

Project no. 269977

APARSEN

Alliance for Permanent Access to the Records of Science Network

Instrument: Network of Excellence

Thematic Priority: ICT 6-4.1 – Digital Libraries and Digital Preservation

D32.1 REPORT ON COST PARAMETERS FOR DIGITAL REPOSITORIES

Document identifier:	APARSEN-REP-D32_1-01-1_0
Due Date:	2013-02-28
Submission Date:	2013-02-28
Work package:	WP32
Partners:	BL, CERN, DNB, DPC, ESA, KNAW-DANS, STFC
WP Lead Partner:	BL
Document status	Final
URN	urn:nbn:de:101-20140516227

Abstract: This report provides a high level analysis of published cost models as well as the initial findings of a review of their cost parameters in relation to the trusted digital repositories standard, ISO 16363. The results of a survey carried out in collaboration with WP36 are also presented and reviewed in relation to cost models.

Delivery Type	Report
Author(s)	Kirnn Kaur (KK) - BL, Patricia Herterich (PH) – CERN, Suenje Dallmeier-Tiessen (SD) – CERN, Karlheinz Schmitt (KS) – DNB, Sabine Schrimpf (SS) – DNB, Heiko Tjalsma (HT) - KNAW-DANS, Simon Lambert (SL) – STFC, Sharon McMeekin (SM) - DPC
Approval	Heikki Helin (HH) - CSC
Summary	The purpose of this report is to present our analysis of published cost models. A review of cost parameters for the selected models focused on how cost parameters mapped to the International Standard for Trusted Repositories (ISO 16363) which enabled us to assess the areas of; Organisational Infrastructure, Digital Object Management and Infrastructure and Security Risk Management. The purpose was to assess whether cost models were measuring, through their parameters, the relevant activities for a trusted digital repository. It is important to note that this report provides the preliminary findings from this review which is due to be completed for the next deliverable, D32_2 (M30). We also provide a cost model focussed review of a survey carried out in collaboration with the Business Cases work package (WP_36) in relation to cost models.
Keyword List	Cost models, cost parameters, ISO16363 Audit and certification of trustworthy digital repositories, digital preservation
Availability	<input checked="" type="checkbox"/> Public

Document Status Sheet

Issue	Date	Comment	Author
0_1	11.2.13	DRAFT	KK , PH, KS, HT, SL, SS
0_2	15.2.13	FINAL DRAFT FOR INTERNAL REVIEW	KK, PH, SS
0_3	15.2.13	INTERNAL REVIEW	HH
1_0	27.2.13	FINAL REPORT	KK, PH, SD, KS, SS, SL

Project information

Project acronym:	APARSEN
Project full title:	Alliance for Permanent Access to the Records of Science Network
Proposal/Contract no.:	269977

Project coordinator: Simon Lambert/David Giaretta

Address:	STFC, Rutherford Appleton Laboratory Chilton, Didcot, Oxon OX11 0QX, UK
Phone:	+44 1235 446235
Fax:	+44 1235 446362
Mobile:	+44 (0) 7770326304
E-mail:	simon.lambert@stfc.ac.uk / david.giaretta@stfc.ac.uk

EXECUTIVE SUMMARY

This report provides a high level analysis of published cost models. The analysis features the scope and sources of models; access and availability of tools; and the benefits and challenges faced. Overall, OAIS principles form the basis for all digital preservation cost models surveyed so far. Based on the same reference model, the cost models are then tailored to the needs of the model creator, depending on the data that is available to the institution and the purpose served which makes them difficult to re-use.

The preliminary findings of a review of cost parameters for the selected models, has also been provided. This review focuses on how cost parameters map to the International Standard for Trusted Repositories (ISO 16363) which enabled us to assess the areas of; Organisational Infrastructure, Digital Object Management and Infrastructure and Security Risk Management, in relation to the cost parameters within each model. The purpose was to assess whether cost models were measuring, through their parameters, the relevant activities for a trusted digital repository. The final report is due to be completed for the next deliverable, D32_2 (M30).

We also provide a review of a survey to investigate the level of preparedness of institutions as a prerequisite to ensure economically-sustainable digital preservation (DP) which was carried out in collaboration with the Business Cases work package (WP_36). The general purpose of the survey was to review financial approaches undertaken by research libraries and the review presented in this report is in relation to cost models only.

CONTENT

EXECUTIVE SUMMARY	4
1 INTRODUCTION	7
1.1 APPROACH AND METHODS	7
1.1.1 Analysis of cost models	7
1.1.2 Cost parameters for digital archives	8
1.1.3 Mapping cost parameters against the trusted digital repositories standard	9
1.1.4 Survey	11
1.2 FRAMEWORK FOR COST MODELS FOR DIGITAL PRESERVATION	11
2 FINDINGS FROM ANALYSIS OF COST MODELS	15
2.1 MODELS SURVEYED	15
2.2 SCOPE AND SOURCES OF MODELS	15
2.3 ACCESS AND AVAILABILITY OF TOOLS	17
2.4 BENEFITS AND CHALLENGES	18
2.5 OVERVIEW OF COST MODELS SURVEYED	18
3 COST PARAMETER MAPPINGS – PROGRESS TO DATE	20
3.1 INITIAL FINDINGS OF MAPPING EXERCISE	20
3.1.1 Cost parameters mapped	21
3.1.2 Cost parameters partially mapped – exercise still to be finalised	22
3.1.3 Cost parameters not mapped – require further investigation	22
3.2 GAPS IDENTIFIED TO DATE	23
3.2.1 Organisational Infrastructure	23
3.2.2 Digital Object Management	24
3.2.3 Infrastructure and Security Risk Management	24
3.2.4 Summary of gaps identified to date	25
3.3 AREAS FOR FURTHER INVESTIGATION AND DEVELOPMENT	26
3.3.1 Mapping exercise - next steps	26
3.3.2 Synthesis of cost models – commonalities and differences	26
4 COSTS AND DIGITAL PRESERVATION FINANCING IN PRACTICE – SURVEY RESULTS	28
4.1 SURVEY RESEARCH COMMUNITY	28
4.2 ANALYSIS OF THE FINANCIAL APPROACH	28
4.2.1 Costs and funding	29
4.2.2 Management	29
4.2.3 Reported benefits and added values	30
4.2.4 Controlling costs of DP activities	32
4.2.5 Cost model requirements	33
4.3 SUMMARY OF COST MODEL SURVEY FINDINGS	34
5 CONCLUSIONS AND PROSPECTS	36
5.1 WHERE NEXT FOR COST MODELLING?	36
5.2 CHALLENGES	36
5.3 CONNECTIONS TO COMMON VISION AND VCoE	37
5.4 RELATIONSHIPS WITH NEW CO-ORDINATION ACTION (4C)	37
5.5 NEXT DELIVERABLE – D32.2	37
6 REFERENCES	39
ANNEX 1: SUMMARY OF COST MODELS	41
ANNEX 2: ANALYSIS OF COST MODELS	45
ANNEX 3: ISO16363 – EXTRACT RELEVANT TO MAPPING EXERCISE	46
ANNEX 4: SURVEY QUESTIONS RELEVANT TO COST MODELS	55
ANNEX 5: OVERVIEW OF MAPPING EXERCISE – DRAFT	59

LIST OF FIGURES

Figure 1: Analysis of cost models.....	8
Figure 2: Mapping cost parameters to ISO16363	10
Figure 3: Extract from ISO 16363	11
Figure 4: A general framework for cost models	13
Figure 5: Organisational infrastructure	23
Figure 6: Digital object management.....	24
Figure 7: Infrastructure and security risk management	24
Figure 8: Areas of Digital Preservation, in which decision-makers are involved	30
Figure 9: How do you think your organisation is likely to benefit from digital preservation?	31
Figure 10: Reported added values.....	31
Figure 11: Preservation mandate as proper DP facilitator – all answers shown.....	32
Figure 12: How do you determine the costs of Digital Preservation within your organisation? (N=81)	33
Figure 13: Reasons for using a cost model	33
Figure 14: On what basis would you select a cost model? (N = 73).....	34

LIST OF TABLES

Table 1: Summary of current status	20
Table 2: Summary of gaps identified.....	25
Table 3: Funding sources by organisation type	29

1 INTRODUCTION

Any organisation that spends and receives money needs to have some awareness of its future expenditure and income. Depending on the size and complexity of the organisation, and the requirements on its governance, this might be a simple and informal forecast, or it might involve elaborate estimation and modelling of future events.

Estimating costs or even the precise calculation of the cost of a specific activity is for many responsible persons and organisations a difficult task. This is complicated further if costs of IT services must be taken into account. IT services are usually viewed as critical to the business. The increases in the demands for new technologies and complexities of networked systems have frequently caused the costs of IT services to grow faster than other costs, even if at some level the costs of the basic hardware are falling. As a result, organisations are often unable or unwilling to justify expenditure, to improve services, or to develop new ones. The use of cost models could help to justify such expenditure.

The situation is even more complicated in the case of digital preservation activities. Apart from anything else the focus is primarily on ensuring the long-term nature of the services, requiring long-term funding. To demonstrate that expenditure is well spent, or at least under control, to its respective funders or backers, a repository or archive has to both understand the true cost of providing a digital preservation activity and manage those costs professionally. Especially given that digital preservation activities do not necessarily produce immediate and tangible benefits to an organisation. The understanding of costs will also allow a repository to make informed choices about investments by considering where such investment would generate the greatest impact. Thus there is a demand for approaches to understanding and estimating the costs of preservation, and this motivates the work of APARSEN to survey what has been done and identify strengths, weaknesses and areas where more development is needed.

1.1 APPROACH AND METHODS

The objectives of this work package are to evaluate and test cost models for the preservation of digital objects. Within this field a number of cost models have been published which cover different elements of costs associated with repositories. The first area of work focusses on a review of these published cost models. This review was undertaken to investigate which cost models would be included as part of the cost parameter mapping exercise. This exercise maps cost parameters against the International Standard for the audit and certification of trustworthy digital repositories (ISO 16363) [1]; the purpose being to show how cost model parameters are concentrated or where areas of activity are not included within a particular cost model. Finally, the results of the survey, as undertaken by research libraries, within the Business Cases work package, WP_36, is reviewed in relation to cost models.

1.1.1 Analysis of cost models

The methodology adopted in order to analyse the different elements of costs associated with repositories involved a review of published cost models, blogs, articles, papers and presentations. The approach was narrowed down to focus on specific published cost models for further detailed research. A good source for this exercise was the Open Planets Foundation website pages [2] on cost models, Annex 1 provides details of the models analysed. Once the models had been identified a series of questions was drafted, see Annex 2 for details.

This focussed the analysis into key areas as follows:

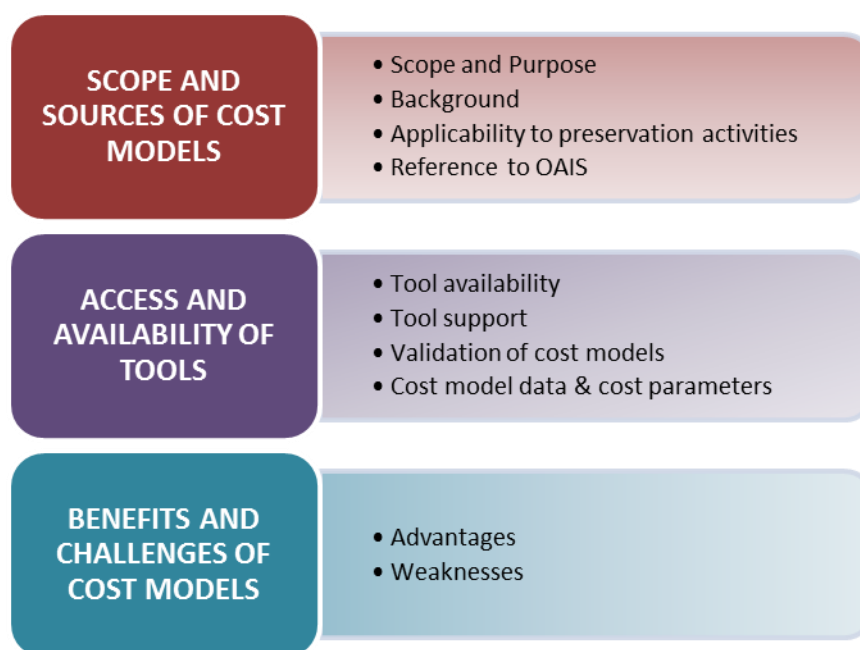


Figure 1: Analysis of cost models

The purpose of this approach was two-fold:

1. To provide a clear structure for the analysis to ensure consistency
2. To provide information as part of the Virtual Centre of Excellence (VCoE) website so that users are able to be directed to relevant cost models for use

In this report we focus on the first, the results of which are provided in Section 2. The second will be reviewed as part of the next deliverable D32_2 due M30.

1.1.2 Cost parameters for digital archives

We recognized that all cost models are different for varying reasons. Initially it seemed reasonable to compare data between each model to see if there was a good fit, but we found that organisations tended to undergo the cumbersome task of tailoring a cost model to their own needs and purposes; or indeed creating their own when those available did not suffice. The International Standard on trusted digital repositories (ISO 16363), which is based on the Open Archival Information System (OAIS) reference model (ISO 14721) [3] was selected to be used as a common denominator between all cost models and provides a benchmark which allows at least a certain level of comparability between the various cost models.

In selecting this standard we had also considered mapping the cost parameters against the OAIS reference model (ISO 14721) which has been adopted as the standard on which to build digital archives. As most cost models are already more or less based on OAIS, with costs apportioned over various functional entities of the reference model, this would have provided clear gaps highlighting where cost models had not included relevant activities within their models. Thus, we decided that mapping cost parameters to the trusted digital repositories standard would achieve this and in addition cover a wider scope of activities, for example, organisational infrastructure and security risk management. In addition to the standard

selected there are two other options which can be used to evaluate whether your digital archive has preserved items properly, as published within APARSEN's Trust brochure [4]; Data Seal of Approval (DSA) [5] and DIN 31644 [6]. DSA provides an online self-assessment tool and a peer review determines whether an organisation will be granted the Seal of Approval. Overall, the sixteen guidelines provide an entry point level of assessment for smaller organisations in the research data field. As for the DIN standard, this is far more extensive and consists of 34 requirements, with a broader scope from institutions such as archives, libraries and museums to those preserving digital information. The DIN standard was published in 2012 and is currently only available in German. As DSA provides a lower level of assessment and the DIN standard is not yet widely available, we selected ISO 16363 for our mapping exercise.

The methodology adopted in analysing cost parameters was to ascertain and assess whether a digital archive's activities in relation to the standard had in fact measured or taken account of all relevant activities through the inclusion and definition of their cost parameters. The aim was to assess whether within this digital landscape there were similarities, gaps or areas for further investigation and development of cost models. This would be achieved by understanding what the current cost models were measuring at cost parameter level. This level of granularity provided valuable insight into where these parameters are defined within the processes and activities of a certified digital archive.

In order for the mapping exercise to be carried out we had to initially extract the cost parameters and their definitions from the cost models under review. The results of the exercise would show us how far these models have been developed since their inception and formation, having been built upon digital preservation principles.

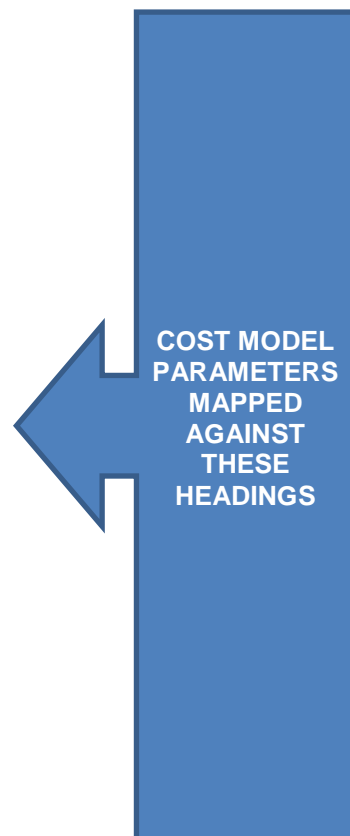
1.1.3 Mapping cost parameters against the trusted digital repositories standard

The International Standard on audit and certification of trustworthy digital repositories is divided into five sections. The first two sections provide the overall document structure and an overview of audit and certification criteria. These two sections are excluded from the mapping exercise.

Sections 3 to 5 provide metrics or measures which allow for the audit and certification of a trusted digital repository, namely; Organisational Infrastructure, Digital Object Management and Infrastructure and Security Risk Management. These sections as well as associated sub-sections are shown in Figure 2.

ISO16363: Audit and certification of trustworthy digital repositories - extract

3. Organisational infrastructure:
 - 3.1. Governance and organisational viability
 - 3.2. Organisational structure and staffing
 - 3.3. Procedural accountability and preservation policy framework
 - 3.4. Financial sustainability
 - 3.5. Contracts, licenses and liabilities
4. Digital Object Management
 - 4.1. Ingest: Acquisition of content
 - 4.2. Ingest: Creation of AIP
 - 4.3. Preservation planning
 - 4.4. AIP preservation
 - 4.5. Information management
 - 4.6. Access management
5. Infrastructure and security risk management
 - 5.1. Technical infrastructure risk management
 - 5.2. Security risk management

**Figure 2: Mapping cost parameters to ISO16363**

There are two levels of mapping: at sub-section level e.g. 3.1, 3.2 or at the level below this e.g. 3.1.1, 3.1.2. For this report we focussed on the mappings at the sub-section level, see Figure 3 for an extract of the standard. For each measure in the level below, the cost parameters will be reviewed further and the results will be provided in the next deliverable.

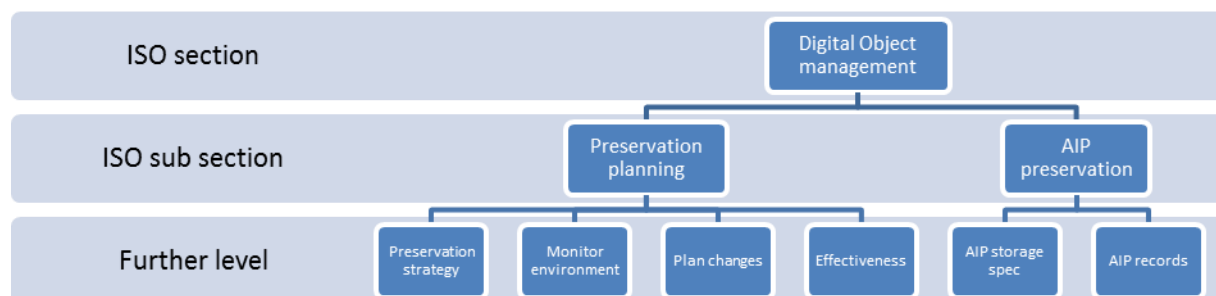


Figure 3: Extract from ISO 16363

A requirement of the exercise was that all parameters should be mapped. Where parameters could be categorised across a number of sections, details were provided in the comments column and re-entered against the relevant heading. Where mappings were implied or ambiguous details were also provided. Annex 3, provides a table of the detailed sections extracted from the standard for the mapping. The evidence column provides examples which were used for the alignment of cost parameters by assessing the activity being undertaken under each section.

1.1.4 Survey

The results of a survey are presented in Section 4, which was conducted in collaboration with the Business Cases work package (WP_36). This survey reflects the level of preparedness of organisations with respect to known costs, financing models and existing management instruments, for example, the use of cost models. We provided input into the survey by preparing questions specifically related to cost models. These questions are provided in Annex 4. The survey results were reviewed and a separate analysis to that presented in D36_1 is provided in this report which is related to issues relevant to cost models. The findings related to the criteria for cost models will help with the further investigation and development work we will be finalising in the next deliverable.

1.2 FRAMEWORK FOR COST MODELS FOR DIGITAL PRESERVATION

Before moving on to examine and compare particular cost models, we clarify what we mean by a cost model in the context of digital preservation. For this purpose, a cost model is taken to be:

A representation of the activity of digital preservation that can be shared, examined and critiqued and whose purpose is to shed light on the costs entailed in the activity of digital preservation.

This definition is deliberately chosen to be broad, for reasons that will become apparent below. Two of its elements are worth further comment. A *representation ...that can be*

shared, examined and critiqued implies some kind of formal and structured approach that is open to study, evaluation and refinement — it is more than just intuition or rule-of-thumb. *To shed light on the costs* implies simply that the purpose is to give some information about costs that presumably was not available before — it does not go so far as to say that a model must produce numerical values in euros or dollars.

It is self-evident that the costs of digital preservation are associated with particular activities or functions. These costs are incurred at all stages of the lifecycle of the digital material that is being preserved, and may include:

- costs of ingesting material, for example of validating it on receipt;
- costs of file format migrations as they become necessary;
- costs of supplementing representation information;
- costs of periodic hardware upgrades and refreshes;
- costs of retrieval and production of Dissemination Information Package;
- costs of disposal.

It is clear that different repositories and archives will be concerned with only a certain subset of all possible costs, depending on their mandate, objectives and the nature of the material that they preserve. For example, an archive of image files is likely to be particularly interested in costs of format migration and not in costs of preserving semantics, as a science data repository would be.

A corollary is that models of costs will tend to be activity-based — that is, they will derive estimates of costs from aggregating the estimated costs of particular activities that are expected to be required for preservation. Activity-based models are not the only type imaginable, however. In a narrow domain with many similar instances of repositories with slightly varying characteristics, it would in principle be possible to produce a purely statistical model, comparing the case in hand to historic cases to produce a cost figure. But even here one would need at least to break down the costs into types (staff, hardware, etc.) and by year for budgeting purposes. A black box model that produced a single figure, no matter how reliable, for the total cost of preservation over a certain number of years would be of little practical use.

Sometimes different vocabulary is used to describe such models. The LIFE model (discussed below) relies on a lifecycle methodology; however this ultimately breaks down into particular activities. Sometimes the term *process model* is used.

During the studies made in APARSEN on extant cost models, it quickly became apparent that, within the broad definition given above, they differ in important ways.

1. **Scope of coverage.** How comprehensive does the model intend to be? Does it aim to cover a wide range of preservation activities or is it very focused on say format migration and validation?
2. **Degree of development.** Some models have undergone extensive development and validation, whereas others are relatively new.
3. **Domain-specific or generic.** A key distinction: is the model applicable to any domain or focussed on one such as earth observation data or scanned images?
4. **Users' responsibility for supplying data.** Does the model itself embed key parameters that drive costs, or is it the users' job to supply these? Contrast a model that can advise the user on staff effort needed for validation of migrated data against one that simply reminds the user that this step is needed and asks them to estimate themselves for their own domain.

It is therefore difficult to compare models systematically, because in some ways they diverge quite profoundly in their aims and bases. In particular the hope of being able to test and compare models has proved elusive.

This divergence caused the APARSEN team to produce a general framework within which specific cost models may be positioned so as to allow high-level comparison in terms of scope and intentions. The framework is outlined here. A key recognition underlying the framework is the distinction, when considering activities required for digital preservation, between *what must be done* and *how much it will cost*. Within the spirit of the definition given at the beginning of this section, a model that does not explicitly deal with monetary figures, but only with activities that will ultimately cost money, is a cost model nonetheless: it is shedding light on where there will be costs, and it may well be useful in some circumstances and to some audiences, though no figure of euros or dollars drops out at the end.

The framework is shown diagrammatically in Figure 4:

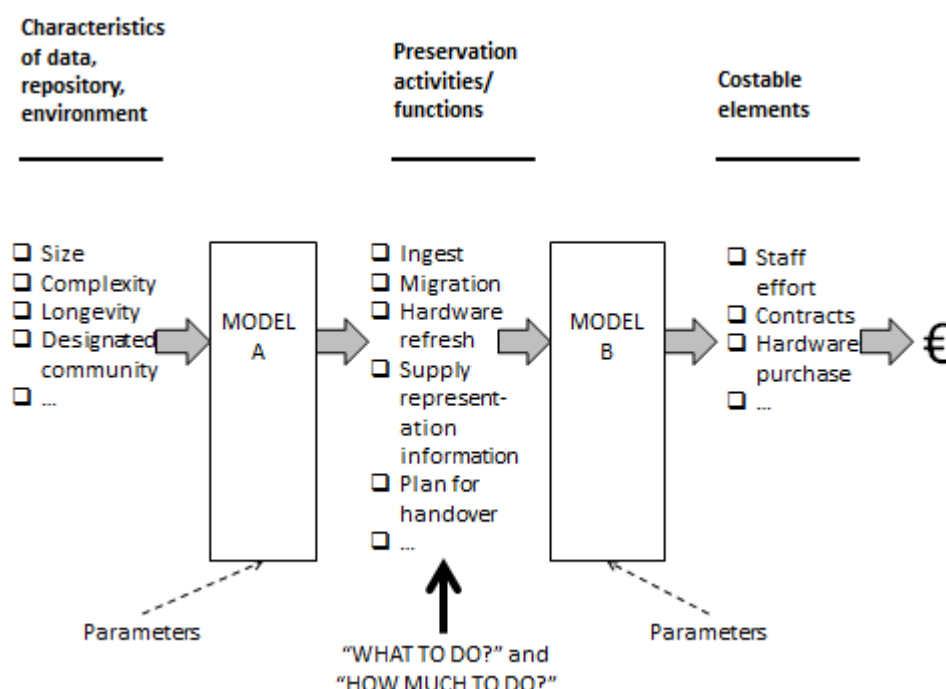


Figure 4: A general framework for cost models

The two sub-models alluded to above are identified as Model A and Model B. Inputs and outputs are indicated by broad arrows from left to right; the lists in each column are indicative of the factors taken into account (as input) or generated (as output) by the models. A model will perform some operations on the input (not necessarily numerical) to produce information about the outputs that is relevant to costs. Thus for example Model A might perform some calculations on the size of the repository (amount of data) and expected longevity to yield a statement about how often format migrations or checks of representation information will be required. These calculations might be based on purely general principles, or they might be founded in experience of repositories in the same domain.

Model B would then convert this information on ‘what to do’ and ‘how much to do’ into costable elements, that is, anything that directly incurs a cost, be it in staff time, training, purchase of hardware or whatever. It is not necessary that a value of euros or dollars drops

out; perhaps the model stops at an estimate of staff-months, leaving it to the user to convert that to salary costs plus overheads.

It is however useful to acknowledge and represent the distinction between models that do lead to some quantification in their output, and so each of the two sub-models may work with a set of parameters. This rather overused word is intended here to stand for values, often but not necessarily numerical, that enable quantification to be introduced into the outputs based on values of the inputs. A simple example of a parameter from another context would be a mathematical model of a swinging pendulum: reasoning only in the abstract, it is possible to draw conclusions about the relationship between the period of oscillation and the length of the cord; the introduction of a definition for the parameter g , the acceleration due to gravity, enables a precise value to be calculated.

Simply by looking at the diagram of the framework, we can see that several different subsets may be presented as cost models, and they may all have their uses. We can thus differentiate between the scope and intent of models within this framework.

1. **Model A without parameters.** Typically such models will be at an early stage of development, as they simply relate characteristics of the data repository and environment to the preservation activities required, without proceeding to translate those into costable elements. The absence of parameters means that only enumerative or indicative statements can be made.
2. **Model A with parameters.** Allows more precise conclusions about ‘how much’ of particular activities are needed.
3. **Model B without parameters.** The starting point is that the user knows what to do in terms of preservation activities; the model will lead to conclusions about the costable elements that are needed to implement those activities.
4. **Model B with parameters.** Allows more precise conclusions to quantify the costable elements.
5. **Models A and B without parameters.** Offers a complete guidance on what must be taken into account when estimating costs, but without quantification.
6. **Models A and B with parameters.** Provides the final and most complete type of model.

This framework does not differentiate on the ability of models to deal with uncertainty, for example by presenting an envelope of values. That is assumed to be part of the sub-models A and B, if it exists.

It will be useful to keep this framework in mind when reading the rest of this report.

2 FINDINGS FROM ANALYSIS OF COST MODELS

2.1 MODELS SURVEYED

To start analysing cost models in line with the approach given in Section 1.1.1, we had a closer look at the list of published cost models provided by the Open Planets Foundation [1]. From the models mentioned there, we analysed the following in detail:

- Lifecycle Information for E-Literature, LIFE3 [7]
- Cost Model for Digital Preservation, CMDP [8]
- Keeping Research Data Safe 1 and 2, KRDS [9]
- PrestoPRIME cost model for digital storage [10]
- Cost Estimation Toolkit, CET [11]
- Digital Preservation for libraries, DP4lib [12]
- Activity based costing of archiving datasets, DANS [13]
- Costing cloud based preservation systems, ENSURE [14]

Furthermore, we looked at the OECD model, and models available to partners engaged in the APARSEN project. Thus, we could include the ISIS model (STFC) [15]. At the point when the work was started, the Cost Model for Small Scale Automated Digital Preservation Archives by Strodl and Rauber [16] was already part of the list. As it is an attempt to implement the LIFE3 model for an automated preservation service, we decided not to include it but to focus on the LIFE3 model. Although this cost model may be reviewed for the second deliverable within this work package. The models were analysed based on the questions given in Annex 2.

2.2 SCOPE AND SOURCES OF MODELS

Most of the analysed models aim to be generic and usable by all kinds of institutions that preserve objects in digital form. Even if they were developed based on institutions' experiences with research data like the DANS cost model or KRDS, they claim that with some adjustments, they could be applicable more widely. However, the PrestoPRIME model covers only the long term preservation costs of audio visual material, the CET model can be used by other institutions handling scientific data and the ISIS model was specifically created for STFC's ISIS facility and is tied to the file formats, software, etc. used there.

It can be said that cost models usually arise from projects, either about digital preservation in general or projects specifically related to the costs of digital preservation, or are developed by institutions with specific goals in mind. Depending on the project or the developing institution, the models tend to be tailored specifically for their needs and therefore, the scope can be too narrow. The projects ENSURE (2007-2013) and PrestoPRIME (2009-2012), European Commission funded, under the 7th Framework Programme, developed cost models as part of the task to provide solutions for long term preservation. ENSURE focuses on companies and organisations in the healthcare, clinical trials and financial sectors and the cost model is developed as a component of a cost-benefit analysis. PrestoPRIME worked on research related to the development of practical solutions for the long-term preservation of digital media objects, programmes and collections.

Several cost models analysed resulted from national projects. The Lifecycle Information for E-Literature (LIFE3) project was a collaboration between the University College London and the British Library and was funded by JISC [17] and RIN [18]. They developed a generic preservation model that was extended through the project phases and now covers all kinds of

materials. A similar project called Cost Model for Digital Preservation (CMDP) was started in Denmark through cooperation between the Danish National Archives and the Royal Danish Library and was funded by the Danish Ministry of Culture. Based on their own case studies, they developed a cost model for the cultural heritage sector “to increase cost effectiveness of digital preservation activities and to provide a basis for comparing and estimating future cost requirements for digital preservation” [19]. Another JISC funded project was keeping Research Data Safe (KRDS) which, carried out by Charles Beagrie Limited, collected and analysed cost drivers and benefits of best practice examples for research data preservation from UK Higher Education Institutions.

Some institutions considered the available cost models as not feasible for their needs and developed their own models. For a cost model for their ISIS facility, STFC worked together with an MSc student at Cranfield University. The Data Archiving and Networking Services (DANS) in the Netherlands are developing their own model that allows accountability towards funders. The German National Library (DNB) and partners are developing a cost-by-service model (DP4lib) for long term preservation with the help of an economist. The aerospace service company Stinger Ghaffarian Technologies Inc. (SGT) has developed the so called Cost Estimation Toolkit (CET) for NASA as part of a Levels of Service and Cost Estimation Study since 2002.

Cost models are usually developed for the purpose of understanding and estimating costs for the future. The DANS and the DP4lib model are developed for budgeting and accounting. The KRDS model provides two application possibilities. It can be used for life cycle costing, but when there is no cost data available, it provides approaches for institutions to analyse the benefits, added value and impact assessments of research data management and preservation activities.

As most of the cost models aim to be generic, they are applicable for either the digital preservation life cycle as a whole (depending on how far they are developed already) or only certain activities. Pre-archive, archive and access are covered by all of the cost models, and are usually split into more detailed activities. The CMDP and the PrestoPRIME models include migration, whereas the DP4lib model does not explicitly include this area within the services analysed. The LIFE3 model can be used to measure costs for outsourcing activities to a third party. Other models might be feasible for this kind of costing as well, although it is not explicitly mentioned. CMDP for example was partly developed based on data from a contract with a third party.

The OAIS model forms the basis of many cost models for digital preservation. All of the model creators reviewed OAIS before starting to develop a new cost model. When appropriate, the models follow the structure of OAIS and use OAIS definitions and terms. The PrestoPRIME project reviewed OAIS and came to the conclusion that it is not specific enough to monitor objects and storage systems. Therefore, there is no obvious influence of OAIS on this model.

Besides reviewing the OAIS model, cost model owners put their work in the context of other cost models and discuss different approaches. KRDS for example studied the LIFE and the CET models in quite a lot of detail. The LIFE3 model refers to KRDS and the CMDP model and the ENSURE and PrestoPRIME models are analysed before starting to develop their own solutions for costing preservation. The DP4lib model took an approach that is based on best practice approaches and located more in the economics and business sector.

Apart from the ENSURE cost model, which is based on the assumption of a cloud storage solution, all the other analysed cost models either do not specify the storage solution or assume that it is hardware based and either the own institution or an external provider takes care of it.

Overall, in terms of scope and sources, cost models arise from projects on digital preservation in general or initiatives on cost models in particular and mostly aim to cover the digital preservation life cycle as a whole with all referring to the OAIS reference model.

2.3 ACCESS AND AVAILABILITY OF TOOLS

Some of the analysed models such as the DANS model, the ENSURE model, are still under development and therefore, there are no tools available as yet. Information on the models is published in several papers though and details are available in Annex 1.

For the DP4lib cost model, documentation and a calculation spread sheet are available (see Annex 1). The DP4lib cost model and the respective spread sheet tool reflect the current state of the DP4lib activities within the German National Library. Because development is still ongoing, the DP4lib cost model cannot be considered as complete.

Most of the models produced a spread sheet that can be used with an institutions own data. The CMDP tool is available in version 4 [20], which is unfinished. The project ran out of funding and it is unclear if the model will be further developed. Thus, except for the literature available, no further support for model re-use can be provided.

The LIFE3 project produced a spread sheet [21], too, and in addition, a web tool is available [22]. The web tool is still in prototype stage and hosted at the Humanities Advanced Technology and Information Institute (HATII) at the University of Glasgow which will provide support in addition to the documentation already available.

The Cost Estimation Toolkit [11], CET, is based on a spread sheet as well. As it has been under development for about 10 years now, the model and the toolkit are one of the most advanced models with tools available. Though, the most current version of the open source software is version 2.4 from 2008. There is no obvious support available for the model, but an e-mail address for bug reports and a general contact is provided.

PrestoPRIME offers several tools that include costs. The browser based storage planning tool [23] is a long-term planning tool which can support decision making on which storage strategy to use. Furthermore, they provide an interactive simulation tool which enables detailed investigation and supports more operational levels of decision making. They are currently developing more digital preservation tools and a platform that integrates them and thus covers the whole of the “plan-do-check-act” cycle. PrestoPRIME offers information on how to use their tools in guidelines, “how to” documents, FAQs and a contact mail address for feedback and support is provided.

KRDS offers the Activity Model which provides and defines activities of the digital preservation lifecycle in order to allocate costs. Furthermore, a Benefit Analysis Toolkit is provided. It consists of two worksheets – the KRDS Benefits Framework and the Value-chain and Benefits Impact tool – as well as a Toolkit Guide [9]. In addition to the tools, there are extensive guides available as well as contact information for project partnerships or consultancy.

Not every model is available though. The ISIS model is not published as it is not intended for re-use and the data used to develop it is confidential.

The models were tested with the data from the case studies they were built on. This data is usually the data arising from preservation projects which the institutions developing the cost model are involved in. None of the models were validated by an external institution or tested with other data than those that were used to develop it.

In general, cost models for digital preservation try to provide a tool that could be used by others. However, re-use and testing with external data rarely happens.

2.4 BENEFITS AND CHALLENGES

When it comes to benefits, it is not as easy to group the cost models as their strengths are very different. The CET model for example is fully functional and can be further adapted if needed. It is hosted on a web server and users can improve it by giving feedback. In addition, it can be used to compare cost estimates for alternatives based on workload or levels of service. The CET toolkit helps to assess cost drivers. LIFE3 is also simple and immediate to use as the template offers automatic completion for data that is not available within an institution. It furthermore provides a checklist to make sure that all costs are covered. If needed, it can be refined and edited according to the institution's needs. Additionally, LIFE3 provides scenarios users can go through. The ISIS model also is scenario based. It gives "best case" as well as "worst case" scenarios. ISIS provides full coverage of costs for the whole digital preservation life cycle.

The DP4lib model gives profit as it is universally applicable and can be individualised according to an institution's needs. Also the ENSURE model aims to be generic. It is the only model applicable to cost cloud storage preservation. In the future, it shall also allow the comparison of cost techniques and provide guidance for the handling of uncertainties. As the model is still under development, this is not defined precisely as yet.

When completed, the DANS model will be able to indicate the most labour intensive and therefore costly activities of digital preservation. Furthermore, it can give guidance on how to distribute overhead costs. The CMDP model is suitable to identify efficiencies in digital preservation costing. PrestoPRIME allows a detailed analysis of costs and factors influencing the digital preservation of audio-visual material. It enables users to see the effects of changing the storage strategy and calculates costs on the risk of loss of assets.

The KRDS model has a very straightforward benefit orientated approach. The model allows cost allocation at any level for different purposes. Two versions of the models are available; an easy to use, 'lite' version and a more detailed version. KRDS tools can be used separately. The benefit worksheet can be used to develop an organisation's own list of possible benefits with the value and impact tool providing help to identify quantitative and qualitative metrics and indicators.

Unlike the benefits, which are very diverse, the analysed cost models face the same challenge: it is not clear how usable they are for other cases and institutions than the ones they were developed for as they are not tested by other institutions or with other cost data. They claim to be generic and usable, but cost data usually has to be adjusted to be able to use the models or the models have to be tailored, which are both very time consuming tasks. The CET, LIFE3 and DP4lib models try to estimate future costs. These estimates have to be treated with care as they are very likely to be inaccurate and will need further clarification and updates.

CMDP and PrestoPRIME furthermore are only applicable to certain materials, which limits the number of institutions that could use the models.

The analysed cost models have different strengths and share the same weakness: they are not tested with external data and are difficult to re-use.

2.5 OVERVIEW OF COST MODELS SURVEYED

OAIS is the starting point for all digital preservation cost models surveyed as part of the work in this work package so far. Based on the same reference model, the cost models are then tailored to the needs of the model creator, depending on the data that is available to the institution and the purpose it shall serve. This is one reason that makes published cost models difficult to re-use and leads to even more tailored solutions as the existing ones cannot be easily applied by others.

In the 4C Project [24], which started in February 2013, one of the aims is to try to find the common ground between these models and, building upon that, create a universal exchange mechanism between the different models.

3 COST PARAMETER MAPPINGS – PROGRESS TO DATE

This exercise, although not fully completed as yet, was carried out for the cost models listed in Table 1. A summary of these cost models is provided in Annex 1.

The mapping exercise is still to be carried out for the ENSURE and ISIS cost models and we may add further models to the review if found to be relevant. These results will be provided in the next deliverable.

Table 1: Summary of current status

Cost Model	Mapping Status	Further Work
CET	To be reviewed	To check against detailed technical guidelines
CMDP	Part completed	To complete cost parameter mapping
DANS	Completed	To be reviewed
DP4lib	Completed	To be reviewed
KRDS	Part completed	To complete cost parameter mapping
LIFE3	Completed	To be reviewed
OECD	Found to be not relevant	n/a
PrestoPRIME	Part completed	To complete cost parameter mapping

Due to the status of this work the results presented should be treated as preliminary findings until further review and the availability of a full analysis in the next deliverable.

3.1 INITIAL FINDINGS OF MAPPING EXERCISE

In starting this exercise the presumption was that all cost parameters would match, even partly, to certain categories of the trusted repositories standard. Gaps would be identified as a result of the mapping task which will be analysed further during the investigation and development phase of this work package. However, in some instances the cost parameters were not clearly defined and more significantly these parameters did not easily align to the standard's defined areas.

It was found that there were certain areas within the standard which could not realistically be measured and defined as a cost parameter; a good example is staff roles and responsibilities. Although this can be easily audited it may not be reasonable to expect a cost parameter to measure this activity within the framework of a cost model.

It became evident over the course of the analysis and mapping exercise that involving cost model owners was an effective way to assess and investigate these models in detail. Where cost model owners were not partners within the project it was difficult to make accurate assessments and obtain guidance when needed. An effective way of developing cost models going forward would be to involve the creators and owners of cost models.

This report provides the initial findings from which we have drawn preliminary conclusions which will help to plan out the remaining work to be undertaken over the coming months. An overview of the mapping exercise is provided in Annex 5.

3.1.1 Cost parameters mapped

The mapping exercise to date proved to be effective in that all cost parameters were mapped (although some partially) for the following cost models:

- **DANS** – it was difficult to apportion the costs across some sections of the standard as some parameters only partially mapped and others not at all, especially for Organisational Infrastructure. Approximately half of the standard headings were not covered and do not seem to fit into the structure of the cost model. These areas may not be relevant to this cost model and will be reviewed further.
- **DP4lib** – the mapping was not easy and some sections of the standard have no cost parameters mapped to them. However, the main categories are covered except for Preservation Planning and Security Risk Management, although Technical Risk Management is covered. A detailed specification of the underlying services, actions and processes will be undertaken and cost parameters are currently in development for Security Risk Management. Overall, the development of the DP4lib cost model is still far from complete with key gaps and expansion options in two areas. Firstly, in some areas the missing measurement methods should be mentioned. Currently, estimation methods predominate over concrete measurement procedures. This is not only due to the lack of technical realisation, but mainly because of the lack of cost-benefit considerations. The benefits of concrete measures of costs must justify the effort to implement the methods of measurement but currently, the cost of implementation far outweighs the benefits. Since a cost allocation is sought, specific measurement methods need to be implemented. The second major gap in the model is that it does not cover all DP activities, particularly Preservation Planning. A cost model can and may only cover those activities that have been implemented within an organisation and those which are of relevance to the model. Within DP4lib services no activities were implemented. The same applies to the gap in the Security Risk Management. Extensions of these functionalities, within the cost model, are planned in the coming years. Overall, the DP4lib model is recommended for most areas of the standard.
- **LIFE3** – The cost parameters were difficult to map against the standard as it is very detailed and there are a number of cost parameters. Overall, the cost parameters mapped well against the standard. A few areas were either unclear or not covered, for example, Financial Sustainability; however, this area may be included within project management duties. The other areas which do not seem to be fully covered are Preservation Planning, although an automatic calculation is included and Security Risk Management, although disaster recovery is measured. Assumptions have been made where the detail is not available within the parameter definition but seems to cover the ‘activity’ as stated within the standard. The cost parameters have been assigned mainly at lifecycle stage level. Where a direct map was possible, a more precise mapping against the parameter lifecycle elements and sub-elements has been made, this was generally where a detailed cost parameter definition was available. The only area not covered by the model is Governance and Organisational Viability, with three other areas requiring further investigation. The LIFE3 model seems to provide the best coverage of the standard and is generally recommended for all areas.

3.1.2 Cost parameters partially mapped – exercise still to be finalised

A further three cost models have been partially mapped with the exercise still to be finalised. Although a requirement of the exercise was to map all cost parameters or at least match them as closely as possible to the most relevant category, a number of parameters remain unmatched and require further investigation:

- **CMDP** - A lot of cost parameters could be mapped whilst others could not as they were quite detailed when compared to the standard, for example, those covering migration costs. Parameters that partly match the standard are mapped to the nearest relevant heading. The cost parameters not mapped, for example, relating to storage system properties and managing storage hierarchy, are very complex and could be split into several sub-parameters as they include software, technical and staff costs. In terms of the standard, the areas where no cost parameters have been mapped are Governance and Organisational Viability, Organisational Structure and Staffing, Contracts, Licenses and Liabilities and Financial Sustainability. Also, some cost parameters are not sufficiently defined to complete this exercise and so these parameters are not mapped at this stage, for example, develop packaging designs, migration plans and storage system properties. Similar parameters are used in subcategories but the CMDP model does not specifically cover overall policies and staff costs (activity based), from the material that is available so far and it seems unlikely that these areas will be covered by the model in the future. However, it should be noted that this cost model is still under development and there may be other areas which are in fact covered in the future. The CMDP model is recommended for costs in the fields of Digital Object Management and Infrastructure and Security Risk Management.
- **KRDS** - Most of the parameters could be mapped although a dozen still remain to be mapped. Therefore, some areas are not currently covered and require further analysis. The areas not covered by the cost model as yet are, Ingest: Creation of AIP, AIP Preservation and Information Management as well as Governance and Organisational Viability and Security Risk Management although these last two may be covered by the benefit orientated tool. The KRDS model covers a lot of areas of the standard but none of them overly extensively. Therefore, KRDS might be a good model to get a general overview without too much of a detailed insight.
- **PrestoPRIME** - Most of the parameters could be mapped, although some matches were the best estimate, others have not been matched as yet. As the PrestoPRIME cost parameter set is rather small there are a number of areas within the standard which are not covered by the model. These are Governance and Organisational Viability, Organisational Structure and Staffing, Contracts, Licenses and Liabilities, Ingest: Acquisition of Content, Preservation Planning, AIP Preservation, and Information Management. Some other sections are only covered by one cost parameter which again might be due to the rather small set of cost parameters provided by the model. A number of cost parameters have not been mapped and so require further investigation. The PrestoPRIME cost model focuses on costs for hardware and software; staff costs are not included or only indirectly in terms of costs per GB. This model can be recommended for calculating the technically-based costs for archiving audio-visual material.

3.1.3 Cost parameters not mapped – require further investigation

For the following cost models the exercise proved to be ineffective and requires further investigation in the next phase of this work package:

- **CET** - It was not possible to perform the mapping exercise, as the cost parameters listed do not correspond at all to the metrics of the standard. The parameters listed are components of the overall cost of preservation, but are expressed in a way that does not allow matching to the metrics. For example: management staff (FTE), system purchase cost, general supplies cost. These cost elements are what would be expected as the output of a model, rather than parameters used as input or for calibration; they do not correspond directly to preservation-related activities. An initial review of this model has found that the information required to input to the tool requires pre-work at a more granular level. However, relevant cost parameters may be available within the technical guideline documents published and this will be investigated further.
- **OECD** - Mapping the OECD Standard Cost Model parameters to these standard headings was not relevant to this exercise as the scope of the model is not digital preservation. The cost model determines administrative costs and provides transparent measures which may be of use when using another cost model.

3.2 GAPS IDENTIFIED TO DATE

To date the following gaps have been identified when mapping cost models to the standard on audit and certification of trusted digital repositories. However, as the mappings are only partially completed for some of the cost models, these gaps should be treated as preliminary findings. Further review and investigation will complete the results and will be provided in the next deliverable.

Taking the standard as the starting point, each of its three sections are reviewed below. It should be noted that of the six cost models mapped, three still contain parameters which have not been fully mapped as yet. Those fully mapped are, DANS, DP4lib and LIFE3.

3.2.1 Organisational Infrastructure

Cost parameter coverage within this section was inconsistent across the six cost models.

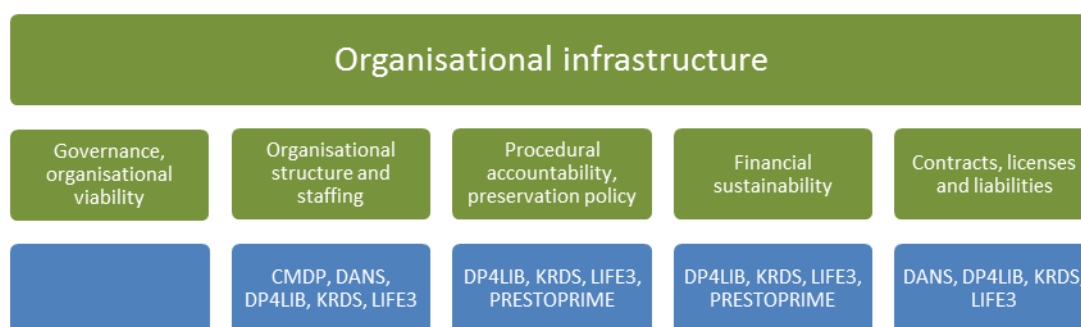


Figure 5: Organisational infrastructure

The Governance and Organisational Viability section consisting of mission statement, preservation strategic plan and collection policy, was not covered by any of the cost models. Although LIFE3 does provide references to repository administration and management, the parameters are not explicit and so the cost model has not been mapped against this section. The reason that cost parameters cannot be provided for this area could be that measurable

parameters are not able to be provided for this type of activity. Overall, LIFE3 and DP4lib provide the most breadth of coverage in this section. The CMDP cost model has the least parameters in this section but it should be noted that this model is still under development. Organisational Structure and Staffing provides the best coverage of cost parameters across all models.

3.2.2 Digital Object Management

All six cost models have some level of cost parameter coverage in this section.

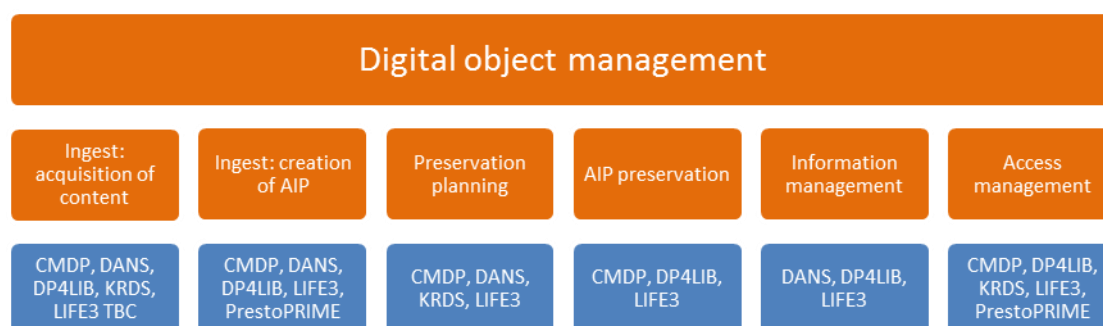


Figure 6: Digital object management

LIFE3 has the best coverage with parameters across all areas, with DP4lib across all areas except for Preservation Planning. CMDP also has good coverage across all areas except for Information Management, although not all parameters have been mapped and this cost model is still under development. DANS has coverage except for in two areas, AIP Preservation and Access Management. Although KRDS and PrestoPRIME show the least coverage not all cost parameters are currently mapped and this may change when the review is completed. Overall both Ingest processes and Access Management have the best coverage across the cost models.

3.2.3 Infrastructure and Security Risk Management

This section seems to have good, although limited, coverage across the six cost models mapped.

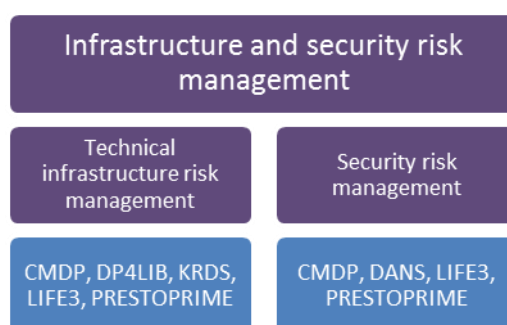


Figure 7: Infrastructure and security risk management

Infrastructure and Security Risk Management is covered well with all cost models having some cost parameters within this section of the standard. CMDP, LIFE3 and PrestoPRIME have parameters across both areas. Cost parameters within Security Risk Management are

currently under development within the DP4lib model which will ensure parameters across both areas. Although Security Risk Management is covered by the CMDP, DANS, LIFE3 and PrestoPRIME cost models, the coverage is limited. Notably, DANS and KRDS only have parameters across one of the areas and although not all cost parameters have been mapped for the KRDS model there do not seem to be any relevant parameters in this area.

3.2.4 Summary of gaps identified to date

Table 2 provides a summary of the gaps identified within cost models against the standard on trustworthy digital repositories and clearly shows that Governance and Organisational Viability is not covered by any cost model. This may be due to the fact that this area is either difficult to measure, in terms of a cost parameter, within the framework of a cost model; or that this area is outside of the scope of the model in question. These gaps will be reviewed further in the next stages of this work package.

Table 2: Summary of gaps identified

ISO HEADING	Organisational Infrastructure				Digital Object Management				Infrastructure and security risk management																	
ISO Sub-heading	Governance, organisational viability		Organisational structure and staffing		Procedural accountability, preservation policy		Financial sustainability		Contracts, licenses and liabilities		Ingest: acquisition of content		Ingest: creation of AIP		Preservation planning		AIP preservation		Information management		Access management		Technical infrastructure risk management		Security risk management	
Cost Model																										
CMDP			X								X		X		X		X				X		X		X	
DANS			X						X		X		X		X				X						X	
DP4lib			X		X		X		X		X		X				X		X		X		X			
KRDS			X		X		X		X		X				X						X		X			
LIFE3			X		X		X		X		X		X		X		X		X		X		X		X	
PrestoPRIME					X		X						X								X		X		X	

3.3 AREAS FOR FURTHER INVESTIGATION AND DEVELOPMENT

3.3.1 Mapping exercise - next steps

The cost models chosen for the mapping exercise (see Table 1) will be reviewed further on three levels in the next phase of WP_32; firstly, all mappings will be reviewed to the level below sub-section within the standard; secondly, research will be carried out to see whether relevant cost parameter information is available to complete the mapping exercise. Finally, an assessment will be made as to the legitimacy of carrying out the mapping to the complete standard. It may be that this standard is of limited relevance to some cost models given their scope and purpose. There may be some specific areas within the standard which we may be able to exclude from the next stage of the mapping exercise e.g. staff roles and responsibilities – quantifying these resources within a cost model as a cost parameter in relation to this may not be relevant.

Where parameter definitions are not clear we could do further research to see whether published documentation is available which provides the cost parameter detail needed to complete the mapping exercise. This may be the case for the CET model as the inputs to the available tool were used and it was not possible to perform this exercise; in this case it may be that the technical guidelines provide the detailed parameter definitions which will allow for the mapping to be completed. Otherwise, we would recommend that where further development is possible these definitions be clarified and be made more ‘meaningful’ to the potential users and stakeholders of these cost models.

Some models are still under development – for these the focus should be to assess whether it would be relevant to cover costs through the inclusion of new parameters within the areas not mapped against the standard. This would ensure that digital archives can be fully costed for all relevant activities. Working with cost model owners would facilitate this process but is out of the scope of this project.

3.3.2 Synthesis of cost models – commonalities and differences

Our preliminary findings have raised some questions which will hopefully be answered once the mapping exercise and further investigations have been completed.

In terms of similarities between the models, it may be that the same activities or processes are being measured in a number of different ways. The cost parameter coverage across the standard, once all parameters have been mapped, will show the common areas between the cost models. It may also be worth reviewing whether the cost parameters, although mapped to similar areas of the standard have varying definitions, thereby measuring a different aspect of cost within the area in question. We may be able to ascertain best practice for certain activities or processes. For those cost models still under development the findings could provide guidance on the areas which should be focused on in the future.

For the gaps identified, with the main area being Governance and Organisational Viability, this provides areas for further investigation and development. It may be that there are legitimate reasons for these gaps, in which case we could limit the scope of the mapping against the standard for trusted digital repositories and exclude areas which realistically cannot be measured as a cost parameter i.e. areas which are not quantifiable as cost parameters. There may be areas within the standard which relate to qualitative rather than quantitative areas or measures. We may be able to suggest cost parameters which are within the scope of the cost model when considering it in the context of a certified trusted digital repository.

We aim to provide answers to these questions over the continuing work in this project which will be reported in the next deliverable.

4 COSTS AND DIGITAL PRESERVATION FINANCING IN PRACTICE – SURVEY RESULTS

This section represents results of a survey which was carried out in conjunction with the Business Cases work package (WP_36). The general purpose of the common survey was to investigate the level of preparedness of institutions as a prerequisite to ensure economically-sustainable digital preservation (DP). For this a web-based questionnaire was created covering issues from five areas:

- The organisation profile
- Importance, values and benefits
- Current preservation activities
- Financial approaches
- Future involvement and collaborations

In this section, the analysis is concentrated on the area of financial approaches which provides the survey results which are of most relevance to the development and use of cost models in practice. This review is independent to that carried out in the deliverable, D36_1. For a comprehensive overview, we refer you to the D36_1 Business Preparedness Report.

4.1 SURVEY RESEARCH COMMUNITY

Given that APARSEN is focused on DP challenges for the records of science, we focus on the issue of economic sustainability in the context of Research Libraries. For this reason the LIBER community of Research Libraries was selected as one of the greatest representatives of this area in Europe. 101 organisations from the currently 455 members of this community participated in the survey. Due to this large sample, the results of the survey can be considered as representative for the entire LIBER community. The detailed composition of the obtained sample is available in the report, D36_1.

4.2 ANALYSIS OF THE FINANCIAL APPROACH

Financial management provides the main stewardship of monetary resources for an organisation. It supports the organisation in planning and executing its business and ensures consistent application of principles and processes throughout in order to achieve economic sustainability.

Within an organization this is usually achieved through three main processes:

- Budgeting
- Accounting
- Charging

For each of these processes, the existence of experienced staff and the use of tried and tested processes are necessary. The use of cost models would facilitate organisations in the financial management and planning of such processes, therefore the availability of such tools are key to the understanding and management of DP activities for any organisation.

4.2.1 Costs and funding

To justify the required funding to donors or sponsors, or to prove that an organisation has managed resources efficiently within the provisions of its funding, a precise knowledge of the costs and related cost elements are required. Providing cases for funding opportunities along with the possible benefits of DP activities are a key element of obtaining such resources. The use of well tested cost models would serve to address these issues with the benefits element being addressed by the use of the KRDS model which provides a benefits analysis tool which would be of relevance in this situation.

Generally, the individual cost elements used to measure DP activities could be grouped into a few main categories. One possible category is the grouping of all personnel costs into the group: *Staff*. This grouping is often seen in cost models, as these costs are relatively easy to determine for an organization overall although apportioning these costs over specific DP activities may be extremely time consuming and difficult. From the survey it was found that 25% of organisations indicated that there was no information on annual budgets available. If that is indeed true, as opposed to the option that the person responding to the survey was simply not aware of it, this would mean that these organisations seem to have no financial control over their activities in which case the need for cost models could be seen as vital.

When considering the respective sources of funding, see Table 3, it was found that most organisations have a mixed funding model in place. Internal institutional funding represents the main funding source in 37% of all organisations. Government funding is also a main source of sustainable funding, given that for research institutes, it makes up a third of their total funding. This seems to be the case across organisations in Eastern Europe, where the lowest number of organisations receiving government funding can be found. Another third of the available budget comes from project grants and awards. Sponsorships and revenues seem to be of least importance.

In all of these instances, and especially in the case of project funding, it may be particularly important to provide cost predictions and there are a number of cost models which can be used to make these predictions specifically for digital preservation activities.

Table 3: Funding sources by organisation type

	Institutional budget	Regional- Governmental Funding	Projects funding	Sponsorship	Revenues	Others
National Library	36%	18%	31%	13%	3%	0%
Research Institute	29%	35%	22%	6%	6%	0%
Government University Library	75%	0%	13%	0%	0%	13%
State Library	40%	26%	23%	5%	5%	2%
Other	33%	27%	27%	7%	7%	0%
Total	27%	9%	36%	0%	0%	27%
	37%	23%	26%	7%	4%	3,4%

4.2.2 Management

It is crucial for the success of the introduction of DP activities to convince decision-makers of the necessity and usefulness of such activities. Only if decision-makers fully support all

necessary activities and fully understand the costs for all DP processes, can the sustainability of DP processes be guaranteed. This is supported by the survey results shown in Figure 8.

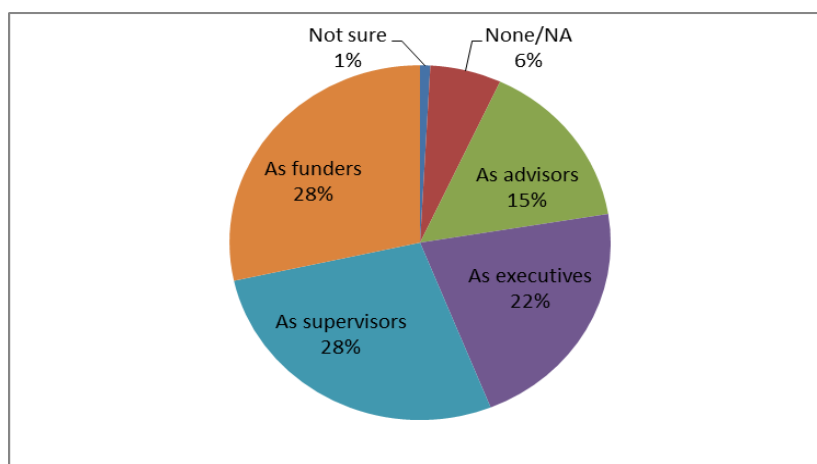


Figure 8: Areas of Digital Preservation, in which decision-makers are involved

The majority of respondents reported that the decision makers are mostly involved in managerial activities as supervisors (28%), executives (22%) or advisors (24%); however, interestingly 28% of the respondents reported that decision-makers are also financial funders. Especially in the role of financial funders, the benefits and expected added values of DP must be convincing. This group must be involved in the future development of cost models and are key to the future of the practical implementation of such models. As a result it would be advantageous to include this group on a collaborative level in any future cost model development activity.

4.2.3 Reported benefits and added values

Within organisations the understanding of DP activities as well as the likely benefits may help us to envisage the future requirements of cost models as well as their relevance. In the context of the survey, LIBER members were asked how their organisations were likely to benefit from digital preservation.

The reported perceived benefits were:

- Increased use of content as a result of better availability and fundability
- Ensuring research results integrity
- Improve organisation and staff reputation (visibility, citations, recognition of being at the forefront of digital preservation etc.)

It is interesting that the benefit of ‘Long term budget saving’ was ranked only in fourth place. The direct financial aspect seems to be of secondary importance in the benefit analysis. Here the KRDS cost model provides toolkits which may be useful for organisations to carry out benefits analysis and value as well as impact assessments for proposals and evaluation and planning activities.

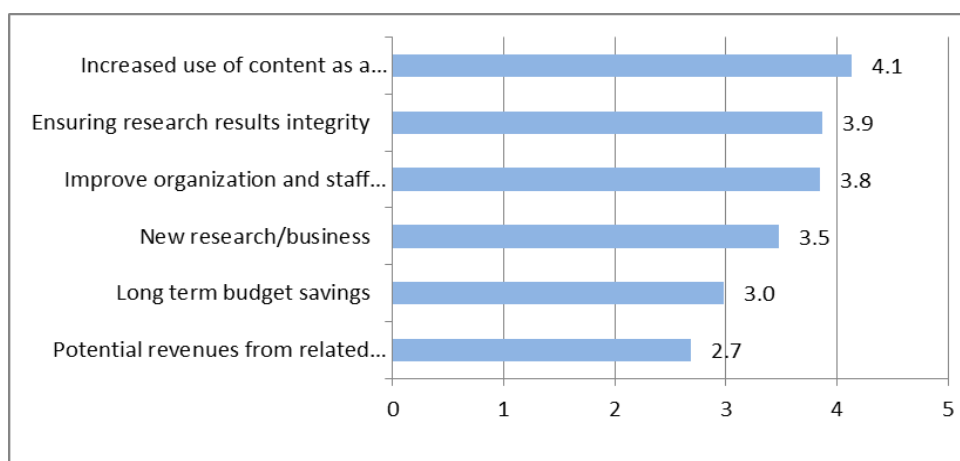


Figure 9: How do you think your organisation is likely to benefit from digital preservation?

(5 being the most likely, arithmetical mean is shown)

The preservation and improvement of current content seems to be of utmost importance in the perception of benefits. The way DP adds value to digital content management is affected by the extent to which operational and strategic processes are embedded in the organisation. The survey identifies that, overall, in organisations active in DP, the two most common ways that digital preservation provides high value to digital content are through identification, interpretation and retrieval of digital objects (63%); followed by optimization of digital content lifecycle management (53%), see Figure 10.

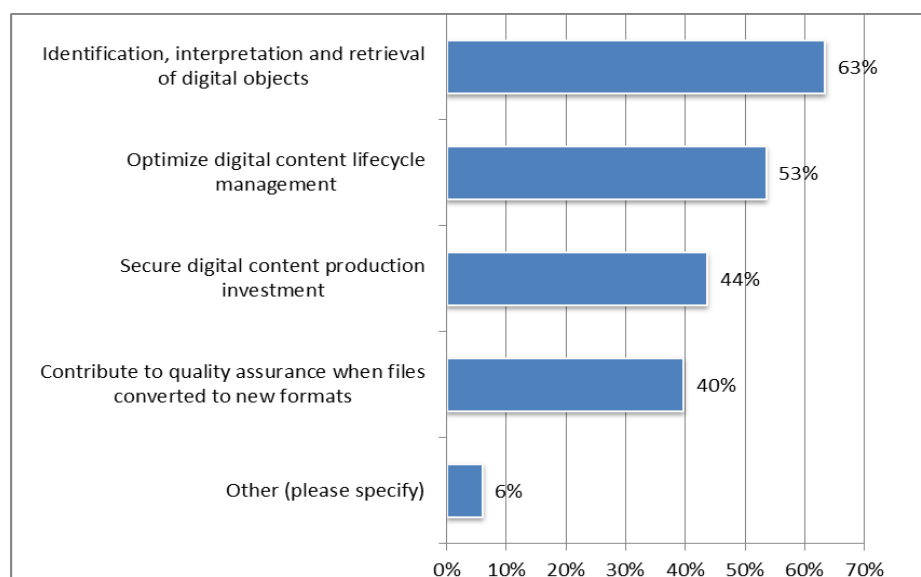


Figure 10: Reported added values

The decision as to whether DP activities can be implemented and to what extent depends certainly on the expected benefits and the possibilities of the organisations to generate their respective added values. A comparison of the benefits and added value, on the one hand and of costs on the other, i.e. cost-benefit analysis, is a prerequisite for any decision-making

process. Ultimately, the arguments contained in cost-benefit analysis must be sufficient to allow decision-makers to decide for or against DP activities. Often the analysis is just one aspect that plays a role in such decisions. In the survey we therefore asked whether a preservation mandate is a proper facilitator of digital preservation. Overall 78% of the organisation either strongly agree or agree with this statement, see Figure 11.

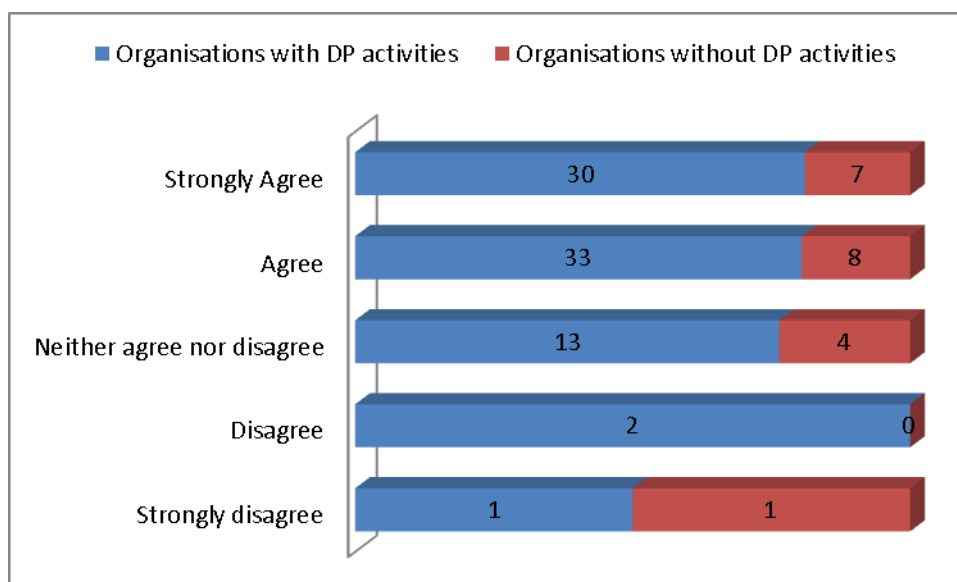


Figure 11: Preservation mandate as proper DP facilitator – all answers shown

4.2.4 Controlling costs of DP activities

The determination of the costs of DP are carried out in a limited way, with only 5% of organisations reporting that they have a cost model in place, and a further 11% having monitoring systems to measure actual costs. The vast majority (44%) of organisations estimate costs on the basis of their long-term experiences, see Figure 12. Most surprisingly, in 40% of all cases, costs were not tracked at all. Given that cost models are not used widely, the need to develop models and provide them in an understandable and ‘usable’ way will certainly help to address the monitoring and measuring aspect of DP costs.

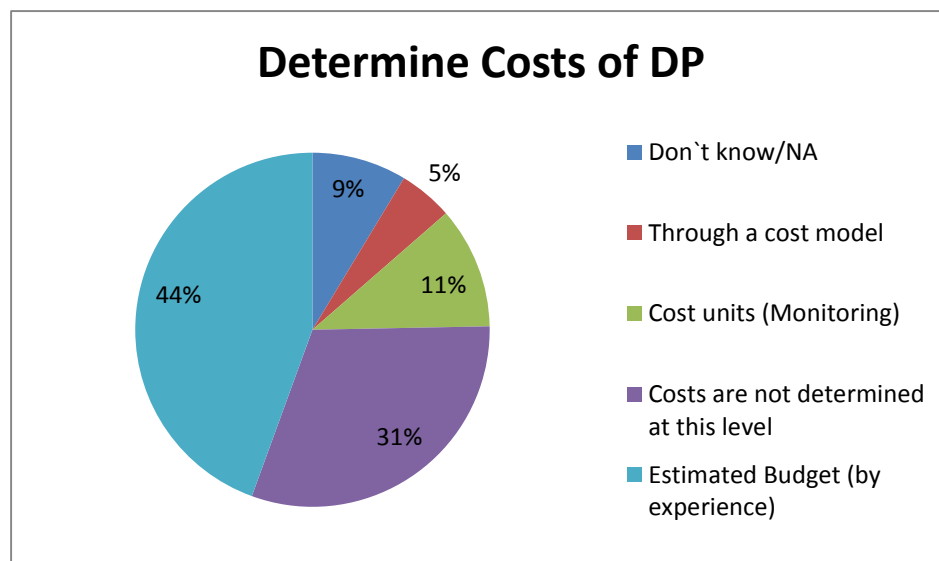


Figure 12: How do you determine the costs of Digital Preservation within your organisation? (N=81)

The survey results clearly show that the majority of organisations would like to use an operational cost model, not only to inform decision-makers but to be able to cost their own DP activities, see Figure 13. The reasons for wanting to use a cost model are extensive and include a number of financial management principles (decision making, cost calculations at activity based level, efficient use of resources, options appraisal), and are key to the decision-makers in any organisation.

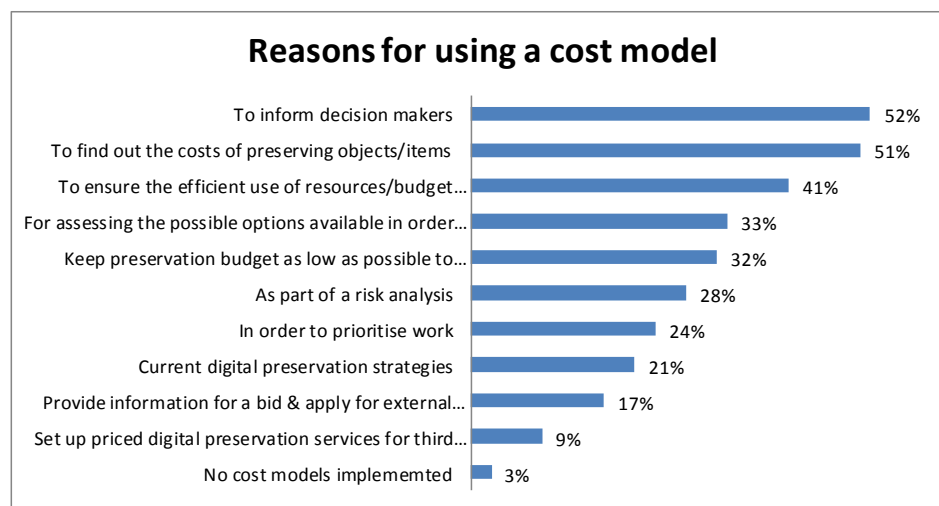


Figure 13: Reasons for using a cost model

4.2.5 Cost model requirements

The survey results, as provided in D36_1, state that organisations active in DP have reported that new and developing cost models are needed not only to support funding requests but also for reporting to high level management, as well as for calculating the real costs of preservation and ensuring efficient use of resources.

A cost model for providing clear cost-benefit justification for DP initiatives should address the following criteria:

- Sound model validated by a similar organisation
- Scope of the model; for example taking into account the complete digital content life-cycle
- Easy to use and adaptable
- Cost of the model

Clearly these criteria need to be addressed in all future development work and indeed some cost models published already address all of these issues, for example LIFE3. It is interesting to note that given that cost model tools are freely available they do not seem to be in use by this community.

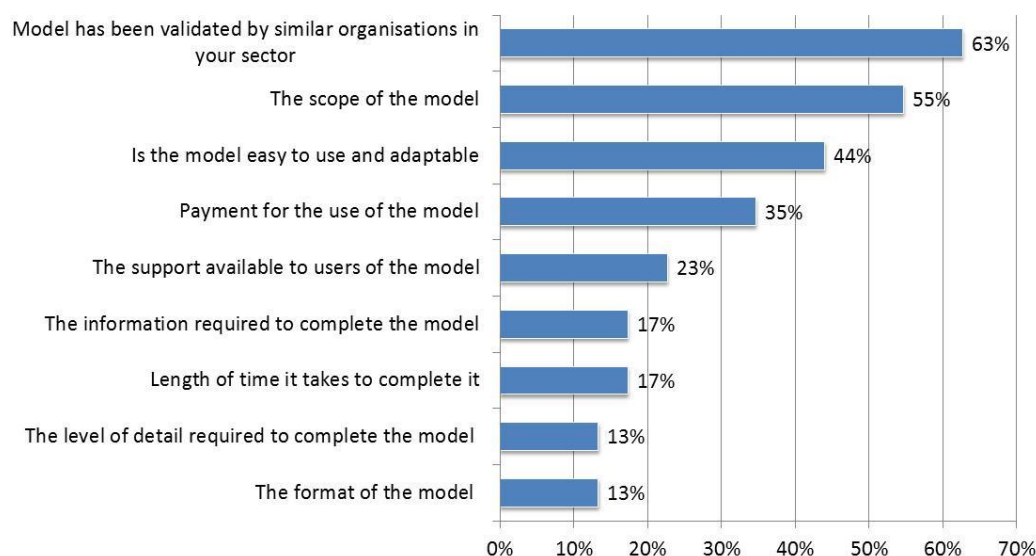


Figure 14: On what basis would you select a cost model? (N = 73)

The need for a tried and tested cost model is by far the greatest with the scope of the cost model being of secondary importance, see Figure 14. The scope should cover the existing processes in an organisation as closely as possible. This report provides details of cost models which have been validated as well as their current availability. The mapping of cost parameters to the trusted repository standard also clearly highlights the activities currently covered by these models. Given that the principles of OAIS have been implemented by cost models the matching of organisations DP activities should be straightforward. Certainly attempting to use a model and looking at the cost parameters required would highlight any areas which have not been identified by organisations and would assist them in costing their DP activities more accurately. However, convincing users of these benefits may be a challenge.

4.3 SUMMARY OF COST MODEL SURVEY FINDINGS

In summary it can be stated that most of the surveyed organisations do not use cost models. Even though the advantages of cost models are widely accepted, significant gaps exist in the knowledge of cost parameters and seemingly in the influence of cost model creators or

owners. Only some cost parameters, for example, staff, seem to be used. Cost parameters related to infrastructure are perceived only as secondary, but even they will have a major impact on overall costs. Nevertheless, it remains that there is a general movement towards the need for cost models; however, current models are not developed enough for organisations to be confident in their use. These results of this survey will help with the future development of cost models and ensure that model creators and owners consider the availability and ease of use of their tools to assist the key stakeholders within the DP community.

5 CONCLUSIONS AND PROSPECTS

5.1 WHERE NEXT FOR COST MODELLING?

APARSEN research has shown that the development of cost models for DP activities is far from being complete. Although cost models are used by organisations in a limited capacity, in most cases they have been used to assist with the budgeting process. Organisations, whose financing model is based mainly on a more commercial basis, need to invest much more effort in developing their accounting and charging processes by assessing their DP activities to a very high level of accuracy compared to those who seek only a simple apportionment of costs across various functional areas.

Although cost models are used in a limited way, where they are used, current leading practice seems to be with the aim of budgeting. This is essential to ensure that resources are used effectively across all desired functional areas in line with the limited funding available over a given period. Where shortfalls are likely to occur, organisations need early warning signs and accurate information to enable effective decision making. Apart from the use of a cost model in practice, perhaps by tailoring it to the organisations specific needs, it is important that the information obtained is processed and acted upon in a timely and effective manner. Therefore, it is necessary to implement a process whereby the information is made available to appropriate individuals, i.e. decision-makers within the organisation, in a timely way. Equally important in the future is the establishment of change control processes for those cost models in use. Only when these exist will a cost model be used at its optimum and the costs of introducing and maintaining such a model can be justified.

The ultimate goal of the cost model is that it is able to support all three objectives of Financial Management: Budgeting, Accounting and Charging. Although the development and general use of cost models are in their early stages, it is important to note the next steps of development must be to establish change management processes in order to test and improve cost models in general. The establishment of these processes will be mandatory because of the changes in DP activities which will take place over time. The corresponding cost models need to be adapted to these changes. Similarly, errors or gaps in cost models, which are revealed by the findings in this WP, should be addressed. Even though cost models are used in a limited way, the need for continuous adaptation must be formalised through a change management process which would allow for their continued use where already adopted.

5.2 CHALLENGES

There are a number of cost models in the field of digital archiving available at the moment. Organisations seem to have found it easier to develop their own and tailor them to their specific needs rather than to use those published. A universal standard model is not generally available which is of relevance. There may be several reasons for this. Digital preservation often takes place in totally different organisational contexts, for example in libraries, data archives, heritage institutes, and it is difficult to abstract common activities. Furthermore digital preservation is a relatively new development and awareness of costs and cost models for digital preservation, is even newer. The urgency to have cost models is however rapidly increasing, given the current economic climate where budgets seem to be ever shrinking. Finally, the organisations concerned often do not have the relevant knowledge and skills on cost models within their organisations, even within financial departments, therefore confidence in the use of cost models becomes even less attractive.

Another observation is that the use of cost models is certainly not widespread at the moment. Case studies would be extremely useful: testing cost models with real data is of course the

best way forward. Some case studies have been completed. The problem here is that there is a certain, understandable, reluctance for organisations to publish these. Organisations do not always want to make detailed cost data public due to the confidential nature of financial information.

Another challenge is that in particular in digital archiving, technological and organisational developments tend to move very rapidly. The creation of cost models on the other hand often do not follow at the same pace and these developments make them irrelevant to organisations who may wish to use such models, especially where predetermined financial predictions and calculations have been used which may be out of date.

There are also complications around the definitions of costs within organisations. This is probably one of the main problems in using published cost models. How should the overhead be calculated? Should for example research and development activities be included, or on the other hand, only physical storage costs be included? These are essential questions to answer and every organisation may have reasons to answer them differently. Therefore, how to make cost models and their parameters comparable remains one of the main challenges ahead.

5.3 CONNECTIONS TO COMMON VISION AND VCOE

This report is an important source of information for the VCoE. It provides a summary and an analysis of the relevant existing cost models used in the digital preservation context. Thereby, helping to direct institutions in need of cost models to those relevant or allows them to build upon models partly relevant for their own specific purposes. With the initial mapping of cost models to the standard on audit and certification of trusted digital repositories some gaps have been identified. In developing cost models in order to meet these gaps the models can determine the activities within an organisation which are key to them becoming a reliable and trustworthy digital repository. Thus good financial planning, perhaps through the use of a cost model, is an indicator that the repository is sustainable and thus trustworthy. This will be investigated further in the on-going work of this WP. Indeed how these gaps should be filled may be relevant to other activities within the VCoE.

5.4 RELATIONSHIPS WITH NEW CO-ORDINATION ACTION (4C)

A number of partners working within this project are also working within the 4C, Collaboration to Clarify the Costs of Curation, project. This European Commission co-funded project, under the 7th Framework Programme runs from February 2013 for 24 months with 13 European partners. The main objective of the project is to ensure that all the existing work relating to the costs of digital curation and preservation is known, understood and can be implemented and exploited by a wide range of stakeholders.

The findings published in the final report as well as these preliminary findings should provide some continuity especially where the work in APARSEN aligns with the work undertaken by 4C. Further discussions and collaboration are to be undertaken with the aim of running a joint workshop over the coming months. It is planned that dissemination activities related to cost models and sustainability will be open to 4C participants.

5.5 NEXT DELIVERABLE – D32.2

The next deliverable, D32_2 will provide a report on the testing of cost models and further analysis of cost parameters and is due in M30.

The results of the testing of models (with the limited data we have) will be included within D32_2. As enough cost data has not been available to date, the focus of D32_2 will be on further analyses of the mapping of cost parameters to the standard on trusted repositories in line with all outstanding work identified in this report. Section 3.3 of this report identified areas for further investigation and development. Further review of the results of the survey related to cost models will also be provided as well as further links to the new Co-ordination Action.

6 REFERENCES

- [1] ISO16363:2012 Space data and information transfer systems - Audit and certification of trustworthy digital repositories:
http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=56510
- [2] Open Planets Foundation website pages on cost models: <http://wiki.opf-labs.org/display/CDP/Home> (list accessed 2011/12/14)
- [3] ISO14721:2012 Space data and information transfer systems - Open archival information system (OAIS) Reference model:
http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=57284
- [4] APARSEN's Trust brochure - "Has it been preserved properly?":
<http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2012/09/APARSEN-Trust-Brochure-Low-Res-Web-Version.pdf>
- [5] Data Seal of Approval: <http://datasealofapproval.org/>
- [6] DIN 31644, is the Information and Documentation - Criteria for Trustworthy Digital Archives: <http://data-archive.ac.uk/curate/trusted-digital-repositories/standards-of-trust?index=3>
- [7] LIFE 1, 2 and 3. Projects to explore digital preservation costing, and develop costing models: <http://www.life.ac.uk/>
- [8] Cost Model for Digital Preservation (CMDP) - Project at the Royal Danish Library and the Danish National Archives to develop a new cost model:
<http://www.costmodelfordigitalpreservation.dk/>
- [9] Keeping Research Data Safe 1 and 2 (KRDS) - Cost model and benefits analysis for preserving research data and tools: <http://www.beagrie.com/krds.php>
- [10] PrestoPRIME cost model for digital storage: <http://www.PrestoPRIME.org/>
- [11] Cost Estimation Toolkit, NASA:
<http://opensource.gsfc.nasa.gov/projects/CET/index.php>
- [12] A cost model developed in the project Digital Preservation for libraries (DP4lib):
[DP4lib-Cost-By-Service-CostModel.docx](http://www.dp4lib.org/DP4lib-Cost-By-Service-CostModel.docx)
- [13] DANS approach based on Activity-Based Cost ABC and Balanced Scorecard BSC models: <http://www.springerlink.com/content/v3r1282x328m607m/?MUD=MP>
- [14] ENSURE project for costing cloud based preservation systems: <http://ensure-fp7-plone.fe.up.pt/site>
- [15] Modelling Digital Preservation costs for ISIS Instrument Data:
<http://epubs.stfc.ac.uk/work-details?w=61296>
- [16] Cost Model for Small Scale Automated Digital Preservation Archives by Strodl and Rauber:
http://www.ifs.tuwien.ac.at/~strodl/paper/strodl_ipres2011_costmodel.pdf
- [17] Joint Information Systems Committee: <http://www.jisc.ac.uk/>
- [18] Research Information Network: <http://www.rin.ac.uk/>
- [19] Kejsers, Nielsen and Thirifays, 'Cost Model for Digital Preservation: Cost of Digital Migration', International Journal of Digital Curation, 6(1), 2011, p255:
<http://www.ijdc.net/index.php/ijdc/article/view/177>

- [20] CMDP tool:
<https://sites.google.com/a/costmodelfordigitalpreservation.dk/pub/contact/archival-storage/doc>
- [21] LIFE3 spread sheet: <http://www.life.ac.uk/tool/>
- [22] LIFE3 web tool: <http://lifedev.hatii.arts.gla.ac.uk/>
- [23] PrestoPRIME planning tool: <http://PrestoPRIME.it-innovation.soton.ac.uk/planning-tool/>
- [24] A Collaboration to Clarify the Costs of Curation, 4C project: <http://4cproject.net/>

ANNEX 1: SUMMARY OF COST MODELS

Reference: Open Planets Foundation [2] pages, as updated by Paul Wheatley and Andy Jackson

CET – Cost estimation toolkit

- Estimates life cycle costs for scientific data activities, can potentially be applied to long-term archive systems
- Two excel based tools developed, CET software package is available <http://opensource.gsfc.nasa.gov/projects/CET/index.php>
- Paper published http://www.pv2007.dlr.de/Papers/Fontaine_CostModelObservations.pdf
- Developed by NASA and SGT
- Further reading:
Fontaine, K., Hunolt, G., Booth, A., and Banks, M., 2007, Observations on Cost Modeling and Performance Measurement of Long Term Archives in PV2007 Conference Proceedings:
http://www.pv2007.dlr.de/Papers/Fontaine_CostModelObservations.pdf

CMDP - Cost Model for Digital Preservation

- Estimates the costs of digital preservation (ingest, preservation planning and migrations, and archival storage), covers cultural heritage organisations
- Still under development, tool available
- Available on-line <http://www.costmodelfordigitalpreservation.dk/>
- Developed by the Royal Library of Denmark and the Danish National Archives
- Further reading:
Kejser, U. B., Nielsen, A. B., & Thirifays, A. (2011). Cost Model for Digital Preservation: Cost of Digital Migration, International Journal of Digital Curation, 6(1). doi:10.2218/ijdc.v6i1.186: <http://www.ijdc.net/index.php/ijdc/article/view/177>
Kejser, U., Nielsen, A., & Thirifays, A. (2009). The cost of digital preservation. Committee for Digital Preservation, Project Report v.1.0. Danish Ministry of Culture: <http://www.costmodelfordigitalpreservation.dk/contact/cmdp-1---preservation-planning-and-digital-migrations/documentation>
Kejser, U., Nielsen, A., & Thirifays, A. (2011). Costs of digital preservation. Project Report for Phase 2: <http://www.costmodelfordigitalpreservation.dk/contact/cmdp-2---ingest-and-archival-storage/documentation>

DANS cost model

- Calculates the costs of archiving datasets, based on activity based costing and balanced scorecard, covers research data archives
- Validation to be undertaken
- Paper published on the model <http://www.springerlink.com/content/v3r1282x328m607m/?MUD=MP>
- Developed by DANS, Data Archiving and Network Services, Netherlands
- Further reading:

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. *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>

DP4lib - Digital Preservation for libraries

- Calculates costs by a service model for long term preservation services to third parties, covers any sector
- Validation taking place this year, documentation and a calculation spread sheet are available on the following website:
http://dp4lib.langzeitarchivierung.de/index_downloads.php.de.
- Developed by the DNB
- Further reading:
Paper published on the model
<http://aparsen.digitalpreservation.eu/pub/Main/CostModels/DP4lib-Cost-By-Service-CostModel.docx>

ENSURE project

- Estimates costs of digital preservation activities, assumes cloud storage is used, covers healthcare, clinical trials and financial sector, may be extended to manufacturing sector
- Initial model to be developed further
- Being developed by EC FP7 project, ENSURE (Feb '11 – Jan '14) <http://ensure-fp7-plone.fe.up.pt/site>
- Further reading:
M Badawy, E Shehab, P Baguley, MD Wilson “Towards a cost model for long term digital preservation”, In Proc. 2012 ISPA/SCEA Joint International Conference: Assuring cost efficiency : global solution, Brussels, Belgium, 14-16 May 2012:
<http://epubs.stfc.ac.uk/bitstream/7711/Towards%20a%20Cost%20Model%20for%20Long%20Term%20Digital%20Preservation.pdf>

ISIS facility model

- Applied specifically to long term preservation costs of data from ISIS facility at STFC (scientific research data)
- Not applicable to other areas
- Developed as part of Cranfield University MSc project in collaboration with STFC
- Further reading:
Poster published <http://ensure-fp7-plone.fe.up.pt/site/Poster.pdf>

KRDS – Keeping research data safe (KRDS + KRDS 2)

- As well as activity cost model provides lists of benefits and potential metrics for research data, is applicable more widely.
- Toolkits - benefits analysis, value and impact - for proposals, evaluation and planning

- Published factsheet, user guide <http://www.beagrie.com/krds.php>
- Development of toolkits funded by JISC partners in project include Charles Beagrie Ltd, UKOLN, DCC, UCL, UKDA, ADS, OCLC
- Further reading:
Beagrie, N., Chruszcz, J., & Lavoie, B. 2008. Keeping Research Data Safe. A Cost Model and Guidance for UK Universities, Copyright HEFCE 2008:
<http://www.jisc.ac.uk/media/documents/publications/keepingresearchdatasafe0408.pdf70H>
Beagrie, N., Lavoie, B., and Woollard, M. (2010), Keeping Research Data Safe 2:
<http://www.jisc.ac.uk/media/documents/publications/reports/2010/keepingresearchdatasafe2.pdf>

LIFE3 – Life Cycle Information for E-literature

- Looks at long-term costs of digital preservation for DP repositories
- Third phase of the LIFE Project producing a predictive costing tool (not developed fully), excel version is available for use
- Published excel tool and papers <http://www.life.ac.uk/>
- Developed by UCL and BL, project funded by JISC and RIN
- Further reading:
Wheatley and Hole. LIFE3: Predicting Long Term Digital Preservation Costs. In iPRES2009, September 2009:
<http://www.fsl.cs.sunysb.edu/docs/unesco12/UNESCO2012-storage-econ.pdf>
Hole, Brian, Li Lin, Patrick McCann, and Paul Wheatley (2010), "LIFE3: A predictive costing tool for digital collections," 7th International Conference on Preservation of Digital Objects (iPRES2010), Vienna, September 19-24, 2010:
http://www.life.ac.uk/3/docs/Ipres2010_life3_submitted.pdf

PrestoPRIME – cost model for digital storage

- Provides cost information and long term forecasting for mass digitisation of AV materials
- Tools available and still under development
- Published report <http://PrestoPRIME.it-innovation.soton.ac.uk/planning-tool/accounts/login?next=/planning-tool/>
- Developed within EC FP7 project <http://www.PrestoPRIME.eu/>
- Further reading:
Addis, Matthew and Jacyno, Mariusz (2010) *Tools for modelling and simulating migration based preservation*. PrestoPRIME Consortium, 56pp:
https://prestoprime.ina.fr/public/deliverables/PP_WP2_D2.1.2_PreservationModellingTools_R0_v1.00.pdf
Westerhof, Hans, Ubois, Jeff and Snyders, Marius (2011). *Financial Models and Calculation Mechanisms*. PrestoPRIME Consortium, 35pp:
https://prestoprime.ina.fr/public/deliverables/PP_WP6_D6.3.1_FM_calculation_R0_v1.01.pdf

We will also be looking at:

ESA model – internal review of cost parameters – not published

Cost Model for Small Scale Automated Digital Preservation Archives (Strodl and Rauber):

http://www.ifs.tuwien.ac.at/~strodl/paper/strodl_ipres2011_costmodel.pdf

May be of interest:

OECD – International Standard Cost Model Manual

- Determines administrative costs, provides transparent measures
- Developed by the Standard Cost Model Network
- Published manual <http://www.oecd.org/regreform/regulatorypolicy/34227698.pdf>

ANNEX 2: ANALYSIS OF COST MODELS

In analysing the cost models under review the following questions were asked under set areas:

Scope:

- Which sector does this cost model relate to?
- Which type of organisation does the model apply to?

Purpose:

- What is the purpose of the cost model? (i.e. budgeting, accounting, cost allocation, raising funds for projects)

Background:

- What are the origins of the model?
- Where did the model come from?
- How was it started?
- Why was it needed?

Cost model data:

- What data was used to build/validate the model?

Applicability:

- Which preservation activity does this model relate to? e.g. 3rd party providing diverse services to different customers, one off activity, cost recovery approach, long term preservation arrangement

Reference to OAIS:

- Is there any reference to OAIS within the model?

Benefits:

- What are the benefits of the model?

Challenges:

- What are the challenges to the model itself (areas where considered weak/untested)?

Cost parameters:

- Are cost parameters clearly defined?
- Can the cost parameters be set out against the headings in the table given (i.e. aligned to the ISO16363/OAIS)?

Validation of model:

- Has the cost model been validated? If so, provide details.

Tool support:

- If a tool is available within the model, is support available?

Availability:

- Has testing been completed?
- Is the model available for use? If not, when will it be available?

References:

- What other work was cited/referenced?

Other:

- Any other information which may be of relevance

ANNEX 3: ISO16363 – EXTRACT RELEVANT TO MAPPING EXERCISE

ISO 16363 - EXTRACTS			
ISO REF	HEADING	SUB HEADING	EVIDENCE
3	ORGANISATIONAL INFRASTRUCTURE		
3.1	GOVERNANCE AND ORGANISATIONAL VIABILITY		
3.1.1	Mission Statement on the preservation of digital information		Statement, mandate
3.1.2	Preservation Strategic Plan	3.1.2.1 succession plan, contingency plan, escrow arrangements, 3.1.2.2 organisational environment to execute 3.1.2.1	Preservation Strategic Plan; meeting minutes; administrative decisions; written plans; steps taken to ensure continuity; escrow of critical code, software, and metadata sufficient in the event of repository failure; escrow and/or reserve funds for contingencies; explicit agreements for the repository's digital content and related assets, and granting of rights to ensure continuity. Active monitoring and preparedness.
3.1.3	Collection Policy		Collection policy and supporting documents; Preservation Policy, mission, goals and vision of the repository.
3.2	ORGANISATIONAL STRUCTURE AND STAFFING		
3.2.1	Staff with adequate training and skills	3.2.1.1 clear duties, 3.2.1.2 number of staff to support all functions and servs. 3.2.1.3 professional development	Plans, job descriptions, certificates, repository reviews, staffing charts, Development plans, training budgets, certificates

ISO 16363 - EXTRACTS			
ISO REF	HEADING	SUB HEADING	EVIDENCE
3.3	PROCEDURAL ACCOUNTABILITY AND PRESERVATION POLICY FRAMEWORK		
3.3.1	Defined Designated Community and knowledge base		Written definition
3.3.2	Preservation Policy to ensure Preservation Strategic Plan met	3.3.2.1 mechanisms for review, update, ongoing development	Policy, plans, procedures, workflows, reviews, surveys and feedback, system logic, functionality
3.3.3	Audit trail of changes to operations, procedures, software, hardware		Capital equipment inventories; acquisition documents, disposal schedules
3.3.4	Transparency and accountability in all actions supporting the operation and management of the repository that affect the preservation of digital content over time.		Financial and technical audits and certifications, disclosure of governance documents, independent program reviews, and contracts and agreements with providers of funding and critical services.
3.3.5	Information integrity measures to be defined, collected and tracked		Repositories integrity measures, computed checksum or hash value, procedures, indicate digital content is at risk, audit process
3.3.6	Self assessment, external certification		Self-assessments, 3 rd p audits, certificates, budget allocations for future certn.
3.4	FINANCIAL SUSTAINABILITY		
3.4.1	Business planning processes		Business plans, financial statements, forecast, budgets, market analysis

ISO 16363 - EXTRACTS			
ISO REF	HEADING	SUB HEADING	EVIDENCE
3.4.2	Financial practices and procedures, 3rd party audits		Standards, practices, audits
3.4.3	Analyze and report on financial risk, benefit, investment, and expenditure (including assets, licenses, liabilities)		Risk mgt documents, risk register, investment planning, cost/benefit analyses, licenses, contracts, asset mgt
3.5	CONTRACTS, LICENSES, AND LIABILITIES		
3.5.1	Contracts or deposit agreements for digital materials	3.5.1.1 Transfer of preservation rights 3.5.1.2 Depositor agreements 3.5.1.3 Acceptance of preservation responsibility for data objects 3.5.1.4 Liability and challenges to ownership/rights	Agreements, licences, policies, service levels, copyright resolution, risk assessments, Review and maintenance
3.5.2	Intellectual property rights mgt		Policy, agreements, process, o'ship, monitoring metadata to capture rights info
4	DIGITAL OBJECT MANAGEMENT		
4.1	INGEST: ACQUISITION OF CONTENT		
4.1.1	Content Information and the Information Properties	4.1.1.1 Information properties procedures 4.1.1.2 Content Information and Information Properties	Statements, agreements, gift deeds, documents, procedures, manuals, surveys, plans
4.1.2	Content Information on deposit		Requirements, agreements, workflows to produce AIP

ISO 16363 - EXTRACTS			
ISO REF	HEADING	SUB HEADING	EVIDENCE
4.1.3	Specifications enabling recognition and parsing of the SIPs.		PI/RI specifications, standards
4.1.4	Verification of Producer identity		Submission agreements/ deposit agreements/deeds of gift, measures , logs
4.1.5	Ingest process to verify SIP		Policies, plans, system log files, registers from transfer and ingest process, procedures , registries
4.1.6	Preservation control over the Digital Objects		Physical control, database, catalogue, metadata
4.1.7	Responses to producer/depositor		Agreements, documents, procedures, correspondence
4.1.8	Content acquisition records		Recorded decisions/actions, preservation metadata, confirmation receipts
4.2	INGEST: CREATION OF THE AIP		
4.2.1	Definition for parsing AIP and for long term preservation	4.2.1.1 Definition of AIP 4.2.1.2 Definition of AIP for long term preservation	Definitions, documentation, provenance
4.2.2	AIP construction form SIP		Documents
4.2.3	Document SIPs	4.2.3.1 SIP discarded or not incorporated	System processing files, disposal records, tracking system, processes

ISO 16363 - EXTRACTS			
ISO REF	HEADING	SUB HEADING	EVIDENCE
4.2.4	Persistent unique identifiers for AIPs	4.2.4.1 Unique identifier for AIP 4.2.4.1.1 Unique identifier for repository 4.2.4.1.2 Repository assigns persistent identifier for AIP 4.2.4.1.3 Identifier change documents 4.2.4.1.4 Identifier list provided by repository (checked for duplications) 4.2.4.1.5 Repository identifiers fit for current and future requirements 4.2.4.2 Linking/resolution service in repository to find objects	Documentation and logs
4.2.5	Provide Representation Information (RI) for all digital objects	4.2.5.1 Identify file type of DO 4.2.5.2 RI needed to make DO understandable to Designated Community 4.2.5.3 Access to RI 4.2.5.4 RI associated with DO	Registries, records
4.2.6	Preservation Description Information (PDI) for Content Information (CI)	4.2.6.1 Documented process 4.2.6.2 Execute process 4.2.6.3 PDI persistently associated with CI	Procedures, checksums, digests. Consult Designated community
4.2.7	AIP Content Information understandable to Designated Community	4.2.7.1 Documented process 4.2.7.2 Testing process for AIP CI 4.2.7.3 Understandability testing	Test procedures, records
4.2.8	AIP verification		Procedures, logs

ISO 16363 - EXTRACTS			
ISO REF	HEADING	SUB HEADING	EVIDENCE
4.2.9	Repository integrity verification		Agreements, logs
4.2.10	AIP creation records		Decisions, actions, metadata
4.3	PRESERVATION PLANNING		
4.3.1	Preservation strategy		Documents, risks
4.3.2	Monitor preservation environment	4.3.2.1 Monitor, notify when RI not adequate	Surveys, subscriptions to RI registry service or technology watch service, processes
4.3.3	Preservation plan changes	4.3.3.1 Create, identify, gather extra RI	Technology watch, frequent reviews, updates, subscriptions as above
4.3.4	Preservation activity effectiveness		Metadata collection, usability tests, track records, Designated Community polls
4.4	AIP PRESERVATION		
4.4.1	AIP storage spec	4.4.1.1 Preserve AIP CI 4.4.1.2 Monitor AIP integrity	Documents, procedures, fixity logs
4.4.2	AIP storage and preservation records	4.4.2.1 AIP actions procedures 4.4.2.2 Compliant actions	Documents, logs, conformance audit
4.5	INFORMATION MANAGEMENT		
4.5.1	Discovery and identification of material by Designated Community		Discovery metadata,
4.5.2	AIP descriptive info		Persistent, unique identifiers, agreements, policy documents, process workflow

ISO 16363 - EXTRACTS			
ISO REF	HEADING	SUB HEADING	EVIDENCE
4.5.3	Bi-directional linkage between AIP and descriptive info	4.5.3.1 Maintain associations between AIP and descriptive info	Persistent, unique identifiers, policy documents, process workflow, maintenance logs, technical architecture
4.6	ACCESS MANAGEMENT		
4.6.1	Access policies	4.6.1.1 Log and review access failures/ anomalies	Statements, logs, audit trails, automates system monitoring tools and notes of actions taken
4.6.2	Authenticity of Digital Objects	4.6.2.1 Error reports or user queries recorded/acted upon	Syst. design docs, samples, work inst., logs, error rpts, actions
5	INFRASTRUCTURE AND SECURITY RISK MANAGEMENT		
5.1	TECHNICAL INFRASTRUCTURE RISK MANAGEMENT		

ISO 16363 - EXTRACTS			
ISO REF	HEADING	SUB HEADING	EVIDENCE
5.1.1	Identify and manage risk management to preservation operations and system infrastructure	5.1.1.1 Tech watch or tech monitoring notification system 5.1.1.1.1 Appropriate hardware 5.1.1.1.2 Monitor and receive hardware technology changes 5.1.1.1.3 Evaluate h/w changes 5.1.1.1.4 Funding h/w changes 5.1.1.1.5 Appropriate software 5.1.1.1.6 Monitor and receive software technology changes 5.1.1.1.7 Evaluate s/w changes 5.1.1.1.8 Funding s/w changes 5.1.1.2 Backup support 5.1.1.3 Bit corruption or loss 5.1.1.3.1 Record and report all corruption/loss 5.1.1.4 Risk–benefit assessment security updates 5.1.1.5 Refresh/ migration 5.1.1.6 Critical processes & mandatory requirements 5.1.1.6.1 Change mgt 5.1.1.6.2 Test and evaluate changes	Component inventory, technology assessments, export of records to independent system, supported software, re-creation of archives form back-up Designated Community expectations, feedback on system adequacy, hardware inventory, capacity audits, review of error rates and bottlenecks Staff expertise Levels of service, budgets Back up audits/testing, inventories, disaster recovery Bit error detection and correction, risk analysis Error logs, escalation process Risk register, updates System lifecycles, migration policies Traceability matrix Change mgt procedures, logs Testing procedures
5.1.2	Manage number and location of copies of Digital Objects	5.1.2.1 Synchronise multiple copies	Workflow, system analysis

ISO 16363 - EXTRACTS			
ISO REF	HEADING	SUB HEADING	EVIDENCE
5.2	SECURITY RISK MANAGEMENT		
5.2.1	Analysis of security risk factors associated with data, systems, personnel, and physical plant		Codes of practice
5.2.2	Security risk controls		Lists, analyses, controls
5.2.3	Staff roles and responsibilities		Standards, org charts
5.2.4	Disaster recovery plans, one off-site backup and offsite recovery plan		Plans, codes of practice
END			

ANNEX 4: SURVEY QUESTIONS RELEVANT TO COST MODELS

Source: D36_1 – EXTRACT OF RELEVANT QUESTIONS



WP_36 – APARSEN SURVEY – BUSINESS PREPAREDNESS -

This questionnaire is being undertaken for the APARSEN Project that runs under the ICT directorate of the EUROPEAN COMMISSION. We aim to produce a study which evaluates digital preservation in terms of importance, value, benefits, currents activities, costs and future involvement.

We would like to inform you that the survey results will not reflect any particular organisation perspective as gathered results will be treated in an anonymised way in compliance with the Data Protection Act.

PROFILING AND EVALUATING DIGITAL PRESERVATION

1. *Could you please select from the following list the sector where your organisation best fits? (Single choice)*

- ☐ National Library
- ☐ State Library
- ☐ University Library
- ☐ Corporate
- ☐ Government
- ☐ Research Institute
- ☐ Other (please specify)

2. *Please tell us the country where your organisation resides? (Single choice)*

- | | | | |
|---|--|--------------------------------------|--|
| <input type="checkbox"/> Albania | <input type="checkbox"/> Denmark | <input type="checkbox"/> Luxemburg | <input type="checkbox"/> Slovakia |
| <input type="checkbox"/> Armenia | <input type="checkbox"/> Estonia | <input type="checkbox"/> Malta | <input type="checkbox"/> Slovenia |
| <input type="checkbox"/> Austria | <input type="checkbox"/> Finland | <input type="checkbox"/> Montenegro | <input type="checkbox"/> Spain |
| <input type="checkbox"/> Azerbaijan | <input type="checkbox"/> Former | <input type="checkbox"/> Netherlands | <input type="checkbox"/> Sweden |
| <input type="checkbox"/> Bangladesh | <input type="checkbox"/> Yugoslav | <input type="checkbox"/> Nigeria | <input type="checkbox"/> Switzerland |
| <input type="checkbox"/> Belgium | <input type="checkbox"/> Republic of Macedonia (FYROM) | <input type="checkbox"/> Norway | <input type="checkbox"/> The Netherlands |
| <input type="checkbox"/> Bosnia and Herzegovina | <input type="checkbox"/> France | <input type="checkbox"/> Poland | <input type="checkbox"/> Turkey |
| <input type="checkbox"/> Bulgaria | | <input type="checkbox"/> Portugal | <input type="checkbox"/> Ukraine |
| | | <input type="checkbox"/> Romania | <input type="checkbox"/> United Kingdom |

- | | | | |
|---|------------------------------------|---------------------------------|---|
| <input type="checkbox"/> Canada | <input type="checkbox"/> Germany | <input type="checkbox"/> Russia | <input type="checkbox"/> USA |
| <input type="checkbox"/> Croatia | <input type="checkbox"/> Greece | <input type="checkbox"/> Serbia | <input type="checkbox"/> Other (please specify) |
| <input type="checkbox"/> Cyprus | <input type="checkbox"/> Hungary | | |
| <input type="checkbox"/> Czech Republic | <input type="checkbox"/> Iceland | | |
| | <input type="checkbox"/> Ireland | | |
| | <input type="checkbox"/> Italy | | |
| | <input type="checkbox"/> Latvia | | |
| | <input type="checkbox"/> Lithuania | | |

5. What is your annual budget for creating/acquiring/archiving digital content (including operational costs)?

(Choose one of the following intervals)

- ☐ <50.000 €
☐ 50.001 - 100.000 €
☐ 100.001-500.000 €
☐ 500.001-1.000.000 €
☐ >1.000.001 €
☐ Don't know

8. The following alternatives describe how we believe Digital Preservation adds value to digital content management. Please select the 2 that you think provide the highest value:

- ☐ Contribute to quality assurance when files converted to new formats
☐ Secure digital content production investment
☐ Optimize digital content lifecycle management
☐ Identification, interpretation and retrieval of digital objects
☐ Other 1 (please specify)

9. How do you think your organisation is likely to benefit from digital preservation?

	Not at all likely	Slightly likely	Moderately likely	Very likely	Completely likely	Not sure/ Don't Know
New research/business opportunities based on content re-use	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Improve organisation and staff reputation (visibility, citations, recognition of being at the forefront of digital preservation etc.)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Long term budget savings	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Increased use of content as a result of better availability and fundability	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Potential revenues from related services provided to third parties	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ensuring research results integrity	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Other (please specify)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

10. The following alternatives describe how we believe Digital Preservation adds value to digital content management. Please select the 2 that you think provide the highest value:

- ☐ Contribute to quality assurance when files converted to new formats
- ☐ Secure digital content production investment
- ☐ Optimize digital content lifecycle management
- ☐ Identification, interpretation and retrieval of digital objects
- ☐ Other (please specify)

11. Using a scale from 1 to 5, where 1 is “strongly disagree” and 5 “strongly agree”, how much do you agree with the fact that a “preservation mandate will be a proper facilitator of digital preservation”?

- ☐ Strongly Agree
- ☐ Agree
- ☐ Neither agree nor disagree
- ☐ Disagree
- ☐ Strongly disagree

DIGITAL PRESERVATION

22. What is the main funding source for your digital preservation activities? [Multiple choices]

- ☐ Institutional budget
- ☐ Regional/Governmental funding
- ☐ Project(s) funding
- ☐ Sponsorships / Donors
- ☐ Revenues (fees from products or services)
- ☐ Other (please specify)

25. *How are decision makers involved in Digital Preservation activities in your organisation?*
[Multiple choices]

- ☐ As funders
- ☐ As supervisors
- ☐ As executives
- ☐ As advisors
- ☐ None / Other (please specify)

26. *From the alternatives below, please select the 3 main reasons for you/your organisation using a cost model:*

- ☐ To inform decision makers
- ☐ To find out the costs of preserving objects/items
- ☐ For assessing the possible options available in order to carry out DP activities
- ☐ Keep preservation budget as low as possible to enable collection development while performing digital preservation
- ☐ To provide information for a bid to apply for external funding
- ☐ As part of a risk analysis
- ☐ In order to prioritise work
- ☐ To ensure the efficient use of resources
- ☐ To set up priced digital preservation services for third parties
- ☐ Other (please specify):

27. *On what basis would you select a cost model? [Multiple choices]*

- ☐ Model has been validated by similar organisations in your sector
- ☐ The scope of the model e.g. covering the digital preservation lifecycle
- ☐ Length of time it takes to complete it
- ☐ The information required to complete the model
- ☐ The format of the model (e.g. online tool, paper based)
- ☐ Payment for the use of the model
- ☐ The support available to users of the model
- ☐ The level of detail required to complete the model (high level with limited information requiring a breakdown in costs as specified by the model)
- ☐ Is the model easy to use and adaptable
- ☐ Other (please specify):

28. *How do you determine the costs of Digital Preservation within your organisation? [Single choice]*

- ☐ Through a cost model
- ☐ Estimated Budget (by experience)
- ☐ Cost units (Monitoring)
- ☐ Other (please specify):

ANNEX 5: OVERVIEW OF MAPPING EXERCISE – DRAFT

