University of California
CDL
California Digital Library

California Digital Library
UC Office of the President

Peer Reviewed

Title:
Cost Model for Digital Curation: Cost of Digital Migration

Author:
Kejser, Ulla Bøgvad, Royal Library, Denmark
Nielsen, Anders Bo, Danish National Archives
Thirifays, Alex, Danish National Archives

Publication Date:
10-05-2009

Series:
iPRES 2009: the Sixth International Conference on Preservation of Digital Objects

Publication Info:
iPRES 2009: the Sixth International Conference on Preservation of Digital Objects, California Digital Library, UC Office of the President

Permalink:
http://www.escholarship.org/uc/item/4d09c0bb

Multimedia URL:
http://www.cdlib.org/services/uc3/iPres/video.html?file=ipres/KejserThirifays&title=Ulla%20B%F8gvad%20Kejser%20and%20Alex%20Thirifays%3A%20Cost%20Model%20for%20Digital%20Preservation%3A%20Cost%20of%20Digital%20Migration

Abstract:
The Danish Ministry of Culture is currently funding a project to set up a model for costing preservation of digital materials held by national cultural heritage institutions. The overall objective of the project is to provide a basis for comparing and estimating future financial requirements for digital preservation and to increase cost effectiveness of digital preservation activities. In this study we describe an activity based costing methodology for digital preservation based on the OAIS Reference Model. In order to estimate the cost of digital migrations we have identified cost critical activities by analysing the OAIS Model, and supplemented this analysis with findings from other models, literature and own experience. To verify the model it has been tested on two sets of data from a normalisation project and a migration project at the Danish National Archives. The study found that the OAIS model provides a sound overall framework for cost breakdown, but that some functions, especially when it comes to performing and evaluating the actual migration, need additional detailing in order to cost activities accurately.

Supporting material:
Presentation

eScholarship
University of California

eScholarship provides open access, scholarly publishing services to the University of California and delivers a dynamic research platform to scholars worldwide.

Copyright Information:

eScholarship
University of California

eScholarship provides open access, scholarly publishing services to the University of California and delivers a dynamic research platform to scholars worldwide.

iPRES 2009

THE SIXTH INTERNATIONAL CONFERENCE ON THE PRESERVATION OF DIGITAL OBJECTS

Proceedings

October 5-6, 2009
Mission Bay Conference Center
San Francisco, California

CDL
California Digital Library

Cost Model for Digital Curation: Cost of Digital Migration

Ulla Bøgvad Kejser,
The Royal Library, Denmark
P.O. Box 2149
DK-1016 København K
ubk@kb.dk

Anders Bo Nielsen,
The Danish National Archives
Rigsdagsgården 9
DK-1218 København K
abn@ra.sa.dk

Alex Thirifays
The Danish National Archives
Rigsdagsgården 9
DK-1218 København K
alt@ra.sa.dk

Abstract

The Danish Ministry of Culture is currently funding a project to set up a model for costing curation of digital materials held by national cultural heritage institutions. The overall objective of the project is to provide a basis for comparing and estimating future financial requirements for digital curation and to increase cost effectiveness of digital curation activities.

In this study we describe an activity based costing methodology for digital curation based on the OAIS Reference Model. Within this framework, which we denote "Cost Model for Digital Curation" we then focus especially on costing digital migration activities. In the terms of the OAIS Model digital migration includes both transfer (no alteration of data) and transformation (alteration of data). In order to estimate the cost of digital migrations we have identified cost critical activities by analysing the OAIS Model, and supplemented the analysis with findings from literature. We then deconstructed the activities in measurable components, analysed cost dependencies, and made equations, which have been expressed in a spreadsheet.

In order to verify the model it has been tested on two sets of data from different migration projects at the Danish National Archives. The study found that the OAIS model provides a sound overall framework for cost breakdown, but that some functions, especially when it comes to performing and evaluating the actual migration, need additional detailing and precision in order to cost activities accurately. Running the two sets of empirical data showed among other things that the model underestimates the cost of man-power intensive migration projects, while it reinstates an often underestimated cost, which is the cost of developing migration software. The model has proven useful for estimating costs of digital migrations. However, more work is needed to expand the equations to the other functional entities of the OAIS Model, including Ingest and Access. The Cost Model for Digital Curation is about to enter its second iteration, where it will be tested on data from more cultural heritage institutions. These tests will enable us to adjust the theoretical model further.

Introduction

Frameworks for costing digital long-term preservation have been proposed concurrently with the development of digital curation strategies and the evolution of repository systems and processes. A recent report on sustainable digital preservation and access gives a comprehensive review of costing methodologies and notes that comparisons of cost data remain difficult because the majority of studies have been specific rather than generic, i.e. aimed at specific types of institutions or materials, or based on special ways of measuring and adjusting costs (Blue Ribbon Task Force, 2008, pp. 36-37).

It is characteristic that cost models for digital preservation take a lifecycle approach, as exemplified in an early study on preservation methods and cost models (Hendley, 1998 (based on Beagrie & Greenstein, 1998)). The reason is the recurring nature of preservation cost and the fact that they are difficult to separate from other lifecycle cost such as creation and access (Granger, Russell, & Weinberger, 2000, pp. 2, 4). Furthermore, they are highly dependent on the range of services an institution offers (Ashley, 1999). However, no consensus has yet been reached on how the lifecycle for costing digital curation should be structured; or on how the individual lifecycle phases should be broken down and detailed. In response to this issue Sanett suggested developing a framework for costing preservation of electronic records, and advocated for mapping cost on a well-defined function model, and for applying generally accepted accounting principles (Sanett, 2002). As an example of such mapping the author referred to the InterPARES project[1] (Ibid. p. 394), in which the OAIS Reference Model (Consultative Committee for Space Data Systems, 2002) had been used for this purpose. The OAIS Model also forms the backbone of the activity model (function model) proposed in a thorough study on costing preservation of research data (Beagrie, Chruszcz, & Lavoie, 2008). Another outstanding challenge is developing formulas for costing the breakdown products.

One area within the lifecycle remains particularly difficult to cost, namely the cost of functional preservation, i.e. the costs of keeping digital resources accessible and understandable in the long term; and very little empirical data is available on the subject. NASA CET is another example of a model that does not handle

[1] The InterPARES project. Retrieved September 15 2009, from http://www.interpares.org/

costing of digital migration.[2] One attempt to fill this gap is the cost model for digital preservation developed by the Digital Preservation Testbed (Nationaal Archief, 2005). It models cost of the digital repository system, development or procurement of preservation software, performance of preservation actions, and labour costs. The costs are expressed as formulas in a spreadsheet. Likewise the LIFE project has not only developed an advanced lifecycle cost model, but also investigated cost of functional preservation in detail and developed the Generic Preservation Model (McLeod, Wheatley, & Ayris, 2006, version 1.0, pp. 90-107; Ayris, et al., 2008, version 1.1, pp. 34-37). This model is also expressed in a spreadsheet, and provides means of estimating cost of digital preservation, including formulas for preservation action frequency and file format complexity.

The overall purpose of this study is to design a framework for costing digital preservation, including a breakdown methodology with sufficient detail to give an accurate outline of required resources, a set of equations that will transform these resources into cost data, and a description of the applied accounting principles. We investigated the usability of the OAIS Model to provide the functional breakdown in measurable components. As a first step we have aimed at breaking down and costing activities related to functional preservation, and more specifically to the migration strategy. The soundness of the proposed model is tested on empirical data from case studies.

Methods

We have applied an activity based costing methodology, structured around the functional breakdown provided by the OAIS Reference Model, which consists of six functional entities: Ingest, Archival Storage, Data Management, Administration, Preservation Planning and Access. Besides we have included the support functions Common Services and Management from the OAIS Model. Each of these entities comprises a series of functions, which are further described in the OAIS documentation. In order to identify what we term cost critical activities, i.e. tasks which take more than 1 person week (pw) to accomplish, we have analysed the functional descriptions, divided these into measurable components, and established equations. The basic formula for an activity is the effective time required to complete an activity (measured in pw) times wage level (including overhead) plus purchases (monetary value). Each activity is adjusted to account for specific cost implications, such as format documentation complexity. Overhead covers indirect costs, i.e. indirect staff, facilities, general management and administration. We make use of different categories of personnel (wages): manager, computer scientist and technician. The overall structure and breakdown methodology of the Cost Model for Digital Curation (CMDC) is shown in figure 1.

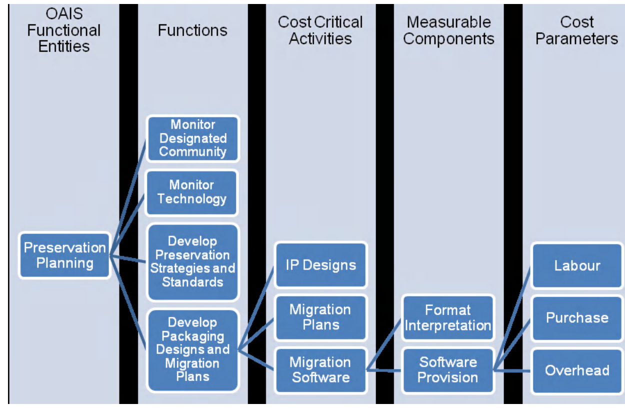

Figure 1 Overview of the breakdown methodology and structure of CMDC.

Costing digital migrations

While the goal is to model the whole lifecycle, the current version of the model only deals with cost of digital migrations. For this purpose we have defined a cost critical flow between the relevant functions within the OAIS archive. The model does not yet include the cost of requesting the content to be migrated from Archival Storage, or the cost of ingesting the new Information Package (IP) version back into Archival Storage. Table 1 shows the OAIS functions, which contain cost critical migration activities.

A central parameter in the model is the Format Interpretation factor, which denotes how difficult a format is to comprehend. The factor depends on identifying and reading format specifications and any other relevant documentation of both the source format (e.g. TIFF) and the destination format (e.g. JPEG 2000). Furthermore, it depends on the amount (number of pages) and complexity level of the documentation. We have estimated that it takes 20 minutes on average to read and understand a page of documentation for a format with low complexity. This number is increased by 25% for a format with a medium complexity, such as TIFF 6.0, and by 50% if it is of high complexity, such as GML.

[2] NASA CET. Retrieved September 15, 2009, from http://opensource.gsfc.nasa.gov/projects/CET/CET.php

Functional Entities and Functions	Cost Critical Activities
Preservation Planning: Monitor Designated Community	Monitor community Report on monitoring
Preservation Planning: Monitor Technology	Monitor technology Report on monitoring
Administration: Manage System Configuration	Monitor systems Report on monitoring
Preservation Planning: Develop Preservation Strategies and Standards	Develop strategies and standards Recommend system evolution Provide advice
Preservation Planning: Develop Packaging Designs and Migration Plans	Develop IP designs Develop migration plans Develop migration software
Administration: Establish Standards and Policies	Review migration package
Administration: Manage System Configuration	Implement migration tools
Administration: Archival Information Update	Update content (migration action) Quality assurance

Table 1: Summary of the cost critical activities in each OAIS function.

The Monitor Designated Community and Monitor Technology functions each consists of two cost critical activities, namely monitoring user community and technology, and reporting on the findings of this monitoring. We assume that Monitor Designated Community depends on how much influence the archive has on production and use of formats: The more influence, the fewer costs. Monitor Technology depends on the general technology development and on the complexity level of the formats preserved by the archive and on those monitored. If the archive uses preservation formats with a high degree of complexity the result is a high cost for monitoring them.

The Develop Preservation Strategies and Standards function assembles the reports received from the monitoring functions and develops and recommends strategies and standards (including profiles) to meet any new challenges.

The Develop Packaging Designs and Migration Plans function includes the cost critical activities of developing Information Package (IP) designs, Migration Plans, and Migration Software. IP Designs denote the structure of the container of the content in the archive. The cost of the activity is based on the total Format Interpretation factor and the frequency of the need to create new IP designs. For simplicity we assume that new IP designs are required when migration is necessary. The frequency of migration is based on average estimated lifetime of formats, which we again for simplicity have set to be 10 years. Due to variation in remaining format lifetime we estimate that migrations take place every 5 years (thereby migrating 50% of the content of the archive). The activity Migration Plans includes development of general and detailed plans for migration, including test plans, community review plans and implementation plans. The cost of the activity is based on the cost of developing new IP designs and thereby indirectly of the Format Interpretation factor.

The activity Migration Software comprises the cost of developing migration tools, including design, development and test. The cost of the activity is based on the Software Provision factor, which depends on the Format Interpretation factor, and thus the complexity of the formats. The Software Provision factor consists of three elements, namely the development of a reader tool, a writer tool and a translator tool for each format. We assume that there is a base software development time of two person-days for each module. Furthermore we estimate that the development time is approximately twice the Format Interpretation factor. If the migration tool is purchased we believe that the cost is reduced to one third.

The Manage System Configuration function under Administration develops and implements plans for system evolution and it implements migration packages, including tools, in the archive systems. The Archival Information Update function consists of the cost critical activity to perform the actual migration process. In accordance with OAIS we assume that the tools and the content at this stage are flawless ensuring an almost automatic process. In order to calculate the time it takes to execute the actual migration, we have introduced a Processing factor. It depends on the Interpretation factor, the amount of data, the computer power, and on the number of computers.

We have expressed and combined all equations in a spreadsheet.[3]

Case Studies

The model has been tested on two case studies. The first consists of data from a large migration project performed from 2005 to 2008 at The Danish National Archives (DNA), where digital materials were migrated to the current preservation standards. A detailed registration of the incurred costs was performed. The migration project dealt with 3 different migrations, all of them handling registries and filing systems: A-archives mostly hierarchical, from 1968-1998; B-archives from 1999-2000; C-archives from 2001-2004. The A-archives represented a heterogeneous mass of data, while the B-archives were in concordance with recent preservation standards and C-archives almost complied with the DNA's present preservation standards. In order to make the transformation process as inexpensive, i.e. automatic, as possible, a normalised description (using XML) was made for each digital archive. Simultaneously a system

[3] The spreadsheet and other documentation is available on the project web page: http://www.costmodelfordigitalcuration.dk

was developed which could then read the normalised descriptions and transform the many variants of data structures, data types, character sets, etc., to the standard. On average 8 persons worked full time for three years describing the data, while 2 persons spent 2½ years developing and maintaining the system.

The second case is a current migration of 6 TB of PDF documents containing property registry data to the JPEG2000 format. Several purchased tools were evaluated. In this case it was necessary to develop tools on top of purchased tools since they were not up to the task by themselves.

Results

Case 1

Table 2 shows the cost in person weeks (pw) of developing IP designs, Migration Plans and Migration Software in case 1. It includes activities related to digital migration from the moment the monitoring functions have issued a migration alert until the migration tools are ready for processing. Thus the table does not show the costs related to the monitoring functions or the processing of the migration itself. The first set of columns (Case 1) gives the actual figures from case 1. The second set (CMDC) shows case 1 simulated in CMDC. The third set (CMDC-Case 1) shows the differences between case 1 and its simulation. The B&C-archives are also combined in a separate row for analytic purposes, as we will see. At the bottom of the table, the three activities are added up under Preservation Planning.

	Case 1		CMDC		CMDC - Case 1	
	pw	%	pw	%	Δ pw	%
Develop IP Designs	**44**	**12**	**50**	**24**	**6**	**14**
A (1968-1998)	29	66	20	40	-9	-31
B (1999-2000)	15	34	16	32	1	7
C (2001-2004)	0	0	14	28	14	n.a.
B & C	15	34	30	60	15	100
Develop Migration Plans	**150**	**42**	**39**	**19**	**-111**	**-74**
A (1968-1998)	105	70	15	38	-90	-86
B (1999-2000)	30	20	14	36	-16	-53
C (2001-2004)	15	10	10	26	-5	-33
B & C	45	30	24	62	-21	-47
Develop Migration Software	**164**	**46**	**116**	**57**	**-48**	**-29**
A (1968-1998)	101	62	48	41	-53	-52
B (1999-2000)	50	30	36	31	-14	-28
C (2001-2004)	12	7	32	28	20	167
B & C	62	38	68	59	6	10
Preservation Planning (total)	**358**	**100**	**205**	**100**	**-153**	**-43**
A (1968-1998)	235	66	83	40	-152	-65
B (1999-2000)	95	27	66	32	-29	-31
C (2001-2004)	27	8	56	27	29	107
B & C	122	34	122	60	0	0

Table 2 Results and comparison between case 1 and simulation of case 1 in CMDC

Generally the comparison indicates that the CMDC underestimates the cost of the illustrated part of the Preservation Planning functions. Case 1 cost 358 pw, while the simulation outputs a cost of 205 pw – there's a deviation of 153 pw. The main reason is that the migration of A-archives should be classified as a normalisation. Even though the migrated archives did not come from Producers but from within the archive, the migration should have taken place years before.

If we therefore disregard the A-archives (see B&C) from our analysis and take a look at the three chunks Develop IP Designs, Migration Plans and Migration Software, we see that the CMDC is capable of estimating all of them with less deviation; albeit it pinpoints certain weaknesses pertaining to Develop IP Designs and Migration Plans (respectively 15 and 21 pw deviation), the explanation may be that the C-archives did not require IP designing, because they were almost ready for processing from the beginning. However the CMDC is designed to allocate a certain number of pw to the IP designing process. This teaches us that if data comply with the IP design at hand, the model should exclude this cost.

Regarding the Migration Plan phase, the deviation is explained by the fact that the CMDC does not presently reflect the size of migration projects well enough: There is a scalability issue here, especially when the migration project uses much man-power, which requires more management. Another interesting fact is that case 1 shows us that it is equally expensive to make migration plans and develop software while the CMDC underestimates the cost of the Migration Plans step.

Case 2

Table 3 shows the results from using the model on data from case 2 (the PDF-JPEG2000 migration). The model shows a cost of 33 pw per migration. Half of this cost in the model is due to development of migration software. In the case only 5 pw were used for the software development. A part of the difference between the model and the case is most likely due to the model overestimating the cost of developing software migration tools; even though we have taken into account that purchasing tools only cost approximately 1/3 of developing it yourself. Another part of the difference is most likely due to a difference in development culture between the model (based on OAIS) and the case. In the case the development was made with very little reporting and controlling. For example there were no official prototypes made for review by administration, nor any lengthy documentation.

In OAIS and therefore also in the model the function Archival Information Update performs the actual migration using the migration tools developed in Preservation Planning. The model estimates the cost of man power monitoring the process to 10% of the machine processing time. OAIS apparently assumes that once the tools have been approved by administration they are almost flawless as are the data to be migrated. In the case the migration has been performed with less than 10% man power for monitoring. One explanation for this difference is the extremely long machine processing time in the case compared to the model. In the model we estimate a machine process speed of 2.5 MB/s for the migration of the formats on the specific hardware in use,[4] but in the case the first 601 GB of data were processed at the very slow speed of 0.2 MB/s. So far 601 GB (2046 files) out of 6 TB have been migrated. One reason for the very slow process speed is probably the very large size of each file, making the process much slower than the total amount of data would indicate.

PDF-JP2000 Year	1	2	3	4	5	6	10	15	20
Monitor Designated Community	9	9	9	9	9	9	9	9	9
Monitor Technology	20	20	20	20	20	20	20	20	20
Dev. Preservation Strategies & Stand.	17	17	17	17	17	17	17	17	17
Dev. Pack. Designs & Migration Plans	0	0	0	0	33	0	33	33	33

Table 3 Simulation of case 2 over time. Units in pw

Compared to case 1 it is important to emphasize the minimal amount of manual work to monitor the migration. The almost flawless migration process is most likely due to a high degree of compliance with the specification, i.e. very few invalid formats in the data. We estimate the compliance in the case to be above 99%. In case 1 concerning the A-archives a massive amount of man power has been used during migration due to a very low rate of compliance (approximately 20%).

Discussion

The Cost Model for Digital Curation is designed to provide a consistent approach for estimating full economic costs of providing digital curation to digital materials preserved within normally efficient and OAIS compliant archives. Envisioned users are practitioners and experts in digital curation. The model is structured on the functional breakdown described in the OAIS standard. While we agree that the abstraction level is not the same for all functional entities (Eggers, 2006), it is our experience that OAIS in relation to digital migrations provides a level of detail equaling or exceeding that of other functional models used for costing. While using the OAIS breakdown, the focus of CMDC is cost based. As such a number of OAIS components are not relevant for the CMDC and have thus been excluded; others have been combined.

The CMDC is applicable for measuring actual baseline costs, i.e. cost based on experiences (ex-post), but the activity based approach also allows tracking costs over time. The precision of the model is low, but regarding the exact degree we dare not make any conclusions, and we invite the readers to simply consult Table 2, which compares case 1 with its simulation. When used for estimating future cost (ex-ante) the precision is even more uncertain due to the challenges posed by handling the predictive element, which influence various aspects of the model. One is the life expectancy of formats, which will influence the required migration frequency. Another is estimating how much software will be available in the future, either as open source or for purchase, and how much has to be developed. A third is estimating the complexity of future formats.

Concerning the complexity of formats, several attempts to define it have failed, and Planets' conclusion seems to be widely accepted: The notion of "digital object complexity" has been disregarded as non-objective and non-scientific (Planets, 2007). Yet we believe that establishing differentiated complexity factors is necessary. These should, however, be based on measurable components: The Format Interpretation factor is thus based on the amount of documentation and its complexity. The LIFE Costing Model operates with a linear scale, dividing format complexity in 10 (McLeod et al, 2006, p. 96). We have then assumed that the complexity of migrations depends on the complexity of both the source and the destination formats' documentation. This has provided us with a factor expressing the migration complexity.

Estimations of how much time software development takes is also based on the Format Interpretation factor. Other software cost estimation tools, such as COCOMOII (Boehm, et al., 2000), use experience from similar projects and qualitative parameters or count function points for estimating the cost. This approach was however not viable for our purpose, because of lack of similar projects and uncertainty of what to develop (e.g. migration tool for an unknown destination format).

Regarding the processing factor it is the norm to assume that the migration process is automatic. The cost of an automated process *is* quite low, but if the data to be migrated does not comply with its contemporary preservation requirements because of lack of quality control (at Ingest or previous migrations), the cost of

[4] HW: Pentium 4 530 Prescott 3GHz, 2GB RAM, 3x7200 rpm SATA disk. Benchmarks for this machine can still be found at Tom's Hardware and other hardware sites. We have 20 machines for this type of migration, but only one was used giving these numbers.

processing the data may rise exponentially due to countless hours of manual fixing. The Dutch Testbed operates with the time it takes to repair or modify records and concludes that "This [repair] can be a slow and labour-intensive process that accounts for the majority of the costs." (Nationaal Archief, 2005, p. 11). A deconstruction of the processing phase in case 1 revealed that on average it took 1 person day to correct 1 faulty file. This example demonstrates the huge importance of compliance with preservation standards.

Migration frequency is also delicate. Formats may be migrated one at a time as they become exposed to the risk of obsolescence. However, this risk typically increases gradually. At the same time there are economies of scale in compiling format migration due the cost of developing IP designs, migration software, changing work processes and system setup. Depending on the IP design the cost of retrieving, updating and re-ingesting an IP also has important economies of scale, even though this is supposed to be fully automatic. We therefore assume that it is more likely that institutions compile format migrations to save cost.

We have assumed that on average preservation formats will be usable for 10 years, and every 5 years a migration is performed, migrating half of the content. This is of course a very rough estimate, considering the many different types of formats and the uncertainty of the technological evolution. Currently the model is not capable of varying this parameter, but we plan to enable this in future versions. For comparison The LIFE Costing Model estimates that the mean life expectancy for formats is 8 years, increasing with 0.1 year for every year that passes (McLeod et al., 2006, p. 93). Even though it is optimal not to migrate formats one by one every single year, case 1 shows that one should not wait too many years, as this become even more costly.

The test of the model on empirical data described in the case studies reveals that a very detailed and nuanced model is imperative. To exemplify this we will briefly discuss some of the most important points from running the model on test data.

A generic model should be able to handle migrations of many; highly complex formats as well as a few, simple ones. It should also be able to correctly reflect the cost of projects with small or larger staffing. Presently, the model cannot do this. This scalability issue does however exist on other levels too, for example concerning processing large or small files, as shown by case 2 where big PDF files process slowly, and small ones quicker. Case 1 also demonstrates that the model cannot yet correctly calculate the cost of a migration that most of all resembles a normalisation. The model also needs more parameters to reflect that not all preconditions are fulfilled. For example in case 1 where A archives complied poorly to their own IP Design and therefore cost many pw to correct manually, while in case 2 the content complied almost fully to the IP design, and the migration was performed with minimal manual corrections. Furthermore, the model has to handle dependencies better, because no cost critical activities stand alone. Their mutual implications are difficult to account for, but highly cost sensitive. The most obvious example from case 1 was the model's difficulty of calculating the high cost of the Migration Plan phase: In our formulas this phase is dependent on the IP Design phase, but not nearly enough on the Interpretation factor (i.e. format complexity).

Conclusions

The ambition is that the Cost Model for Digital Curation becomes sufficiently precise and generic to calculate the cost-critical migration activities performed by an OAIS compatible archive, providing estimates that are consistent across repositories.

The functional entities in OAIS have been analysed in detail and we have found the cost critical elements and the flow between them. Dividing the entities into cost elements, we have estimated those using measurable components. The central parameters are the Format Interpretation factor, the Software Provision factor and the Processing factor, that in general are based on the amount of person weeks needed to read the documentation in order to understand the formats and provide the migration software. Using the model on two cases shows that the model is usable, but needs further development, especially in handling deviations from the OAIS model's preconditions. An example is a low rate of compliance to the IP design, as seen in case 1, causing a massive quality control cost during the migration process.

The CMDC holds the implication that costing models are inaccurate. The degree of precision needs defining, but its importance is probably dependant on the purpose of the model. Nonetheless methods to increase precision are of high value, and one of the main objectives of CMDC is to pursue ways of fulfilling this goal. Thus, we will continue deconstructing cost critical activities until they become measurable and well-defined components. A manner of achieving this objective is to test the model on empirical data and iteratively improve the model. However, we also lack theoretical studies on for example how to establish parameters for migration frequency and format life expectancy.

Future work will focus on extending the model to include all archiving functions and handle various preconditions and dependencies, thus increasing the overall precision of the model. The downside of increasing the level of detail is that it will inevitably complicate the usability of the model. Therefore, we are aware that considerations should also be given to provide a more user-friendly interface to the model.

Understanding the nature of digital curation cost is prerequisite for increasing the overall efficiency, and thus achieving first quality for preservation of cultural heritage materials.

References

Ashley, K. 1999. Digital archive costs: Facts and fallacies. Proceedings of the DLM-Forum on electronic records, pp. 121-126. Retrieved September 15, 2009, from http://ec.europa.eu/archives/ISPO/dlm/dlm99/dlm_proceed99_03.pdf

Ayris, P., Davies, R., McLeod, R., Miao, R., Shenton, H., & Wheatley, P. 2008. The LIFE2 Final Project Report. Retrieved September 15, 2009, from http://eprints.ucl.ac.uk/11758/

Beagrie, N., Chruszcz, J., & Lavoie, B. 2008. Keeping Research Data Safe. A Cost Model and Guidance for UK Universities, Copyright HEFCE 2008. Retrieved September 15, 2009, from http://www.jisc.ac.uk/media/documents/publications/keepingresearchdatasafe0408.pdf

Blue Ribbon Task Force on Sustainable Digital Preservation and Access. 2008. Sustaining the Digital Investment: Issues and Challenges of Economically Sustainable Digital Preservation, Interim Report, pp. 36-37. Retrieved September 15, 2009, from http://brtf.sdsc.edu/biblio/BRTF_Interim_Report.pdf

Boehm, B., Abts, C., Winsor Brown, A., Chulani, S., Clark, B. K., Horowitz, E., Madachy, R., Reifer, D.J., Steece, B. 2000. Software cost estimation with COCOMO II. Englewood Cliffs, NJ: Prentice-Hall, ISBN 0-13-026692-2

Consultative Committee for Space Data Systems. 2002. Reference Model for an Open Archival Information System (OAIS), CCSDS 650.0-B-1, Blue Book, (ISO14721:2003). Retrieved September 15, 2009, from http://public.ccsds.org/publications/archive/650x0b1.pdf
Egger, A. 2006. Shortcomings of the Reference Model for an Open Archival Information System (OAIS). *TCDL Bulletin Vol. 2*(2). Retrieved September 15, 2009, from http://www.ieee-tcdl.org/Bulletin/v2n2/egger/egger.html

Granger, S., Russell, K., & Weinberger, E. 2000. Cost Elements of Digital Preservation. *Working paper from the CEDARS project*. Version 4.0 October 2000. Retrieved September 15, 2009, from http://www.webarchive.org.uk/wayback/archive/20050410120000/http://www.leeds.ac.uk/cedars/pubconf/pubconf.html

Hendley, T. 1998. Comparison of Methods & Costs of Digital Preservation. *British Library Research and Innovation report 106*. Appendix 3 includes a table of digital preservation cost elements compiled by Beagrie & Greenstein Retrieved September 15, 2009, from http://www.ukoln.ac.uk/services/elib/papers/tavistock/hendley/hendley.html

InterPARES project. Retrieved September 15 2009, from http://www.interpares.org/

McLeod, R., Wheatley, P., & Ayris, P. 2006. Lifecycle information for e-literature: full report from the LIFE project. *Research report*. Retrieved September 15, 2009, from http://eprints.ucl.ac.uk/1854/

NASA Cost Estimation Toolkit (CET). Retrieved September 15, 2009, from http://opensource.gsfc.nasa.gov/projects/CET/CET.php

Nationaal Archief of the Netherlands. 2005. Digital Preservation Testbed. Costs of Digital Preservation. Retrieved September 15, 2009, from http://www.digitaleduurzaamheid.nl/bibliotheek/docs/CoDPv1.pdf

Planets. 2007. Report on tool and service approach, pp. 1, 32-33. Retrieved September 15, 2009, from http://www.planets-project.eu/docs/reports/Planets_PA4-D1_ReportOnToolAndServiceApproach-Final_Public.pdf

Sanett, S. 2002. Toward Developing a Framework of Cost Elements for Preserving Authentic Electronic Records into Perpetuity. *College & Research Libraries 63, 5*, pp. 388-404.