



**The Project NUMERIC:
Statistics for the Digitisation of the European
Cultural Heritage**

Roswitha Poll
Münster, Germany

Meeting: **92. Statistics and Evaluation, Information Technology and
Preservation and Conservation**

WORLD LIBRARY AND INFORMATION CONGRESS: 75TH IFLA GENERAL CONFERENCE AND COUNCIL
23-27 August 2009, Milan, Italy
<http://www.ifla.org/annual-conference/ifla75/index.htm>

Abstract:

Archives, libraries and museums are busily converting their huge „analogue“ collections into digital format. Their main purpose is to facilitate access to the collections for the various potential user groups. Most digitising activities concentrate on the national cultural heritage, but statistics on digitisation have hardly ever been collected on a national scale.

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

The central task of the project is to develop a framework for the collection of statistical data that are most suitable to give a national overview of digitisation. For this, various aspects had to be considered:

- *The materials that are digitised (print material, audiovisual material, manuscripts, museum objects)*
- *Formats and standards of digitisation*
- *Costs of digitising*
- *Accessibility of the digitised items*
- *Use and users of digitised items*
- *The relation of cultural heritage objects that have already been digitised to those that are eligible for digitising*

Special emphasis was laid on assessing not only quantities in digitisation, but also the value of the projects for learning and research and cultural identity.

After testing the survey in a number of archives, libraries and museums, the project team collaborated with nominated experts in each country in order to apply the survey throughout Europe. The paper gives an overview of the survey results and of possible next steps for the projected framework.

1. Introduction

Archives, libraries and museums are the main institutions that collect and preserve the cultural heritage of a country. Since decades already, they try to convert part of their large „analogue“ collections into digital format. 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 aims at preserving the original of an item without restricting access to it.

Most digitising activities concentrate on the national cultural heritage. This includes texts, pictures, and sound, but also artefacts or natural objects. Though digitisation activities are manifold, often supported by national or regional funding programs, it is nearly impossible to get reliable data about what has been achieved. Statistics of digitisation are in most cases only collected in the individual institution or within a funding programme, but have hardly ever been collected on a larger scale. The data that are collected and the collecting methods differ considerably between regions, countries and types of institutions. Therefore, even if data exist, they cannot be grossed up for a national overview, and comparison between institutions or countries will not be possible.

This unsatisfactory situation gave reason for a European Commission project that aimed at finding measures for digitisation activities that could be used for a European overview and that might also be permanently used in European cultural institutions: NUMERIC¹

2. The project NUMERIC

As a European Commission project, NUMERIC addresses the digitisation issues in the European countries, and especially the digitisation of the national cultural heritage. Its goal is to define and test measures and methods for assessing and describing the current state of digitisation in Europe's cultural institutions. This includes the issue of financial input as well as the output of digitised items.

The questions that NUMERIC had to answer look simple at first sight. Governments, foundations and other funding institutions yearly spend considerable sums for digitisation projects. What the funders and the public want to know is:

- What has been done in digitisation until now?
- What did that cost?
- What remains to be done?
- What will that cost?

In order to answer these questions, the following facts have to be determined:

- Number and type of the “analogue” collections in the institutions
- Percentage of those that have been digitised at a specified point of time
- The resources (funds and staff time) spent up to that point
- Percentage of the analogue collections that should be digitised in future
- The resources required for that future digitisation

For an individual institution, it might be comparatively easy to assess such data. For collecting data that could be added up to a meaningful and reliable national overview, it is necessary to identify clear definitions of what should be counted and how it should be counted.

¹ <http://www.numeric.ws/>

The project was managed by Phillip Ramsdale of IPF (Institute of Public Finance, Chartered Institute of Public Finance and Accountancy). The research team consisted of 9 experts in the field of cultural digitisation and statistics in libraries, archives and museums.

3. The phases of NUMERIC

During the two years of the project (May 2007 until May 2009), the following steps were taken:

- During the first half-year, the team evaluated the existing reports and websites of digitisation projects and identified concepts, statistics and definitions. A first set of definitions was chosen, relying as far as possible on international standards.
- After that, a “pathfinder” survey was designed and tested in a sample of archives, libraries and museums. A recall of 60 answers allowed to judge on the survey structure and to make some important changes.
- A workshop in Luxembourg in April 2008 assembled almost 60 participants from 26 of the 27 EU member states. Issues for discussion were the contents and definitions of the intended survey, and especially the question how to choose an adequate sample of cultural institutions in each country. It was decided to have coordinators in each country that would 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 institutes.

A special issue at the meeting concerned the digitisation of monuments. As this would extend the scope of the survey that was intended for movable cultural heritage, it was decided to have a special smaller survey for agencies responsible for monuments.

After the institutions had been selected for each country and the questionnaire had been translated into 14 languages, the survey started about July 2008.

Over 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 differentiated as to types of institutions:

	Relevant institutions	Sample	Responses	Response rate
Archives	848	262	133	51 %
AV/film/broadcasting institutes	109	60	41	68 %
Libraries	2.754	690	222	32 %
Museums	1.932	457	332	73 %
Others	109	70	60	86 %
	5752	1539	788	51 %

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

The response rate was evidently higher where there is only a small number of institutions of a certain type in a country, e. g. AV/film/broadcasting institutes.

² NUMERIC. Study report. Study findings and proposals for sustaining the framework. May 2009, p.25

4. The questionnaire

As this IFLA programme deals specifically with statistics for the cultural heritage, this paper focuses on the choice of measures that would be best adapted to give a national overview of digitisation.

During the first phase of NUMERIC, the desk research, more than 30 reports and studies on digitisation activities were analysed in order to find measures that had proved useful. The result of this research was that the majority of surveys had concentrated on qualitative information, e. g. descriptions of digitised collections, so that hard data were missing. Most studies were one-time projects describing the state of digitisation at a certain point of time, not aiming at systematic data collection over years. And lastly, most projects considered only one type of institution.

Institutions that yielded most information about possible statistics for digitisation were:

- 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

In order to find definitions for the data to be collected, the adequate ISO standards were the main source to be consulted.⁶ They contain a wide range of definitions and counts for:

- Types of material in library collections
- Forms of electronic usage
- Costs (with and without staff costs and depreciation)

Definitions and counting procedures for archive material and museum objects still need standardising.

The expert group decided that the following aspects would be considered in the survey and would therefore need specified definitions:

- The analogue *materials* that are digitised (print material, audiovisual material, manuscripts, museum objects)
- The *number* of digitised items (problem of counting)
- *Formats* of digitisation (e. g. TIFF, OCR...)
- The *costs* of digitising (as well for past as for planned projects)
- The *funding sources* of digitisation
- The *accessibility* of the digitised items
- The *usage* of digitised items
- The *remaining task* (relation of cultural heritage objects that have already been digitised to those that are eligible for digitising)

In addition, the types of cultural institutions that should be surveyed had to be clearly defined. The following groups were identified:

- Archive/records office
- Audio-visual or film institute
- Broadcasting institute
- Museum of art, archaeology, or history
- Museum of science and technology (or ethnology)

³ <http://www.cenl.org/>

⁴ <http://www.egmus.eu/index.php?id=139>

⁵ IMLS. Technology and digitization survey. Available at:
http://www.imls.gov/publications/TechDig05/Archives_Survey.pdf

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

- Other type of museum
- National library
- Higher education library
- Public library
- Special or other type of library
- Other type of organisation

5. The types of analogue and digitised materials

The first question was which analogue materials should be counted separately. 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.

The main problem for assessing the number of digitised items was the question, in what units the items should be counted. Print material can e. g. be counted in volumes, issues, pages or sheets, audio or film material could be counted in physical carriers or hours of duration. Table 2 shows the measures that were decided on.

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

Table 2: Types of material and units of measurement

The results of the survey showed, that a great part of the institutions could not deliver data as to the number of specified materials in their collections. "...it is clear that for many institutions the quantification of their analogue collections remains as problematic as tracing their digital outputs".⁷ Yet, this information will be needed in order to calculate the costs of future digitisation that certainly differ widely between types of material.

⁷ NUMERIC. Study report. Study findings and proposals for sustaining the framework. May 2009. p.66

6. The sources of funding for digitisation

The questionnaire asked whether the institutions have earmarked a special part of their budget for digitisation. Only 48 % replied that they possess such a digitisation budget, which – over all institutions – constitutes only a very small part of the general budget, namely 1.1 %. The survey also tried to find out what part of the 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 are necessary for calculating the true costs of digitising.

The majority of respondents were able to name their sources of funding. Over all institutions, digitisation was funded as follows:

Source of funding	%
own resources	62.1
government programmes	29.9
private donations	3.6
other	4.5

Table 3: Sources of funding in %

In every type of institution, the own resources constituted the main funding for digitisation, while government programmes supporting digitisation projects made up 30 %.

7. Cost per item digitised

The survey asked for the currently planned digitisation projects and the calculated costs for these projects. In order to make cost data for print 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.

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

Table 4: Cost per page digitised

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

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

8. Accessibility of digitised items

Former digitisation studies had for the most part not considered the outcome of digitisation, namely the accessibility of the digitised items for users and the actual usage. NUMERIC

asked for the accessibility of digitised material 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?
Not all respondents have an online catalogue for their collections, and still less show the digitised item beside the analogue item in the catalogue. As was to be expected, libraries have well developed online catalogues. Of national libraries for instance, 95.5 % responded that they have online catalogues, and of those 88.9 % distinguish the digitised items.
Over 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 institutions, but differed considerably between the types of institutions. Libraries (70 %) and archives (48.5 %) have already a large part of their digitised collection available on the Internet.
- What is the access policy of the institution? Does it offer its digitised collections free (without payment or restriction), or with payment or restricted access, e. g. only in-house access?
About 50 % of all institutions and 75 % of all libraries answered that they allow free and unrestricted access. There may be restrictions for specified parts of the digitised collections.

9. Usage of digitised collections

When designing the questionnaire, the NUMERIC team realised that it would be difficult for the institutions to answer this question, and that answers might not be reliable enough to be grossed up. Usage data for electronic library resources are still a problem in national library statistics, even concerning commercial publications for which vendors supply COUNTER-compatible data. It was to be expected that many cultural institutions have not yet found reliable methods for counting digital usage.

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. The data as to online requests varied too much to give a reliable picture. Apparently, single requests and longer "virtual user visits" were not clearly separated. More reliable data were delivered for offline usage, where over all institutions a sum of 27.222.732 requests was counted.

Though this first attempt at assessing usage was not quite successful, usage data are indispensable for showing the benefit of digitisation.

10. The present state of digitisation and the remaining task

This issue, together with the topic of unit costs, is probably the most interesting one for all funding institutions. It is understandable that they want to know what state has been reached in the digitisation of the national cultural heritage, and what remains to be done. The questionnaire therefore asked three questions:

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

"Needs to be digitised" is explained either by preservation reasons and/or because the material is sufficiently relevant to justify digitisation to improve open accessibility to 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 which does not form an important part of the

national cultural heritage, which is duplicated in comparison with 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 is highest in libraries with their large collections of duplicate copies and lowest in museums where most objects are unique.

Over all responding institutions, the percentages were:

- Already digitised 19.3 %
- Does not need to be digitised 30.2 %
- Outstanding digitisation 50.5 %

These proportions differ widely between types of institutions, but it is apparent that everywhere much remains to be done. The survey also included a question as to the main purpose of digitising, namely either preservation reasons or open access for a broad public. But the answers showed that respondents could not differentiate between these two aspects.

11. Further development

NUMERIC ended in May 2009, but a number of actions have been proposed to further improve and utilise the measures and data collection methods developed in the project.

The NUMERIC team has used the response data for an estimate of the digitisation situation in Europe. This seems to be possible, if grossing up is done with caution and with consideration of different national circumstances. But a one-time view will have to be followed up by further surveys in order to assess developments in the countries.

The experience of the survey showed that there is still need to refine some terms and procedures. ISO will take up this issue in its adequate committee.⁸ Another important issue is that the “relevant institutions” for the cultural heritage should be defined more clearly and identified in all countries following these definitions.

A general comment to the survey by the respondents was that it might be shortened. For a first survey it was probably necessary to try for a full view. But a follow-up questionnaire could be restricted to those questions that best show the input and output of digitisation in Europe, including questions as to the outcome (usage and users) of the digital cultural heritage. Special emphasis was laid on assessing not only quantities in digitisation, but also the value of the projects for learning and research and cultural identity.

The paper gives an overview of the survey results and of possible next steps for the projected framework.

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