, a lot of the required information is at a curation specialist level.
Model Example: CMDA is intended for trusted repository staff.
Model Example: DP4Lib supports a level of detail that would make it suitable for preservation specialists.
Model Example: It would be difficult for non-specialist users to collect and analyse the technical information needed for CMDP.
Model Example: PP-CMDS is intended for the managers of AV archives.

Model Example: KRDS doesn't specify a specific role, but recommends that "Dedicating a person to be responsible for collecting the cost information will save you effort and deliver results of better quality. The person should be responsible for checking the progress of the survey. Use someone who will be seen as independent and trusted by all staff".
Model Example: CDL-TCP modified OAIS terminology to expand applicability and increase understanding suggests an intention to include non-specialists.
Model Example: T-CMDP assumes familiarity with the field of digital preservation.
Model Example: CMDA assumes that users are employees of a trusted digital repository.
Model Example: LIFE3 is fairly straightforward but may need input from a digitisation team for specifics.
Model Example: DP4Lib can be applied quickly if a user can identify the cost types, elements and activities.
Model Example: with KRDS progress could be made in a day in a day, but actual calculations of costs would require developing a tool.

Models and tools which are challenging to use provide a significant barrier to adoption. The challenges of user expertise and learning curve for the model can to some extent be offset by usability testing of tools and good quality documentation (see section 4.7).

5 Gateway Specification Template

This gateway specification template supports a standardised approach to developing cost models. A degree of standardisation of approach (see section 4.1) and in supporting documentation provides a firm foundation for cost model developers, a common reference point for those selecting and implementing cost models, and consistency for those analysing the outputs of several curation costing approaches.

Cross-references apply to the *'Developing a Model'* (see section 4) material where detailed guidance and relevant questions that developers should address are provided.

> Example information for each category has been supplied below and is displayed in bordered text. These have been taken from the gateway specification for the 4C Cost Comparison Tool (beta release).

5.1 Model Metadata

Title

Provide a clear, unique title and abbreviation. This will support resource discovery of the model and of references to the model, including future work to compare and contrast cost models.

> Cost Comparison Tool (CCT)

Subject

Supply key words or phrases to support resource discovery.

> Activities, Add, Compare, Cost data, Cost input, Cost per GB, Cost unit, Current cost analysis, Financial accounting, Framework of comparable costs, FTE, Human resources, Overheads, Procurement, Share, Staff

Version

Provide a version number for this release of the cost model.

It is recommended that model versions are aligned with version numbers for associated documentation and tools.

> 2.5

Status

Clarify whether the cost model is complete or is currently under development. Make it clear if the model is experimental or ready for production use in a curation environment.

> Beta-testing throughout August and September 2014
> Candidate release October 1st
> Subject to minor changes beyond that point

Release Date

2014.10.01

Source

Reference any pre-cursor cost models from which the work is derived in whole or in part, include any relevant identifiers.

The CCT is based on the OAIS Reference Model, results from the Blue Ribbon Task Force, studies conducted within the 4C-project as well as user consultations undertaken during 2013 and 2014

Creators/Funders/Contributors/Maintainers

Provide the relevant person, organisation or service names and roles.

Creators: 4Cproject.eu

Language

Identify the language(s) of the cost model and associated materials.

English

5.2 Purpose

See section 4.2

Briefly describe the use cases (See 4.2) that the model supports.

I want to get an overview of my organisation's current spending on digital curation. I want compare my organisation's cost data with other organisations. Any details of the cost data, actors, calculations, comparisons and conclusions implied by the uses cases must be expanded upon in later sections.

5.2.1 Type of Tool

If a supporting tool is used for data entry and/or analysis, document the tool and describe how it should be used.

The CET is a web based system

5.3 Organisation Context

See section 4.6.1

Define any specific types of organisation supported by the model.

The CCT can be used by a broad range of institutions including, but not limited to, Universities, Government agencies, Big data science, Digital preservation vendors, Small or medium enterprises, Memory institutions or content holders, Publishers or content producers, Research funders, Industries and possibly others.

Clearly identify the variables related to the organisation which are required to support the model.

Organisation
- Organisation name*
- Organisation type*

University, Government agency, Big data science, Digital preservation vendor, Small or medium enterprise, Memory institution or content holder, Publisher or content producer, Research funder, Industry, Other.
- Description, purpose and mission
- Country*

Collection(s) profile
- Scope of the cost information* (organisation, department, collection or project)
- Size of staff working within the scope*
- Data volume*
- Number of copies*

Asset types*
- Unformatted text, Word processing, Spreadsheet, Graphics, Audio, Video, Hypertext, Geodata, E-mail, Database,

Currency definitions
- Euro, dollar*

*=mandatory

Lifecycle Coverage

Define the lifecycle stages supported by the model with reference to a clear lifecycle model. See *Lifecycle Scope*

Simplified OAIS lifecycle:

Production

Digital content production involves any activity related to the preparation of digital assets for archiving. This might encompass digitisation, extraction of data from databases, metadata enrichment, migration of production formats to preservation formats, etc.

Ingest

This activity covers processes related to receiving digital assets from an external source and preparing them for storage. Examples of activities that could fit into this activity category are: appraisal, submission agreement, validation of digital assets, metadata enrichment, preparing digital assets for storage within the archive.

Archival Storage

This activity covers processes related to storing, maintaining and retrieving the digital assets. Examples of activities that could fit into this activity category are: error checking, media migration, storage hierarchy management, providing disaster recovery capabilities.

Access

This activity covers processes related to accessing the stored digital assets. Examples of activities that could fit into this activity category are: providing access to digital assets; providing order mechanisms for digital assets; providing conversion between stored formats and delivery formats; producing objects for delivery whose content derives from many different stored assets and describing them meaningfully by adding relevant metadata to them.

5.4 Incentives, Risks and Benefits

Detail any relevant approaches to incentives (risk or benefit) which are addressed by the model (see section 4.3.1).

The model does not incorporate a benefit component

5.5 Cost Data Collection

Define the subjects of data collection.

Resources (see section 3.2)

Resources submitted in currency or FTE's. Resources normalised into the following 'resource' categories: Procurement categories: Hardware, software, external services Staff roles: Producer, IT-developer, Operations, Preservation specialist, Manager, Overhead

Services/Activities (see section 3.1)

Pre-Ingest, Ingest, Archival Storage, Access.

Information Assets (see section 3.1.5)

Unformatted text Word processing Spreadsheet Graphics Audio Video Hypertext Geodata E-mail Database

Include any relevant financial and service adjustments needed to support the required calculations, comparisons and conclusions.

Financial adjustments: None Service adjustments Asset adjustments Data complexity addressed by asking for asset types (formats) Data volume addressed, both in total and per asset type System/service adjustments Quality of activities. Not addressed. Quality of repository. Not addressed.

The concept of quality is only indirectly addressed 1) salary levels and 2) the cost comparisons that the CCT allows for. However thorough analysis is necessary to conclude anything regarding levels of quality. It is probably even necessary to contact the organisation(s) you are comparing yourself with in order to evaluate quality parameters properly.

5.6 Calculations, Comparisons and Conclusions

Clearly define the calculations (see section 4.3) being made on the collected cost data, any comparisons made, and any conclusions offered. Describe whether past or future costs are supported. Be clear on the levels of cost data accuracy necessary to support these functions and on any limitations on accuracy in outputs (see section 4.5).

The **calculations made** are simple additions of submitted cost units after they have been normalised into procurement, staff and activity categories.
Results are in total costs, costs per category, cost per category item, percentage of raw costs, costs per Gigabyte, average, distribution in percentages.
Comparisons can be made to

- A global average
- A filtered average
- A peer

There are **no conclusions or recommendations** offered, except to contact your peers if you need clarification about the comparisons.
It is possible that some later version will incorporate simplistic and automatically generated recommendations based on the user input.

5.7 Resources

Documentation should use the descriptions and definitions from the *Implementation Guide* section where applicable. Document the tool and its user interface including any technical interoperability with other systems. Clarify how long it will take to input data and derive outputs. Document any pre-defined settings and customisation options available to end users.

List and include, or link to, resources, tools and documentation which support the model.

Provide relevant URLs and identifiers for software and documentation.

Make it clear which audience the material is directed towards (see section 4.7).

Provisional link: http://4c.keep.pt/ccex/
Tool is **documented** online http://4c.keep.pt/ccex/ and the assumption is that it is self-explanatory with only a few accompanying texts as well as help texts appearing when mouse hovers.
The time it takes to submit data varies, but when you have provided the raw cost units from your financial department, it shouldn't take more than ½-3 hours.
Outputs are immediate.

6 A Nested Model for Digital Curation

Here, we introduce the *Nested Model for Digital Curation* which represents an implementation of the Core Concepts Model being applied within a more detailed model reflecting the contextual considerations to help judge the value of curation investments. Placing the core cost concepts into wider economic models surrounding digital curation activities, including the context of business models in terms of both costs and benefits, will place any cost model on a clear foundation.

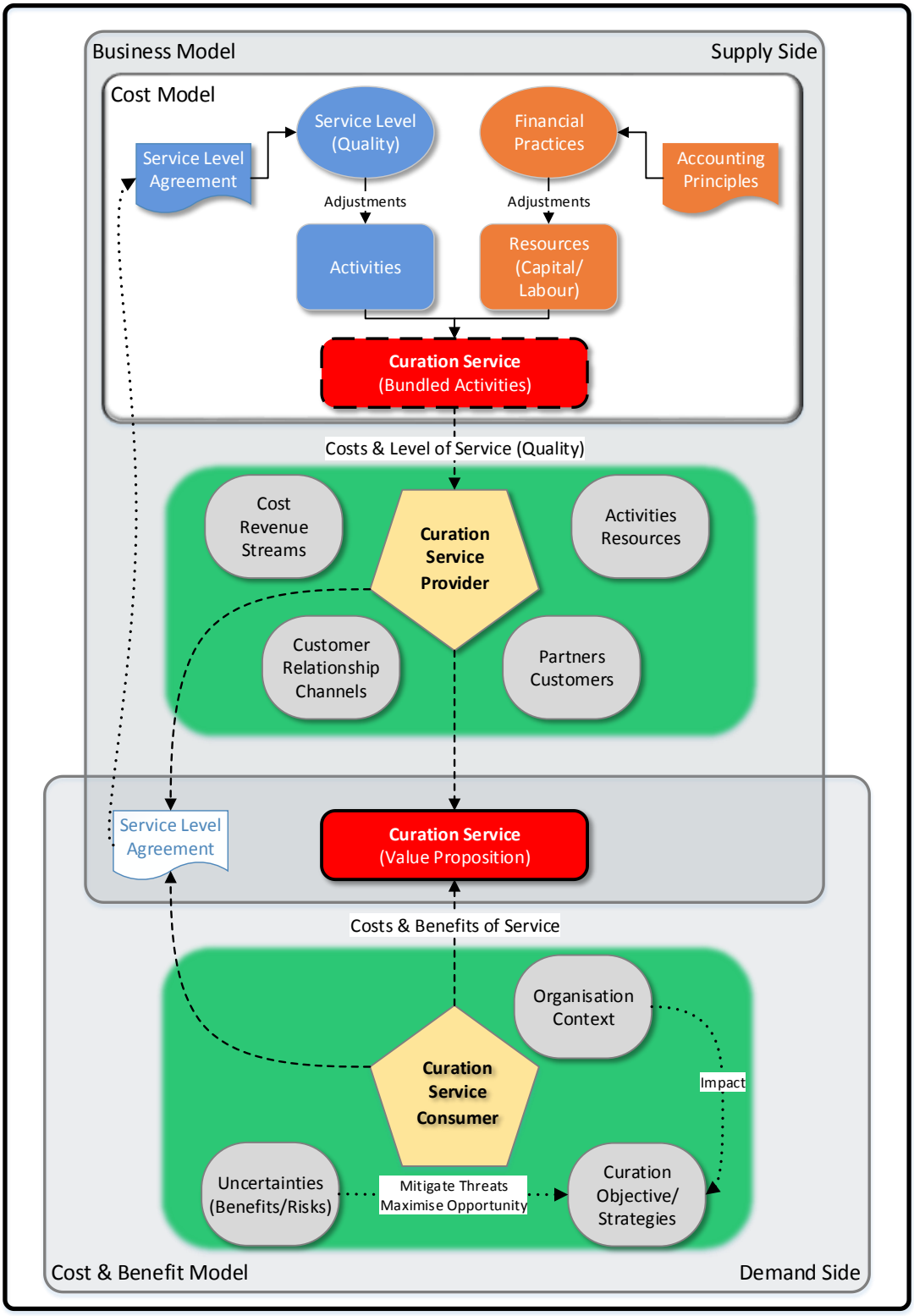

Figure 4—Nested Curation Models

6.1 A Cost Concept Model for Curation

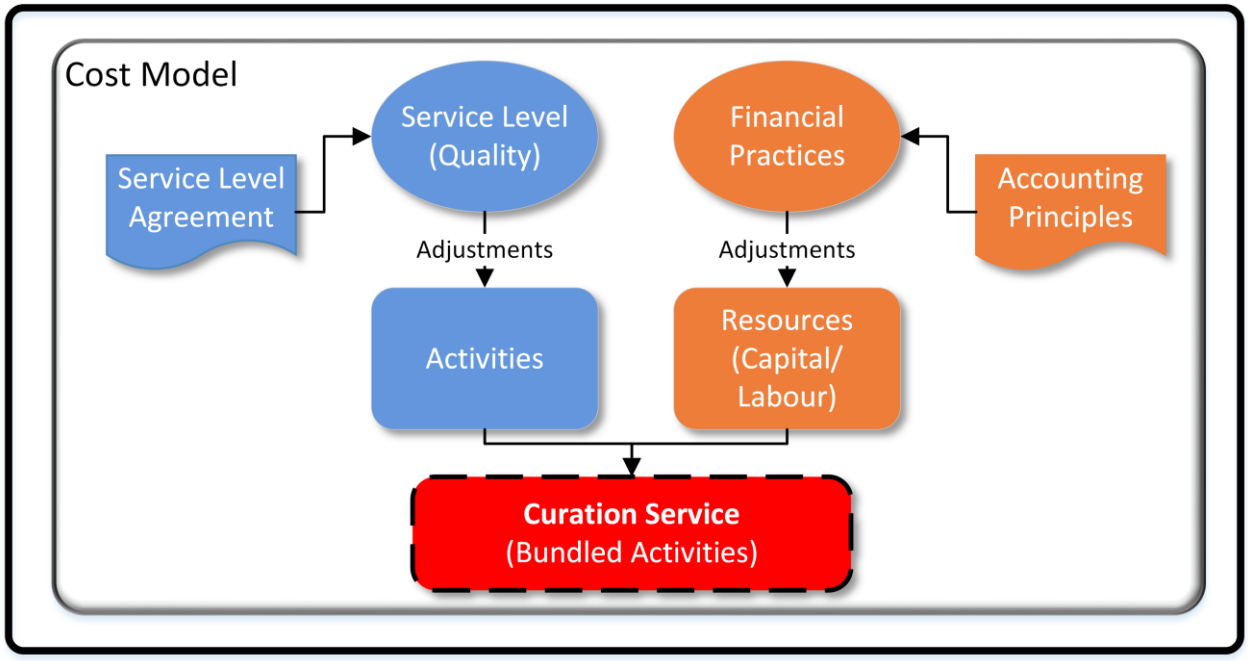

Figure 5—From cost concepts to raw curation services

Cost models apply and extend the resource and activity concepts (see section 3) to meet a given set of use cases. Direct and indirect costs are defined in terms of a specific approach to capturing capital and labour costs for a set of activities.

Past costs may be counted or future costs estimated with a view to:

- Comparing actual costs over time in a standardised way
- Comparing past costs to estimated future costs
- Comparing the projected cost of two or more curation options
- Comparing one organisations costs to another's

Local *Financial Practices* are used (in line with prevalent *Accounting Principles*) to apply appropriate adjustments (inflation, deflation, depreciation, interest etc.) which support the calculations relating to the expenditure of resources in terms of capital and labour.

Curation activities are clearly defined down to an agreed level of granularity. For 'Activities' adjustments may be made to ensure an agreed level of quality. A Service Level Agreement (SLA) may be developed to formally manage quality levels between the provider and consumer (see *A Business Model for Curation*).

Together these interactions between the core cost concepts form a cost concept model for curation. A given set of activities undertaken with a given quantity of resources provide us with a *Raw Curation Service* with an associated cost.

6.2 A Cost and Benefit Model for Curation

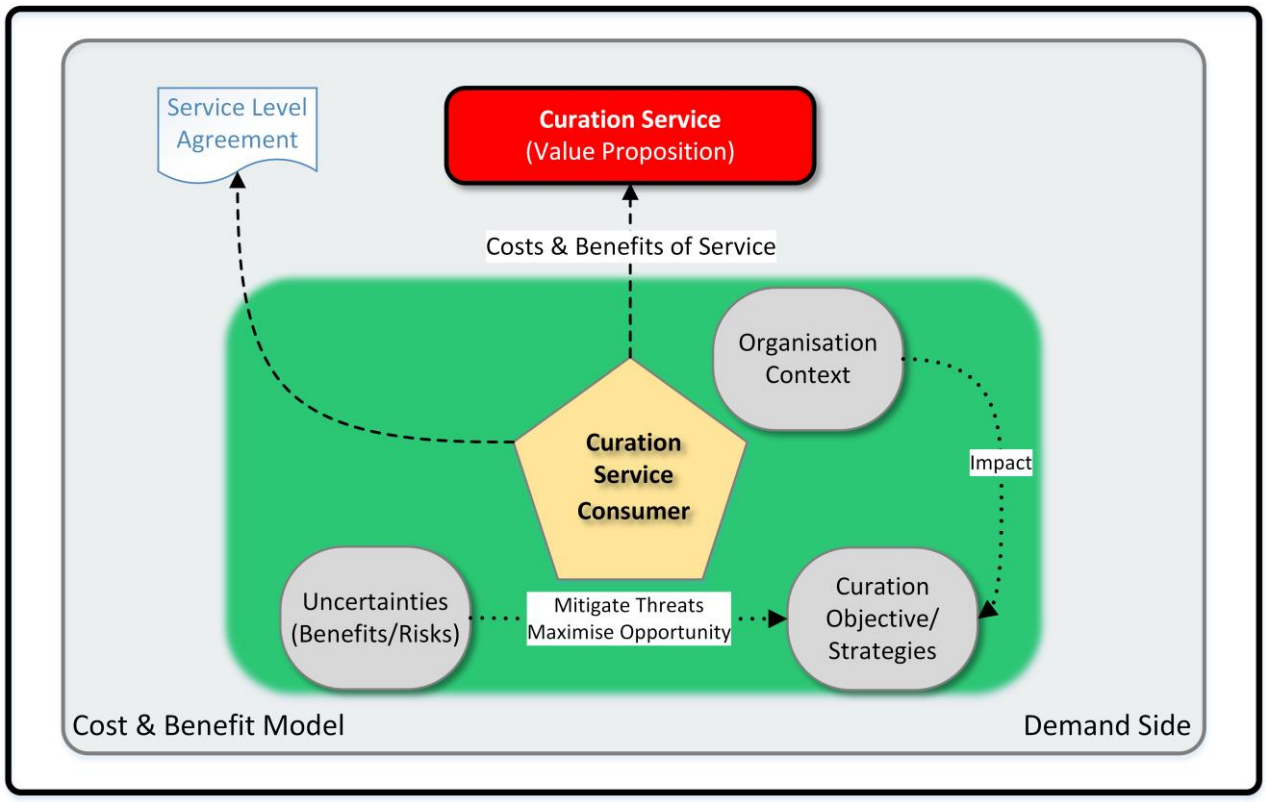

Figure 6—Cost and Benefit Model for Curation

Value may be expressed as a willingness to pay for or fund, a new or changed service. Benefits are subjective in the sense that, the value that a curation service delivers is judged by the service consumer. A repository service provider considering whether to invest in a benefit such as increased trust will factor in the customers (if the service is supplied by an external provider) or the funders (if the service is supplied in-house) judgements about the value of the certification before seeking trusted digital repository status.

Some benefits have a market price. For example, the benefits of a music service that offers streaming of songs based on user fees or licenses, or the benefits of migrating a collection of image files to a format, which takes up less space and thus potentially saves costs on storage. These benefits are also called financial or economic benefits. Others have easily identifiable costs, but more difficult pricing. For instance, greater replication of content over a more widely-distributed area brings additional costs that are easy to calculate. It also reduces the risk of loss by a factor which can be calculated with a high degree of accuracy. The value of that reduced risk will depend on the customer, the content and the service model.

If there is no conventional market on which a benefit can be traded, no market price can be applied. Europeana.eu which aggregates European memory institutions' cultural heritage assets to make them more easily accessible to the general public provides an example of just such a scenario. Even though such non-financial or non-economic benefits do not have a direct market price, they still represent 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 analyse past behaviours and stated preferences (also known as contingent valuation) which asks hypothetical questions, for example about willingness to pay for a service or a given level of service. Benefits can also be categorised in other ways (see *Integrating*

Benefits) and encompass uncertainties (threats and opportunities) which may be subjected to some level of risk analysis.

The consumer analyses requirements for curation, evaluates requirements against the costs and benefits of current and proposed curation services, and makes decisions about which services to select.

Information derived from the stakeholder ecosystem is integrated into the *Organisational Context* including the desired benefits, variables such as the mission, people and systems in place and the information assets being managed (including their quantity and quality); all these influence curation objectives and consumer decisions about curation services.

Note that the conceptual model distinguishes between the quality of a curation service and the benefits of the service. The costs and the quality of a service can be assessed uniquely and independently of the consumer, whereas the benefits of a service are relative in the sense that they depend on the consumer's service requirements.

The cost and quality information feed into the Business Model, which forms the basis for generating a value proposition for the curation service. Information about the proposed curation service, including a specification of the quality and the costs (possibly with profit added) feed into the Cost and Benefit model. This nested model enables a comparative analysis of the costs and benefits of the curation service in relation to the consumer's service requirements, which have been deduced from the organisation context via the stakeholder ecosystem. The current generation of cost models are only in the early stages of integrating benefits into the cost of curation.

6.3 A Business Model for Curation

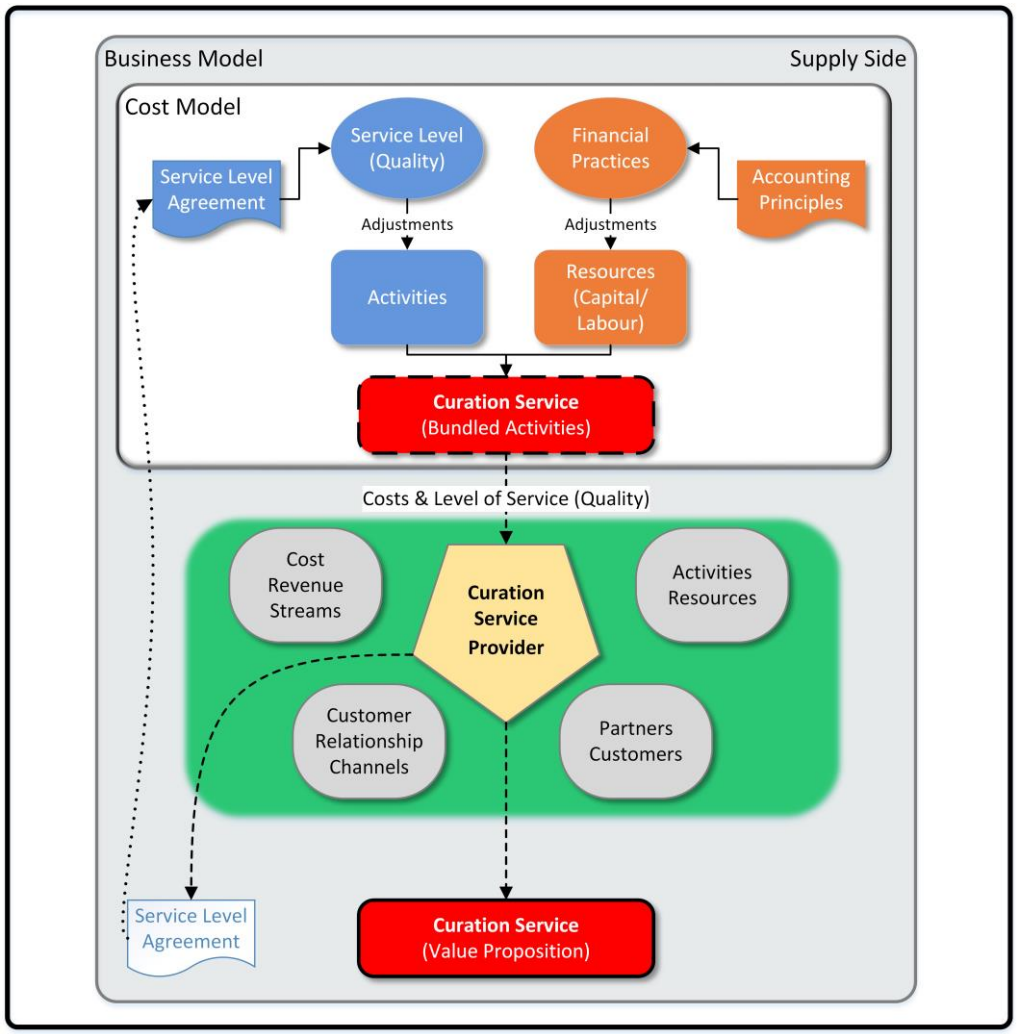

Figure 7—Applying a business model to develop a curation service

Business Models allow organisations to illustrate and understand how they create, deliver and capture value. Day to day curation operations and their raw services do not exist in isolation. The raw curation service must be presented as a full service and provide consumers with a clear value proposition through the generation of a business case. Depending on the business case and whether it is an in-house or external service provider different relationships may exist between the price and the underlying costs of the service.

It may be possible for the consumer to specify the required service, while in other cases, such as some outsourced solutions, it may only be possible to select from one or more predefined services. The *Curation Service* can be defined in an agreement between the provider and the consumer, also known as a *Service Level Agreement* (SLA). Such agreements may be legally binding or have a more informal or ad hoc character.

Business models for digital curation describe the role digital curation plays within an organisation and how curation creates value within the organisation. They will vary depending on objectives and context, but organisation of the same type will have similar business models.

Development of a business case may be undertaken through some standard process such as a Business Model Generation Canvas[5] which analyses *Customers* (consumers), *Partners, Activities, Resources, Customer Relationships, Channels, Cost Structure* and *Revenue Streams* as well as integrating information about the quality and cost of the raw curation service.

A business model canvas employs a standardised template consisting of 9 building blocks (see *Business Model Canvas*). Each block is analysed using concise questions to support a clear overview of the business model applicable to a wide variety of organisations.

To fully cover the needs of the consumer the value proposition of the service and the associated SLA negotiated between the demand and supply side must address the benefits derived as well as the costs incurred.

6.4 Economic Model

Definition: Economic Model

A representation that describes how economic processes around digital curation work; including the flow of resources (costs and revenues) within the economic lifecycle of digital information assets, and stakeholders (from the demand, supply and management side) interaction with this lifecycle.

Figure 4 shows an overview of the relation between economic models and cost and benefit models [4C, MS9, 2013, p. 41].

One method of understanding the broader economic model is through the Economic Sustainability Reference Model (ESRM) which is best understood as a strategic tool for planning at executive and managerial level rather than for operational staff. It provides a foundation for the development of successful sustainability strategies for digital curation. It does this by defining the issues; 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.

The ESRM provides a framework that assists in thinking through sustainability issues over the complete lifecycle for digital assets. The related self-assessment questionnaire steps through that framework and offers planners a chance to consider each component of the reference model against a local context, thereby identifying areas where change, improvement or implementation may realise benefits and/or mitigate risks.

Figure 4 presents the cost model as providing detail on the time and effort involved in managing digital assets through a series of activities and processes. In contrast the benefit model presents the curation process as a ‘black box’ where the focus is on the outcomes of curation which provide some perceived value or benefit to the curation system or curation service.

[5] The Business Model Canvas (http://www.businessmodelgeneration.com/canvas) is a generic business model used by WP4 T5.4 to analyse and generate business cases around curation services.

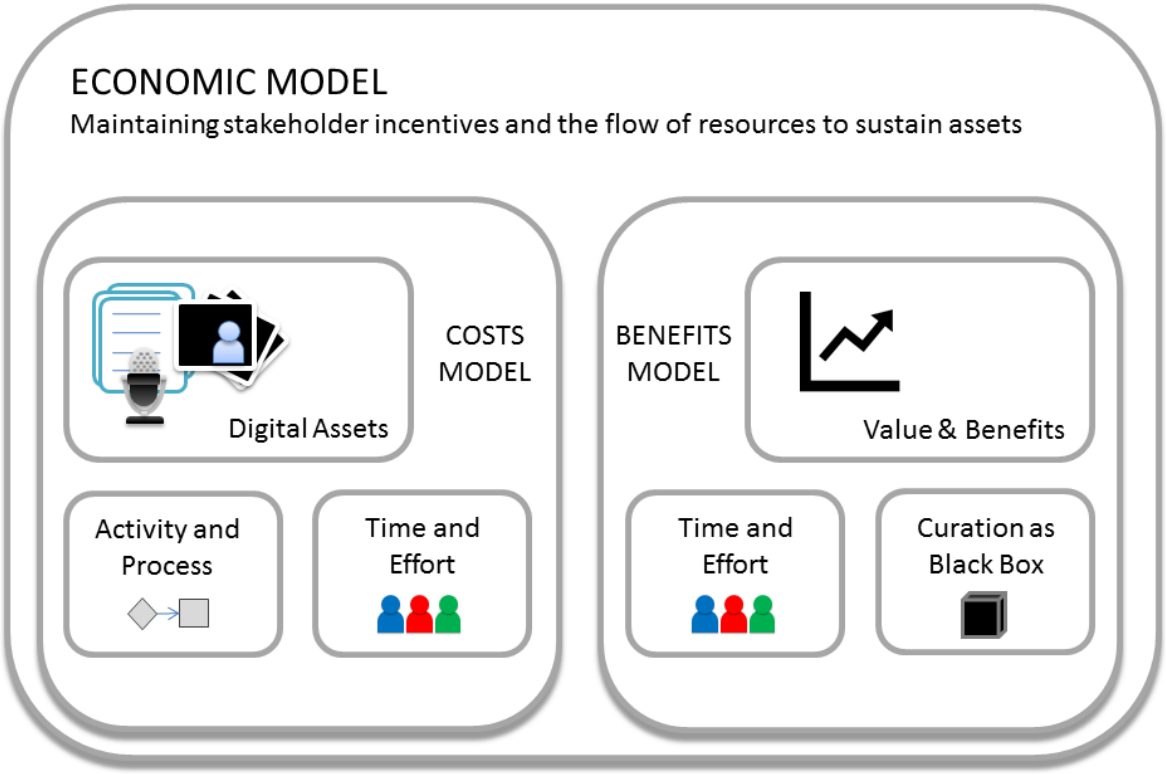

Figure 8—Benefits Model from the senior management perspective

The black box analogy is designed to indicate that at the executive and managerial level of the stakeholder ecosystem (shaded in green on the nested diagrams) as well as indirect economic determinants (IED) such as trust, efficiency, transparency and so on are evaluated against the outcomes of curation, not against its detailed internal processes.

But costs and benefits cannot be easily separated at the operational level where most current cost models focus. The demand for curation services, associated quality criteria and strategic IEDs must all be identified through stakeholder interaction. The organisation must work to convert the IEDs to more concrete, actionable statements of 'benefit' which can be considered alongside costs (see section 6.3). In this service-oriented approach to costs and benefits the *Curation Service* constitutes the heart of the nested model—it is the value proposed by the *Curation Service Provider* (supply side) to the *Curation Service Consumer* (demand side).

A service may cover the whole digital curation lifecycle or selected parts of the lifecycle, such as an ingest service. In practice an organisation with a demand for a curation service may assemble this service using any mix of in-house and outsourced services but the conceptual separation between provider and consumer allows us to make their different incentives for curation explicit.

7 Conclusions and next steps

This document introduced a Framework against which current and future cost models for curating digital assets can be benchmarked. As noted earlier, the goal of this task was not to create a single, functionally implementable cost modelling application but rather to design a model based on common concepts and to develop a generic gateway specification that can be used by future model developers, service and solution providers, and by researchers in follow-up research and development projects.

The core cost concepts, implementation guide, gateway specification, and nested model will be distilled and delivered via the Curation Costs Exchange (CCEx) as a means of enabling organisations to develop and implement new—or refine existing—cost models that reflect their organisation context while still facilitating comparability with peers. These outputs will, of necessity, require updating and refinement as our understanding of curation costing matures and Framework users will have the ability to feed into the refinement of the core concepts through our ongoing engagement work.

Take up of the 4C Framework will help to ensure that greater standardisation relating to the capture and sharing of curation costs is realised. As the ***resources*** and ***activities*** required to support the curation lifecycle are more universally understood and more comparable across institutions, we anticipate that emerging curation and preservation solutions and services will better reflect the wide range of ***organisation contexts*** where curation activities must be undertaken. Ultimately, we envisage that the improved clarity around the costs of curation and the context in which they occur will help to define requirements for more cost-effective and efficient systems that will help to ensure that sustainable services and solutions can be realised.

Annex 1: The role of Indirect Economic Determinants in relation to risks and benefits

Organisations' decisions incorporate benefits of a non-financial character as well as costs. These non-financial variables in the decision-making process can be characterised as 'indirect economic determinants' (IED) which act as conceptual controls to support discussions and strategic thinking about curation.

IEDs are characterised as one word statements of concepts that are important in digital curation. Each IED statement has an implied set of related uncertainties that will influence whether any associated curation cost should be viewed as a sound investment. Addressing these uncertainties has a cost and will realise a specific benefit or mitigate a particular risk. These are the main incentives for incurring any costs associated with IEDs. Applied at organisational management level by the Curation Service Consumer, the IEDs and related risks and benefits help to inform decisions making about curation investment and help to shape business cases and sustainability strategies. The types of digital asset, the purpose of preservation, the intended target audience and the type of organisation all determine how these IEDs can benefit the organisation. And the costs of investing into IEDs are balanced against the benefits they can yield or the risks they will mitigate. The degree of importance that is placed on an IED will influence the amount of effort and resources that will be invested to ensure that it is adequately addressed. The attitude towards IEDs will be reflected in policies and strategies and impact the requirements for Curation Services and Service Level Agreement.

To effectively integrate the issue of benefits into cost calculations the broad goals defined as 'Indirect Economic Determinants' at the stakeholder level must be operationalised into the organisational context.

For example Trustworthiness is an IED that factors into decision making about curation aims and strategies, uncertainties and needs for curation services at the Service Consumer management level. Prioritising trust within curation could be translated into a policy of working towards certification as a trusted digital repository. This has implications for the Curation Service and the Service Adjustment made by the Service Provider including which activities are performed and their quality criteria. This will impact the price of running the curation service. At the same time investing into trustworthiness would yield benefits like authority, a good reputation within the community and trust from funders and external stakeholders.

The 4C project has identified a list of IEDs that are considered significant in curation today:

- Authenticity
- Benefit
- Confidentiality
- Efficiency
- Flexibility
- Impact
- Innovation
- Interoperability
- Quality
- Reputation
- Risk
- Sensitivity
- Skills
- Sustainability
- Transparency
- Trustworthiness
- Value

This list provides examples, IEDs are not a fixed set of concepts, their interpretation is not universal and they are of limited value without context including to whom the IED is relevant and the incentives they imply. If for example, an investment results in mitigation of a certain risk, this only represents value

proportional to the stakeholder's incentive to reduce this risk. If controlling the risk is not critical for the business case the likelihood of investment is reduced. Evaluation of these determinants is shaped by the organization's objectives.

In order to integrate the non-financial variables into the costing of digital curation it is useful to demonstrate how these IEDs can be evaluated and expressed as non-financial incentives which, alongside cost data, are used in decision making. Because the IEDs are subjective and their interpretation is context and stakeholder dependent it is impossible to make generally applicable translations of IEDs into something measurable. It is only possible to make some overall considerations about how IEDs can be taken from one-word statements into something that can be evaluated and integrated into decision-making alongside cost data and benefits assessment. Using the list of IEDs above as a reference point we may consider IEDs in broad types depending on how they can be evaluated with the caveat that some may be measured in several ways.

Subjective measures

The first group of IEDs include for example the desire for a good reputation, or flexibility, or innovation which can only be evaluated through subjective measures. For example Reputation can be expressed as a 'general feeling' within a particular group of stakeholders about an organisation or system. It may be possible to agree a measurement standard for reputation, for example a survey of stakeholders, but it remains at a somewhat subjective level of measurement and evaluation. Subjective IEDS tend to be customer/end user related and reflect investments that may be made to enhance quality of service or increase service demand. Related IEDs include transparency, authenticity, trustworthiness, quality, reputation, and innovation.

Context dependent objective measures

The second group are IEDs which are objectively measurable if some additional context is defined. This group includes amongst others integrity, confidentiality and quality. With appropriate context defined around the business processes these IEDs can be directly measured. Confidentiality can for example be measured once we have said that this is derived from authentication, authorisation, training on data disclosure and so on. Context dependent objective measures IEDs tend to relate to investments that realise operational excellence and contribute towards achieving organisational objectives.

Standardised measures

The third group of IEDs are the ones that are subject to audit and standardisation. Some of the IEDs in the list have management standards and practices are supported by formal standards. These measure governance and management practices at the repository level rather than evaluating every sub-process undertaken or every digital object managed and include:

- ISO 16363 for Trustworthiness
- The ISO31000 for risk governance
- ISO 27000 for Information Security touches on authenticity, confidentiality and course risk
- ISO9000 for process quality

Standards offer a broadly accepted way of expressing and measuring IEDs. But even a ISO certification can only be an indicator of good practice which identifies the organisation as seeking continuous improvement; these measures do not (and cannot be) total guarantees of quality, risk avoidance or information security. Standardised measures IEDs tend to relate to the organisational culture and the

skills of the workforce. IEDs in this category include skills, trust, transparency and quality. In each of the IED types there are various challenges in moving from the one-word statements to context-specific expressions of incentive. The biggest challenge lies with the context and stakeholder dependency itself. Therefore it is necessary to clearly define in relation to what and whom the IEDs must be evaluated and demonstrated.

Annex 2: Maturity

In describing the interactions of core concepts and the movement from stakeholder-driven IEDs to operationalised approaches to integrating benefits alongside costs it is clear there are organisational challenges in implementing a cost and benefit driven approach which will drive decisions about adopting a curation cost methodology. Once adopted the conclusions derived from the process must then be reintegrated into the business processes of the organisations and such changes must be managed.

From these statements it is clear that a certainly level of organisational maturity is necessary to implement methods and act upon results. Less mature organisations may use simpler cost approaches to drive improved practice but the more ambitious and granular the cost collection goals the greater the need for a mature infrastructure. Maturity implies greater accuracy and greater impact from costing efforts.

Concepts of organisational maturity will be familiar to organisations which undertake formal information governance, records management or risk assessments.

The original capability maturity model[6] focussed on software development but already included the familiar five levels of maturity which are also referenced in the Prince 2 Maturity Model.

1. Initial (chaotic and ad hoc)
2. Repeatable (processes documented sufficiently to be repeatable)
3. Defined (processes standardised)
4. Managed (with appropriate metrics for measurement)
5. Optimising (reaching a level of continuous improvement)

Carnegie Mellon University (http://www.sei.cmu.edu/) broadened the focus of the work with the Capability Maturity Model Integration whose current version 1.3 updates all three of the current maturity models: Development, Services and Acquisition but retains the 5 levels.

It is important to keep in mind that the 'organisation' involved in delivering curation services may actually be a single project, or a system made up of several organisations with different goals and expectation and different understanding of the level of service to be attained. If curation costing efforts cover several lifecycle stages delivered by multiple organisation then the maturity of the overall system may only be that of the least mature participant. Similarly an organisation with a formal approach to maturity may identify different levels of maturity within different functions, groups or processes, which it undertakes.

Organisations at level 1, with reactive and undocumented processes may find most meaningful curation costing activities are challenging to implement and unreliable in their results. A lack of process stability will make cost and related measurements less accurate. A lack of organisational infrastructure stability may imply that taking the correct action in response to curation cost conclusions will be challenging. The ad hoc nature of level 1 organisations means that they are unlikely to have a sufficient understanding of the stakeholder ecosystem and organisational profile to integrate clear risk and benefit criteria into cost decisions.

This does not imply that cost models have no value to a less mature organisation. Collection of costs beyond those undertaken for standard accountancy processes may be complex and time consuming, but

[6] Capability Maturity ModelSM for Software, Version 1.1 (http://www.sei.cmu.edu/reports/93tr024.pdf)

this can be justified if the cost information collected, such as indicative 'snapshots' for limited scenarios can be used to drive improved practice.

Using curation cost information to deliver improvement will be less challenging for organisations at level 2 as with documentation sufficient to support repeatable processes any cost benchmarking can be repeated with some level of meaningful comparison possible over time.

At level 3 defined processes which are consistent across the organisation will support greater curation cost granularity across clearly defined functions and activities.

For organisations moving from level 3 to 4 it may be useful to integrate the issue of curation costing directly into plans for developing appropriate metrics for managed processes. Organisations which already have managed processes are likely to find that significant information necessary to support curation cost models is already part of their analysis and reporting structure. As well as the potential for collecting more granular business process metrics about activities the stakeholder identification and management are likely to be more developed and a coherent approach to risks and benefits becomes practical at level 4.

For most organisations reaching an optimising level 5 is likely to be extremely costly and challenging but the concept of 'continuous improvement' is familiar to organisations undertaking an ISO audit process where ongoing surveillance audits ensure that current practice is examined, optimised and improved.

The issue of organisational maturity is relevant to curation costing from several perspectives. An understanding of maturity levels will allow cost model developers to understand the level of information management necessary to support their cost methodology. For cost model adopters an understanding of their own organisations maturity impacts the scope and ambition of their curation cost work. For the wider community the availability of curation cost data from a range of mature organisations will provide a more meaningfully comparable and accurate evidence base.

Of course these levels will never be perfectly aligned across the whole of an organisation and there may be, for instance, detailed definition, management and measurement of curation processes such as ingest while stakeholder identification and management remain less mature. The maturity issue is introduced here to support critical evaluation of what the likely accuracy and impact of curation cost efforts will be for different implementers.

Annex 3: Lifecycle Descriptions

Before addressing the detail of business process activities it is useful for model developers and users to have a common understanding of what parts of the digital object lifecycle are in scope. Selecting and using a common lifecycle model also provides a good general structure and vocabulary for communications.

It is likely that those adopting curation costing methods are already involved with a variety of models intended to communicate complex issues more simply. The models may be delivered at different levels of abstraction to different audiences with different purposes. Like the more detailed models of activities, lifecycle models are at their most useful if they become maintained artefacts which form part of an organisations management approach but managing multiple activity and lifecycle models presents a number of challenges.

The variety of lifecycle model approaches below are all at a higher level of abstraction than that required to monitor business processes. Though they present varied design approaches none of them have been designed with the integration of a costing approach in mind.

The options and challenges around developing more detailed descriptions of curation functions and activities for curation cost methods are covered in *Structuring Activities.*

Some approaches to integrating the full lifecycle such as the DCC Lifecycle will focus on the cyclical and layered nature of curation processes

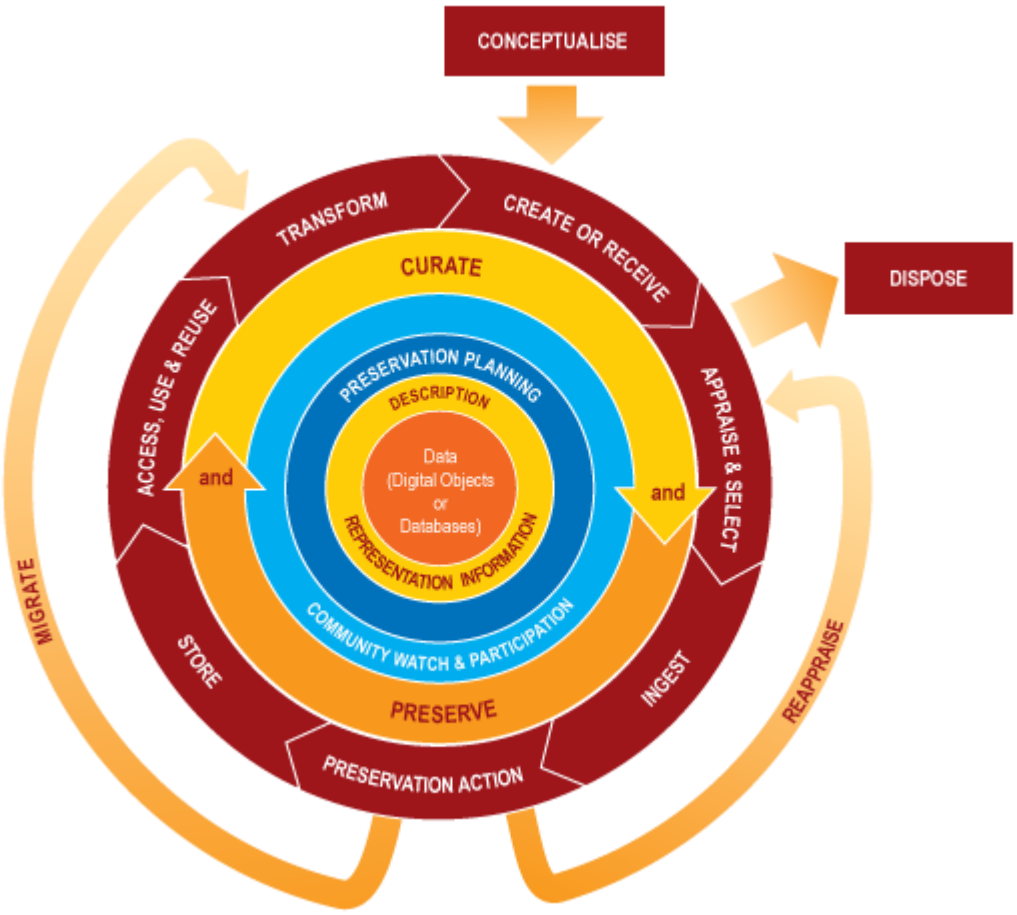

Figure 9—The Digital Curation Coalition Lifecycle

Others such as DCM consult (http://dmconsult.library.virginia.edu/) have a clear project lifecycle focus and do not see the archival process as a necessary precursor to data sharing.

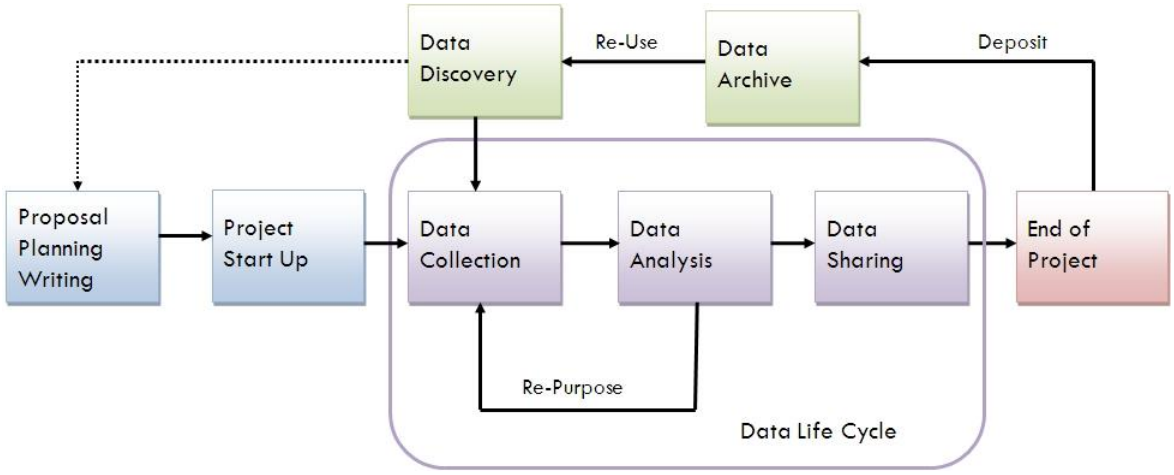

Figure 10—Data Management Consulting Group (DMConsult) Research Lifecycle

The Data Documentation Initiative offers a similar structure but assumes the 'study' as the basic unit within the lifecycle and explicitly describes where the Initiative see the lifecycle interaction with the scope of the OAIS model.

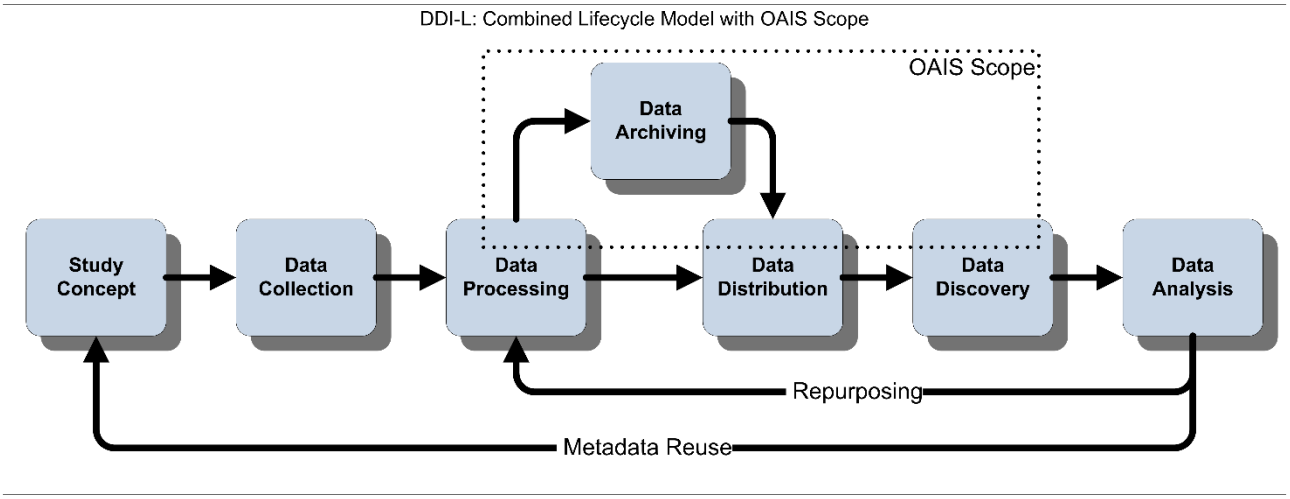

Figure 11—The Data Documentation Initiative (DDI) Combined Lifecycle Model with OAIS Scope.

The Generic Longitudinal Business Process Model (GLBPM) represents an effort from the DDI community to represent longitudinal dataset lifecycles more accurately. It uses a relatively linear approach to describing the stages but clarifies within the documentation that events are by no means strictly sequential.

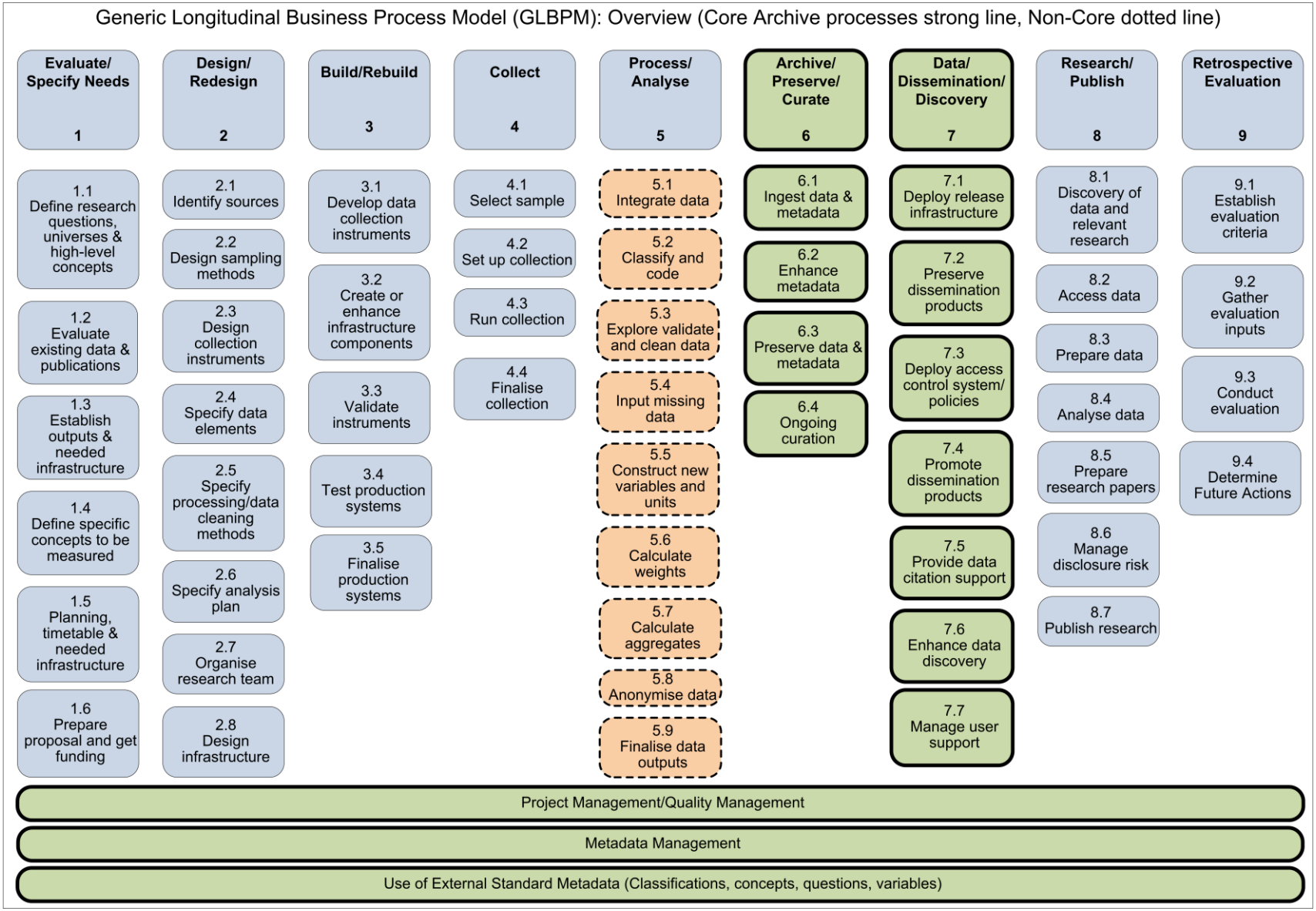

Figure 12—The Generic Longitudinal Business Process Model (GLBPM)

In the diagram above, items presented in green are the traditional remit of archives with long term digital preservation responsibility, but for many archives offering data validation, cleaning and quality assurance the items in orange remain significant.

Developers and users must maintain communication channels and share clear documentation to ensure a common view of the lifecycle is maintained. Much like the OAIS model, these lifecycles have not been explicitly developed with curation costing in mind and continue to evolve to address the challenges met within digital curation.

During the development of guidance on improved metadata quality for European Infrastructures the DASISH project (T5.3) identified that most data lifecycles are understandably centred on the data object which is the subject of research. The focus is on the integrity and fixing of that data and while it may be validated and enriched in some ways the data itself is seen as somewhat 'fixed'. For those engaging in curation (who of course overlap with those collecting, creating and using data) the metadata remains more dynamic than the original data. This more metadata focussed lifecycle may contain some elements useful to those considering full lifecycle curation costs.

Metadata design, redesign and implementation continue to be 'live' issues for curators and access providers even when the data remains unchanged. Those managing metadata, or using metadata to manage, continue to update to new standards and re-enrich metadata to meet the changing needs of the user communities. This contrasts somewhat with the research data lifecycles tendency to assume a fairly 'static' data object (barring preservation/admin metadata etc.) from the time of ingest into an Archive to the next Access/User/Re-use event. The metadata lifecycle below is based on OAIS as a central reference point for Archives and the DCC model as a familiar approach for general curation. It takes into account full lifecycle processes, processes which can be designed once and applied at several times during the lifecycle, and the traditional sequential lifecycle approach.

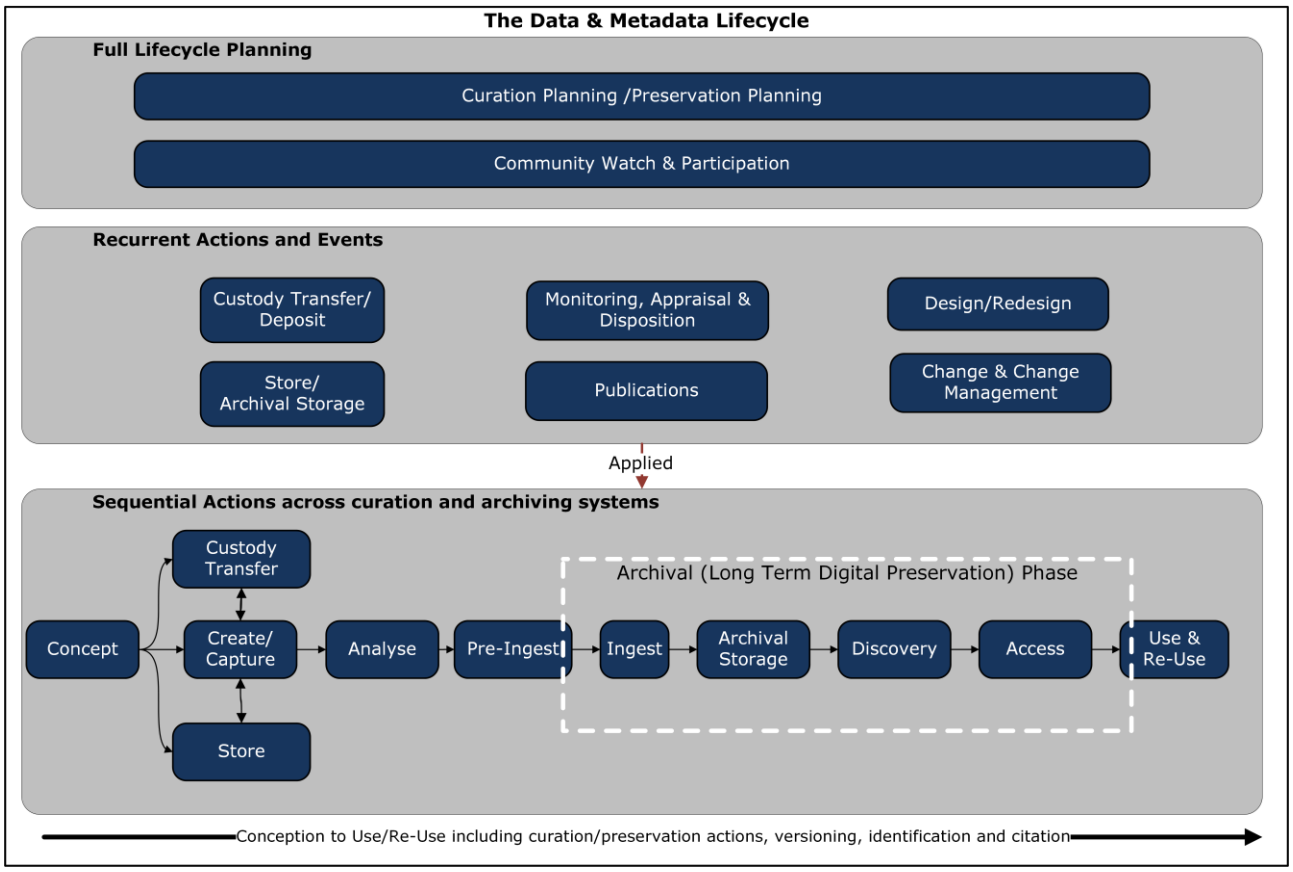

Figure 13—DASISH Project Metadata Quality Lifecycle

Annex 4: Amending or Extending the OAIS model

Since mapping processes to the OAIS functional model is not a straight forward process different other (cost) models have adapted OAIS. The Keeping Research Data Safe 2 (KRDS2) activity model and the activity model from the Bavarian State Library (BSL) and the University of the Federal Armed Forces Munich are two examples that demonstrate how curation activities can be specified and defined.

KRDS2 Activity Model

The Keeping Research Data Safe 2 (KRDS2) study (Beagrie et al., 2010) enhanced the KRDS2 activity model based on a review of its predecessor study.

The KRDS2 activity model can describe the lifecycle costs of research data. There are three components of the cost model:

- **Key cost variables and units—**affect the cost of preservation activities. There are two groups:
    - Economic adjustments
    - Service adjustments
- **An activity model**—describing cost relevant preservation activities. They are divided into:
    - Pre-Archive
    - Archive
    - Support Services
- **A resources template**—with categories (e.g. staff) and duration (year 1, year 2, etc.) of the costs.

In principle the necessary or used resources are identified using the activity model. The economic adjustments will distribute and maintain them over time and by using service adjustments resources are identified and adjusted to tailor the model to requirements. Finally the elements are assembled in the resources template and implemented as a TRAC-based cost model (Beagrie et al., 2010).

The KRDS2 activity model needs to be tailored to the end-users requirements. Similar to the OAIS reference model it is also generic and uses the same terms and definitions which support understanding and communications.

Depending on their requirements organisations need to decide on the appropriate level of detail for the definition of activities. Operations planning and process improvement need more granularity than cost management (Ibid.). There are two KRDS2 activity model versions with different level s of details. A simpler version provides a quick overview of the main phases and a more detailed version supporting operations planning and process improvements. We will describe the overview of the main phases of the KRDS2 activity model.

Pre-Archive Phase

This phase is related to research at universities their produced data where data is prepared for an archive.

> **Outreach—**Depositors and data producers receive guidance and training on best practices by the archive to support researchers with funding proposals and research data creation.

Initiation—Describes the initiation of the research activity which creates the data. Significant implications for preservation cost downstream.

Creation—Project activities related to the creating research data. Significant implications for preservation costs or archive access/use downstream.

Archive Phase

These activities are required for long-term archiving of research data.

Acquisition–This activity describes the process of acquiring research data.

Disposal–The disposal describes all processes related to transfer to a different archive or the controlled destruction. This applies to material which has not been selected for long-term curation and if policies, guidance or legal requirements demand it the disposal needs to be done securely.

Ingest–This description affects all activities dealing with receiving, reading, quality checking, cataloguing, of incoming data to the point of insertion into the archive (manual/electronic).

Archive Storage–This activity comprises services and functions relevant for storage and retrieval of Archival Information Packages (AIPs).

Preservation Planning–This activity describes services and functions relevant for monitoring, providing recommendations, and taking actions in order to ensure accessibility of the stored information.

First Mover Innovation–This describes the phase of implementation and R&D development of first tools, standards and best practices. This is a highly variable cost and up-front investments will affect downstream preservation costs.

Data Management–These services and functions are needed for populating, maintaining, and accessing descriptive information of archive content and administrative data used to manage the archive.

Access—These services allow access to the archive holdings for the consumers of the data.

Supporting Services

Administration—The administration is responsible for the operation of other functional entities.

Common Services—These are shared supporting services such as operating system services, network services etc.

Estates

These activities are not directly related to the preservation activities. Estate management and attendant costs are treated as a separate cost element.

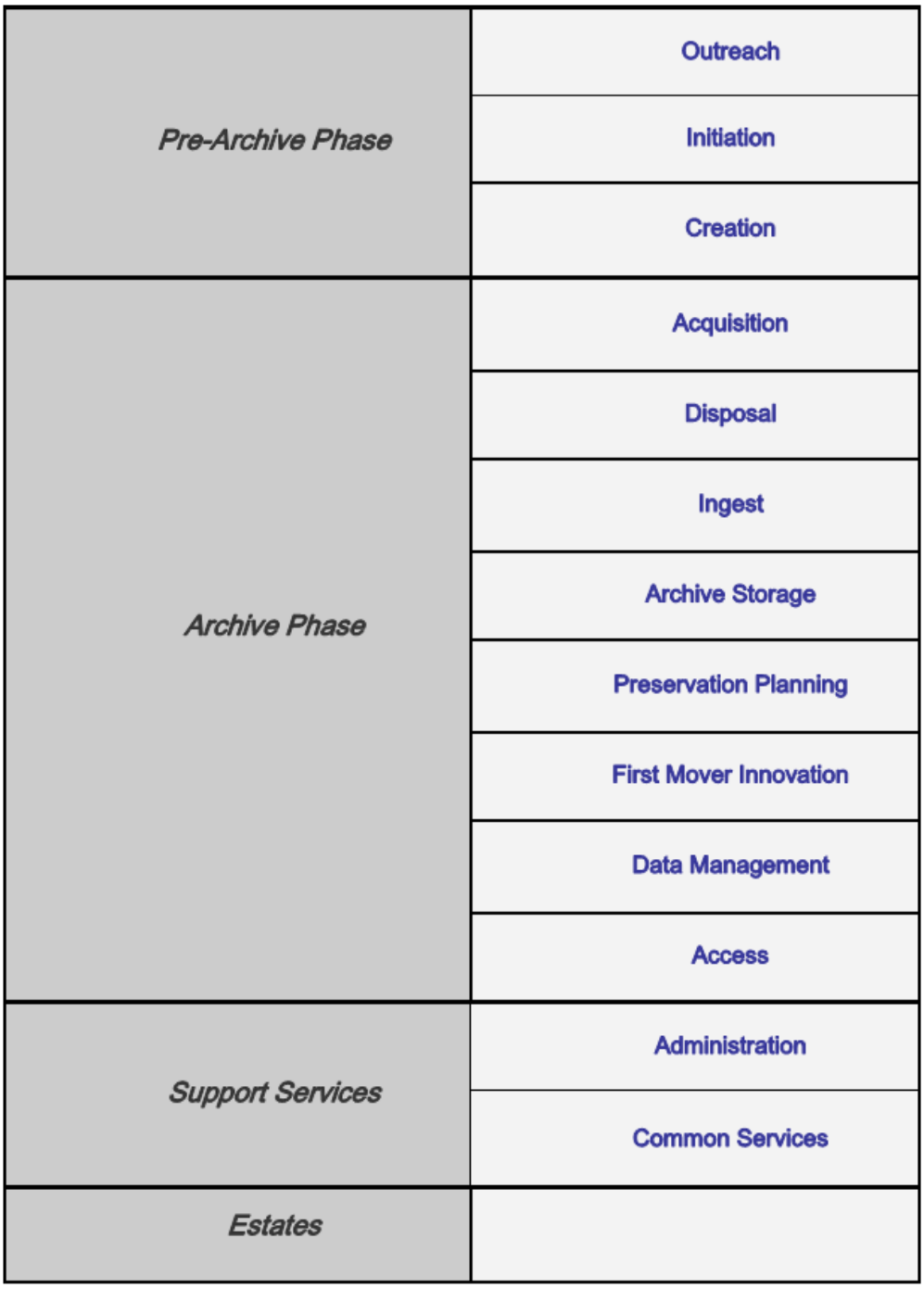

Figure 14—Overview of the main phases and activities in the KRDS2 Activity Model

Development of Organisational and Business Models for the Long-Term Preservation of Digital Objects

Another example for an activity model based on OAIS is the model developed a study of the Bavarian State Library (BSB) and the University of the Federal Armed Forces Munich named "Development of Organisational and Business Models for the Long-Term Preservation of Digital Objects"(Beinert et al., 2009). The activity model is based on the OAIS reference model (Book, 2002) and it is designed to be applicable beyond the boundaries of memory institutions or other service providers by using a customer-provider concept. Therefore other memory institutions and external project partners are also described by the roles of producer and consumer. The functional entity Preservation Planning is changed into

Preservation Planning & Action. It monitors, plans updates of the AIPs. Additionally it directs preservation actions and coordinates these preservation actions between the functional entities.

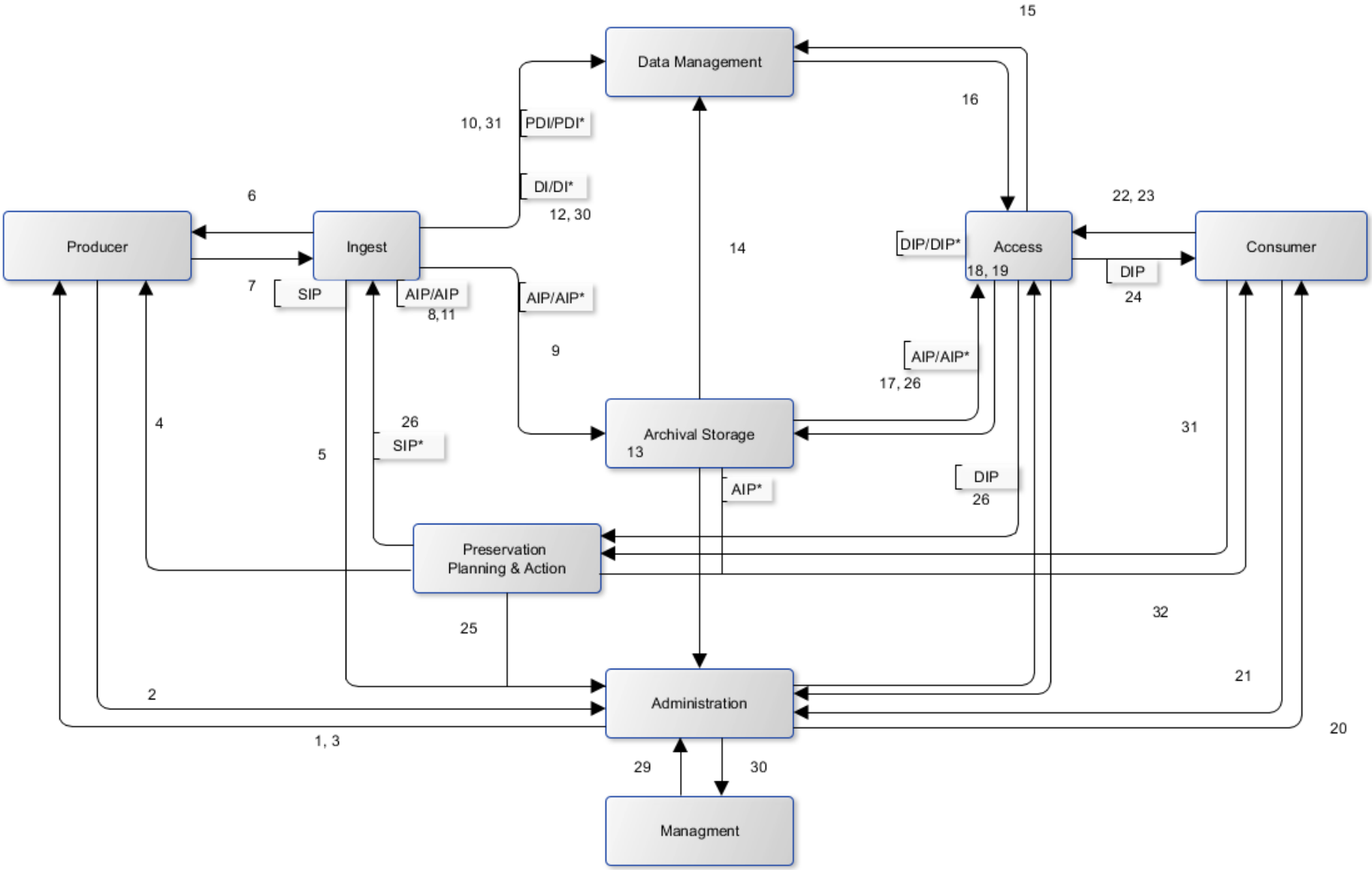

Figure 15—Process model based on extended OAIS

Description of the elements of the extended OAIS reference model

The numbered activities of the process model in Figure 15 are listed below.

1. **Request**
   A publisher addresses his request to a long-term digital archive to deposit content. Example: a small library stating a request to a deposit library to integrate its digitized assets.
2. **Offer**
   As a response to a request the administration creates an offer including a list of services and requirements for the assets to be submitted. Example: The archive describes the details of the storage and the capacities/capabilities of the access. The limitations and requirements regarding acceptable file formats are defined, the possibility of accessing high resolution copies on offline storage media, the offer to include metadata of the digital object to a centralized directory of references and the cost for this additional services.
3. **Agreement (Negotiation)**
   The agreement for the inclusion of the digital object is based on the offer and can be a contract or any other kind of agreement between both parties. Parts of the agreement are the specification of the preserved digital objects (data representing the content as well as data describing the access/representation of the objects).
4. **Submission of assets/SIP (receive submission)**
   The publisher delivers the assets to the archive or they are prepared to be fetched (push or pull mechanism). Example: the assets are transferred from the library to a work / submission area of the archive.
5. **SIP creation**
6. **Receipt confirmation, resubmit request to the producer**
7. **Quality assurance**
8. **Generate AIP**
9. **Transferring AIP to the archival storage**
10. **Transfer Preservation Description Information (PDI) to Data Management (Descriptive info)**
11. **Generate Descriptive Information (DI) (Generate Descriptive Info)**
12. **Transfer DI to Data Management**
13. **Archival Storage (Bitstream Preservation)**
14. **Updates of Archival Storage** (Database update request)
15. **Query Request** to Data Management
16. **Result List** (Query Response)
17. **Archival Storage transfers AIP** (Provide Data)
    Only parts which have been requested by the DIP are sent (configuration, reconstruction of AIP)
18. **Generation of DIP** (Generate DIP)
19. **Upgrade of DIP**, Augmentation
    The Upgrade includes all changes and extensions which increase the usability. Parts of the DIP are specialized processed or linked to other information resources. Example: image processing.
20. **Billing information** transferred to Consumer
21. **Consumer Access Permission**
    First time access may require a prior agreement to the user rights
22. **Consumer Query Request**
23. **Result Selection** based on query response
    Consumer selects data from the query response. Data request is transferred to Access

24. **Deliver Response**
    DIP is delivered to the Consumer
25. **Preservation Planning**
    Developing Preservation Strategies (brief description of the Preservation Planning Functional Entity)
26. **Generating AIP**
27. **Extending DI**
    DI is extended during preservation
28. **Extending PDI**
    PDI is extended during preservation
29. **Management policies, budgets**
    (Establish standards and policies)
    The management provides planning, installation, operation, control and budget policies. The management receives periodic reports.
30. **Reports**
    Administration sends reports to the management.
31. **Consumer Requirements**
    Consumers can suggest changes for the Access. Examples: different file format, including specific computing platforms.
32. **Response to Consumer Requirements**
    To improve customer relationship appropriate response to expressed requirements is needed. This also includes smaller memory institutions who are customers of deposit libraries.

Annex 5: Business Model Canvas

4C project task T4.5 investigates business models for digital curation, and it identifies and describes common basic business models for curation activities for each of the stakeholder groups defined in the 4C project. These basic models provide a starting point for the development process of a suitable business model for organisations and give insight into the business drivers, incentives and value relevant for different stakeholder groups.

Brief description of motivation and usage of the Business Model Canvas

The Business Model Canvas (BMC) is a management tool to support the creation of business models. These are descriptions of the strategic components of an organisation and it captures the essence of how on organisation creates, delivers and captures value (Osterwalder et al., 2010)

Business models of research related stakeholders, universities, memory institutions, and vendors are described with the BMC. These stakeholders are interviewed and asked to describe their organisation with the BMC. The resulting BMCs are evaluated and common basic business models for each analysed group will be derived. These basic models provide a starting point for the development process of a suitable business model for the organisations. As these basic business models cannot contain detailed descriptions examples of the captured business models will provide further guidance.

As most public organisations' main revenue consists of public funding, budget cuts have a negative impact on long-term curation projects. Additional revenue alternatives could decrease the dependence on public funding and increase the economic stability for planned projects. The canvas approach offers a way to discover alternative revenue models besides many well-known models such as subscription, usage, or royalty models.

Because some of the collected business model examples describe projects or services which are still in a planning phase the usages of the BMC could support the organisations in the process of creating a business model for new services.

The motivation to use this management tool is that it allows the description of organisations by analysing each building block with clear and concise questions. It uses concepts that are easy to understand and can describe a wide variety of organisations with the same basic building blocks. The following descriptions of the building blocks are from the book "Business Model Generation" (Osterwalder et al., 2010).

Description of the nine building blocks

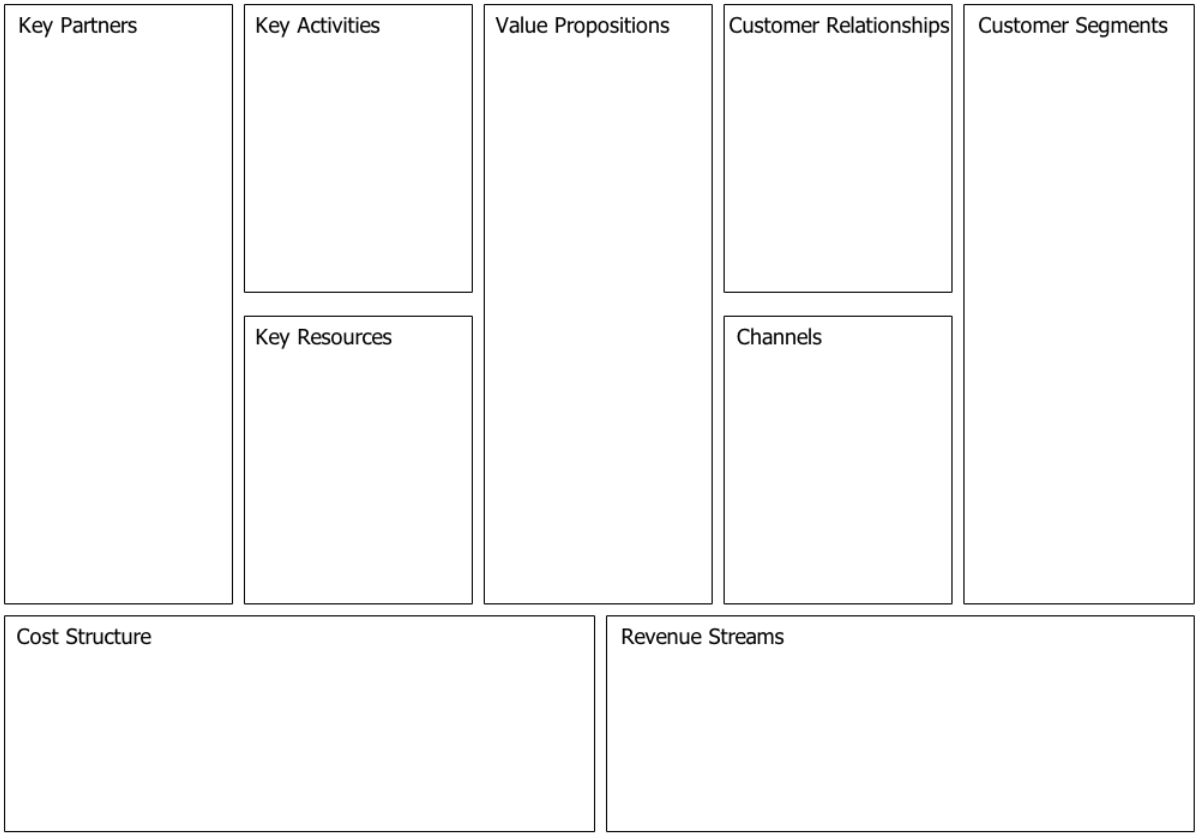

Figure 16—Business Model Canvas and its nine building blocks

Value Propositions

- What value do we deliver to the customer?
- Which one of our customer's problems are we helping to solve?
- What bundles of products and services are we offering to each Customer Segment?
- Which customer needs are we satisfying?

Characteristics (examples)

- Newness
- Performance
- Customization
- "Getting the Job Done"
- Design
- Brand/Status
- Price
- Cost Reduction
- Risk Reduction
- Accessibility
- Convenience/Usability

Customer Segments

- For whom are we creating value?

- Who are our most important customers?
    - Mass Market
    - Niche Market
    - Segmented
    - Diversified
    - Multi-sided Platform

Customer groups represent different segments if:

- Their needs require and justify a distinct offer
- They are reached through different Distribution Channels
- They require different types of relationships
- They have substantially different profitabilities
- They are willing to pay for different aspects of the offer

Channels

- Through which Channels do our Customer Segments want to be reached?
- How are we reaching them now?
- How are our Channels integrated?
- Which ones work best?
- Which ones are most cost-efficient?
- How are we integrating them with customer routines?

Channels can cover up to five distinct phases. There are different type of channels (own/partner, direct/indirect)

Channel Phases

1. Awareness—How do we raise awareness about our company's products and services?
2. Evaluation—How do we help customers evaluate our organization's Value Proposition?
3. Purchase—How do we allow customers to purchase specific products and services?
4. Delivery—How do we deliver a Value Proposition to customers?
5. After sales—How do we provide post-purchase customer support?

Customer Relationships

- What type of relationship does each of our Customer Segments expect us to establish and maintain with them?
- Which ones have we established?
- How are they integrated with the rest of our business model?
- How costly are they?

Examples

- Personal assistance
- Dedicated personal assistance
- Self-Service
- Automated services
- Communities
- Co-creation

Revenue Streams

- For what value are our customers really willing to pay?
- For what do they currently pay?
- How are they currently paying?
- How would they prefer to pay?
- How much does each Revenue Stream contribute to overall revenues?

Types

- Asset sale
- Usage fee
- Subscription fees
- Lending/Renting/Leasing
- Licensing
- Brokerage fees
- Advertising

Fixed Pricing

- List price
- Product feature dependent
- Customer segment dependent
- Volume dependent

Dynamic Pricing

- Negotiation (bargaining)
- Yield Management (price depends on inventory and time of purchase, e.g. used for perishable resources such as hotel rooms or airline tickets)
- Real-time-market (dynamic price estimation based on supply and demand)
- Auctions (price as outcome of competitive bidding)

Key Resources

- What key resources do our value propositions require?
- Our distribution channels?
- Customer relationships?
- Revenue streams?

Types of resources

- Physical
- Intellectual (brand parents, copyrights, data)
- Human
- Financial

Key Activities

- What key activities do our value propositions require?
- Our distribution channels?
- Customer relationships?
- Revenue streams?

Categories

- Production
- Problem solving
- Platform/network

Key Partnerships

- Who are our Key Partners?
- Who are our key suppliers?
- Which Key Resources are we acquiring from partners?
- Which Key Activities do partners perform?

Motivation for partnerships

- Optimization and economy
- Reduction of risk and uncertainty
- Acquisition of particular resources and activities

Four types of partnerships:

1. Strategic alliances between non-competitors
2. Coopetition: strategic partnerships between competitors
3. Joint ventures to develop new businesses
4. Buyer-supplier relationships to assure reliable supplies

Cost Structure

- What are the most important costs inherent in our business model?
- Which Key Resources are most expensive?
- Which Key Activities are most expensive?

Is your business more

- Cost Driven (leanest cost structure, low price value proposition, maximum automation, extensive outsourcing)
- Value Driven (focused on value creation, premium value proposition)

Sample Characteristics

- Fixed Costs (salaries, rents, utilities)
- Variable costs
- Economies of scale
- Economies of scope

References

4C, D2.1—Baseline Study of Stakeholder & Stakeholder Initiatives, 2013, http://4cproject.eu/community-resources/outputs-and-deliverables/d2-1-baseline-study-of-stakeholder-stakeholder-initiatives [Accessed 31 Jan 2014].

4C, D4.1—A prioritised assessment of the indirect economic determinants of digital curation, 2013, http://www.4cproject.eu/community-resources/outputs-and-deliverables/d4-1-a-prioritised-assessment-of-the-indirect-economic-determinants-of-digital-curation [Accessed 31 Jan 2014].

4C, MS9—Draft Economic Sustainability Reference Model, 2013, http://4cproject.eu/community-resources/outputs-and-deliverables/ms9-draft-economic-sustainability-reference-model [Accessed 31 Jan 2014].

APARSEN, D32.1, Report on cost parameters for digital repositories, 2013, http://www.alliancepermanentaccess.org/index.php/knowledge-base/member-resources/documents-and-downloads/?did=123 [Accessed 31 Jan 2014].

APARSEN, D32.2, Report on testing of cost models and further analysis of cost parameters, 2013 http://www.alliancepermanentaccess.org/index.php/knowledge-base/member-resources/documents-and-downloads/?did=150 [Accessed 31 Jan 2014].

Beagrie, N., Chruszcz, J. and Lavoie, B. Keeping Research Data Safe. A Cost Model and Guidance for UK Universities, Copyright HEFCE 2008, http://www.jisc.ac.uk/media/documents/publications/keepingresearchdatasafe0408.pdf [Accessed 31 Jan 2014].

Beagrie, N., Lavoie, B.F., Woollard, M., 2010. Keeping Research Data Safe 2. HEFCE.

Beinert, T., Hagel, H., Minkus, M., Rödig, P., Lang, S., Schoger, A., 2009. Entwicklung von Organisations- und Geschäftsmodellen zur Langzeitarchivierung digitaler Objekte aus DFG-geförderten Digitalisierungsprojekten.

Book, B., 2002. Reference Model for an Open Archival Information System (OAIS).

Book, M., 2012. Reference Model for an Open Archival Information System (OAIS), Recommended Practice, CCSDS 650.0-M-2 (Magenta Book) Issue 2.

BRT—Blue Ribbon Task Force on Sustainable Digital Preservation and Access, “Sustainable Economics for a Digital Plant: Ensuring Long-Term Access to Digital Information,” Final report, 2010, http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf [Accessed 31 Jan 2014].

CCSDS (Consultative Committee for Space Data Systems), Audit and Certification of Trustworthy Digital Repositories, CCSDS 652.0-M-1, Magenta Book, 2011, (ISO 16363:2012), http://public.ccsds.org/publications/archive/652x0m1.pdf [Accessed 31 Jan 2014].

CCSDS (Consultative Committee for Space Data Systems), Producer-Archive Interface Methodology Abstract Standard (PAIMAS), CCSDS Magenta Book, 2004, (ISO 20652:2006, reviewed and confirmed in 2009) http://public.ccsds.org/publications/archive/651x0m1.pdf [Accessed 31 Jan 2014].

CCSDS (Consultative Committee for Space Data Systems), Reference Model for an Open Archival Information System (OAIS), CCSDS 650.0-M-2, Magenta Book, 2012 (ISO14721:2012), http://public.ccsds.org/publications/archive/650x0m2.pdf [Accessed 31 Jan 2014].

Du Plessis, J., McConvill, J., Bagaric, M., 2005. Principles of Contemporary Corporate Governance

Garrett, J. (co-chair) and D. Waters (co-chair), Preserving Digital Information: Report of the Task Force on Archiving of Digital Information, The Commission on Preservation and Access and The Research Library Group, 1996, http://www.clir.org/pubs/reports/pub63/reports/pub63watersgarrett.pdf [Accessed 31 Jan 2014].

Lunghi, M, N. Grindley, B. Stoklasová, A. Trehub, and C. Egger, "Economic Alignment," in Aligning National Approaches to Digital Preservation, ed. N. McGovern (Educopia Institute Publication, 2012), 195-268, http://educopia.org/publications/ANADP [Accessed 31 Jan 2014]

Osterwalder, A., Pigneur, Y., Clark, T., 2010. Business model generation: a handbook for visionaries, game changers, and challengers. Wiley, Hoboken, NJ.

Palaiologk, A. S., Economides, A. A., Tjalsma, H. D., & Sesink, L. B. (2012): "An activity-based costing model for long-term preservation and dissemination of digital research data: the case of DANS" in International Journal on Digital Libraries, 12(4), 195–214. doi:10.1007/s00799-012-0092-1: http://link.springer.com/article/10.1007%2Fs00799-012-0092-1

Post, James (2002). Redefining the Corporation: Stakeholder Management and Organizational Wealth.

Schumann, N., Recker, A., 2012. De-mystifying OAIS compliance: Benefits and challenges of mapping the OAIS reference model to the GESIS Data Archive. IASSIST Q. 36.

Watson, J., The LIFE project research review – Mapping the landscape, riding a life cycle, 2005, http://discovery.ucl.ac.uk/1856/1/review.pdf [Accessed 31 Jan 2014].