

	<p><b>ENUMERATE Specialist Meeting 1</b></p> <p><b>Size and Growth of Digital Collections</b></p>
<p><b>Place and Date</b> 27 January 2012, The Hague</p> <p><b>Venue</b> Digitaal Erfgoed Nederland (DEN)</p>	

### Participants:

Name	Organisation
Annelies van Nispen	Europeana,NL
Ingeborg Verheul	IFLA,NL
Rand Eppich	ICOMOS/EU-CHIC,ES
Gerrit de Bruin	APEX,NL
Monika Hagendorn-Saupe	EGMUS,DE
Lizzy Jongma	Rijksmuseum Amsterdam,NL
Annemieke de Jong	FIAT/ Sound & Vision,NL
Ivan Boserup	CERL,DK
Chantal Keijsper	Univ. Library Leiden, NL
Sjoerd Bakker	ENUMERATE,NL
Gerhard Jan Nauta	ENUMERATE,NL
Marco de Niet	ENUMERATE,NL

### Background to the meeting

The first Specialist Meeting of ENUMERATE was held at 27 January 2012 in The Hague. The topic was *“Improving the methodology for measuring digital heritage collections.”*. The primary objective of the meeting was to find the methodology to measure the size and growth of digital heritage collections.

During the meeting four main issues were discussed:

- Usefulness of measuring size and growth of digital heritage collections,
- Options to create a generic typology of heritage collections,
- How to deal with heritage collections that cannot be counted easily. (e.g. newspapers, monuments)
- Recommendations for the ENUMERATE Thematic Survey.

## Agenda and Minutes

Agenda item	Minute
<p><b>Usefulness of measuring size and growth of digital heritage collections</b></p>	<p>Before the usefulness of measuring size and growth was discussed those present thought it necessary to reflect on the varying interpretations of the word “digital heritage collections”. A few issues did arise:</p> <ol style="list-style-type: none"> <li>1. A common understanding of what is cultural heritage is necessary: in the Mediterranean area the tendency is to use the concept primarily for monuments and sites. In Northern European countries in particular museums and archives are associated with it. Common ground is in the idea that objects of cultural heritage are man-made things (as opposed to natural heritage), worthwhile preserving for future use. Curatorial care is part of the mission of cultural heritage institutions.</li> <li>2. What is meant by digitisation? Collections of digitised cultural heritage objects are not necessarily “digital heritage collections”, i.e. collections that are being built in order to safeguard the objects for future use, while offering access for present day users. A lot of digitisation is not meant for access. For example: in the domain of monuments and sites heritage objects are being digitised in order to make work easier for those preserving or repurposing the physical objects. Such digitised resources are not necessarily “digital heritage collections”. They are tools in the service of something else.</li> <li>3. But the ideas about what is and what is not cultural heritage is subject to change. Prints were once seen as mere reproductions, they are considered cultural heritage objects now. Books can evolve from mere vehicles of scientific research and scholarly debate, into objects of cultural heritage, expressing the convictions and beliefs of a certain era. Twitter messages are now often valued as worthless, but who knows how future generations will perceive tweets? As a consequence there may be diverging opinions about what to include or exclude in these measurement operations. It is part of the cultural heritage profession to come up with new /adapted selection criteria.</li> <li>4. Quality needs to be addressed here too. It is not useful to measure size and growth if nothing is known about the quality standards applied.</li> <li>5. Monitoring size and growth of digital heritage collections as such is only part of the story. As one of the participants said: “It is important to know what you have, but it is more important to know what you can (afford to) keep in the long run.” In other words: measuring the actual status of digitisation should be related to understanding and managing the total costs of ownership.</li> <li>6. Speaking of the costs of digitisation and long-term sustainability of digital heritage collections, some felt that the target groups in digitisation efforts should somehow come into focus: “What costs you want to spend on digitisation is dependent on your goals. Access may be an important goal. New research questions are an important goal for us [a research library] too, not [even] preservation in the first place.” Clearly the aims and missions of the organisations digitising cultural heritage may vary a lot.</li> </ol> <p>All in all the question of the usefulness of measuring size and growth of digital heritage collections was met with a resounding “Yes.” It was considered useful for various reasons:</p>

	<p><i>a. Efficiency of business</i></p> <p>Information about the size and growth of digital collections is valuable for an institution's decision-making process. Plans based on a reliable baseline of data can be carried out more efficiently and, as a result, cost effectively.</p> <p><i>b. Accountability for investments done</i></p> <p>Digitisation efforts are often financed with the aid of public money, grants or external investments. Reliable information about the resulting growth of digital collections is an invaluable instrument when reporting to these financiers.</p> <p><i>c. Projecting future investments</i></p> <p>Digital collections cost money to create, but also to sustain. Information about the size of digital collections can be used to create reliable estimates of the costs of long-term sustainability. This has implications for the selection process as well. If there is a limited budget choices must be made, but: "In order to choose you have to know: how much material is there? How much needs to be digitised? And this is not known yet."</p> <p><i>d. Understanding the transformation of digital information into digital cultural heritage</i></p> <p>The ideas about what constitutes cultural heritage change over time. We are still in the early stages of the emergence of digital heritage. It is unclear which born digital materials will be considered as cultural heritage in the future. In that respect it is worthwhile to note that in the digital world, everything is information. In the physical world, the distinction between an object and information about an object is clear. In the digital world, where complex digital objects emerge, the distinction gets blurred.</p> <p><i>e. Collaboration</i></p> <p>Harmonised data about the size and growth of digital collections is essential to efforts like Europeana, where the ambition is to digitize cultural heritage on a large scale. If that is the ambition, it is usefull to know how much has been digitised: "these figures/numbers of digital and digitized heritage materials are very important for Europeana."</p> <p>One of the participants stated: "It is very important to have a mutual vocabulary." And the aspiration is not new. In earlier research/surveys measuring the size and growth of collections was done by way of meticulously counting actual numbers of analogue objects (paintings, prints, books, photographs, three-dimensional objects, etc.) and relating these to the digitized outputs on the basis of some sort of common measurement unit. In reality there are no stable once and for all assessments of the relationship between the analogue and the digital realm. This is being complicated by the introduction of working digital. Is there a way out of this?</p> <p>In Part 2 of the expert meeting the focus was on feasibility of creating one typology for all sorts of heritage domains and what is needed to realise this.</p>
<p><b>Creating a generic typology of heritage collections</b></p>	<p><b><i>"Is it possible to create one typology that transcends the boundaries between the different types of heritage institutions?"</i></b></p> <p>To investigate the initial ideas about this, the group was asked to exchange initial thoughts on the question above. On the whole the participants had the idea that it will</p>

be possible to devise some sort of standard list of [digital heritage] object types: "Defining the object is a possibility across sectors." Some even take a radical stance: "You should not ask: is a typology feasible? It is a necessity. We need that typology."

In discussing the feasibility of asking staff in institutions to work with the typology a pragmatic consideration was expressed: it cannot be too refined. One attendee said: "If it could be on a high level it might work: "pieces of art" is ok... prints, engravings etc. is much too detailed." We should not ignore the fact that such a generic typology does not exist yet, even though professional collection management has been around for a long time. There must be a reason for that...

Another problem was recognized in the differing realities in the analogue and digital realm. It is clear that determining object types is not a simple issue in the analogue domain. As one remarked: "Some complex manuscripts are in effect archaeological sites... They are compound objects with different versions, interpretations, improvements, and commentaries. It is difficult to count under these circumstances. There are lots of layers." But in the digital realm interrelationships between (parts) of documents become even more complicated. Digital objects are compiled of different types of data. "In the audiovisual domain: it's all-in wrappers." In the library world one speaks of "complex objects".

A related complication is in the imbalance between the analogue objects and their digitized manifestations/reproductions. One analogue heritage object - the example was Rembrandt's painting of the Night watch in the Rijksmuseum - can be digitized in many different manifestations: high resolution images, details, spectrographic images, etc. Even tabular data, covering specific technical measurements may be seen as a digitised version of the artwork. If all of these are taken together as a conceptual **digital unit**, then a generic typology might be possible. So the real challenge seems to be the assessment of such digital units.

One might think that this does not count for simpler objects, like analogue and digital photographs, but this does not seem true. The digital heritage object may of course be restricted to immediate copies, more or less reproducing the features of the original. This way, if you have one painting, you can create one digital image of it. But it is more complicated than that: "even a photograph may have a negative and several differently made prints of it (dark/light/retouched etc.)."

**The participants agreed that trying to establish an unambiguous link between the analogue world of cultural heritage objects and the world of digitally reproduced heritage objects is too hazardous. For that reason it was proposed to leave the link between the analogue and the digital realm for what it is: "I think it might be necessary [to have some typology]. And it might be feasible, [but] only if we focus more on the digital experience than on the analogue experience." The digital unit is more or less the same across sectors.**

An additional advantage to this approach, focussing on the digital, is that it will not be necessary to create a separate typology for born digital materials.

To conclude: all participants underlined that it is possible to create one typology transcending the boundaries between domains if we start from the digital end, not the physical one. Big advantage is that we also have the born digital materials on board.

The next question posed was more pragmatic in nature:

***“Can a generic typology be more than a bureaucratic construct that can also serve institutions, e.g. with collection management?”***

As could be expected participants in the meeting thought that any typology should be relevant to the people working in the heritage institutions. One might even go as far as to set the ambition to create a “a definite qualifier list in the Dublin Core descriptor set”. Developing the typology should be a bottom-up approach, rooted in and with relevancy for the daily work in the institutions. But it was noted that there is a danger in this as well. From DEN comes the experience from research into born digital heritage (born digital art) that institutions have a tendency to avoid collecting “problematic” object types. It will also be necessary to look at the developments in information management outside the cultural heritage institutions.

Another complication was identified, while returning to the Night watch/Rijksmuseum example. If the full package of digital materials that in some way reproduce or document a masterpiece is considered as a digital unit, it may prove to be impossible to determine the boundaries of what a digital unit is.

Furthermore, it was agreed that museums are organized very differently: some focus on local history, others on art, others on archival materials. When different museums count objects, it happens in different ways. So it will not help to try to harmonize counting procedures by ignoring the different characteristics of institutions. The focus should be on (the value of) the objects. The unit to define is the individual object type.

***“What would be the uses of a generic typology? Who would benefit (most) from it?”***

In the discussion on specific uses of a generic typology those present noted the difference between benefits for the individual institutions and benefits in a national or European context. The group agreed that it is probably easier to define the usefulness on a collaborative level than on the level of the individual institutions. One participant said: “From the perspective of Europeana knowing how much there is in collections, is very useful information. [...] Europeana sends monthly reports to the EU commission about how much there is in the Europeana database. But it is also important to have data about: what is around there, what is in the various institutions.”

And what about the value for policy makers? If a generic typology be deployed it may be useful for strategy development on a national scale. It is an instrument that can be used to determine where there are flaws in national heritage programs. Born digital materials are mentioned as a striking case. If a typology is used to measure the progress being made in collecting born digital materials - or the lack of it! - that would certainly help policy makers in state bodies developing national action plans.

To resume, the following uses of a generic typology were identified:

- a. questionnaires, such as the ENUMERATE one
- b. collections management per institution
- c. collection and selection policy on a national scale
- d. improving search strategies (e.g. in Europeana)
- e. (reporting to) policy makers, collaboration activities, funding

***“Which level of detail can realistically be expected from a typology that can be used by all heritage institutions?”***

The group discussed a born digital typology produced by DEN during a Dutch survey

	<p>on born digital heritage (in 2009). One of the outcomes of this particular research project was that lists of object types that are too detailed, are disadvantageous for the response rate. More in general the scope of this is confirmed by one of the experts: "If there is too much detail: people in the (...) museum will sigh. (...) Every museum in the Netherlands has its own classification of materials and any museum would have difficulty counting objects using a general typology."</p> <p>Interestingly the DEN list of born digital object types elicited quite a few responses from the group of experts. Specific categories of born digital materials - e-mail files, AutoCad files, etc. - were missed, since these are in the DEN list classified under high level headings (personal archives, digital 3D designs or reconstructions of objects and buildings, etc.). A suggestion to overcome the problems identified is to develop a classification instead of a typology. The hierarchy of a classification can be used to control the different dimensions and relationships that exist between digital units. It was agreed during the meeting that "Classification" is the better term to use from now in the ENUMERATE project.</p> <p>After a short break an inventory was made of other research initiatives that can be used as a foundation for the proposed classification of object types. Those present are invited to continue sharing additional references with the ENUMERATE Team. References will be published on the ENUMERATE Delicious list at: <a href="http://delicious.com/enumeratesources">http://delicious.com/enumeratesources</a> .</p>
<p><b>How to deal with heritage collections that cannot be counted easily</b></p>	<p>In the NUMERIC project and during the formative process of the first ENUMERATE Core Survey four types of heritage provided difficulties: archival records, newspapers, monuments and sites, and born digital heritage. The problems with the former three are caused to a significant extent by the lack of harmonisation and standardisation in collection management. This makes it very difficult to ask questions regarding the size and growth of the physical collections. The solution for this problem lies in the approach discussed earlier in the meeting to count from the digital end, not the physical one. What is needed, is a generic concordance from the digital to the physical (x numbers of scans relate to x meters of shelves or x issues of newspapers). Such generic concordances do not exist yet, and ENUMERATE needs to look into the options to develop them in the project.</p> <p>About born digital heritage, the additional question that emerged was: When is something considered to be heritage? Aside from digitising their heritage collections or collecting born digital heritage, many institutions also possess digital research data and other forms of documentation that provide context for heritage collections. This data will be lost when not properly preserved, but...</p> <p>The attendees of the meeting had no reservations about the importance of these contextual data, however, two questions arose: Is the contextual documentation surrounding an object of cultural heritage part of the digital (or digitised) object? Is the contextual documentation cultural heritage in itself and, if so, should it be preserved by the institution? No definitive answer could be given to these questions, but at least for research data a recommendation was made. The attendees recommended that research data should be part of future renditions of the ENUMERATE survey. In this manner institutions can decide for themselves whether they consider this to be cultural heritage.</p>

<p><b>Other suggestions for future ENUMERATE surveys</b></p>	<p>In the tour de table the participants gave suggestions to the ENUMERATE Team. These varied from a warning for oversimplification of questions on cost data, to remarks about survey fatigue and the need to clearly explain what is the use of doing the survey: “It should be clear what’s in it for us”, to emphasising the need to raise awareness of the long term sustainability in heritage organisations.</p> <p>Some final points:</p> <p>We are only witnessing the first generation of digital output. Understanding the planning of long term preservation is vital to move forward with digital collections.</p> <p>To approach the topic of measuring size and growth of digital heritage collections from the digital perspective is the way to go forward.</p> <p>You will hardly get any response if you just send questionnaires around. Go to the museums when doing the survey and assist them in person with responding ! That way you get very consistent output.</p> <p>Through future questionnaires it would be valuable to learn: How do institutions organize their workflow? How do they implement policies? How does management take it up? At present digital asset management is not considered to be a core business!</p> <p>More attention should be paid to the outreach side of digital heritage collections. How much benefit do you get? How many people know about your collection? Is your government happy?</p> <p>Investigate the barriers caused by legal issues: how much work does it take to settle these?</p> <p>Keep questions about the built environment separated from those dealing with heritage objects in museums, libraries and archives. The new EU_project EU-CHIC would be a good partner to liaise between the two worlds.</p> <p>Take storage problems (may change), public private partnerships in relation to the ownership of images; and the different demands of different target groups into consideration.</p>
<p><b>Recommendations</b></p>	<p><b>Definitions</b></p> <ol style="list-style-type: none"> <li>1. Promote a common understanding of what cultural heritage is.</li> <li>2. Define criteria or guidelines of what information objects can be considered as digital cultural heritage.</li> </ol> <p><b>Context</b></p> <ol style="list-style-type: none"> <li>3. For heritage wide harmonisation, only generic, high level solutions can be applied. The practice across sectors is too diverse to change.</li> <li>4. Measuring the actual status of digitisation should be related to understanding the total costs of ownership.</li> <li>5. Understanding the developments in information management outside the cultural</li> </ol>

heritage institutions is needed in order to create new selection and acquisition policies for digital heritage

**Survey methodology**

6. The starting point for improving the methodology of measuring digital heritage collections should be the digital realm, not the physical one.

7. Measuring digital collections requires a better determination of digital units (which may be simple or complex)

8. Determining digital units can best be done on the basis of a hierarchical classification that expresses dimensions and relationships.

The cultural heritage field is invited to share references to projects, publications and other initiatives in measuring digital heritage collections with the ENUMERATE Team (e-mail: [answers@enumerate.eu](mailto:answers@enumerate.eu)). References will be on the ENUMERATE Delicious list at: <http://delicious.com/enumeratesources/>