



Digitisation in European Libraries: Results of the NUMERIC Project

Roswitha Poll

Universitäts- und Landesbibliothek Münster, Germany,
pollr@uni-muenster.de

Abstract

NUMERIC, a project of the European Commission, started out to define units of measurement and methods for assessing the current state of digitisation in Europe's cultural institutions (archives, libraries and museums). The aim was to show on the one side the financial input into digitisation and on the other side the progress achieved in digitising the national heritage.

The article describes methods and results of the project, with special consideration to libraries.

Key Words: digitisation; statistics; cultural heritage; libraries

1. The NUMERIC Project

NUMERIC was a European Commission project that dealt with the digitisation of the national cultural heritage in archives, libraries and museums. The project aimed at developing measures and methods for assessing and describing the current state of digitisation in Europe.

During the last decades, archives, libraries and museums have accepted the conversion of their large 'analogue' collections into digital format as an important task. Their main purpose is to facilitate access to the collections for the various potential user groups, e.g., researchers, teachers, or the general public. A second objective of digitisation projects is to preserve the original of an item without restricting access to it.

Although digitisation activities are manifold, often supported by national or regional funding programmes, it is nearly impossible to obtain reliable data about what has been achieved as yet. Statistics about digitisation projects and digitised collections can be found in the individual institution or within a funding programme, but not on a national scale. The definitions for what is counted and the methods of counting differ considerably between regions, countries and types of institutions. Therefore, even if there are statistics for digitisation, they cannot be grossed up for a national overview, and comparison between institutions or countries is not possible. This unsatisfactory situation was the background for the project NUMERIC.¹ NUMERIC intended to develop and test a dataset for assessing the status of digitisation in Europe. The goal was on the one hand to obtain a one-time overview, and on the other hand to produce an instrument that might be permanently used in European cultural institutions. Such an instrument had to include both data for input and output.

Governments, foundations and other funding institutions yearly spend considerable sums on digitisation projects. What those stakeholders and also the general public want to know is:

- What has been achieved in digitisation as yet?
- What did that cost?
- What remains to be done?
- What will that cost?

For an individual institution, it might be comparatively easy to produce such data, but for assembling statistics that could be added up to a meaningful and reliable national overview, it was necessary to establish clear definitions of what should be counted and how it should be counted.

The project was managed by Phillip Ramsdale of IPF (Institute of Public Finance, Chartered Institute of Public Finance and Accountancy, UK). The research team consisted of nine experts in the field of digitisation and statistics in libraries, archives and museums.

2. The Phases of the Project

The project ran from May 2007 until May 2009 and included the following phases:

1. The team evaluated existing websites and reports of digitisation projects and identified concepts, methods, statistics and definitions.

A first set of definitions was chosen, relying as far as possible on international standards.

2. Next, a 'pathfinder' survey was designed and tested in a sample of archives, libraries and museums. Based on a recall of 60 answers the structure and content of the survey were revised.
3. In a workshop in Luxembourg in April 2008 nearly 60 participants from 26 EU member states discussed the projected survey. A main issue was how to choose an adequate sample of cultural institutions in each country. The final concept was that coordinators in each country should identify 'relevant' institutions and select a sample of at least 30 such institutions per country. 'Relevant' institutions for the study were defined as those whose collections would add considerable value to the nation's digitised cultural heritage. Besides archives, libraries and museums, the selection should include film/audiovisual and broadcasting institutions.
4. After the 'relevant' institutions had been selected for each country and the questionnaire had been translated into 14 languages, the survey started about July 2008. Across all countries, 5.752 institutions had been identified as 'relevant' for the digitisation of the cultural heritage. Of these a sample of 1.539 had been selected that were asked to fill out the questionnaire. The response rate was 51%. Table 1 shows the samples and response rates for the different types of institutions:

Table 1: Number of relevant institutions, sample size for the survey and response rate².

	Relevant institutions	Sample	Responses	Response rate
Archives	848	262	133	51%
AV/film/broadcasting institutes	109	60	41	68%
Libraries*	1.932	690	222	32%
Museums	2.754	457	332	73%
Others	109	70	60	86%
	5.752	1.539	788	51%
* National libraries	64			
Higher education libraries	774			
Public libraries	778			
Special or other libraries	317			

3. The Questionnaire

Many details as to possible statistics for digitisation and methods of data collection were gathered from institutions that had already engaged in digitisation surveys:

- CENL (Conference of European National Librarians) that had started to collect digitisation statistics in 2007³
- EGMUS (European Group on Museum Statistics)⁴
- IMLS (Institute of Museum and Library Services).⁵

Especially for the library sector, there are ISO standards that helped to find adequate definitions for the data to be collected.⁶ They offer a wide range of definitions and data collection procedures for materials in library collections, different forms of electronic usage, and calculations of costs.

The types of cultural institutions that should be surveyed were defined as follows:

- Archive/records office
- Audio-visual or film institute
- Broadcasting institute
- Museum of art, archaeology, or history
- Museum of science and technology (or ethnology)
- Other type of museum
- National library
- Higher education library
- Public library
- Special or other type of library
- Other type of organisation

In order to assess the status quo of digitisation, the following aspects had to be considered in the survey:

- The number of analogue materials in the collections (print material, audiovisual material, manuscripts, museum objects)
- The number of digitised items
- File formats of digitised objects (e.g., TIFF, OCR ...)
- The costs of digitisation

- The sources of funding for digitisation
- The execution of the digitisation project (in-house, external contractors, partners)
- The accessibility of the digitised items for users
- The usage of digitised items
- The remaining task (relation of cultural heritage objects that have already been digitised to those that are eligible for digitisation).

A difficult question for the survey was to what extent analogue materials should be subdivided. This was easiest to answer for libraries, as library statistics are traditionally very detailed as to numbers and types of materials in the collection. Some categories like photos, posters, maps, and even paintings can be found in both museums and libraries and even archives. Archival records were not differentiated further in the questionnaire. Museum objects, if not classified as works of art, were subdivided into man-made artefacts and natural world specimens.

When it came to counting the number of digitised items, it was important to define the units of measurement. Printed material can, e.g., be counted in volumes, issues, pages or sheets, audio or film material can be counted in terms of physical carriers or hours of duration. Table 2 shows the measurements that were chosen:

Table 2: Types of material and units of measurement.

Type of material	Counted as
Archival records	Metres, volumes, or number
Books, serials	Volumes
Newspapers	Issues
Manuscripts	Number
Sheet music	Number
Microforms, microfilms	Number
Maps, photographs, engravings, prints, drawings, postcards, posters	Number
Paintings	Number
Any other 2-dimensional objects	Number
3-Dimensional works of art	Objects
Man-made artefacts	Objects
Natural world specimens	Objects
Other objects in collections	Objects
Film, video recordings	Hours
Audio (music and other recorded sound)	Hours

4. Survey Results

4.1. General

As this was the first time that a digitisation survey was tested across all EU member countries, it is not surprising that the results need interpretation and must be used with caution. The sample of 'relevant' cultural institutions for the survey was selected independently in every country, and though selection criteria had been defined by the project, they may have been interpreted differently.

The following problems were mentioned by participants or identified in the data analysis:

- The questionnaire is rather long and complex.
- The completion rates for specific questions differ. A number of questions could not be answered by the majority of participants.
- In spite of the definitions supplied with the survey, the interpretation of certain terms may have differed.

Nevertheless, NUMERIC has used the response data for an estimate of the present state of digitisation in Europe in order to get a first overview.

4.2. The Sources of Funding for Digitisation

The majority of respondents were able to name their sources of funding for past digitisation projects. Across all institutions, digitisation was funded as shown in Table 3.

To some degree, libraries show a deviating picture. Digitisation programmes, donations and 'other' sources (e.g., revenue from commercial arrangements) seem to be somewhat more important for some library types, as shown in Table 4.

Table 3: Sources of funding in %.

Source of funding	%
Own resources	62.1
Government programmes	29.9
Private donations	3.6
Other	4.5

Table 4: Sources of funding for libraries, in %.

	Own resources	Government programmes	Private donations	Other
National library	44.6%	36.0%	16.6%	2.8%
Higher education library	39.6%	30.7%	6.1%	23.5%
Public library	67.8%	21.5%	7.4%	3.3%
Special and other library	53.3%	41.1%	1.0%	4.6%

The questionnaire also asked whether the institutions have earmarked a special part of their budget for digitisation. Only 48% replied that they have such a digitisation budget, which — across all answering institutions — constitutes only a very small part of the general budget, namely 1.1%.

4.3. Costs of Digitisation

The survey asked for data on the median costs per unit in past digitisation projects as well as the calculated costs for currently planned digitisation projects. The second question is more relevant for estimating the costs of future digitisation, i.e., of the ‘remaining task’. In order to make the data for printed and manuscript materials to some degree comparable, units like ‘volumes’ or ‘metre of archival records’ were converted into ‘pages’. For the pages, unit costs were then calculated out of the projected resources for future projects (Table 5).

Costs calculated for audio and film material varied greatly between institutions and projects. Across all answering institutions, the following costs per unit were given:

- Audio: 30.00 € per hour
- Film: 55.20 € per hour
- Video: 34.29 € per hour

Table 5: Cost per page digitised in the calculation of planned projects.

Unit	Number of pages	Cost per page in €
Volume (book)	250	0.45
Volume (serial)	350	0.30
Newspaper issue	14	0.91
Manuscript	45	8.74
Sheet music	23	0.68
Metre of archived records of government/admin.	768	0.74
Metre of archived records of historic importance	300	0.80
Metre of all other archived records	1.868	0.80

When calculating such averages, it is of course a problem that not only the types of material can vary, but also the processes and technology used and especially the labour costs in different countries. The survey tried to find out how much staff is engaged in digitisation and what costs this staff time (calculated in full-time equivalent) would represent. Most institutions could not answer this question. Nevertheless, these data will be necessary for calculating the total costs of digitising.

4.4. Who Carries out Digitisation?

This question referred to all completed projects. Across all answering institutions, about one-third of the projects had been executed by external contractors and 62% as in-house work, the remainder by partner institutions. As to libraries, digitisation projects were carried out as shown in Table 6.

Table 6: Execution of digitisation projects in libraries.

Digitisation in %	In-house	External contractors	Partner institution
National library	48.1	48.8	3.2
Higher education library	34.3	20.5	45.2
Public library	70.1	27.8	2.2
Special and other	63.8	30.4	5.8

4.5. Accessibility of the Digitised Items

Previous studies had, for the most part, not dealt with the impact of digitisation projects on the population, namely the accessibility of the digitised items and the amount of actual use. NUMERIC asked questions about the accessibility of digitised materials via online catalogues and via the internet and for the institution's access policy (free, restricted etc.). The following questions were asked:

- Does the institution possess an online catalogue for its collections, and are digitised items distinguished in this catalogue?
Libraries have well developed online catalogues. In the responding higher education libraries for instance, 97.9% have an online catalogue, and 75.9% show the digitised items in the catalogue. Across all responding institutions, 67.4% have an online catalogue, of which 62.2% show the digitised materials.
- What proportion of the digitised material is publicly available on the internet? The proportion of digitised material available via the

internet showed a median of 20% for all responding institutions, but differed considerably between the types of institutions. The responding libraries reported that a considerable part of their digitised collection is available on the internet (70%).

- What is the access policy of the institution? Does it offer free and unrestricted access to its digitised collections, or do restrictions apply such as payment or, e.g., only in-house access?

About 50% of all responding institutions and 75% of the responding libraries allow free and unrestricted access. There may be restrictions for specified segments of the digitised collections.

4.6. Use Statistics for Digitised Materials

The questionnaire asked for the number of user requests for digitised material, either online via the internet or offline, e.g., on CD-ROM inside the library. It was to be expected that the reported data would not give a reliable picture. Usage data for electronic resources are still a problem in all institutions, even for commercial publications in libraries about which vendors supply COUNTER-compatible⁷ data. Therefore, the answers to this question were too inconsistent to be summed up. Especially as to online requests, the data indicated that single requests and longer 'virtual user visits' were not clearly separated. Though this first attempt at assessing the use of digitised material was not successful, usage data will be indispensable in future for showing the benefit of digitisation. Quantitative and qualitative data that could demonstrate benefits might be:

- The amount of use
- The types of users (e.g., researchers, teachers, specified groups)
- Projects based on digitised collections
- Research based on digital cultural heritage
- User opinions about benefits received by using the digitised collections.

5. The Present State of Digitisation and the Remaining Task

The question what remains to be done and what this may cost is probably the crucial one for all funders of digitisation projects. The survey therefore asked three questions:

- What part of your analogue collections has already been digitised?
- What needs to be digitised?
- What does not need to be digitised?

'Needs to be digitised' refers to preservation needs and/or to the collection's importance, justifying digitisation for better access by a larger clientele. 'Does not need to be digitised' refers to material that is insufficiently relevant for open access to a wider clientele. This concerns material that does not constitute important national cultural heritage, that is duplicated in other collections, or material that has already been or will be digitised by other institutions. The proportion of material that does not need to be digitised proved to be highest in libraries with their large collections of duplicate copies and lowest in museums where most objects are unique.

The percentages differ widely between types of institutions, but it is apparent that everywhere much remains to be done. Across all institutions, answers indicated that 19.3% of the collections have been digitised, that 30.2% do not need digitising and that 50.5% wait for digitisation. The percentages shown (Table 7) may have been compromised by varying interpretations of the term being 'digitised'.

Table 7: Progress of digitisation by type of library.

	No need to digitise (%)	Digitisation completed (%)	Digitisation outstanding (%)	Valid responses
National library	45.6	3.0	51.4	26
Higher education library	64.7	6.1	29.2	42
Public library	56.8	10.8	32.4	54
Special or other type of library	43.9	9.2	46.9	50

The survey also included a question as to the main purpose of digitisation, namely either preservation reasons or better access. But the answers showed that respondents did not differentiate between these two aspects.

6. Further Developments

NUMERIC ended in May 2009, but a number of possible measures have been proposed to further improve and utilise the results, and a working group will take up these issues. The main result of NUMERIC is certainly that data collection methods and statistics were developed and tested for assessing the state of digitisation in cultural institutions. Another important outcome of the project is that it has raised awareness of the importance of gathering data on digitisation.

The experience of the survey showed that there is still a need to refine and explain some terms and procedures. ISO will take up this issue in one of its committees.⁸ Another important challenge is that the definition of 'relevant institutions' for cultural heritage should be refined and applied consistently in all countries.

Many participants in the survey suggested that it should be shortened. For a first survey it was probably necessary to try for a full view. A follow-up questionnaire might be restricted to those questions that seem best suited to show the input and output of digitisation in Europe. Yet, even a shorter survey should not be restricted to counting numbers and costs of digitisation, but should also try to assess the impact and outcome of the digital cultural heritage on individual users and society, especially on learning, research and cultural identity.

Notes

¹ <http://www.numeric.ws/> (accessed February 4, 2010).

² NUMERIC. *Study report. Study findings and proposals for sustaining the framework*. May 2009, p. 25 [Not yet published].

³ <http://www.cenl.org/> (accessed February 4, 2010).

⁴ <http://www.egmus.eu/index.php?id=139> (accessed February 4, 2010).

⁵ IMLS. *Technology and digitization survey*. Available at: http://www.imls.gov/publications/TechDig05/Archives_Survey.pdf (accessed February 4, 2010).

⁶ ISO 2789 (2006), Information and documentation — International library statistics. — ISO 11620 (2008) Information and documentation — Library performance indicators. — ISO 5127 (2001) Information and documentation — Vocabulary.

⁷ COUNTER *Code of Practice for Journals and Databases* (2008); *COUNTER Code of practice for books and reference works* (2006). Available at <http://www.projectcounter.org/index.html> (accessed February 4, 2010).

⁸ ISO TC 46 SC 8 Information and documentation — Quality, statistics and performance evaluation.