ENUMERATE:
Measuring costs of digitisation

JCDL/IMLS-workshop
Models for Digital Cost Analysis

Washington, 14 June 2012

A family of projects

NUMERIC (EU; 2007-2009)

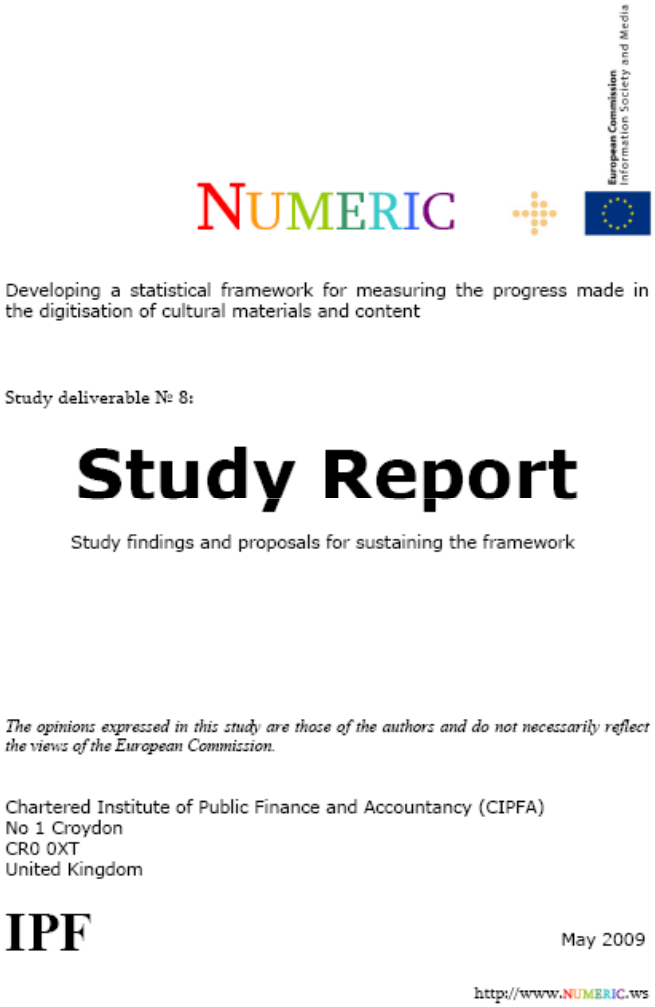
The Digital Facts (NL; 2008-2009)

More Digital Facts (NL; 2009-2010)

SIG-STATS (EU; 2009-2010)

ENUMERATE (EU; 2011-2014)

Cost study Comite des Sages (2010)



NUMERIC

Developing a statistical framework for measuring the progress made in the digitisation of cultural materials and content

Study deliverable № 8:

Study Report

Study findings and proposals for sustaining the framework

The opinions expressed in this study are those of the authors and do not necessarily reflect the views of the European Commission.

Chartered Institute of Public Finance and Accountancy (CIPFA)
No 1 Croydon
CR0 0XT
United Kingdom

IPF

May 2009

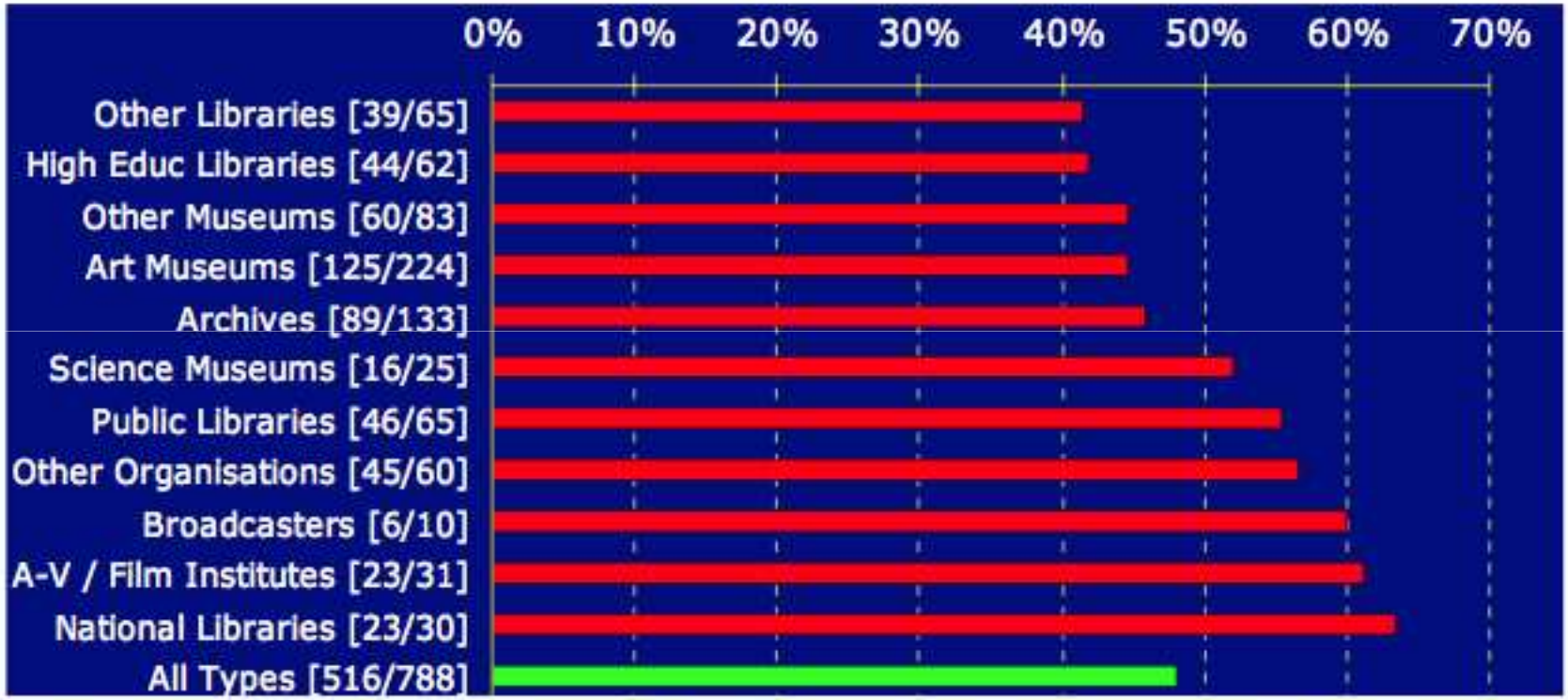
http://www.NUMERIC.ws

788 respondents from 26 countries

Main subjects:

- digitisation budgets
- costs of digitisation
- status information policies
- staffing for digitisation
- role of third parties (companies)
- progress of digitalisation
- most popular formats
- use and access rights

Figure 3 Institutions possessing a specific budget for digitisation activity



[The first figure in brackets indicates the number of institutions responding to the question. In this instance, the figure includes those who responded saying they had a budget heading, although it contained no resources. The bars in the diagram only represent the institutions with allocated resources.]

NUMERIC Study Report (2009), p.44

Table 10 Financial resources identified in budgets

Total for: Type:	Institution € millions [1]	Digitisation € millions [2]	% [2] / [3]
Archive/records office	223.0	10.0	4.5
Audio-visual or film institute	150.1	3.1	2.1
Broadcasting institute	2,578.9	6.6	0.3
Art museum (archæo/hist)	1,087.8	6.4	0.6
Science tech museum	171.3	2.1	1.2
Other type of museum	294.0	2.8	0.9
National library	676.0	25.0	3.7
Higher education library	517.5	2.0	0.4
Public library	250.1	2.0	0.8
Special or other library	530.9	3.9	0.7
Other organisations	543.1	16.2	3.0
All types	7,022.5	80.0	1.1

NUMERIC Study Report (2009), p.45

Table 20 Median reported digitisation costs

Type of cultural institution:	Images alone €/page [1]	Text & images €/page [2]	Text alone €/page [3]	Audio €/hour [4]	Film[18] €/hour [5]	Video €/hour [6]
Archive/records office	3.00	0.58	0.77	29.11	100.00	100.00
A-V or film institute	5.36	0.72	0.32	17.92	128.89	7.00
Broadcasting institute	...	...	...	78.84	678.10	100.00
Art museum (archæo/hist)	5.00	1.00	0.80	24.23	54.00	35.02
Science tech museum	1.50	0.25	0.45	16.71	16.37	13.02
Other type of museum	6.00	1.98	0.32	38.33	26.67	26.67
National library	1.00	0.50	0.48	58.42	16.76	55.00
Higher education library	0.72	0.81	0.10	6.42	8.00	5.46
Public library	0.96	0.65	0.70	9.70	...	...
Special or other library	1.03	0.59	0.19	35.00	1,040.00	120.00
Other organisation	2.00	0.87	0.55	39.00	22.22	35.13

NUMERIC Study Report (2009), p.59

NL- Digital Facts / More digital facts

Financed by the Dutch Ministry of Culture

Dutch contribution to Numeric (2008-2009)

Plus additional research on 3 topics (2009-2010):

- born digital heritage (methodology)
- costs of digitisation (cost model)
- webstatistics (guidelines)

Domain oriented approach

Collections Trust

The Cost of Digitising Europe's Cultural Heritage

A Report for the Comité des Sages of the European Commission

Prepared by Nick Poole, the Collections Trust

November 2010

"Digitisation is a process, and as with any process the actual cost depends both on the organisational context, the complexity of the material and the sophistication of the output."

"The estimated total cost of digitising the collections of Europe's museums, archives and libraries, including the audiovisual material they hold is approximately €100bn"

"The cost of preserving and providing access to this material over a 10-year period after digitisation would be in the order of €10bn to €25bn, provided that centralised repository infrastructure is made available for the purpose."

ENUMERATE Aims

1. The development of a vibrant and sustainable European **community of practice**.
2. The creation, promotion and development of a **statistically valid open methodology** for surveying the digitization, use, preservation and associated costs of cultural heritage materials in Member States.
3. The implementation of a multi-annual programme of coordinated **surveys** based on this methodology.
4. The creation and maintenance of an open, sustainable **data platform** to collate, analyse and promote the use of the normalized data and intelligence arising from these surveys.

Key topics in ENUMERATE

1. Size and growth (supply):
#analogue; #metadata; #digital reproductions; #planned; #born digital

2. Access and impact (demand):
access policies; usage statistics; methodologies

3. Costs (economics):
production; management; preservation; inhouse vs. external

4. Preservation (sustainability):
preservation plans; budgets; tools; partnerships

ENUMERATE Surveys

Year 1: CORE SURVEY

A high-level survey aimed at establishing a baseline of statistical information about Digitisation, Digital Preservation, Digitisation Costs and Online Access to Digital Cultural Heritage.

Year 2: THEMATIC SURVEY

A detailed and in-depth survey aimed both at enhancing the quantitative data and supporting it with qualitative information.

Year 3: CORE SURVEY

A reiteration of the initial survey, aimed at improving the quality of the data and providing trend information and analysis to inform policy development.

Preparing the surveys

Overview of Harmonisation Tools (to assist the respondents):

1. Terminology
2. *Cost Models*
3. Collection Type Analysis
4. Guidelines for Web Statistics

Five cost models:

1. DiCoMo: An Algorithm Based Method to Estimate Digitization Costs in Digital Libraries (2005)
2. Prestospace: Preservation Project Cost Calculator (2007)
3. DEN: Calculation model for digitisation costs (2009)
4. CollectionsTrust: Digitisation Costs Calculator (2010)
5. IMPACT: Digitisation Cost Estimator (2011)

www.enumerate.eu > Guidance

DEN's calculation tool

An extended version (incl. storage calculator) & basic version

Extended version (2009):
-Preparations
-Transport
-Scanning & photography
-Metadata
-Quality control
-Online storage
-Archival storage
-Development website
-Management website
-Promotion
-After care
-Other
-Risk percentage
-Graphics

Basic version (2012)
-Basic parameters
-Workflow steps
-Calculation report

Notes:
It was difficult to find enough institutions willing to test the extended version

The basic version is based on the IMPACT calculator for text digitisation

Extended version

Microsoft Excel - Gen_rekenmodel1_excl07.xlsx [Alleen-lezen]

1. Totale kosten per hoofdonderdeel tijdens en na de projectperiode

Kostenpost	Projectposten (incidentele kosten)	Kosten na looptijd pfj (structurele kosten)
Voorbereiden	72.259,89	
Vervoer	4.089,79	
Scanning en fotografie	545.764,34	
Metadata	60.065,51	
Kwaliteitscontrole	127.214,26	
Opslag online	7.356,01	37.607,87
Archiefopslag	10.125,24	139.967,86
Webdienst ontwikkelen	21.814,85	
Webdienst beheer		16.968,70
Promotie	14.036,30	
Nazorg	13.076,19	
Overige kosten	35.643,56	5.000,00
Subtotaal	911.445,94	199.544,43
Risico-inschatting	5%	5%
Risico-inschatting in geld	45.572	9.977
Totaal	957.018,24	209.521,66

Totale kosten genormaliseerd naar 1000 opnames

Kostenpost	Projectposten (incidentele kosten)	Kosten na looptijd pfj (structurele kosten)
Voorbereiden	35,40	
Vervoer	2,00	
Scanning en fotografie	267,37	
Metadata	29,43	
Kwaliteitscontrole	62,32	
Opslag online	3,60	18,42
Archiefopslag	4,96	68,57
Webdienst ontwikkelen	10,69	
Webdienst beheer		8,31
Promotie	6,88	
Nazorg	6,41	
Overige kosten	17,46	2,45
Subtotaal	446,51	97,76
risico-inschatting	5%	5%
risico-inschatting in geld	22	5
Totaal	468,84	102,64

2. Personele en niet personele kosten tijdens projectperiode (Incidentele kosten)

Kostenpost	Personeel	Niet-personeel	
Voorbereiden	63.029,12	9.230,77	
Vervoer	3.949,79	140,00	
Scanning en fotografie	354.288,31	191.476,03	
Metadata	48.334,74	11.730,77	
Kwaliteitscontrole	96.701,44	30.512,82	
Opslag online	5.509,86	1.846,15	
Archiefopslag	5.509,86	4.615,38	
Webdienst ontwikkelen	5.314,85	16.500,00	
Promotie	5.536,30	8.500,00	
Nazorg	3.845,42	9.230,77	
Overige kosten	6.643,56	29.000,00	
Totaal incl risico	628.596,40	328.421,83	957.018,24

Incidentele kosten genormaliseerd naar 1000 opnames

Kostenpost	Personeel	Niet-personeel	
Voorbereiden	30,88	4,52	
Vervoer	1,93	0,07	
Scanning en fotografie	173,56	93,80	
Metadata	23,68	5,75	
Kwaliteitscontrole	47,37	14,95	
Opslag online	2,70	0,90	
Archiefopslag	2,70	2,26	
Webdienst ontwikkelen	2,60	8,08	
Promotie	2,71	4,16	
Nazorg	1,88	4,52	
Overige kosten	3,25	14,21	
Totaal incl risico	307,95	160,89	468,84

3. Personele en niet personele kosten na projectperiode (structurele kosten per jaar)

Structurele kosten genormaliseerd naar 1000 opnames

Webdienst_beheer | Promotie | Nazorg | Overig | Risico-inschatting | Totalen | Grafieken

Basic version

1. DEFINE BASIC PARAMETERS	
Total number of scans or exposures	2,495,000
Storage space needed for master files (TB)	74
Yearly cost of storage per terabyte for master files	450 €
Storage space needed for access files (TB)	22
Yearly cost of storage per terabyte for access storage	350 €
working days in a year for FTE	220
working hours in a day for FTE	8
Overhead costs	10%

2. SET YOUR WORKFLOW STEPS (add steps if necessary)							
	Step 1	Step 2	Step 3	Step 4	Step 5	Step x	Workflow totals
Total amount of minutes needed for each digitised unit or object	0.1	0.6	0.5			0	1.2
FTE needed (destinct from digitised units)	0.1			0.1	2	0	2.2
Personnel available (in FTE)	2	10	10	0.1	2	0	24.1
The annual gross salary per FTE (in €).	42,000 €	18,000 €	36,000 €	24,000 €	24,000 €	0 €	
Total FTE needed	2.46	14.18	11.81	0.10	2.00	0.00	31
Total working days (time cost in man-days)	271	312	260	220	220	0	1,283
Personnel costs	103,433 €	255,170 €	425,284 €	2,400 €	48,000 €	0 €	834,288 €
Personnel costs including overhead	113,776 €	280,688 €	467,813 €	2,640 €	52,800 €	0 €	917,716 €
Non-personnel costs (Hardware etc)	5,000 €		20,000 €	50,000 €	2,000 €		77,000 €
Outsourcing / 3rd party costs				500,000 €			500,000 €

3. CALCULATION REPORT	
Planning	
Total man-days needed	1,283
Minimal turnaround of project in days	312
Minimal turnaround of project in years	1.42
Incidental costs	
Total personnel costs including overhead	917,716 €
Total outsourcing costs	500,000 €
Total non-personnel costs	77,000 €
Total incidental costs	**1,494,716 €**
Costs per scan or exposure	**0.60 €**
Structural costs	
Total storage costs master files per year	33,300 €
Total storage costs access files per year	7,700 €
Total structural costs per year	**41,000 €**
Costs per scan or exposure	**0.02 €**
Total incidental and structural costs (per year)	**1,535,716 €**
Costs per scan or exposure	**0.62 €**

We hope that the use of this tool will be self-explanatory, however, if you would like any further support in its use, please contact den@den.nl

Although great care was taken in producing this cost model, there is always a possibility of errors or omissions that have been overlooked. Please help us improving this tool by reporting any mistakes or omissions that you encounter in using the tool. You can send your feedback to den@den.nl

The user is responsible for the use of the figures produced by the cost model.

ENUMERATE CORE SURVEY #1
(Jan. – March 2012)

ENUMERATE Core Survey Questions

9.What is your institution's annual revenue budget?

26.Please estimate your annual expenditure on Digitisation

27.What is included in the budgets specified in the previous question?

28.What is the total number of paid staff engaged in the digitisation activities of your institution on an annual basis?

29.What is the total number of volunteers (in full-time equivalent) engaged in the digitisation activities of your institution?

30.From what sources are your digitisation activities funded?

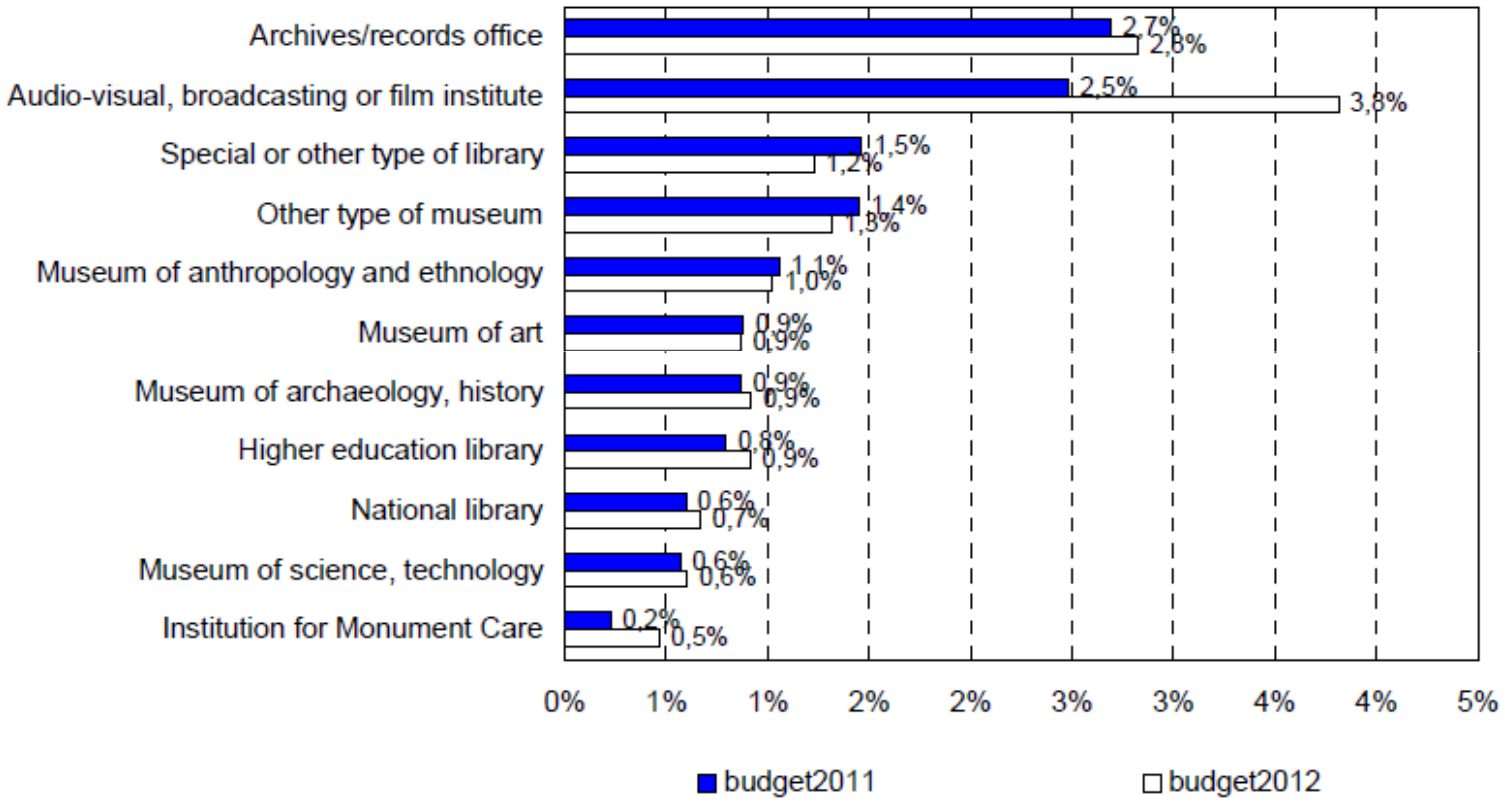


Figure 20: Annual budget for digitisation activities in % of total institutional budget (n=703)

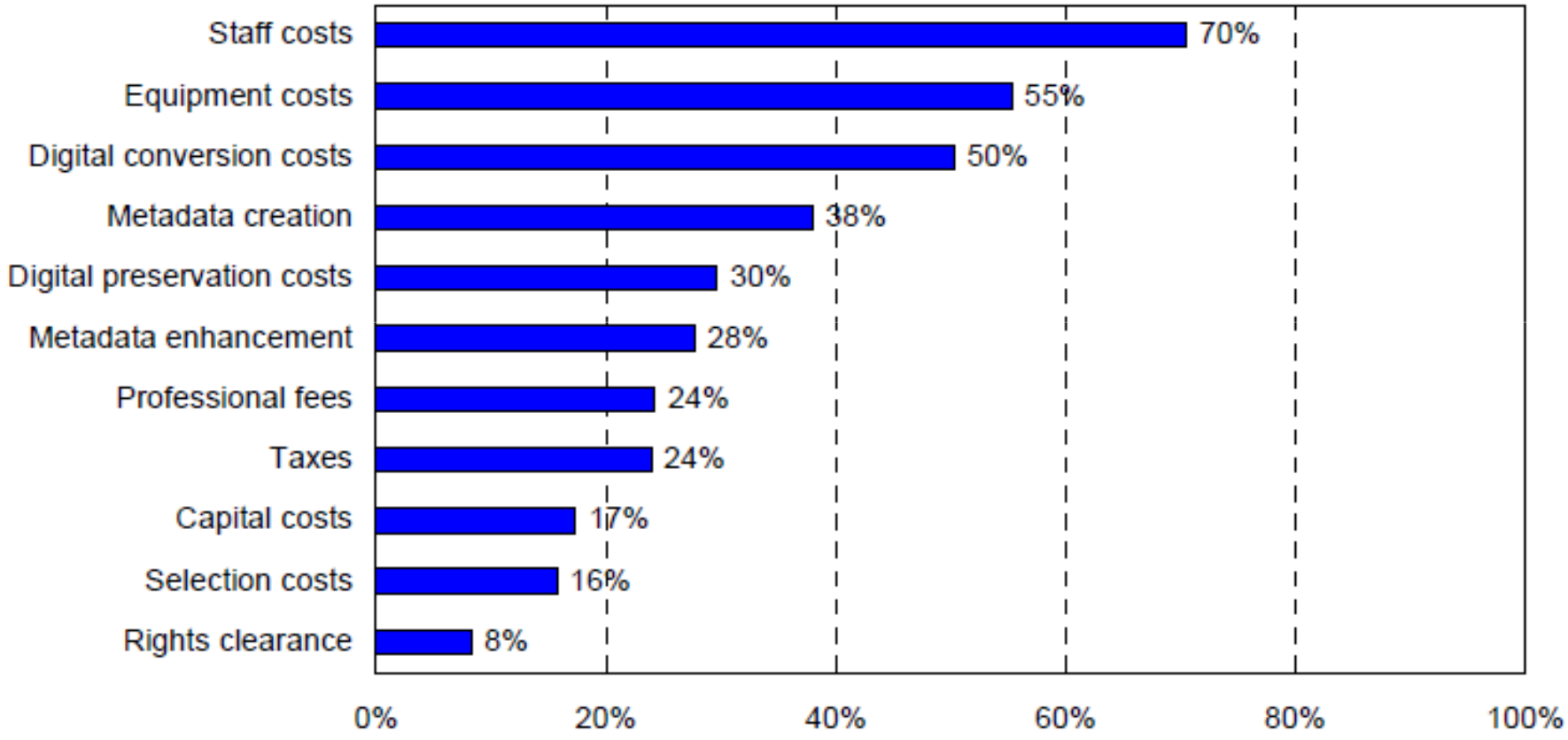


Figure 21: Included in budget estimation (n=1584)

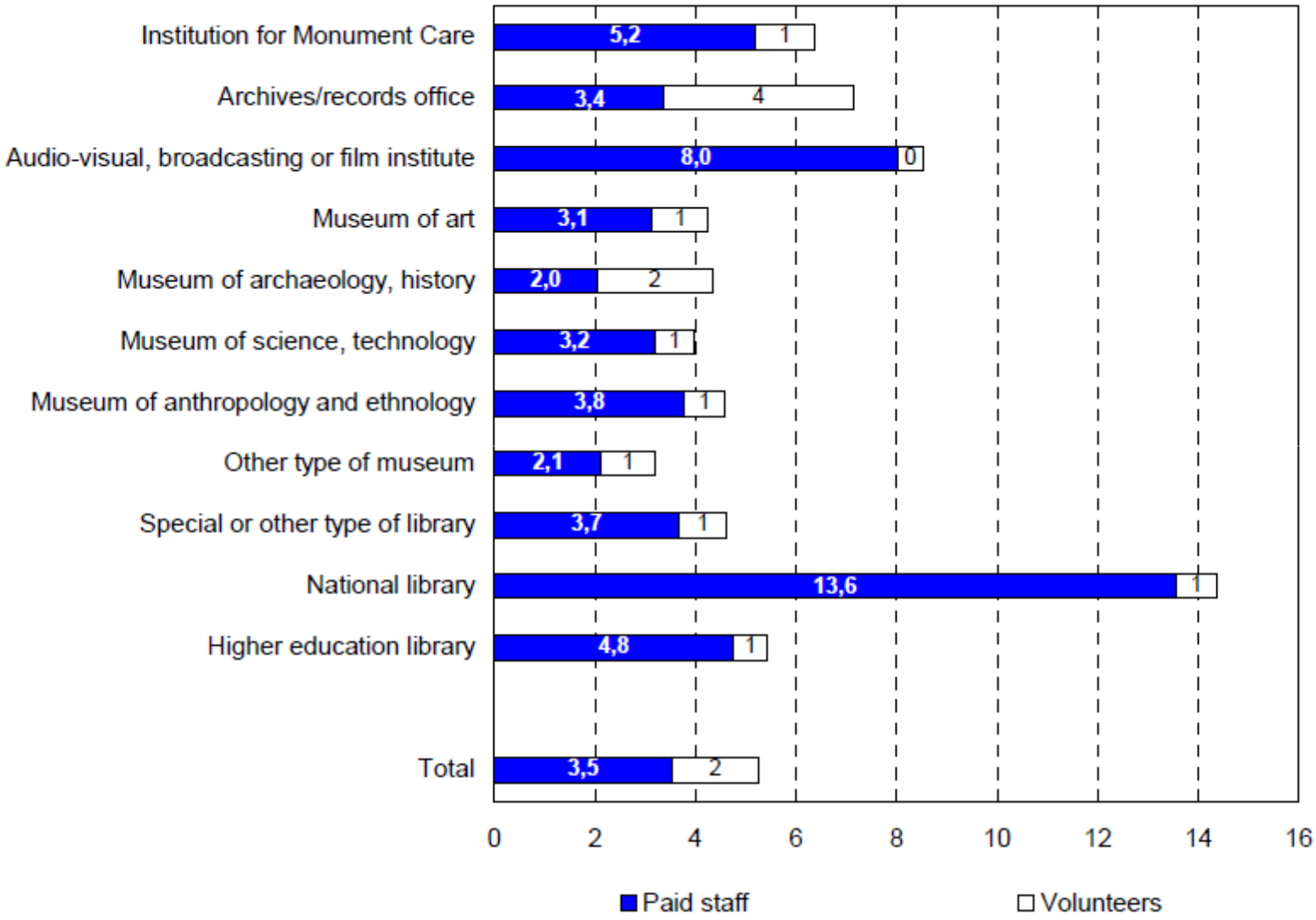


Figure 22: Staff involved in digitisation activities, in FTE (n=1584)

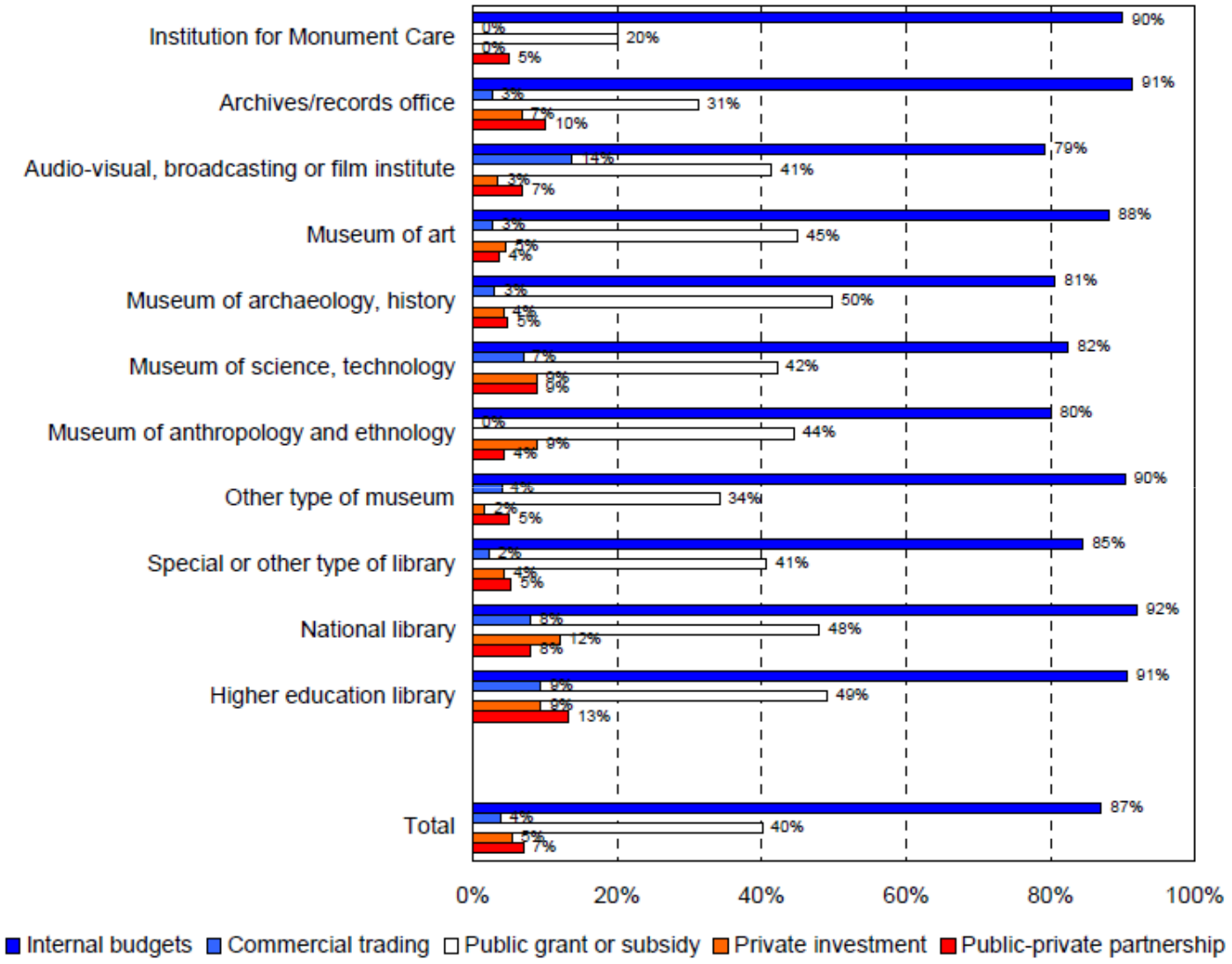


Figure 23: From what sources are the digitisation activities funded? (n=1584)

ENUMERATE THEMATIC SURVEY (Nov-Dec 2012)

Four specialist meetings to prepare Thematic Survey

30 March 2012: Meeting in London on measuring costs of digitisation

1. What is the usefulness of measuring costs of digitisation of cultural heritage? (USEFULNESS)

2. Is a common methodology to calculate costs feasible? (FEASIBILITY)

3. Can we build on existing cost models to predict future costs? (TOOLS)

Usefulness:

Five levels were distinguished: the institution itself, the heritage domain, national governments, the European Commission and funding agencies.

Three motivations:

1. Accountability: increased awareness of the importance of reliable cost. Did we spend the (extra) money well?
2. Policy planning: how much time and money do we need for future digitisation projects? How much digital heritage will become (potentially) available for Europeana? However: less attention for 'hard statistics' and more attention to value creation is needed.
3. Benchmarking: are our investments and quality criteria comparable to what others are doing? On the whole, there is a lack of understanding of the digital transition. We often do not fully understand the consequences of our digitisation activities.

Feasibility of a heritage wide methodology:

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.

Extrapolation is useful as 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.

The degree of **digital preparedness** is considered as an important parameter 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**.

Feasibility of a heritage wide methodology:

Variable cost factors that were considered to be decisive for the quality and thus the cost of digital collections:

- Experience
- (Available) infrastructure
- Material types (old/new, fragile)
- Planned outcome of digitisation (image, 3D, OCR, enrichted text etc.)
- Intended use (incl. rights management; long term / short term availability; level of quality)

For defining these, two perspectives can be considered:
the organisational-centric perspective, and the object-centric perspective.
The **object-centric approach** is thought to provide more specific information about the costs

Feasibility of a heritage wide methodology:

Four models were considered as starting point:

- Total Cost of Ownership
- Digital Content Life Cycle
- Workflow Analysis
- Supply-Chain Model for Digital Cultural Content

The majority of the group favoured **Workflow Analysis**.
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 remains to be seen how **born digital content** relates to existing workflows. Again: object driven approach is preferred.

Usefulness of cost models:

A general concern with the existing models was that they don't presuppose **existing infrastructure**. One needs to understand the logic of the entire model first and then map that to the practice in the institutions.

Most cost models do not cover all aspects of digitisation. A cost model based on **work flow analysis** would probably work best.

An issue that needs closer attention is capital cost in relation to equipment. Because of rapid changes in technology, investments in equipment for digitisation and storage of digital output can be a risk.
A cost model should include a **risk analysis element** to normalise this.

Usefulness of cost models:

A balance between the specific and the generic can be found, when a model 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 a model to collect costs

Most current cost models are AV or text based, and they **are less usable for an object approach**, such as a museum object, a building or archaeological find. The result is that costs of digitisation of such objects can be misrepresented in surveys

Not all institutions would be willing to **reveal cost factors** to the outside World. This may compromise reliability

Other work on calculating costs of digitisation in the Cultural Heritage domain

1. Zinaida Manzuch (Lithuania)
2. ITHAKA
3. JISC / Curtis+Cartwright

Groundwork for NUMERIC

Zinaida Manzuch, *An analysis of the state-of-the-art in measuring the progress of digitisation of cultural materials* (2007):

An analysis of 32 research reports and questionnaires until the year 2007.

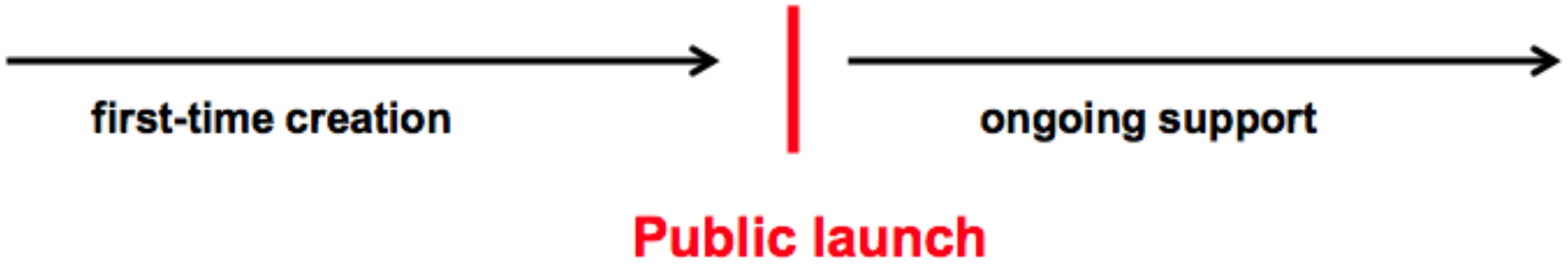
- How is digitisation funded?
- What are the digitisation expenditures?
- What are the major costs of digitisation?
- What are the components of costs and cost factors influencing the amount of expense?

ITHAKA S+R Survey: Up-front versus ongoing costs

•“Expenditures for first-time creation”: Costs incurred for the creation of the digitized collection up to the point of public launch.

•“Expenditures for ongoing maintenance, enhancement, and preservation”: All the costs incurred for the ongoing maintenance, enhancement, and preservation post-launch.

Estimates for these figures are okay!



JISC/ Curtis+Cartwright: "Understanding the costs of digitisation" (2009)

"Digitisation projects are distinct, and it is not possible to provide a formula (or even approximate figures) to cost a project."

"This report highlights areas which are likely to require significant resource to complete, and suggests approaches which will minimise the risk of a project running into difficulties, and maximise the efficiency with which it can be conducted." These are:

- Project management
- Content capture
- Metadata generation
- Procurement
- IPR
- Service delivery