Europeana – Core Service Platform

MILESTONE

MS20: Ingest plan for adding new content to the ENUMERATE Observatory

Revision	6
Date of submission	August 31, 2015
Author(s)	Berry Feith, Wietske van den Heuvel, Gerhard Jan Nauta, Marco de Niet, (DEN Foundation), Gordon McKenna (CollectionsTrust)
Dissemination Level	Public

Funded by
the Connecting Europe Programme
of the European Union

REVISION HISTORY AND STATEMENT OF ORIGINALITY

Revision History

Revision No.	Date	Author	Organisation	Description
1	10/7/2015	Gerhard Jan Nauta	DEN Foundation	Document structure
2	4/8/2015	Berry Feith, Wietske van den Heuvel, Gerhard Jan Nauta, Gordon McKenna	DEN Foundation / Collections Trust	Body text fragments
3	1/9/2015	Berry Feith, Gerhard Jan Nauta, Marco de Niet	DEN Foundation	Internal review
4	4/9/2015	Gordon McKenna	Collections Trust	Review and annotations
5	4/9/2015	Berry Feith, Gerhard Jan Nauta, Marco de Niet	DEN Foundation	Edited version
6	8/9/2015	Berry Feith, Gerhard Jan Nauta	DEN Foundation	Final version

Statement of originality:

This milestone contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Table of Contents

1 Introduction

The purpose of this report is to make plans for the development of the ENUMERATE Observatory, preceded by a description of its raison d'être. The report builds on an earlier ENUMERATE milestone document, namely Europeana Version 3, Milestone 7, *Draft plan on future topics to be included in the ENUMERATE framework*. It also reverts on the original Description of Work of the ENUMERATE Thematic Network dated 2010, an ambitious document containing project goals that were substantially slimmed down at the time the ENUMERATE Thematic Network was realised.

Effective decision-making in the domain of digital cultural heritage depends on reliable data. That is one of the reasons why efforts started about a year ago to integrate the project based ENUMERATE data and research framework into the operational Europeana Pro platform. Early in 2015 ENUMERATE organised a Europe wide survey (Core Survey 3) as part of the Europeana V3 project, and legacy data from the ENUMERATE Thematic Network project (2011-2014) has been mounted onto the Europeana Pro platform, be it in rather unsophisticated manner.

The rationale behind these efforts was that although the Europeana Network already had access to *internal* data, i.e. quantitative data about the use of Europeana services. This is made publicly available through the Europeana Dashboard, http://statistics.europeana.eu/welcome There was a lack of reliable data about the world *outside* of the Europeana Network. Of course data and information of that kind are highly relevant for the development of Europeana strategic plans and there is potentially a much broader user group as well.

Therefore it was decided to continue the integration efforts under Europeana DSI. Under the new funding programme (2015-2016) it is planned to create an ENUMERATE Observatory, where freshly collected data will be stored, analysed and published, together with data from earlier surveys and data that was collected in European surveys not associated with ENUMERATE. The Observatory will also be the place where the framework for collecting digital heritage statistics is documented and further developed. Key to the success of such an observatory is that the data collected is meaningful and presented in an intelligible way to the key stakeholders.

Since what is meant by meaningful data is not fixed, the Observatory will also have a mechanism to adapt to changing circumstances. For instance, in the previous ENUMERATE Core Surveys no question was asked about copyright restrictions that prevent institutions to make their digital collections public, whereas this topic has attracted a lot of attention recently.

This report presents an ingest plan for adding new content to the ENUMERATE Observatory.

An ingest plan in this context is a description of:

- The current state of affairs;
- The vision;
- The kind of data that will be collected;
- How the data will flow into the observatory.

Because such a plan presupposes a clear view of the benefits of what needs to be in the observatory, this report also scopes the ENUMERATE Observatory in general terms: the kind of data that will be collected; the way these data will be collected and how it can be a become a flexible service to serve the needs of different user groups within available means. A more precise description of the Observatory and analysis of the relevant data will be the topic of a deliverable later on in the project. In the first half of the WP4 task 4.3 time span actual (new) data will not be ingested. The focus is on scoping.

In Chapter 2 we will describe the success criteria and potential impact: when will the ENUMERATE Observatory be a success and what are the risks involved? In Chapter 3, the current state of affairs is documented. What data do we have and how are these made available (under what conditions) to the public? Chapter 4 defines the scope of the Observatory, taking alternatives into account through the analysis of a selection of relevant websites and applications. Chapter 5 has provided a description of data types and actual data sets which will be analysed more in detail in D4.3. Finally, in Chapter 6 data ingest details, or how these data flow into the Observatory, are stipulated.

2 Definition of Success and Risk Analysis

When will the ENUMERATE Observatory be a success? What risks should be managed? The bottom line with the first question is the impact of the Observatory on the key stakeholders. The greatest risk for the project will be a lack of sustainability, resulting from a misfit between project goals and the available means.

2.1 Stakeholders

The Observatory will be a success if the primary stakeholders are aware of the service and they can find the information they want to support new policy decisions.

The primary stakeholders are the Europeana strategic planners and wider Europeana community / network. Other stakeholders include the European Commission and EU member state governments and ministries; funding bodies; national statistical agencies; national and international umbrella organisations (e.g. IFLA, ICA, ICOM, ICOMOS, ACE, etc.); managers and other professionals at individual heritage institutions; other (cultural heritage) professionals; and companies active in the cultural heritage domain (software developers, service providers).

In an attempt to further specify the interests of these stakeholders we summarise their characteristics and possible benefits they will gain of the Observatory in figure 1.

Stakeholders	Raw statistical data	Visualisations and Analysis	Statistical tools	Methodology documents	References to other sources
1. Europeana staff	✓	✓			✓
2. Europeana Network participants.	✓	✓	✓		✓
3. European Commission		✓			
4. Ministries of Culture	✓	✓		✓	✓
5. Funding bodies		✓			
6. Statistical agencies	✓		✓	✓	✓
7. Umbrella organisations	✓	✓		✓	✓
8. Cultural heritage institutions		✓	✓		✓
9. Cultural heritage professionals	✓	✓	✓	✓	✓
10. Commerce (e.g. vendors)		✓			✓

Fig. 1: *Stakeholders: who are they and what aspects of the Observatory are important to them?*

2.2 Risk Analysis

The primary risk to the Europeana Observatory is the uncertainty future of the financial resources that will support it, if any. Therefore any plan needs to be flexible. This is all the more important where remedial actions need financial means too. So the general principle is to design the observatory in a modular way.

Risks with the greatest impact are presented in figure 2. A full risk analysis is in Appendix 1.

Description of possible risk	Impact	Probability of occurrence: 1=low, 3=medium, 5=high	Summa	Remedial Actions
Lack of participation amongst the principal stakeholders [organization]	5	2	10	Mobilize MSEG; and/or encourage top-down approach (EU).
Budget for carrying out the actual data collection activities is insufficient [organization]	5	4	20	Find additional budget; downsize the scope of the survey; split up (core) survey into smaller units.
(Statistical) Expertise needed not sufficiently represented in WP participants [analysis]	5	1	5	Downscaling the survey. Liaise with experts in the field and/or find additional financial means.

Fig. 2: *Risk analysis.*

3 Current status of ENUMERATE at Europeana Pro and what's available on the Europeana Dashboard

3.1 Europeana Pro

The information on the ENUMERATE project website, www.enumerate.eu was migrated to the Europeana Pro platform in March 2015, http://pro.europeana.eu/enumerate The old project website is still online, but will not be kept up to date and will go offline in 2016.

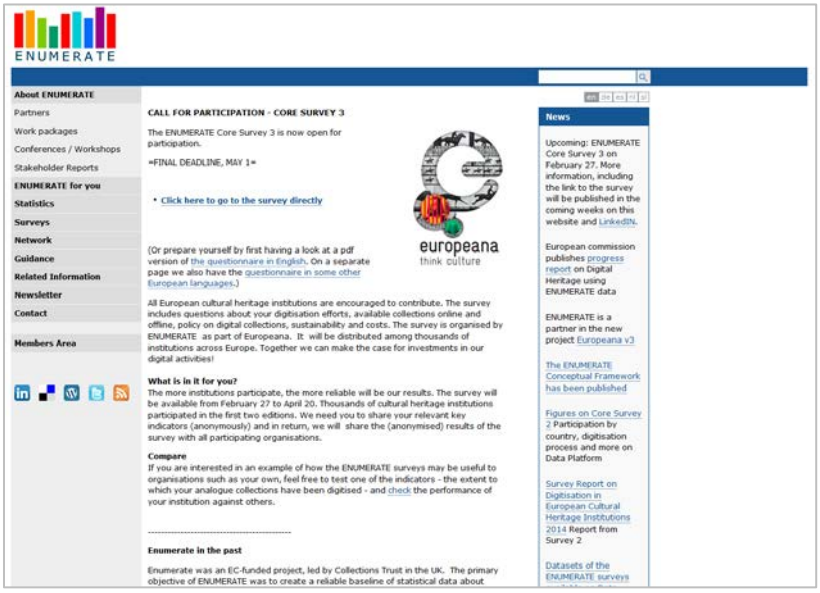

Fig. 3: *Screenshot from the old ENUMERATE project website.*

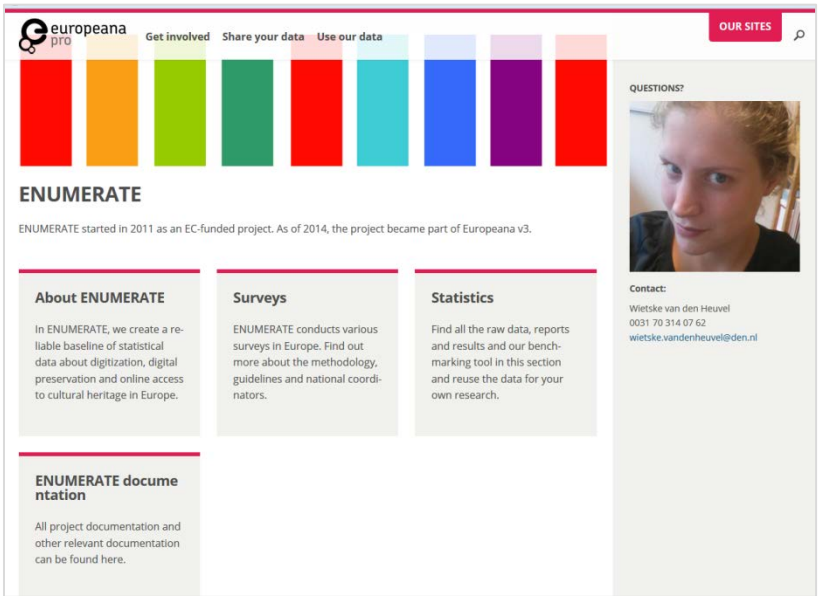

Fig. 4: *Screenshot from the ENUMERATE section in Europeana Pro.*

Overview of the ENUMERATE content in Europeana Pro:

- Description of the project;
- Information and documentation on the surveys, national coordinators, etc.
- Statistics and results;
- All relevant documentation from the current project and the former surveys.

The data from the surveys can be downloaded from the ENUMERATE Data Platform. This is a separate platform which is linked to the website.

3.2 Europeana Statistics Dashboard

The Europeana Statistics Dashboard, http://statistics.europeana.eu/welcome is in the Alpha phase. Currently, there are three data categories: Content, Traffic & Usage, and Data Providers. The Content section visualises statistics related to the number of digital objects available in the Europeana repository. The Traffic & Usage section visualises statistics related to web traffic on the Europeana.eu website together with statistics related to Europeana's reach on social media and Wikipedia. The Data Providers section visualises statistics related to traffic and usage on the Europeana.eu website for a number of data providers.

Data can be downloaded and is presented in graphs that can be adjusted to a certain extent to user's needs. The graphs can also be embedded in other sites.

3.3 ENUMERATE Data Platform

The ENUMERATE Data Platform currently is the online platform for the storage, publication and dissemination of statistical data collected during the ENUMERATE project. The platform is hosted by Digibis: http://dataplatform.enumerate.eu.

The results of each ENUMERATE Survey are published in detail on the ENUMERATE Data Platform in the form of reports, public data sets, multilingual questionnaires, benchmark and a review of tools. The public datasets have been anonymised[1] and can be downloaded and used to a Creative Commons CC0 license. Some data sets can only be downloaded after login.

3.4 ENUMERATE Benchmark Tool

The ENUMERATE Benchmark Tool is still in beta. Currently, institutions can insert their digitisation progress and compare it with the results from one of the surveys, based on the type of institution and the country. The results are presented in a graph. The platform is hosted by Digibis and because of limited budget, the tool has not been developed further since its launch: http://enumeratedataplatform.digibis.com/benchmark

[1] I.e. Personal and organisational names are removed.

4 Scoping the ENUMERATE Observatory service

4.1 Introduction

In general terms the ENUMERATE Observatory can be described as:

- A central access point where the past and present state of *digital* heritage in Europe can be monitored;
- A place where the need for new intelligence in the domain of digital heritage can be articulated, and where tools are available to initiate new research;
- A discussion board where professionals can ask for specific information on the past and present state of digital heritage in Europe.

In the stakeholder analysis in section 2, broad categories of Observatory content types were introduced: raw statistical data; data analysis results (visualisations etc.); statistical tools (e.g. for data collecting, processing and mining); methodological documents; and references.

We propose that the following kinds of content are eligible for ingestion in the ENUMERATE Observatory service:

0. Information needs / Requests for information
    - 0.1. Recognised information needs (e.g. topics covered in existing core surveys)
    - 0.2. New information needs
1. Raw (statistical) data collections
    - 1. 1. Legacy data
        - 1.1.1 Core Survey and NUMERIC survey statistics
        - 1.1.2. Statistical data from other (third party) initiatives
    - 1.2. Newly collected statistical data
        - 1.2.1. Data from subsequent Core Survey studies
        - 1.2.2. Data from miscellaneous (also third party) data collection initiatives
2. Data collection and analysis instruments
    - 2.1. Questionnaires
    - 2.2. Polls
    - 2.3. Data harmonisation and analysis tools
3. Data analysis results
    - 3.1 Reports
    - 3.2 Visualisations
        - 3.2.1. Static visualisations
        - 3.2.2. Dynamic (online) dashboard like visualisations
4. ENUMERATE Framework [methodological stuff]
    - 4.1 Indicator encyclopaedia / Question bank
    - 4.2 Procedures
5. References

Fig. 5: *Contents that are eligible for ingestion in an ENUMERATE Observatory service.*

At the core of all these content types are the *indicators*. The wish to understand the digital heritage domain starts with a loosely defined quest for information. For instance, knowing that heritage institutions are reluctant to put their digital collections online because they foresee copyright claims makes both institutions and funding bodies curious about the percentage of (digital) objects for which rights issues are a constraint. Once the quest for specific information gets sufficient weight, this is an incentive to take action. This action can be twofold: it could be

tracing information sources already available in the field; or, if such sources are non-existent, it could be initiating new research (e.g. data collection) efforts.

Because the usefulness of the Observatory has to be proven continuously and the preparedness to fund the initiative will depend on its achievements, the *financial* scope for the project in the near future remains uncertain. This means that the setup will have to be lean, having the potential to expand, depending on the evolution of needs for statistical data. Below are three scenarios: two are at the extreme, as far as the financial scope is concerned. The third takes a middle position.

4.1.1 Scenario 1

The assumption is that, given the availability of the Europeana Pro platform, it is feasible to maintain a service that hosts the raw data and analysis reports of past and future ENUMERATE Core Surveys with relatively modest financial means. Also an open-ended repository/link service of various third party data sets and reports can be realised without having to spend large amounts of money. Also new ENUMERATE Core Survey results can be accommodated.

4.1.2 Scenario 2

Scenario 2 represents the present situation, where once every two years a separate core survey can be organised, using a modest dedicated project budget. The aim is to build on the baseline of data collected in the three successive ENUMERATE Core Surveys and extend the observatory step by step. The outcomes (raw data) of these surveys, together with varying sorts of dissemination reports are mounted on the Europeana Pro website.

4.1.3 Scenario 3

In this scenario the starting point is Scenario 2 with the aspiration to dynamically grow to a situation where other sources of income can be generated to support the development and use of the Observatory, e.g. by developing business models for trading data sets.

The remainder of this report is built around Scenario 2, which represents the status quo.

In order to better understand the options we have to give shape to the Observatory within the given means, we have analysed a number of similar services that we are involved in or have easy access to.

4.2 Examples of other websites with statistical data

4.2.1 Heritage monitor (Erfgoedmonitor)

Created by the Cultural Heritage Agency - part of the Ministry of Education, Culture and Science in the Netherlands.

The purpose of this Monitor is to understand:

- Effects of heritage policy;
- Effects of tools to support policy;
- Gaps in the knowledge;
- National Heritage Review and Heritage Monitor websites.

http://erfgoedmonitor.nl/indicatoren/musea-collecties-online

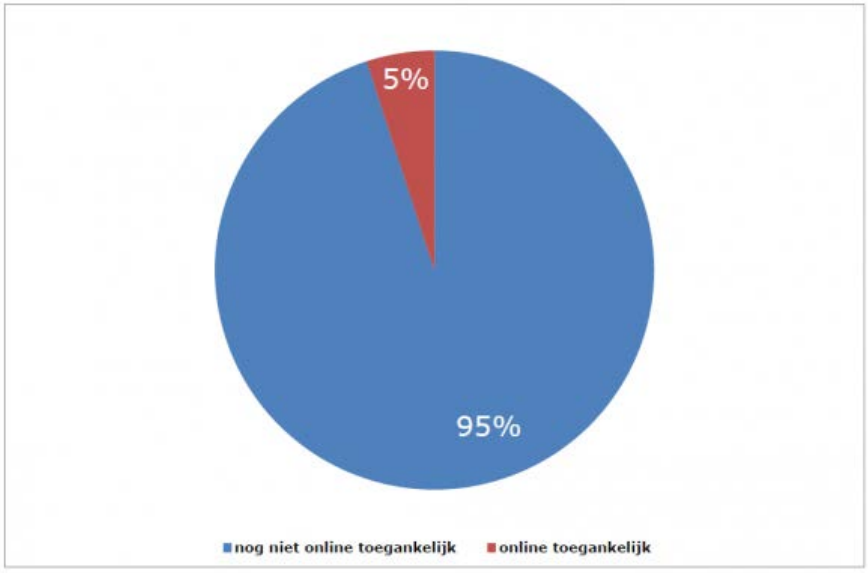

Fig. 6: *On May 1, 2014 there were over 3 million museum objects centrally accessible online, this is 5% of the total size of the Netherlands Collection.*

On the site we find at each topic :

- A resume about the topic;
- Sources;
- Reference date;
- Justification.

The Cultural Heritage Agency regularly collects data to enable the heritage to be 'measured'. This gives an idea of the impact of heritage policy and how the heritage is currently faring.

The program is designed to produce a system for monitoring over time. The system has a comprehensive set of indicators, data and methods to enable sustainable effect measurements. 'Sustainable' means that the system must work in the long term, and that the data must be constant and collected over several years in order to reveal developments, trends and effects. Data collection must also become a mainstream part of the work of the organisation(s) involved.

4.2.2 Arts Index Netherlands (Cultuurindex)

Created by the Boekman Foundation and The Netherlands Institute for Social Research.

The Arts Index Netherlands is initiated by the Boekman Foundation, http://www.boekman.nl/en in co-operation with The Netherlands Institute for Social Research (SCP), http://www.scp.nl/english/Organisation/About_SCP with the aim of collecting a great amount of relevant and available cultural data and presenting the collected data in an accessible and understandable way.

http://www.cultuurindex.nl

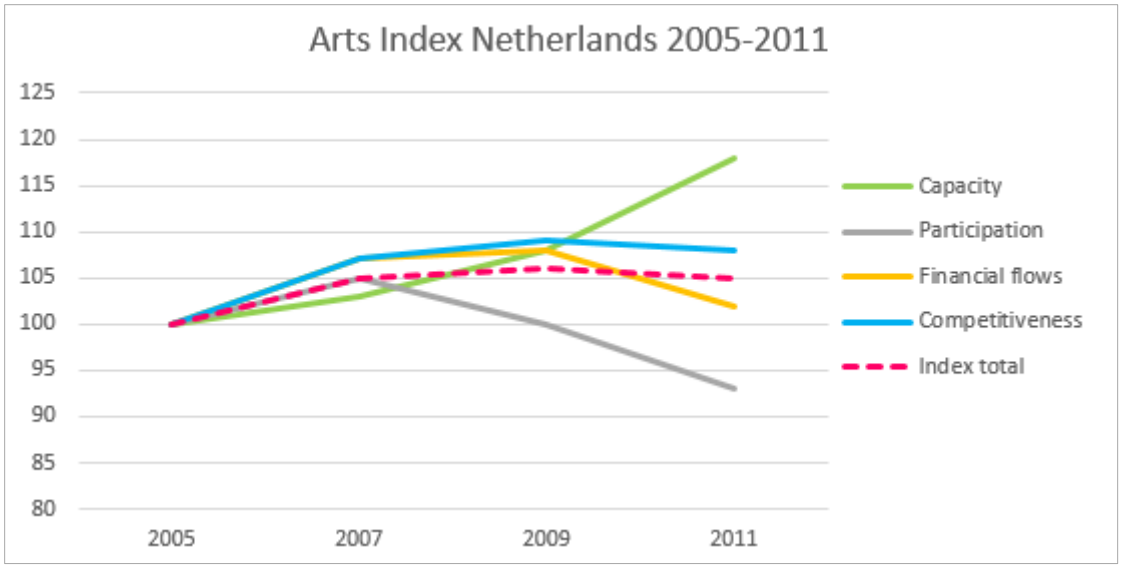

Fig. 7: *Trends in the four pillars of the Arts Index Netherlands.*

The index is built on four 'pillars': Capacity, participation, financial flows and competitiveness and is inspired by the US National Arts Index, http://www.americansforthearts.org. Both subsidised and non-subsidised sectors are covered in the index.

4.2.3 Benchmark Emancipation (Benchmark Emancipatie)

Created by Atria, http://www.atria.nl/atria/eng the Knowledge institute on emancipation and women's history in the Netherlands.

Custom-made tables on women in the areas of labour, income and combining work and family. The tables are broken down by municipality and by economic sector.

http://www.benchmarkemancipatie.nl

Fig. 8: *Local emancipation policy: voluntary work.*

Most of the data is delivered in Excel (Statistics Netherlands and some other organizations). The survey data was converted to Excel from SPSS. New analyses were carried out by Statistics Netherlands and linked to their own files and other data to obtain the right data. They also added municipal codes in their databases for all municipalities, so that the data could be shows in the map of the Netherlands.

The site was created with a one-off grant. The result is a website with interesting but static data.

A more extensive annotated list of examples is on the *enumeratesources* overview on Delicious, https://delicious.com/enumeratesources

4. 3 Lessons learned

What was learned from the examples above is that these websites focus on data and information more in general and are not pretending to be integral or complete observatories, processing all of the content types as mentioned in Fig. 5. The sites are giving detailed explanations on the data. The indicators are shown and explained. The results are placed in time and in a context. The statistical data can be displayed graphically in various ways. Most data can be downloaded to be edited by users.

Within the framework of the collaboration between ENUMERATE and Europeana Pro, we propose to start from Scenario 2 and explore ways to improve and where necessary extend data collection activities based on perceived needs.

To make steps towards a better relationship between data needs to data acquisition procedures, it is essential to further develop the ENUMERATE framework with clearly defined indicators at the core that are supported by the key target groups.

In short, we propose to model the data processing according to seven functions:

1. Antenna (actively processing requirements from the field)
2. Indicator assessment (expressing support from key target groups)
3. Data collection (active and passive)
4. Data processing (e.g. harmonisation and analysis)
5. Data storage
6. Data presentation
7. Administration

Fig. 9: *The seven functions of an ENUMERATE Observatory service.*

Each of these seven components/functions needs to be in place to have a manageable *modus operandi*. As the financial means to support these functions are limited, the key to keeping it lean is not to raise the number of indicators without elaborating financial substantiation. Currently, the ENUMERATE Framework contains 20 indicators that were tested in the previous Core Surveys and that will remain at the heart of future Core Surveys. Extension of this core set of indicators can happen along two lines: either by acknowledging needs for new data or by incorporating indicators from third party surveys. Any extension of the core set will need to be supported in a review process by a panel of specialists, as was done during the ENUMERATE Thematic Network project.

Alongside the analysis of existing data sets that will be performed later on in the project (D4.3), we will further develop this into an indicator assessment procedure that lays the foundation for future ENUMERATE questionnaires.

Other assumptions about the ENUMERATE Observatory in the plans outlined below:

- It should be complementary to the Europeana Dashboard (no overlap);
- It should be scalable [non-exhaustive]; achieve maximum result with minimal effort;
- It should allow for quick-wins (a few - manually construed - high level graphics, and for example some PowerPoint slides with key statistics) and in-depth data (raw data);
- It should be easy to find [any budget for dissemination?] and simple to operate;
- It should be flexible: keeping information relevant based on feedback.

Since this is an ingest plan, in the next chapters we will elaborate the functions 1, 2 3 and 7.

5 Contents to be ingested or what data should go into the Observatory

5.1 Data sets

In MS7 (Draft plan on future topics to be included in the ENUMERATE framework and in Section 4 of this report we noted that few examples exist of projects in the cultural heritage domain where the assimilation of digital cultural heritage statistics from diverse sources has been achieved.

In the examples we discussed here, like the Heritage monitor, a substantial budget for manually harmonizing data was available. In a situation where there is a limited budget for doing statistical research [Scenario 2] these are the contents that are eligible for inclusion in the Observatory. [Refer to Figure 5]:

- Identified and newly traced information needs (e.g. topics covered in existing core surveys);
- Legacy data: Core Survey and NUMERIC survey statistics;
- Newly collected statistical data (Core Survey 4 in particular);
- Static (print) report and visualisations;
- Dynamic (online) dashboard like presentations;
- Updated ENUMERATE Framework documentation (including Indicator assessment);
- References.

5.2 Data types [standards]

As was documented in Chapter 4, in the medium scenario we discern the following types of data:

- SPSS;
- Excel;
- Raw text;
- PDF;
- PowerPoint;
- URI's.

Most of the projects and applications we studied made use of manually harmonised and reformatted data. If data from diverse sources were harmonised in many cases the format used to relate these data was the Excel spreadsheet format. Explorations so far have not resulted in newly found data types and data standards, other than those we identified in previous ENUMERATE research efforts. The *enumeratesources* overview on Delicious, https://delicious.com/enumeratesources is the place where the accumulated references can be consulted. The list is non-exhaustive and will be constantly updated.

A standard for the registration and documentation of indicators needs to be developed.

6 Data ingest details or how these data flow into the Observatory

In Figure 9 the seven functional components of an ENUMERATE Observatory were listed. This section focuses on the first three of these functions: Antenna, Indicator assessment and (survey) Data collection. Administration, the 7th function, is of course an essential part in any process and is also briefly described.

6.1 Antenna

An important goal of this function is to recognised new information needs. Are there specific topics that attract a lot of attention among heritage policy makers and professionals but that lack supporting data or evidence? If so, we will explore ways to incorporate these topics in future ENUMERATE core surveys.

In an earlier ENUMERATE report (Europeana v3/MS7: *Draft plan on future topics to be included in the ENUMERATE framework*) a number of ways to explore information needs in the field was already listed. For this report two methods were applied: desk research and the consultation of digital heritage professionals in two workshops. The report concluded with an overview of potential topics to be included in future ENUMERATE surveys. However, due to time constraints in validating the topics, no new topics could be included in the third Core Survey.

In the current project we build on the results of this report by continuing the desk research and the consultation of experts in the digital heritage domain. In addition, we will use some others sources as well:

- Information requests from other existing surveys;
- A limited number of in depth interviews [expert consultation];
- Consultation of the MSEG group;
- An online call to suggest topics via the Europeana Pro website.

6.2 Indicator assessment

After we have identified the needs for new data by the key target groups, the next step is to define the proper indicators with which data can be collected that address those needs. Key to a solid procedure for adjusting and optimising the set of indicators is the mechanism through which identified information needs are documented and evaluated. This mechanism is needed, as not all information needs can be easily addressed through statistical surveying. Also, an essential characteristic of the ENUMERATE framework is that it addresses core issues related to digital heritage. A review mechanism is needed before the decision can be taken that we will really start collecting data on a specific topic. In the next project phase we will develop an indicator assessment procedure and an indicator assessment base, in order to expand the current set of ENUMERATE (core survey) indicators.

6.3 Data collection

6.3.1 Legacy data: Core Survey and NUMERIC survey statistics

The data collected during the ENUMERATE project are already available on the ENUMERATE Data Platform, which is linked to the Europeana Pro website, including the final reports and (references) to the data sets.

The NUMERIC data (from the period 2007-2009) was distributed per country on CD-ROM by the project coordinator to the national contact points. During the analysis of other surveys, we will

assess whether it is useful to incorporate this data into the data platform. Some of the legacy data of Numeric is also 'hidden' in static PDF reports. We will explore ways to make this data more visible on Europeana Pro, while safeguarding that this data meets the requirements from the indicator assessment.

6.3.2 Newly collected statistical data (Core Survey 4)

Up to now three successive core surveys were organised. According to the plans core surveys will be organised biennially. The next round will be in early 2017. Therefore in the coming months the processing of new core survey data is not an urgent issue, but it is advisable to reckon with these kind of data in any attempt to systematically ingest digital heritage statistics.

It is also something to reckon with that the number of respondents in a four year period steadily regressed from almost 2000 in Core Survey 1 to about 1000 respondents in the most recent Core Survey 3. The regression may partly be due to survey fatigue among respondents (institutions) and the national coordinators, who over the years have participated in the surveys on a voluntary basis.

The main challenge in organising the follow-up to Core Survey 3 in 2016-2017 will be in achieving an acceptable response rate. We will use the current project period to tinker with alternative ways to raise the number of participants in ENUMERATE monitoring practices.

6.3.3 Data from miscellaneous (also third party) data collection initiatives

The idea to process results from other monitoring initiatives is a long standing ambition. In the Observatory well defined indicators are probably the key to relate statistics from various sources. Among the examples discussed in section 4.2 are some initiatives - as for instance the Heritage monitor - where per indicator several sources are used to clarify a state of affairs in the cultural heritage domain.

The exact way in which well-defined indicators could be the pivot on which the Observatory runs is part of the research to be done.

6.4 Administration

References where ENUMERATE is recognised as a major initiative in digital heritage monitoring, the team receives hints and tips about other monitoring initiatives on a regular basis. Up to now a Delicious account was used to keep a record of what is happening, but a more integrated way of record keeping is needed. In the coming months we will explore the strengths and weaknesses of other reference management applications.

6.4.1 Updated ENUMERATE Framework

Most of the methodological refinements that result from the work done in this work package will be documented in the ENUMERATE Framework on Europeana Pro.

7 Appendices

Appendix 1: Risk analyses

Description of possible risk	Impact	Probability of occurrence: 1=low, 3=medium, 5=high	Summa	Remedial Actions
Mobilization of the proposed community is limited [organization]	4	3	12	Plan the project according to the actual size of the (active) community
Lack of participation amongst the principal stakeholders [organization]	5	2	10	Mobilize MSEG; and/or encourage top-down approach (EU)
Re-use of existing surveys in the cultural heritage domain is more problematic than was expected [organization]	2	5	10	Drop objective of reuse per se
National differences in the landscape of cultural heritage institutions are difficult to bridge [organization]	2	4	8	Group EU countries according to encountered similarities; stipulate more than one survey implementation scenario
Budget for carrying out the actual data collection activities is insufficient [organization]	5	4	20	Find additional budget; downsize the scope of the survey; split up (core) survey into smaller units
Cooperation with information providers and others needed to collect data turns out to be laborious [implementation]	3	4	12	Develop backup scenario, e.g. procedures to direct institutions in “problematic” countries to online (automated) surveys
A lack of interest in heritage institutions [implementation]	4	2	8	Increase community engagement efforts if the signals are alarming
Response of institutions in the heritage sector remains poor, due to low estimated benefits [implementation]	4	1	4	Increase community engagement efforts if the signals are alarming
Institutions are unable to come up easily with the necessary data [implementation]	4	5	20	In earlier survey efforts (ENUMERATE Thematic Network) this proved to be a major challenge. Anticipating this “risk” should actually be a structural activity in any digital heritage monitoring project. One of the things done in

				ENUMERATE Core Survey 3 was simplifying the questionnaire, which of course raises the new problem of lacking information
WP lead insufficiently equipped [budget!]	4	4	16	Seek additional funding; strive for alignment with existing survey efforts; consider options to distribute WP lead tasks among another stakeholders
(Statistical) Expertise needed not sufficiently represented in WP participants [analysis]	5	1	5	Downscaling the survey. Liaise with experts in the field and/or find additional financial means
Quality of third party survey data insufficient, preventing the data to be harmonized [analysis]	4	4	16	Preparatory actions are aimed at avoiding this; “outliers” or questionable data will be skipped in the analysis phase
Ingested data not statistically valid (because of…) [analysis]	3	4	12	Subsampling (separating valid from invalid results); report on procedures to tackle this in future data ingest procedures
Barriers (any) prohibit effective use of (“publishing”) survey data (doubts about the validity of data, copyrights, strategic considerations etc.) [dissemination]	2	4	8	Use statistically sound procedures; encourage awareness of the benefits of the dissemination of results; etc.
Work being abandoned after end of funding period (sustainability not accomplished) [dissemination]	2	4	8	The WP will pay much attention to achieving a sustainable service [Observatory]; since by definition this risk can only surface after the end of the current funding period, more cannot be done

Appendix 2: References

Website examples:

http://www.ala.org/alcts/resources/preservation/presstats
ALCTS is the Association for Library Collections & Technical Services, a division of the American Library Association.

http://www.benchmarkemancipatie.nl
Atria, http://www.atria.nl/atria/eng is the Knowledge institute on emancipation and women's history in the Netherlands.

http://www.cultuurindex.nl
The Arts Index Netherlands is initiated by the Boekman Foundation, http://www.boekman.nl/en in co-operation with The Netherlands Institute for Social Research (SCP), http://www.scp.nl/english/Organisation/About_SCP with the aim of collecting a great amount of relevant and available cultural data and presenting the collected data in an accessible and understandable way.

http://erfgoedmonitor.nl
The Cultural Heritage Agency - part of the Ministry of Education, Culture and Science in the Netherlands. The Cultural Heritage Agency regularly collects data to enable the heritage to be 'measured'.

http://www.europeanvaluesstudy.eu
The European Values Study is a large-scale, cross-national, and longitudinal survey research program on basic human values. It provides insights into the ideas, beliefs, preferences, attitudes, values and opinions of citizens all over Europe.

https://www.openaire.eu
Portal includes over 11.5 million open access documents. Also with statistical graphics.

http://www.statista.com
Statista is one of the world's largest (commercial basis) statistics portals.

http://uk.swing.eu
Swing is a commercial party that makes applications to manage, analyse and present geo-statistical data.