



---

**www.enumerate.eu**

---

## Editorial

This is the third issue of the ENUMERATE newsletter informing about latest developments related to the project and its mission: *Developing a sound basis for digitisation statistics concerning European cultural heritage institutions.*

This time we focus on the results of the first ENUMERATE Core Survey. The results are now available online. Here we provide a short summary. ENUMERATE also cooperates closely with others in order to reach a wide audience. In this issue we report about our co-operation with EGMUS, the European Group on Museum Statistics. EGMUS is currently considering using some of the questions which ENUMERATE used in its survey.

In the last few months ENUMERATE organised the specialist meetings on: "Measuring Digital Heritage Collections", "Measuring the Cost of Digital Cultural Heritage", "Measuring the Use and Impact of Digital Cultural Heritage", and will organise a fourth meeting on "Digital Preservation Practice". Short reports from the meetings are also part of this newsletter. Finally we introduce from our project partner member from DEN, Marco de Niet, in the section "Meet the team".

For all those, that cannot wait for the next newsletter, and for all others interested in the most up-to-date information we are also at Twitter and LinkedIn.

**Monika Hagedorn-Saupe, SPK**

Leader of the work package on *People (Community Engagement)*

Contact: [m.hagedorn@smb.spk-berlin.de](mailto:m.hagedorn@smb.spk-berlin.de)

*Contents: Editorial, 1 – Results of ENUMERATE Core Survey online, 1 ENUMERATE inspires, 3 – Specialist meetings on monitoring digital cultural heritage, 3 – Meet the team – Marco de Niet, 6*

---

## Results of ENUMERATE Core Survey online

The ENUMERATE Thematic Network has published its first report on the state of digitisation in Europe's cultural heritage institutions. The report is based on nearly 2000 'core' survey responses, from 29 European countries. With the support of the national coordinators the ENUMERATE Core Survey was carried out across Europe:

[http://www.enumerate.eu/en/about\\_enumerate/partners/national\\_coordinators](http://www.enumerate.eu/en/about_enumerate/partners/national_coordinators)

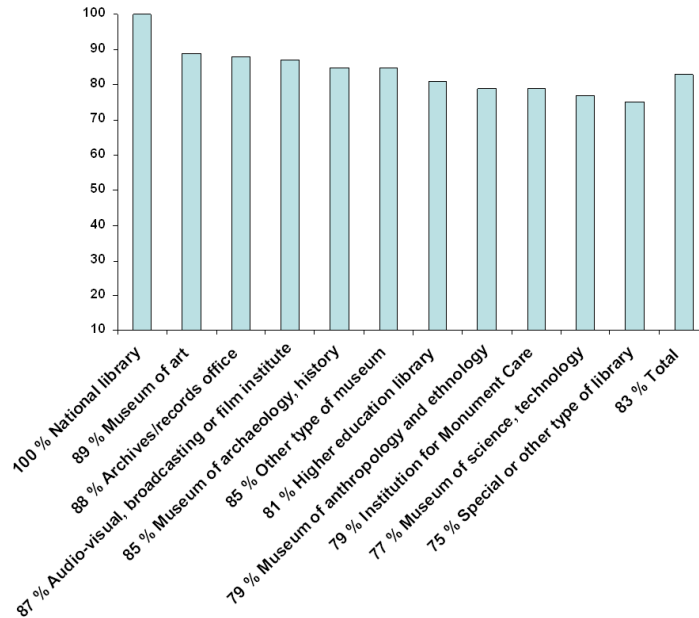
and was online from February 1st to March 31st 2012.

Here are some highlights of the report's findings:

83% of cultural heritage institutions have a digital collection, or are currently involved in digitisation activities. Each national library has digital collections, but only 75% of the special libraries own such a collection.



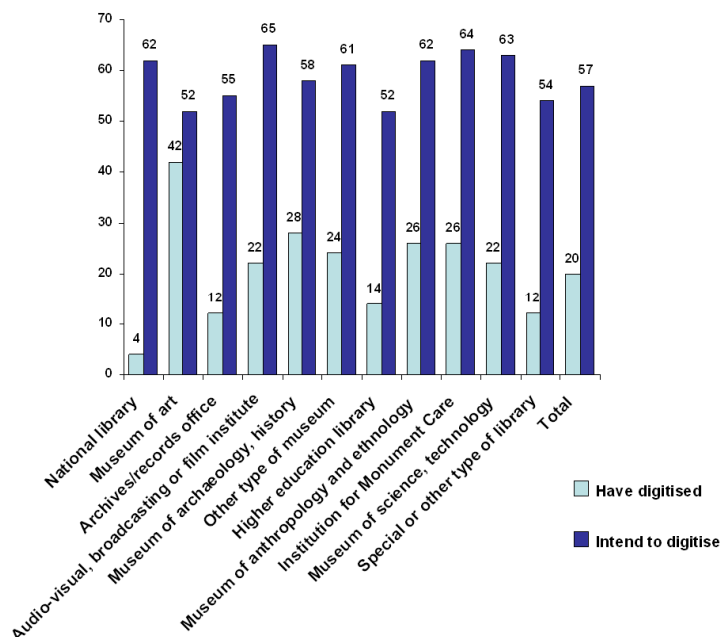
Does your organisation have digital collections or is currently involved in digitisation activities?



Cultural institutions owning digital collections have digitised on average about 20% of the whole collection, highest among them museums of art with 42%. National libraries still have a long way to go with 4% digitised of a target of 62% of the collection.

Of all cultural institutions on average about 57% of the whole collection still needs to be digitised (23% of collections, over all, do not need to be digitised).

Estimated percentage of heritage collection already digitally reproduced and percentage still to be digitally reproduced



52% of cultural heritage institutions collect born digital materials. A born digital object is one originally created digitally, i.e. digital images, video, sound, digital art, games or websites. Not surprisingly audio visual, broadcasting and film institutions mostly have born digital material (89%), followed by national libraries (85%). Museums of anthropology and ethnology and museums of science & technology have middle ranking in the collection of born digital material. 43% of museums of art and museums of archaeology & history have born digital material.

Born digital objects include: photographs; videos; DVDs; CDs; further audiovisual objects; audio-recordings; TV and radio programmes; film; archives and archival records; E-books and E-journals; web pages and websites; computer software and games; datasets; interviews; oral histories in PDF (or other formats).

Up to now only 34% of cultural heritage institutions have a written digitisation strategy. About half of national libraries and museums of art have written digitisation strategies, followed by museums of anthropology and ethnology. Only 25% of higher education libraries have such a written document but they are the ones who most carry out web statistics monitoring.

Web statistics are widespread as a tool to monitor the use of digital collection. About 85% of institutions use Web statistics to measure the use of their digital collections.

The report is the first in a series of three in the lifetime of the project. The full report is available for download at:

<http://www.enumerate.eu/en/statistics>

Later this year there will be an in-depth 'thematic' survey and an update to the 'core' next year.

## ENUMERATE inspires

EGMUS, the European Group on Museum Statistics has monitored surveys on museums in European countries for many years. The EGMUS consortium developed a "Standard

Questionnaire" parts of which are used in several European countries.

The current version of the EGMUS questionnaire (in use since 2008) contains only a few questions relating to digitisation.

At a recent meeting of EGMUS it was decided to expand the digitisation questions of the Standard Questionnaire with some of the questions ENUMERATE raised as a starting point.

## Specialist meetings on monitoring digital cultural heritage

### *Background*

During the first half of 2012 the ENUMERATE Thematic Network organised three specialist meetings dedicated to topics relevant for monitoring the status of digital heritage collections in Europe. The topics of these meetings coincide with the themes of the ENUMERATE Core Survey done earlier in 2012 and the Thematic Survey, planned for the end of this year. They are: the size and growth of digital collections; the cost and use and impact of digital collections, and digital preservation practice.

Taken together the specialist meetings aim to improve the framework for monitoring the overall state of digital cultural heritage in Europe. Each meeting was split up in three parts, addressing the usefulness of measuring, the feasibility of a common (i.e. heritage wide) methodology, and existing methods and tools.

The ENUMERATE consortium invited external professionals with relevant expertise on the topics to be discussed. Representatives of the network were also present.

### *Measuring digital heritage collections (The Hague)*

Measuring the size and growth of digital heritage collections was the focus of the first specialist meeting. The starting point of the discussion were the questions of the NUMERIC Survey of 2008 and the adaptations made to

these in the ENUMERATE Core Survey. All participants in the meeting agreed that it is very necessary to have some sort of detailed monitoring instrument, which can assist individual institutions in getting a grasp of their digital collections, while in the same time offering the possibility to compare the status of digitisation and digital collections across institutions, within the heritage field as a whole. The bottleneck in any survey is the relationship between digital collections and their analogue (physical) counterparts.

During the meeting a summary of monitoring initiatives up to date was given. It was generally felt that more care should be given to the criteria and guidelines defining what information objects may be considered as digital cultural heritage. For example, the view of an insider from the domain of sites and monuments will be different from the perspective of someone working in the domain of special libraries, or audio-visual archives.

Conventionally determining the size of digital collections is done using some sort of generic table, with more or less detailed object types in the rows, while in the columns the size of physical collections can be expressed, in relation to the size - and expected growth - of digital collections.

The problem with these combined physical/digital collections measurements is that the units of measurement are difficult to relate to. One physical object can have one or several digital representations, in varying qualities. Furthermore for born digital materials there is no analogue counterpart at all. A radical outcome of the meeting was that in the Thematic Survey the starting point in measuring digital heritage collections should be the digital realm. Furthermore, measuring digital collections requires a better determination of digital units (which may be simple or complex). It was found that determining digital units could best be done on the basis of a hierarchical classification that expresses dimensions and relationships between digital objects. In other words: the field needs a more thorough classification of digital materials.

Measuring the actual status of digitisation should be related to understanding the total costs of ownership (i.e. including the costs of digital preservation).

## Measuring the cost of digital cultural heritage (London)

The meeting started with a concise overview of related previous and current initiatives on monitoring the cost of digitisation. A recurring theme in the discussions provoked by this was that it would be artificial to separate the cost of the digitisation process as such (*up-front costs*), from the costs of having digital collections (*on-going costs*). This is consistent with the outcomes of the earlier meeting, in The Hague.

Attendees found that it would be equally artificial to discard born-digital materials and all sorts of acquisition of digital materials where others than the individual institutions supply the digital file.

It was agreed that 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 so many variables may have an influence on the figures. But this does not mean that the extrapolation of costs is not worthwhile to pursue. In any statistically sound research effort - in particular in the ENUMERATE Thematic Survey - an annotated overview of clear definitions of costs should be available.

One of the specialists remarked: "*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, as in some cases where you (as an institution) are not allowed to involve a third party contractor.

The suitability for the ENUMERATE framework of four recently developed models was also discussed. These models are: Total Cost of Ownership; the Digital Content Life Cycle; Workflow Analysis; and the Supply-Chain Model for Digital Cultural Content. Taking all into consideration of these four models the majority of the group favoured 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 - see the earlier remarks on up-front and on-going costs - to stick to the traditional digitisation workflow, only mapping out the conversion process. Born digital content may enter a heritage institution at points that were previously unimaginable.

The final topic discussed was the usefulness of cost calculation tools. A general concern with existing tools was that they usually assume an institution starts from scratch with digitisation activities.

Furthermore existing cost calculators tend to cover only a subset of the aspects of creating and sustaining digital collections, whereas heritage professionals are getting more and more aware of the cost 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. Case based working, it was agreed, is the only realistic way forward.

### ***Measuring the use and impact of digital cultural heritage (Madrid)***

Understanding the use and value of digital collections is a key driver to measuring access. Equally important drivers are accountability, making investments worthwhile in the long run, and the direction of future research. One participant in the meeting on use and impact expressed that it is important that cultural organisations change the way they are working, ensuring they are transparent about the use and impact of their digital heritage. He added to this that ENUMERATE should adopt the approach to present data in a very clear way online, this way creating a motivator for organisations to take further action.

With representatives of Let's Get Real and Museum Analytics at the table, much attention was given to recent initiatives in the field of web analytics. These projects are very pragmatic, measuring what can be measured with relative ease, using inexpensive tools and data that are already available on the web, to get at comparisons between institutions as to

the use of their websites and their activity in the field of social media. Projects so far are largely aimed at raising awareness in the institutions.



The counterpart to this is any initiative focussing on monitoring the diverse channels through which an individual institution may reach out to its audience. An example comes from the library domain, where an increasing need is felt to stop "just putting everything online", and to monitor more closely how people are interacting with digital content.

The tools used for this kind of research are e.g. database log files, commercial products like Piwik, and - in the library world - a suite of tools like the Toolkit for the Impact of Digitised Scholarly Resources (TIDSR), developed by JISC.

One of the questions raised was whether there is reason to assume that uses of digital collections from museums are different from libraries, archives or other heritage institutions? In the discussion it was noted that even in the domain of libraries institutions use statistics in various ways. Research libraries are likely to measure statistics on downloads/views of materials whereas public libraries are less likely to do so. It is important to know what is being done with the content: researchers use content within their daily work whereas the general public uses it as more of a leisure activity.

Discussion demonstrated that it is not necessarily about discovering a common cross domain methodology to analyse statistics, but about understanding different types of use of cultural content. In line with this it was put forward that one should not (in the Thematic



Survey) classify types of users, but rather types of use.

It was generally felt that, when it concerns the methodology of measuring use and impact, the digital heritage domain has a lot to learn from other domains. Again Google Analytics (GA) was taken as an example. A suggestion was that ENUMERATE could create very basic rules which demonstrate segmentation techniques within GA to the community.

This should include when to use GA and what to do to get the results needed. One could include best practices of how to implement and use GA, and what kind of reports may be worthwhile to produce.



An important and in many discussions surfacing point for attention is how to boost participation in the ENUMERATE survey(s). Repeatedly the role of the national coordinators and the use of social media was mentioned as being crucial in creating involvement.

### ***Measuring digital preservation practices (The Hague)***

The final specialist meeting in the series will be held in The Hague, on October 5<sup>th</sup>. As for the Thematic Survey, the final part of the meeting will be dedicated to a review of digital preservation should be a consideration in the entire digital heritage life cycle, and as this meeting is the final preparatory meeting full ENUMERATE Thematic Survey methodology.

### ***Wrap Up***

The consortium is most grateful for the contribution of the specialists in the separate meetings. The valuable input collected will be used to further the creation and improvement of tools that are part of the conceptual framework of ENUMERATE. The actual Thematic Survey will run between November 2012 and February 2013.

Detailed reports of the meeting will be available at:

<http://www.enumerate.eu>

\*..\*

## Meet the Team – Marco de Niet



Marco de Niet is the Director of Digitaal Erfgoed Nederland (DEN foundation), the Dutch knowledge centre for digital heritage. DEN is a government funded organisation to promote standardisation and monitor the progress of national digital

infrastructure for cultural heritage. Marco studied Dutch language and literature and History of the book at Leiden University. Before DEN, he worked at the R&D department of the Koninklijke Bibliotheek, the National Library of the Netherlands. In that capacity, he was responsible for Gabriel, the website of all the National Libraries in Europe in the 1990's. Gabriel evolved into The European Library, which laid the foundations for Europeana. Through Gabriel, Marco got involved in European cross-sectoral collaboration between archives, libraries and museums. With that wider perspective in mind, he speaks and writes about topics that concern all cultural heritage institutions, such as standardisation, digital preservation, business model innovation and, of course, accountability and statistics.

### **Role in ENUMERATE**

In ENUMERATE Marco is responsible for work package 2, Methodology. Together with his colleague Gerhard Jan Nauta, he is building the ENUMERATE Statistical Framework, which includes documentation and guidelines, questionnaires and supporting tools (e.g. terminology sources and cost models). In the past year, DEN has organised various specialist meetings, to involve European specialists in the creation of that framework. DEN is currently preparing the Thematic Survey, which will run at the end of 2012.

### **Why ENUMERATE?**

When asked why ENUMERATE is important Marco said:

*"My main driver to participate in ENUMERATE is that it provides a great opportunity to improve the management of digital heritage collections on an international level. It should be common knowledge in every institution how big their digital collections are, what the costs involved are and how often they are used. I see in my day to day work that many institutions are struggling to make the most of their digital collections. ENUMERATE can help them understand and utilize their digital assets better, both on a strategic and on a practical level."*