ENUMERATE Specialist Meeting 2

Measuring of the cost of digital cultural heritage

Place and Date
30 March 2012, London
Venue
CILIP offices (Charter East room)

Participants:

Name	Organisation
Hans Petschar	CENL/ The European Library, AT
Adolf Knoll	Czech National Library,CZ
Alison Macdonald	ICA,UK
Richard Ferguson	National Maritime Museum,UK
Natasha Stroeker	Panteia,NL
René Vogels	Panteia,NL
Richard Wright	PrestoCentre,UK
Max Hammond	MH-Strategy,DE
Trilce Navarrete	University of Amsterdam,NL
Gordon McKenna	Collection Trust,UK
Nick Poole	Collection Trust,UK
Katie Smith	Collection Trust,UK
Gerhard Jan Nauta	DEN,NL
Marco de Niet	DEN,NL
Jesús Domínguez	Digibis,SP
Maribel Campillejo	Digibis,SP
Monika Hagedorn-Saupe	Stiftung Preußischer Kulturbesitz,DE

Background to the meeting:

The second meeting was on 30 March 2012 in London. The reason of the meeting was to discuss the *measuring of the cost of digital cultural heritage*. . The goal was to discuss the need for research, the usefulness of common methodology and the difficulties encountered so far in ENUMERATE and the previous NUMERIC project.

The discussed issues were the following:

- Usefulness of measuring cost of digitisation of cultural heritage,
- Feasibility of a common methodology to calculate costs,
- Evaluation of Cost models to predict future costs,
- Recommendations for the ENUMERATE Thematic Survey.

Agenda and Minutes

Agenda item	Minute
Previous and current related research on the cost of digitisation	The meeting started with a presentation of the ENUMERATE Thematic Network and its ambitions, followed by a concise overview of related previous and current initiatives on monitoring the cost of digitisation. **a)NUMERIC:** In preparation of the NUMERIC Study (2008-2009) a thorough *analysis of the state-of-the-art in measuring the progress of digitisation of cultural materials* stands out. Zinaida Manzuch (Vilnius University) reviewed the major monitoring projects in the years 1997-2007. Manzuch's report covers such questions as: How is digitisation funded? What are the digitisation expenditures of heritage institutions? What are the major costs of digitisation? And what are the components of costs and cost factors influencing the amount of expense? One of Manzuch's findings was that: "None of the reports provided a methodology that was applied for calculating costs. The content of survey instruments sometimes suggests that respondents were faced with the necessity of calculating data themselves." Manzuch identified five (high level) cost factors: • digitisation objectives; • collection parameters; • quality of the digital surrogate; • application of cost reduction mechanisms; • technological infrastructure. Relating further specifications of these cost factors to four digitisation processes (preparation of digitisation, digital conversion and management, post-digitisation activities and project management and operation) enabled her to roughly specify high and low costs. For example the digitisation objective *access* resulted in relatively low digitisation costs as compared to relatively high costs associated with the objective of *preservation*.

The focus in these earlier research reports was primarily on a rather narrow concept of digitisation, as is clear from the cost factors identified by Zinaida Manzuch. These are about the conversion process as such. Little or no attention is paid to models that are currently gaining importance in the management of digital heritage collections, such as Total Cost of Ownership (TCO).

Manzuch's study also made it clear that there is a bias in answering depending on who is asking the question and why they are asking the question.

Regarding the NUMERIC Study itself, it proved to be difficult to assess the reliability of some of its outcomes, as a few example tables with aggregated results from the NUMERIC Study Report clearly show.

- A table about the financial resources identified shows very different figures for digitisation budgets as a percentage of the total budget of heritage institutions for different types of institutions. For example, the digitisation budget in Broadcasting institutes was estimated at 0.3% of the institutional budget, whereas for Archives/Record offices this was estimated to be 4.5%.
- In another table the median reported digitisation costs per broad object type (images alone, text & images, text alone, audio, etc.) varied between improbable extremes. For example the digitisation of one hour of film was estimate to be €128.89 in A-V or film institutes and only €16.75 in National libraries. Equally remarkable: in A-V or film institutes the estimated price of digitising Text & images (per page) was much higher than the price of digitising (per page) Text alone, whereas in Archive/records offices it was the inverse.

b) SIG-STATS

The Special Interest Group on Cultural Heritage Statistics (SIG-STATS) noted that substantial variance was found in the Numeric survey concerning the recorded costs of digitisation. Here are some explanations for this:

- the way questions about expenditure were phrased may have resulted in too much ambiguity;
- the inclusion or exclusion by respondents of 'invisible costs', such as staffing costs, costs of making policy or project plans, etc.;
- there have been different persons from the same institution answering the survey questions, or different types of persons across the institutions;
- the complexity of the organisations could vary considerably.

SIG-STATS concluded: It is not justified to make generalisations about digitisation costs solely based on the responses in the Numeric survey. But SIG-STATS also proclaimed: "calculating costs for digitisation projects is (…) complicated (…), yet for the management of institutions, policy makers and funding agencies a better insight in these costs would be profitable.

c) Curtis+Cartwright Report: *Understanding the costs of digitization*

An indication of the prevailing cost factors in some big digitisation projects comes from a research report entitled "Understanding the costs of digitization", by Max Hammond - one of the invited specialists at the meeting - and Claire Davies. The main cost factors identified in this research report are project management, content capture, metadata generation, procurement, IPR and service delivery. The report is an annotated breakdown of these general costs, with broad indications of how such cost items may

vary according to the particulars of a digitisation project.

CollectionsTrust, The Cost of digitising Europe's Cultural Heritage (2010)

Nick Poole introduced the cost study he has written for the Comité des Sages, whose task was "to make recommendations to the European Commission, European cultural institutions and any stakeholders, on ways and means to make Europe's cultural heritage and creativity available on the Internet and to preserve it for future generations". The underlying questions were: How can we increase the pace of digitisation? And how much would it cost to digitise all of European cultural heritage? CT's task was strictly to answer the second question, but in fact the mission entailed much more than "simply" calculating the cost of creating the digital surrogate. It's not the scanning costs that count - the creation of the digital counterpart -, what counts is the having of digital collections. CollectionsTrust set out to answer the question by posing three sub-questions: How many museums, libraries and archives are there in Europe? How many objects do they have in their collections? And how much would it cost to digitize all these objects? To make answering practicable all objects were assigned to a few very broad object classes: simple and complex books; simple 2D; complex 2D; simple 3D; complex 3D; simple and complex audio and video.

The outcomes of all calculations were presented in three sets of figures, both with low and high-end estimates:

- the costs of digitisation done in-house and project-based;
- the costs of digitisation when it is outsourced;
- the costs of digitisation in public-private partnerships (PPP).

The cost of digitisation in PPP were estimated to be lowest. In-house, project-based digitisation was rated as the most expensive solution. There were estimates based on doing the job at different paces. And although it was not strictly part of the mission, one of the goals of the study was to make cost of ownership visible.

On request of the Comité des Sages the costs of digitising heritage materials were compared with other public investments. In this report the digitisation of a percentage of all available books in EU libraries, or of all historic photographs in EU memory institutions were compared to the cost of delivering 100km of main road in Europe, or the cost of producing a Joint Strike Fighter plane.

An interesting but puzzling figure was what part of the European heritage collections should be considered not appropriate for digitisation. Another interesting topic had to do with "accelerator issues": the longer the digitisation process runs, the faster it gets; your workflow becomes more efficient and the cost will decrease.

Part of the assignment was the design and creation of an online calculator tool. The tool can be accessed from the Collections Link website.

The political message of the study was: digitisation of cultural heritage is big money, but not BIG big.

e) ENUMERATE Core Survey

In the ENUMERATE Core Survey, which ran from January to March 15th 2012, only some broad questions concerning the cost of digitisation were included. The respondents were asked to give an estimation of the total budget spent on digitisation, the total number of staff involved, the main sources of funding and the factors included in

the digitisation budget. For this, the following entries could be selected: Staff costs, Taxes, Capital costs, Equipment costs, Selection costs, Digital conversion costs, Metadata enhancement, Metadata creation, Professional fees, Rights clearance, Digital preservation costs. At the time of the meeting, the results of the survey were not yet available.

f) ITHAKA S+R survey: *Sustainability of Digitized Special Collections*

A final example presented was a recent survey - the results of which still had to be published at the time of the meeting - amongst North American higher education libraries having special collections - the ITHAKA S+R survey on the *Sustainability of Digitized Special Collections*. In this survey the issue of organizational granularity is addressed. The survey is intended to "provide a detailed portrait of how libraries and cultural institutions are managing their digitized collections; offer useful financial data on what libraries and cultural institutions are spending not just to create, but to sustain these collections; shed light on the various sources of support - financial and non-financial - for these activities; [and] highlight some projects that provide valuable examples of strategies and tactics of use to others facing similar issues." The ITHAKA survey was broken down into three separate questionnaires, to be used at three different levels in each institution:

- The Institutional Perspective: to be explained by the director of the institution in a short (10 minutes max.) questionnaire.
- Collections in the Aggregate: a more detailed questionnaire on the institutional collections, taken together (60 minutes).
- Specific Collections: investigated the specifics of single collections. The questionnaire was designed for project leads (45 minutes approximately).

Comments on the previous research	After the presentation of these research activities, the participants felt the need to arrive at clear definition of digitisation.

After the presentation of these research activities, the participants felt the need to arrive at clear definition of digitisation.

NUMERIC used the IMLS definition: "...the process of converting, creating and maintaining books, art works, historical documents, photos, journals etc, in electronic representation so they can be viewed via computer and other devices."

SIG-STATS recommended to "...use the more generic Wordnet definition for 'digitisation' as a general starting point: "conversion of analogue information into digital information", and from there make explicit distinctions between 'digital descriptions' (or 'metadata') and 'digital reproductions' (or: 'representations') in future surveys."

The group agreed that most problematic in both definitions is the notion of "conversion". It is evident that born digital materials are not covered this way. It would be best make a clear distinction between 'digital' and 'digitisation/digitized' from now on. What ENUMERATE is presently trying to do is to provide a cost model for the transition the institutions are in, but the project also need to look at the costs 'post transition'. In that sense, it might be better to consistently talk about the cost of digital collections instead of digitised collections.

In relation to this, the *up-front costs* and the *on-going costs* as distinguished in the

ITHAKA S+R Survey were discussed. Ithaka defines these as follows: Up-front costs are "expenditures for first-time creation: costs incurred for the creation of the digitized collection up to the point of public launch." On-going costs are "expenditures for on-going maintenance, enhancement, and preservation post-launch." Not all participants found the distinction clear or useful. There is also the difference between 'costs' and 'expenses' ('out of pocket').

It was noted that up-front costs do not necessarily equal the actual analogue to digital conversion process ('digitized collection'); still this is on the whole a substantial component of these costs.The Ithaka distinction may be useful in relation to funding agencies: most of them will only support the actual conversion process and not the long term maintenance of the digital collection. However, professionally operating institutions would not separate these two. It was also noted that there are funding agencies that won't provide the up-front costs anymore, since the creation of digital collections are considered to be the regular business of the heritage institutions.

Some other issues related to doing a survey on the cost of digitisation that were addressed in the discussion:

- Expecting only one single response from a larger institution is problematic. Institutions are often segmented and respondents cannot easily cover the large institutions. Some know the costs of staff, equipment etc. Some know all costs. And some cannot specify any costs at all
- If you compare digitisation costs, you will not be able to get sound results over all if one interviewed person is (implicitly) talking about low quality digitisation (e.g. digitisation of text as an image) while another considered the outcomes of digitisation actions aimed at high quality output (e.g. digitisation of text as hand-corrected, machine readable text).
- Be aware that the type of budget employed for digitisation can have an effect on the specification of costs: "Some projects have a specific (recurrent) budget, in others some (one-time) budget - possibly more or less independent of the envisaged digitisation project - is found somewhere in the organisation."
- The digitisation objectives should not be overlooked: In the audio-visual world preservation is the main driver. From a preservation point of view you do other things (differently prized) than when you are merely digitising for access.
- In research on cost issues, there should be a clear distinction between recording actual costs and projecting future costs.

To summarize: on the whole participants agreed that it would be artificial to separate up-front costs from on-going costs and it would be equally artificial to discard born-digital materials and all sorts of acquisition of digital materials where the digital file is supplied by others than the individual institutions.

Furthermore, in any statistically sound follow-up research - in particular in the ENUMERATE Thematic Survey, planned for the second half of 2012 - an annotated overview of clear definitions of costs should be available.

Usefulness of measuring the cost of digitisation of cultural heritage	After this short historical overview, the group was asked to comment on the usefulness of measuring the costs of "digital collections". A few questions were formulated to structure the debate.
	What are the biggest benefits of having more precise information about costs of digitisation (and why not accept costs of digitisation as part of a wider budget)?
	The group identified several entities to which this kind of information would be useful, especially in relation to accountability: : the institution itself, the heritage domain, national governments, the European Commission and (other) funding agencies.
	Common for all entities is the benefit of an increased awareness of the importance of reliable cost estimates in a rapidly changing world: the heritage domain comes from a situation where local (physical) collections were maintained and extended for local audiences. The costs associated with that are fairly simple, at least in theory. With the digital, things get different. The creation of digital collections in a networked environment becomes a joint effort, irrespective of where the actual walls of museums, libraries and archives stand. Access to digital collections becomes global and so may be the audience. The costs associated with this change and are far from easy to specify.
	If we zoom in on the entities, there are several reasons why better data about costs and investments in digitisation are useful.
	Institution
	On the whole managers lack evidence to decide on priorities and what the return on investment could be for the institution. "We need toolkits to get best guess metrics." Organisational changes resulting from the digital developments make it necessary to explain a vision: where do you want to be with your institution?
	In addition to this, it is troublesome, without further specifications or guidance, to assess the costs of activities in wider digital infrastructures and frameworks, like for instance the costs of delivering data to Europeana. More precise information from institutions with experience in the field would be of value.
	In general, there is a lack of understanding of the digital transition. The associated risk is spending money on digitisation without a clear picture of the desired targets. Why are you digitising? You can't do the whole thing, because you cannot afford it. .
	The more you digitise, the more you will have to take future costs into account. We often do not know the consequences of our digitisation activities. Understanding costs now will help decrease future costs, e.g. the cost of digital preservation.
	Finally, the institutions need this kind of data for their accountability to society. Especially in Europe, where museums, libraries and archives are for a large part funded from public money, it is reasonable that they offer insight into how the money is spent.
	Heritage domain
	For the heritage domain as a whole more precise cost information could be a driver for collaboration. There is a clear shift from institutional to national ways of building digital collections and services. Having a harmonised vocabulary on measuring costs and maybe even shared practices is important for strengthening the sharing of resources and exploring opportunities for economies of scale and synergy.

National governments

Attendees noted that at the national level (and international levels for that matter), there is a shallow focus on mere quantities of digitised objects as an indicator of success in national policy. The topic of digitisation is oversimplified. Creating a baseline for evidence-based policy would be an important step forward.

European Commission

On a European level a lack of intelligence about the cost of digitisation is impeding informed decisions about where and how much to invest in (digital) cultural content - in line with the EU 2020 strategy.

This lack of intelligence also obscures statements about return on investment from European research and other programmes. For the European Commission, (as it is for national governments) Europeana is a success if measured in terms of mere quantities, but estimating the cost of building the Europeana database is only part of the problem. The model is/should be changing into a model of value creation. However, it is a well known fact that the Commission doesn't want complexity.

Funding agencies

The final entity that would benefit most from better intelligence about costs of digitisation would be the various funding agencies at all levels (local, regional, national, international). Benchmarking costs would be a good approach to assess value for money for funding agencies and tenders.

What cost figures are relevant?

A general remark, which surfaced several times during the meeting, is that the discussion about costs of digitisation cannot be separated from discussion about the benefits of digitisation. There is a moment of appraisal when you decide whether or not you want to keep cultural heritage materials. This is clear, for instance, in substitution digitisation, where you throw away originals or let them deteriorate and only keep the digital surrogate. What's being balanced here is the cost of digitisation versus the cost of having to preserve an analogue original. So input about the relevancy of specific cost figures cannot be made without paying attention to the value that is being created with the investments.

This is all the more important with born digital materials:Several attendees stressed the fact that the digital heritage landscape is rapidly changing. The old situation, in which individual institutions take care of the digitisation of their local collections and all the workflows and management processes associated with and resulting from that, is becoming less mainstream. From the Austrian National Library comes the example of a recent contract with a particular press agency, regulating the acquisition of off the shelf digital photographs covering specific topics in Austrian politics. It is important to know the costs associated with these lateral collection entries. As stated earlier, it would not be sufficient to characterise these developments as 'digitisation'. "Getting a hard drive containing digital stuff in your museum. Is that digitisation?" To be able to identify costs factors, a wider perspective than the traditional approach to digitisation is needed. A nice illustration of this are crowd-sourcing projects. The audience taking pictures that can be uploaded to some sort of publicly available service - think of Wiki Loves Art - is another example of why the old digitisation paradigm does not suffice anymore to understand cost factors.

With crowd sourcing you have opportunity costs as well. It is not for free. For instance, you need to have a tool to capture the information from the crowd. And once you have captured this kind of "wisdom", you will need systems to guarantee long-term sustainability. Money is needed for that too. The example of the Tate Gallery was given. The Tate needs to invest substantial amounts to enable user-generated content. These kinds of opportunity costs are often missing in cost queries..

When setting up a survey, it is useful to look at these opportunity costs and other costs that are less visible in a traditional digitisation budget, such as:

- Training costs
- Cost of delivering data to aggregators like Europeana
- Copyright costs (clearing)
- The cost of use and reuse of digital collections
- Aggregated costs of digitisation in a narrow sense
- Long-term costs of preservation

Feasibility of a common methodology to calculate costs

Is there reason to assume that cost structures in museums are different from libraries or archives? What are the overlaps, what the differences?

To be able to understand similarities and differences in cost structures across different types of heritage institutions, clear definitions are decisive, as a museum may use a different term than an archive while meaning the same thing. This can even happen between two institutions from the same heritage domain. . Some cost factors may be typical for specific domains, but in general, the conviction is expressed that the processes in museums, libraries and archives are comparable, all the more when object types are the same. Activities like the preparation for digitisation may be slightly different in different types of heritage institutions, but the retrieval and setup process are considered to be similar. Calculating the cost of digitisation is not about domains and subdomains, it is about types of objects and the contexts in which they function.

One of the attendees pointed at the role of commercial parties in digitisation. Are these specialised firms working only for museums, libraries or archives? They are not, is the consensus.

Finally it was remarked that differentiating along domains is the wrong message. Sector is actually a generalisation in itself. The reality is that there are many individual institutions that don't operate as a sector.

What cost factors queried in the individual institution can be extrapolated (in a sound way) to costs on a national/EU scale?

"While significant sums have been spent on digitisation to date, it has proven difficult to obtain figures on what has been spent on a national basis so far."

Caution is needed when extrapolating costs queried in the individual institutions to the national level. It will be difficult to prove that the results are statistically valid, because there are so many variables. But this does not mean that the extrapolation of costs is not worthwhile to pursue. Here too clear definitions and a convincing segmentation model are a prerequisite for further research. What are the characteristics that are decisive for the costs to be comparable? It is clear that not all cost factors can be easily extrapolated.

One of the reasons why extrapolation is useful is that large differences across

institutions and projects become visible. Because of this, prices of doing the digitisation job as charged by commercial companies might begin to level out. In Spain prices of digitisation work have gone down, partly caused by the fact that many companies are active in the field.This is a valid argument to include external parties in future questionnaires. In the current one, they are left out.

Another reason for extrapolation is to support learning and a knowledge flow on a national scale, e.g. from national institutions to smaller institutions. Roles will be defined differently in different types of organisations, and extrapolation may help to harmonize the definitions by creating a full set of characteristics and comparing differences.

This point led to the argument that the degree of digital preparedness is considered as decisive for measuring and extrapolating costs. The level from which you depart is an important variable. There can be so many other variables in the workflow, that comparing the total is less useful than comparing the steps. A suggestion for the thematic survey is to distinguish between beginning - intermediate - and advanced levels of digitisation experience, but all in all it will be difficult to establish some sort of base level across different institutions.

Comparing at the EU level will probably point out that there is a competitive component in the EU market. A question that countries will be confronted with is: how mature is the supply of their digital heritage services? Is there a national narrative about the state of digitisation of cultural heritage?

On the EU level, we should make a distinction between the needs of policy makers, who are only interested in generic figures, and involvement of the industry (which needs more granular figures).The industry is for instance an important party in achieving interoperability of information systems. Benchmarking interoperability on the EU-level would be very useful, as there is a lot of locked-in software. The greater the interoperability, the cheaper the digital collections may become.

An alternative to benchmarking the final cost - for which the necessary data may be troublesome to collect in many institutions - may be benchmarking time. In a digital media project at the National Maritime Museum (UK), for instance, the most expensive components in the digitisation process are the photo studio and the documentation department. In many institutions (in the UK) there is not enough awareness of the time (in terms of full time equivalents) spent on digitisation. It is suggested to introduce one or two questions drawing attention to the amount of time and effort it takes to do the digitisation. Normalisation of the awareness of costs would be a big plus.

It is concluded that extrapolation and benchmarking on a national scale is worthwhile, if only to get staff a bit more business oriented, as a counterpart to the passion about the collection they manage. It is pretty pathetic that few institutions are able to answer the question how much the scaling up of their digital services is going to cost.

Will there be such differences across the domains re. the quality of digitisation that a common methodology will not be possible?

A key issue where differences may occur is the quality of digitisation. Would different approaches to quality hinder a common methodology? The problem is illustrated by one of the findings from the NUMERIC Study Report:

"(…) the typical (median) costs reported by different institutional types for text only documents ranged between €0.10 and €0.80 per page; and audio materials between

€6.42 and €78.84 per hour".

This variety, is it contextual? Or is that just a fact of the data that we happen to have? Such differences are usually attributed to the fact that digitisation quality may vary considerably across digitisation projects, while in survey questions this is usually not taken into account.

Several attendees emphasised again the importance of clear definitions. Questions must be posed in such a way that there is no risk of mixing e.g. the cost of digitising modern documents with the cost of digitising old manuscripts. And because of poor definitions or poor segmentation of the cost factors, the hidden costs - think of capital cost, the cost of equipment, staff time - are often not included by default.

Costs are not comparable if you don't involve the desired outcome of digitisation (images of text versus OCR-ed text for instance). It is suggested that perhaps ambiguities become less troublesome if questions focus on the results of the digitisation process rather than on the input side of the it. This would be in line with the outcomes of the first ENUMERATE Specialist Meeting: measuring/counting the number of analogue objects to be digitised is more troublesome than measuring the number of digital objects after the conversion.

Some variables (cost factors) that were considered to be decisive for the quality and thus the cost of digital collections:

- Experience
- Infrastructure
- Material types (old/new, fragile)
- Intended use (incl. rights management; long term / short term availability; level of quality)

How do you measure these variables? How do you pinpoint an institution on where they are on the scale of these variables? One way would be to distinguish levels of complexity. Here a reference was made to the "Digital Preservation Capability Maturity Model" (DPCMM), as developed by Charles Dollar. Procedures similar to those of the DPCMM could be part of the survey tools of ENUMERATE: a way of specifying or defining an institutions digital maturity.

The discussion then concentrated on differences regarding intended use. The idea that upon the acquisition of an object you may decide that you intend to manage it only for a limited period of time, is not alien to an archive, but to a museum it could be quite a scary idea. The decisions taken about intended use may have large consequences for the cost of digital collections. This effect is amplified in the case of born digital collections. With digitisation you have control over the output, with born digital material the control is much less. An extreme example is born digital art, where the intended use may be defined in terms of perpetuity (improbable), in terms of a life span of 2 to 5 years, or in terms of a more or less thorough documentation of the original work (e.g. a video recording). Such different practices do not make it easy to come up with a common methodology.

It is obvious that there are two perspectives to consider: the organisational-centric perspective, and the object-centric perspective. Discussions about what goes into the process; what state the digital object is in in each phase of the process; the transformations that happen to the individual object within this overall transition; what comes out in the end and what do we do with it: that is the object-centric perspective. It's

a different way of seeing the problem. It is necessary to establish that distinction clearly.The advantage of the object-centric approach is that you can get more specific about the costs.

How decisive are national differences between EU member states?

How determining is the national context, in particular the national economic context, on the actual cost of digitisation within the institutions?

All agree that labour cost is a defining variable. But this variable is modified by another variable, which is efficiency (availability of skills, professionalisation). When measuring the cost of scanning an item, it is evident that values get lower with the passing of time. This can partly be attributed to differences in speed of processing in the digitisation workflows. The variables are the same, but may be different for different (EU) countries. You cannot compare the digitisation of a page in the UK with the digitisation of a page in Serbia. National policy might constrain options that are available in other countries, e.g. allowing institutions to generate own income. So the national legal framework is also a factor to consider. But how do you account for that?

Nick Poole summarized: "So the model we are getting into here, is a sort of catalytic series of layers. We are talking about object variables, we are talking about institutional context variables, and we are talking about the national context variables." The challenge will be to define what these variables are and to quantify them. There will be differences in between digitisation projects. And such differences may be quite surprising and hard to overcome, as in some cases where you (as an institution) are not allowed to involve a third party contractor. But overall, benchmarking can help institutions in the learning curve to become more economic, and showing progress achieved in that respect.

What generic deductions can be made about the differences in costs between digitisation by private companies and in house digitisation?

Private companies provide cheaper services, but there is usually a loss in quality and material handling. Heritage institutions are usually inefficient in their digitisation activities, and thus more expensive. For this reason, the National Library of the Czech Republic has decided to have everything outsourced.

It was noted that there is a difference between outsourcing and public-private partnerships (PPP). In the first there is control of content (by the institution), in the second there may be a loss in control of content, as in partnerships there can be agreements about favoured licensing agreements, which may result in a (temporary) lock down of content.

Institutions tend to fool themselves about in-house digitisation. The actual in-house digitisation cost is often not specified to the full. As mentioned earlier, with outsourced digitisation, the sum total of costs becomes much clearer.

There are various approaches to understanding costs. What model(s) should be the starting point of a cross-domain methodology?

Next, the suitability for the ENUMERATE framework of four models that have recently been developed are discussed. These four models are:

Total Cost of Ownership: a "financial estimate whose purpose is to help consumers and

enterprise managers determine direct and indirect costs of a product or system."

- Digital Content Life Cycle: a widely used comprehensive framework which describes the stages of managing digital content projects and programmes, with the aim of "adopting a good practice approach for each to ensure continued success.". The stages identified in the model are Creating, Describing, Managing, Discovering, and Using and Reusing,
- Workflow Analysis: activity based costing, where the activities can be boxed.
- Supply-Chain Model for Digital Cultural Content : "work in progress" by Collections Trust, framing various actors, activities, and objects/instruments (e.g. publishers, scanning, picture libraries) under broad phases in the supply chain like: Strategy (strategic planning/business models), Extraction (acquisition/selection/etc.), Production (scanning/metadata enrichment/etc.), Management, etc.

The models are in effect different pictures of the same reality.

The group was asked to assess the appropriateness of these approaches.

The TCO and Digital Content Life Cycle models were considered to be the most problematic. Both models do not reckon appropriately with the role of social media, which changes the flow represented. TCO starts from the assumption that the one who digitises materials is also the one who owns the materials. It would be better to relabel the model into "Total Cost of Stewardship".

The Digital Content Life Cycle is rather exclusively focussed on the conversion process. It is unclear where born digital materials enter into the flow.

Workflow Analysis and the Supply-Chain Model are moving away from the objects and collections and are putting the cost issue into the organisational context.

Taking all of these four models into consideration the majority of the group favours Workflow Analysis most. But serious reflection should be given to defining how broad or how narrow the concept of 'workflow' is taken. It would be a mistake to stick to the traditional digitisation workflow, only mapping out the conversion process

It is expected that conventional practice will be overthrown and this is not just limited to the mere creation and management of digital collections. Take for instance the recently developed procedures of pro-actively collecting press photographs as part of the (digital) collection policy of the Austrian Library: in the old situation curators were specialists in certain fields who used to work for two or three years on one collection. In the new system, they will get thousands of photographs every month. - But it is not only a matter of redefining staff roles. There are also issues having to do with the institutional infrastructure. Both will completely change the library. For the Austrian National Library the Supply Chain model is really the model of a digital library at work.

A point was made that proactive collection building activities are often much cheaper than acting afterwards and acquiring or completing collections. The Austrian Library negotiated the right for giving direct access for scientific purposes, and for long-term preservation. "Which is also an advantage for publishers, because they know that in a hundred years they will have the original files if they need them. So this can be also a cost reductive procedure." It marks the transition to a completely digital library and digital collection process.

It is felt that this is really important: there will be less distinction between the point of

publication and the point of archiving. If publication and archiving are getting more and more together, that will become very efficient.

There is a need to give politicians a concept of the changes we are dealing with on an institutional level in relation to the bigger infrastructures. The comparison that comes to mind is the transition from a situation with lots of independent railway networks, into an (inter-)national railway infrastructure. A transition like that takes a one-off/once in a generation, huge capital expenditure. It should be articulated in terms of the benefits of what the integrated rail system will look like from the outside, in terms of economic competitiveness, growth, job creation, etc.

But being convincing is going to be difficult. All the evidence we have is related to a fragmented world. Being able to construct a convincing story, choosing the right model for surveying and monitoring, will be crucial.

Cost calculation tools to predict future costs	The final topic discussed at the meeting was the usefulness of cost calculation tools. Several tools have been developed in the past years to help calculate the costs of digitisation activities in general and digital preservation in particular. In the ENUMERATE report on tools for validation and harmonisation (September 2011), such tools have been identified and described. The ENUMERATE partners would like to hear from the specialists at the meeting if they thought that it would be useful to promote one or more of these cost calculators that may help the respondents to provide valid answers to the questions on costs. Or would it be better to construct a new, more generic calculator that would fit in the ENUMERATE methodology?

A general concern with the existing tools was that they usually assume that an institution starts from scratch with the digitisation activity. One needs to understand the logic of the entire model first and then map that to the practice in the institutions. But in daily practices one builds on previous work. The calculators traced usually don't presuppose existing infrastructure.

Furthermore, there are so many factors involved in digitisation, that most cost calculators do not cover all aspects of digitisation. For instance, the cost tool that the DEN Foundation developed, does not include costs for training staff.

The bottom line is that we need a template, based on the 'experience factor', to reflect the organisational budget and to identify all the factors and effects The meeting agrees that cost calculation based on work flow analysis would probably work best. Based on that workflow analysis, some representative cases can be made and with those in mind, a logical model can be constructed that can be tested on its predictive capability.

An issue that needs closer attention is capital cost in relation to equipment. Capital cost is "the opportunity cost of the funds employed as the result of an investment decision" (definition from Wordnet). Because of rapid changes in technology, investments in equipment for digitisation and storage of digital output can be a risk. Can a cost calculation tool include a risk analysis element and normalise this? It is agreed that identifying risks is important, but there is a difference between project risks and cost risks. If a cost calculator is meant for calculating costs of project based digitisation, the risk assessment will be restricted to project factors, such as will the project be finished on time and within budget. The more complex the programme is, the more you need to develop it from the beginning, including the acquisition process. This is part of good

	programme management.

People are getting more and more aware of the variables, so the models underlying cost calculation tools tend to become more and more complicated. Generalisations and normalisations are needed to compensate for this. A balance between the specific and the generic can be found, when a calculator is used to collect costs, not predict costs in a generic way. It is agreed that case based working is the only realistic way forward and that best guesses are acceptable in such an approach to collect costs. "Become empowered by what you know rather than get frustrated about what you don't know". Monitoring and auditing as projects go along would be good mechanisms to check the validity of the outcomes.

Two final points were made.

Most current cost calculation tools are developed for AV or text based materials, and they are less usable for an object approach, where the object is e.g. a museum object, a building or archaeological find. The result is that costs of digitisation of such objects can be misrepresented in surveys if these cost calculators are used.

Finally, a remark was made that not all institutions would be willing to reveal cost factors to the outside world. If we want a cost model to generate bench marking data, this could be a problem for reliability. |
| **Suggestions for future ENUMERATE surveys** | The final part of the meeting was a summing up of the outcomes of the meeting, and what recommendations can be made for the ENUMERATE Thematic Survey:

Recommendation 1:

Review the terminology and definitions on costs. Use clear, simple, short descriptions in all EU-languages to provide a clear understanding of what you are talking about. However, people usually don't read definitions in questionnaires. So the wording of the questions themselves is vital. Quality assurance of the translations is of the utmost importance.

Recommendation 2:

Distinguish between three levels for the collecting of qualitative information (cf. the Ithaka Survey):

- Collection level
- Institution level
- National level

Recommendation 3:

Scalability is an issue. E.g. maturity models need to have an eye for the right scale. Find a balance between involving both large institutions and smaller institutions. Both are important for understanding costs of digitisation. An online tool will be most convenient for all.

Recommendation 4:

It is vital to have personal relationships with staff in the institutions for advocacy of the ENUMERATE project and its relevancy for understanding and improving your own |

working practice. We need to ensure that when we approach institutions for this detailed information that we are seen as supportive and it is a collaborative process. In relation to that: ask for a person's decision making's responsibility in the organisation. Involve the director and ask for time to spend on the questionnaire.

Recommendation 5:

Think in the long term. ENUMERATE is a framework to document the transition of the sector. We are heading for a data standard for digitisation of cultural heritage, but is it realistic to expect that we can achieve that? More desk research and community involvement will be needed...

Recommendation 6:

Organise a 5th meeting, to bring the various topics that ENUMERATE addresses together such as costs and value.