

# ENUMERATE WP 2 *System (Thematic Survey methodology)*

Gerhard Jan Nauta



Co-funded by the



## Deliverables 2012

- 2.6 Thematic survey schedule
- 2.7 Revised documentation and guidance materials
- 2.8 Thematic survey methodology (incl. reports of the meetings)

## Time frame Thematic Survey Methodology

adaptations

- 2012
- January – October: Specialist meetings,
  - September - **January**: Preparations of methodology
  - **December - March**: Preparing implementation of surveys
- 2013
- **February - April**: Implementation of the survey incl. helpdesk
  - **April**: Analysis
  - April - **May**: reporting & reviewing
  - May - **June** : submitting data to data platform
  - Desk research throughout the TS

## Specialist meetings

## Four Specialist Meetings were organised in the year 2012

During each of these meetings one of the main themes of the ENUMERATE project was discussed.

The meetings were:

- 27 January 2012, The Hague: Size and Growth of Digital Collections
- 30 March 2012, London: Cost of Digital Collections
- 6 June 2012, Madrid: Access/Use of Digital Collections
- 17 January 2013, The Hague: Focus Group Digital Preservation.

## Specialis Meetings: some conclusions

### SIZE/GROWTH

- The methodology for measuring digital heritage collections should be based in the *digital realm*, not the physical one.

### COST

- Comparing the total is less useful than comparing the steps.
- Of various models available, the analysis of the workflow around the objects (incl. born digital) was considered the most useful

### ACCESS/USE

- Bring together both quantitative and qualitative research.
- Educate institutions in collecting the relevant data; do this over a large time period (prolonged collaboration, not one-off questionnaires)

### DIGITAL PRESERVATION

- There is already quite a lot of surveying on DP in Europe, make use of that instead of designing a new methodology.

## Preparing the Thematic Survey

## Building Blocks for Thematic Survey Methodology

- Recommendations by Numeric & SIG-STATS
- Results from the 1<sup>st</sup> Core Survey
- Review by Panteia of 1<sup>st</sup> Core Survey
- Feedback from National Coordinators received
- Comments from Specialists at the Specialist Meetings
- Methodology from national activities (e.g. born digital in NL)



## Thematic Survey Methodology

Focus on specific information we want to obtain within main topics

Decide on research methodology per topic (e.g survey, focus group, case studies, etc.)

Determine target audience (institutions, National Coordinators, other surveys)

Select institutions: how many from what countries?

Determine the actual survey planning and questions

## Role of the National Coordinators:

Contribute to a European address database of cultural heritage institutions for future surveys that allows random sampling

Supply information about national digitisation and digital preservation policies

Contributing to one or more of the four topics of the Thematic Survey (e.g. by selecting institutions to participate)

Supporting with the translation of the Thematic Survey tools

(N.B. Quality assessment of translations)

# SIZE

## Summary

At the core of the questionnaire is a list of object types that may be represented in digital heritage collections.

Since the aim is to assess the size of both digitised analogue and born digital collections (combined) and since such a list could not be traced in earlier research and surveys, the list has to be developed afresh.

The list used here is a combination of the NUMERIC/ENUMERATE Core Survey list of object types and a list of born digital object types that was developed in research on born digital heritage in The Netherlands, initiated by DEN in 2009. Accompanying questions are selected to put the results of assessing collection size and growth in context.

## SECTION 1/5 - ORGANISATIONAL INFORMATION

- [1] Name of institution / organisation
- [2] Type / Domain of institution / organisation
- [3] Your name
- [4] Your telephone number and/or Skype contact details
- [5] Your role in the institution
- [6] Your e-mail address
- [7] What is your institution's annual revenue budget?
  
- [8] Is curational care for the collections of your institution part of its mission?
  
- [9] Does your organisation have digital collections or is it currently involved in collection digitisation and digital born objects collecting activities?

## SECTION 2/5 - COLLECTION REGISTRATION SYSTEM APPROACH

Measure the size of your digital collections based on the quantities of descriptive metadata records in your Collections Registration System(s)

- How many objects are registered in the CRS?
- How many of these are available as digital born and digitally reproduced objects? (Everything that is not in the CRS need not be considered here.)

[13] What percentage of your collections is *not* registered in your CRS?

[15] What is the total # of descriptive metadata records in your CRS?

[17] What is the total # of Gigabytes of your digital heritage collections as recorded in your CRS?

## SECTION 3/5 - DETAILED MEASURING OF DIGITAL SUB COLLECTIONS APPROACH

In this second approach the aim is to measure the size of your digital heritage collections through a detailed quantification, starting from a pre-defined table of digitally reproduced and born digital objects.

[19] Please indicate the size of your digital heritage collections per object type (columns A and B) and assess the need to digitise for objects in your entire analogue collections (columns C and D).

	Term ([digital] objects)	Units (#=Number of) [#Records means: # Metadata records in CRS]	Estimated # of units in entire digital collection (A)	Estimated % born digital of units in entire digital collection (B)	Estimated % of <i>entire</i> analogue collection still to be digitized (C)	Estimated % of <i>entire</i> analogue collection no need to be digitized (D)
<b>DIGITAL INTERACTIVE RESOURCES [EXCLUSIVELY DIGITAL]</b>						
	Databases (containing cultural heritage metadata)	# Records				
	Digital (3D) designs or reconstructions of objects and buildings	# Records				

## **SECTION 4/5 - IMPROVED ENUMERATE CORE SURVEY QUESTIONS APPROACH (a)**

An approach based on estimates of the size of digital heritage collections and the number of objects that still need to be digitised. This is supplemented by two questions about the born digital materials collected.

[20a] Estimate the percentage of your heritage collections that has already been digitally reproduced.

[20b] Please specify the object types and percentages that were included in the previous answer [20a].

[21a] Estimate the percentage of your heritage collections that still needs to be digitally reproduced.

[21b] Please specify the object types and percentages that were included in the previous answer [21a].



## SECTION 4/5 - IMPROVED ENUMERATE CORE SURVEY QUESTIONS APPROACH (b)

[22a] Does your organisation collect born digital heritage materials?

[22b] If Yes, what types of born digital heritage?

*(Specify the object types that were included in the answer.)*

## SECTION 5/5 - EVALUATIVE REMARKS

[23] How much time / effort was needed to fill out the questions for the three different approaches?

[24] How difficult/hard was it to collect the necessary information to answer the questions in each of the three approaches?

[25a] Which one of the approaches would you prefer to fill out in future issues of the ENUMERATE Core Survey?


[25b] Why do you prefer the approach selected?

[26] Which one of the alternative approaches will offer in your opinion the most faithful insight into the actual size of the (digital) collection of your institution? And why?


nebu.stratusbv.nl/scripts/dubinterviewer.dll/frames?Quest=10024&resp=1&password=1234
Reader

---

Panteia: Research voor Beleid EIM NEA IOO Stratus IPM
Definitions | Instructions | Confidential | Helpdesk



**Panteia**  
Research to Progress



**ENUMERATE**

48 %

Testinterview - at Gr19a

### SECTION 3/5 - DETAILED MEASURING OF DIGITAL SUB COLLECTIONS APPROACH

In this second approach the aim is to measure the size of your digital heritage collections through a detailed quantification, starting from a pre-defined table of digitally reproduced and born digital objects.

Again it is important to keep in mind the difference between item level metadata records and digital born and digitally reproduced objects. (See above.) The questions below are questions about the digital born and digitally reproduced (analogue) objects only. So if we ask you to specify the number of metadata records, this concerns only the records with related digital born and/or digitally reproduced objects.

Column A: Units (#=Number of)[#Records means: # Metadata records in CRS]  
 Column B: Estimated # of units in entire digital collection  
 Column C: Estimated % born digital of units in entire digital collection  
 Column D: Estimated % of entire analogue collection still to be digitised  
 Column E: Estimated % of entire analogue collection no need to be digitised

---

	A.	B.	C.	D.	E.
Audio files: Speech & other (excl. digital audio books; incl. oral history files)	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

Back
Next

# COST

## Summary

The starting point consisted of a few questions focussing on the cost of digital collections from the ENUMERATE Core Survey (2012). Indications and suggestions from the Specialist Meeting on Cost (London, March 30) were taken on board as well, and several surveys in the field (e.g. the recent ITHAKA S+R Survey on Sustaining Digitized Special Collections) were examined.

The central part in this survey is a listing of activity related costs that were deduced from research done at the DEN Foundation (2009).

After testing on a small sample of institutions the original questionnaire has been simplified drastically.

## Focus questionnaire

The costs of acquiring and valorising digital collections are divided into incidental (upfront) costs and structural (ongoing) costs.

- Incidental costs are defined as the costs having to do with the initial creation or acquisition of a digital collection.
- Structural costs are the (annual) costs needed for the ongoing maintenance, enhancement and preservation of a digital collection

The survey consists of 14 (was 22) questions. Some questions have an accompanying field for evaluative remarks.

## SECTION 1/2 - ORGANISATIONAL INFORMATION

(...)

[8] Does the most recent annual report of your institution explicitly mention costs related to your institution's digital collections?

[9] If you answered Yes in the previous question what does this most recent annual report of your institution mention?

- *Growth of digital collections*
- *Fte's involved in creating and managing digital collections*
- *Budget related to creating and managing digital collections*
- *Growth of digital storage (in Terabyte)*
- *Other: ...*

## SECTION 2/2 - THE COST OF ONE SPECIFIC DIGITAL HERITAGE COLLECTION (a)

In this section the focus is on one specific major digital heritage collection within your institution...

[12] What quantities of digital object types are in the selected collection?  
*(Table)*



## SECTION 2/2 - THE COST OF ONE SPECIFIC DIGITAL HERITAGE COLLECTION (b)

[13] Activity based costing: determining *Incidental costs*

Estimate how much your institution has spent (in Euros) for the first time creation or acquisition (*incidental costs*) of the selected digital collection (including the cost of staff time) for each of the following activities, wherever these occurred in the institution or have been outsourced.

Leave blank the items that you cannot estimate. Enter 0 if a cost item is not applicable. If you miss any items in the table, please help us and mention these under *Other costs*.

## SECTION 2/2 - THE COST OF ONE SPECIFIC DIGITAL HERITAGE COLLECTION (c)

Cost item	In-house #EURO	Outsourced #EURO
Project management		
Selection of material for digitisation		
Acquisition of digital born material		
Logistics (shipment of collection for digitisation, etc)		
Analogue-Digital conversion (including all technical and staff costs associated with the act of preparing materials for scanning, the scanning itself, and quality control):		
Copyright clearance		
Metadata creation and/or enhancement		
Collection of user generated content (UGC)		
Web design and/or software development		
Other costs, please specify: [...]		

## SECTION 2/2 - THE COST OF ONE SPECIFIC DIGITAL HERITAGE COLLECTION (d)

[14] Activity based costing; determining *Structural costs*

For the past fiscal year, please estimate the *structural costs* for the selected digital collection. Fill in the form for each cost item that can be estimated, wherever it occurs in the institution or is outsourced.

## SECTION 2/2 - THE COST OF ONE SPECIFIC DIGITAL HERITAGE COLLECTION (e)

Cost item	In-house #EURO	Outsourced #EURO
Project management		
Archiving (storage, including backups)		
Activities concerning the (long-term) preservation of the digital collection (storage costs excluded).		
Licences		
Costs related to giving access to the digital collection (e.g.maintenance webservers and web services)		
User outreach and support (including all staff time for efforts to promote the use of the collections)		
Usage analysis (including surveys, interviews, and other activities)		
Editorial (including content selection and updating)		
Other costs, please specify: ...		


COST


nebu.stratusbv.nl/scripts/dubinterviewer.dll/frames?Quest=10023&resp=1&password=1234

Griekse - N...e - woorden VUW Remote ...ace Log In DEN Web Access NEW Hum Web Access Wikipedia Diigolet


COST

Panteia: Research voor Beleid EIM NEA IOO Stratus IPM

 **Panteia**  
Research to Progress

 **ENUMERATE**

Definitions | Instructions | Confidential | Helpdesk

 58 %

Testinterview - at Gr16h

**DIGITAL INTERACTIVE RESOURCES (EXCLUSIVELY DIGITAL)**

What quantities are in the digital collection, and in what qualities.

		HQ	MQ	LQ
Databases (containing cultural heritage metadata)	<input type="text"/> Records	<input type="text"/>	<input type="text"/>	<input type="text"/>
Digital (3D) designs or reconstructions of objects and buildings	<input type="text"/> Records	<input type="text"/>	<input type="text"/>	<input type="text"/>
Digital art objects	<input type="text"/> Records	<input type="text"/>	<input type="text"/>	<input type="text"/>
Digital research files (incl. GIS files)	<input type="text"/> Records	<input type="text"/>	<input type="text"/>	<input type="text"/>
Games	<input type="text"/> Records	<input type="text"/>	<input type="text"/>	<input type="text"/>
Software (customized)	<input type="text"/> Records	<input type="text"/>	<input type="text"/>	<input type="text"/>
Websites (and parts of websites)	<input type="text"/> Records	<input type="text"/>	<input type="text"/>	<input type="text"/>
Other digital interactive resources	<input type="text"/> Records	<input type="text"/>	<input type="text"/>	<input type="text"/>

# ACCESS/USE

## Summary

The starting point consisted of questions focussing on, in this case, the use of digital collections from the ENUMERATE Core Survey (2012).

These questions were amplified by a number of questions that were developed with suggestions from the Specialist Meeting on Use/Access (Madrid, 6th June 2012) in mind. Other sources are recent surveys in the field, e.g. the ITHAKA S+R Survey on Sustaining Digitized Special Collections and survey research that was conducted in the Netherlands by Henk Voorbij (2009).

An online test version of the questionnaire on USE/ACCESS, was developed by Panteia and links to the questionnaire were distributed among about 30 institutions in 5 EU member states.

## Focus questionnaire

The aim is to get a grasp of the access and use of digital collections. In more detail, the questionnaire focuses on:

- Online access: which part of your collection is available online and what are the platforms where content is presented to the public?
- The development of online user services: has your institution developed any websites, apps, etc.?
- The measurement of online user activities: what methods and tools does your institution use for collecting usage metrics?
- Licensing and legal issues regarding the online collection: how much does your institution spend on clearing copyright?



### SECTION 3/6 ACCESS TO ONLINE DIGITAL COLLECTIONS (a)

[12] What object types in the digital heritage collections of your institution are accessible online as digitally reproduced or born digital objects?

[12a] In what quantities are the following objects - as digitally reproduced or born digital objects - accessible online?

*(Table. The table can be folded out, starting from the high level object classes.)*

[13] What kind of descriptive metadata/cataloguing data are you providing online?

*No metadata at all / Metadata about items in online accessible digital collections / Metadata about items in all digital collections, including offline digital collections / Metadata about items that are not digitally available / Other metadata: ...*

### SECTION 3/6 ACCESS TO ONLINE DIGITAL COLLECTIONS (b)

[14] Please indicate the estimated percentage of all the digital objects you have (digitally reproduced and/or born digital) that are and/or will be accessible through mentioned access options: ...

*(Table.)*

## SECTION 3/6 ACCESS TO ONLINE DIGITAL COLLECTIONS (c)

	% of digital objects currently accessible	% of digital objects accessible 2 years from now
Offline		
Institutional website		
National aggregator		
Thematic aggregator		
Europeana		
Memory of the World		
Wikipedia		
Institutional API		
3rd party API		
Social media platforms like Flickr or Facebook		
Other access options (please specify)		

### SECTION 5/6 - MONITORING (a)

[16] Does your organisation measure the number of times digital metadata and/or digital objects are being accessed by your users?

[17] If Yes, which of the following methods does your organisation use to measure the number of times digital metadata and/or digital objects are being accessed?

*Web statistics*

*Database statistics (if not included in Web statistics)*

*User studies*

*Other: ...*

[18] What tool(s) is/are being used in your organisation for measuring the number of times digital metadata and/or digital heritage objects are being accessed?

### SECTION 5/6 - MONITORING (b)

[19] How often do you analyse the results of your access/use monitoring activities? (+Please give an explanation of your answer.)

- *On a weekly basis*
- *On a monthly basis*
- *On a yearly basis*
- *Other: ...*

[20] In what way does your organisation make use of the outcomes of access/use data analysis?

[21a] Does your organisation apply Key Performance Indicators based on your monitoring [access/use] data?

[21b] If Yes, what KPI's does your organisation apply?

[21c] If Yes, how does your organisation report on these KPIs?

ACCESS/USE

nebu.stratusbv.nl/scripts/dubinterviewer.dll/frames?Quest=10022&resp=1&password=1234

Gerhard Jan Nauta – Outlook Web App

Panteia: Research voor Beleid EIM NEA IOO Stratus IPM

**Panteia**  
Research to Progress

**ENUMERATE**

Definitions | Instructions | Confidential | Helpdesk

29 %

Testinterview - at V012

**SECTION 3/6 ACCESS TO ONLINE DIGITAL COLLECTIONS**

What object types are accessible online?

**ARCHIVAL RESOURCES**  
*(such as Archives: Government documents, government archives, other archival records and other archives)*

**TEXT BASED RESOURCES**  
*(such as books, serials and other text based resources)*

**VISUAL (2D) RESOURCES**  
*(such as drawings, engravings/prints, maps and ground plans, paintings, photographs, posters, sheet music and other visual resources)*

**3D MOVABLE OBJECTS**  
*(such as 3 dimensional works of art, archeological man made objects, coins and medals and other 3 dimensional man-made objects)*

**GEOGRAPHY BASED RESOURCES**  
*(such as landscapes, archeological sites, monuments and buildings and other geography based resources)*

**NATURAL RESOURCES**  
*(such as natural inert specimens and natural living specimens)*

**TIME BASED RESOURCES**  
*(such as audio files: music, speech & other, digital audio books, film (cinematic) , video recordings and other time based resources)*

**DIGITAL INTERACTIVE RESOURCES [EXCLUSIVELY DIGITAL]**  
*(such as databases, digital (3d) designs, digital art objects, digital research files, games, software, websites and other digital interactive resources)*

**Back** **Next**

# DIGITAL PRESERVATION

<b><i>Sub-study 4</i></b>	<b><i>Digital preservation (no countries needed, experts / specialists)</i></b>
Subject	Consultation with specialists, since this is a relatively new topic
Method	Focus Group
Needed	No involvement of National Coordinators needed



## Questions / Discussion

