ENUMERATE

… wo Zahlen zählen …

… wo Zahlen zählen …

… oder …

Antwort = f $_{(Frage)}$

1. Enumerate ?

2. Was bisher geschah …

3. Kritik der Fragen …

4. Was kommen wird …

ENUMERATE is a EC-funded project, led by Collections Trust in the UK. The primary objective of ENUMERATE is to create a reliable baseline of statistical data about digitization, digital preservation and online access to cultural heritage in Europe. Currently, statistical data on Europe's digital heritage is tentative and scattered at best. For the European Commission and many of the agencies and actors in the field of culture there is no consistent evidence base for making strategic decisions on investments in digitisation. ENUMERATE will bring about major improvements in the quality and availability of intelligence about digital heritage.

- Ein Projekt („Thematisches Netzwerk")
- EU finanziert (2011-2014)
- Statistik der europaweiten Digitalisierung im Kulturerbe-Bereich im Fokus
- Verlässliche Statistik als Ziel
- Diese kann Grundlage für strategische Entscheidungen zur Digitalisierung sein

Die Partner

- Collections Trust, UK (project coordinator);
- Digitaal Erfgoed Nederland (DEN), Netherlands;
- Stiftung Preußischer Kulturbesitz, Germany;
- Digibís, Spain;
- FARO Vlaams Steunpunt voor Cultureel Erfgoed, Belgium;
- Ministère de la Culture et de la Communication, France;
- Österreichische Nationalbibliothek, Austria;
- Narodna in univerzitetna knjižnica (National and University Library), Slovenia;
- Országos Széchényi Könyvtár (OSZK; National Széchényi Library), Hungary;
- The European Library (hosted by the Koninklijke Bibliotheek, Netherlands).

Partner der Partner

Für jedes Land wurden von politischer Seite (MSEG) sogenannte National Coordinators benannt.

- Diese „Coordinators" sind das Bindeglied zwischen Projekt und Einrichtungen
- Sie sorgen u.A. für Übersetzungen von Fragebögen
- Sie sind Ansprechpartner der Einrichtungen

Darüber hinaus gibt es ein aus drei Personen bestehendes „Advisory Board"

Austria	Irene Hyna	irene.hyna(at)bmukk.gv.at
Belgium	Elena Phalet	elena.phalet(at)stis.belspo.be
Bulgaria	Deiana Danailova	d.danailova(at)mc.government.bg
Cyprus	Marinos Ioannides	gammat(at)cytanet.com.cy
Czech Republic	Yvona Brázdová	Yvona.Brazdova(at)mkcr.cz
Denmark	Mikkel Christoffersen	mikchr(at)kulturstyrelsen.dk
Estonia	Indrek Eensaar	indrek.eensaar(at)kul.ee
Finland	Minna Karvonen	minna.karvonen(at)minedu.fi
France	Sonia Zillhardt	sonia.zillhardt(at)culture.gouv.fr
Germany	Monika Hagedorn-Saupe	m.hagedorn(at)smb.spk-berlin.de
Greece	George Bolanis	gbolanis(at)gmail.com
Hungary	Lajos Vonderviszt	vonderviszt.lajos(at)oszk.hu
Ireland	Annette Kelly	akelly(at)librarycouncil.ie
Italy	Rosella Caffo	rcaffo(at)beniculturali.it
Latvia	Kristine Paberza	kristine.paberza(at)kis.gov.lv
Lithuania	Zinaida Manzuch	zinaida.manzuch(at)gmail.com
Luxembourg	Guy Frank	Guy.Frank(at)mc.etat.lu
Malta	Joseph Bezzina	joseph.bezzina(at)gov.mt
Netherlands	Marco de Niet	marco.deniet(at)den.nl
Norway	Lars Gaustad	lars.gaustad(at)nb.no
Poland	Joanna Potega	j.potega(at)bn.org.pl
Portugal	Helena Simões Patrício	hpatricio(at)bnportugal.pt
Romania	Irina Oberländer-Târnoveanu	irina(at)cimec.ro
Slovak Rep.	Martin Katuščák	martin.katuscak(at)fpv.uniza.sk
Slovenia	Gorazd Vodeb	Gorazd.Vodeb(at)nuk.uni-lj.si
Spain	María Luisa Martínez-Conde	martinez.conde(at)mcu.es
Sweden	Rolf Källman	rolf.kallman(at)riksarkivet.se
Switzerland	Genevieve Clavel-Merrin (ad interim)	genevieve.clavel(at)nb.admin.ch
UK	Nick Poole	nick(at)collectionstrust.org.uk

Erreichbar im Web: www.enumerate.eu

- Informiert über Projekt

- Bietet Zugang zu Ergebnissen

Der Plan … (um zu einer möglichst verläßlichen Statistik zu kommen)

1. Aufbauen auf Vorarbeiten (insbesondere Numeric)

2. Fragenkatalog entwickeln

3. Fragen stellen

4. Antworten analysieren

5. Fragenkatalog verbessern

6. Erneut fragen

Der Plan … (um zu einer möglichst verläßlichen Statistik zu kommen)

1. Aufbauen auf Vorarbeiten (insbesondere Numeric)

2. Fragenkatalog entwickeln

3. Fragen stellen

4. Antworten analysieren

5. Fragenkatalog verbessern

6. Erneut fragen

Der Plan … (um zu einer möglichst verläßlichen Statistik zu kommen)

1. Aufbauen auf Vorarbeiten (insbesondere Numeric)

2. Fragenkatalog entwickeln

3. Fragen stellen

4. Antworten analysieren

5. Fragenkatalog verbessern

6. Erneut fragen

Der Plan … (um zu einer möglichst verläßlichen Statistik zu kommen)

1. Aufbauen auf Vorarbeiten (insbesondere Numeric)

2. Fragenkatalog entwickeln

3. Fragen stellen   (Anfang 2012)

4. Antworten analysieren

5. Fragenkatalog verbessern

6. Erneut fragen

Survey Report on Digitisation in European Cultural Heritage Institutions 2012

by

Natasha Stroeker and René Vogels, Panteia (NL)
on behalf of the EUMERATE Thematic Network
May 2012

EUMERATE is funded under the
ICT Policy Support Programme part of the
Competitiveness and Innovation Framework Programme.
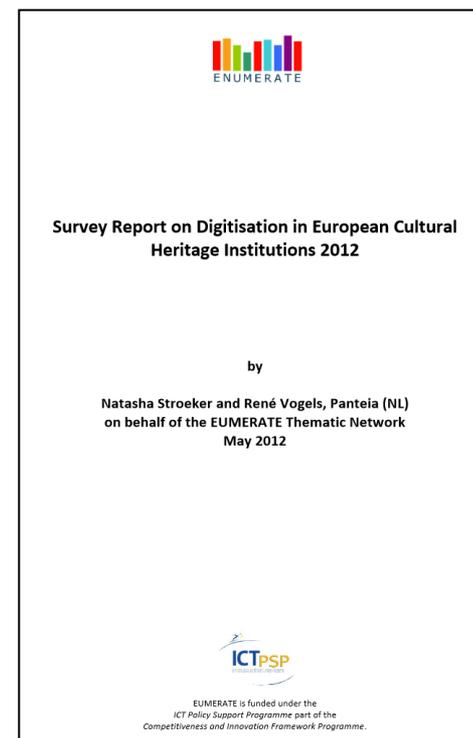
Der Plan … (um zu einer möglichst verläßlichen Statistik zu kommen)

Some highlights of the findings are:
- ~ 83% of cultural heritage institutions have a digital collection;
- ~ 20% of all collections have been digitised and ~ 57% still needs to be digitised (for 23% of collections over all there is no need to digitise);
- More than 50% of cultural heritage institutions collect born digital materials;
- ~ 34% of institutions have a digitisation strategy;
- ~ 85% of institutions use Web statistics to measure the use of their digital collections;
- 2 years from now institutions estimate to make twice as much of their collections accessible through Europeana as compared to today;
- On average 3.3% of paid staff in all cultural heritage institutions is working full time on digitisation.

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

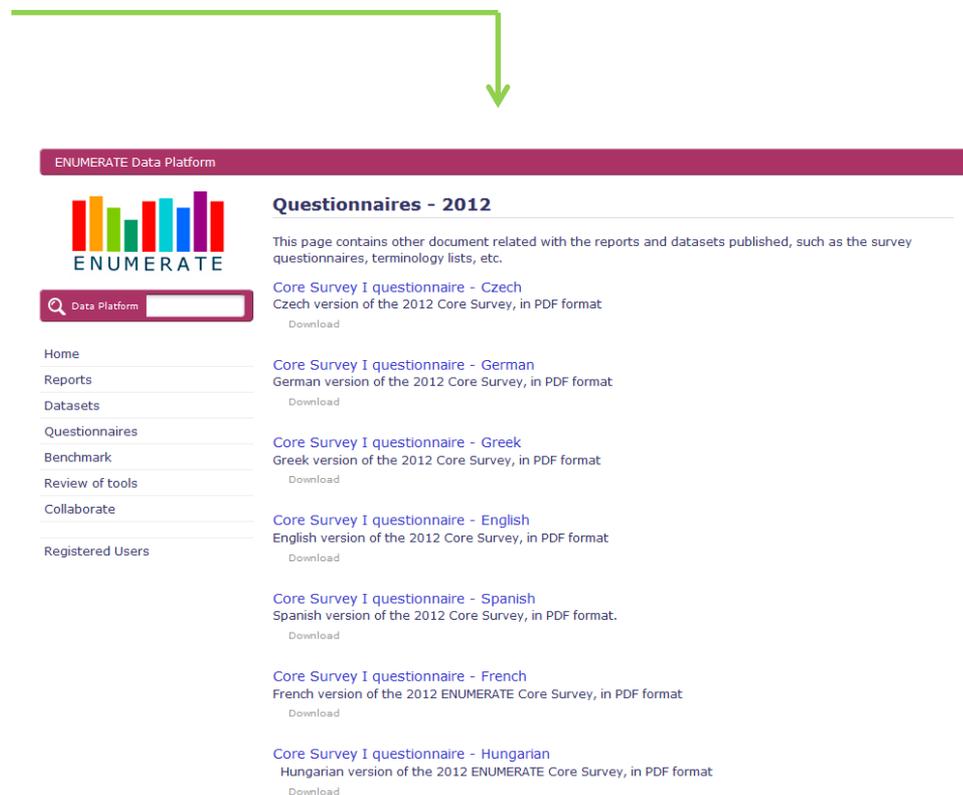**Der Plan … (um zu einer möglichst verläßlichen Statistik zu kommen)**

- 1. Aufbauen auf Vorarbeiten (insbesondere Numeric)
- 2. Fragenkatalog entwickeln
- 3. Fragen stellen
- 4. Antworten analysieren
- 5. Fragenkatalog verbessern
- 6. Erneut fragen

3. Kritik der Fragen …

- ○ Antworten analysiere
- ○ Fragenkatalog verbessern

3. Kritik der Fragen …

Das Vorgehen …

1. Bildung von Themenkomplexen

2. Expertengespräche zu den Themenkomplexen

3. Darauf aufbauend spezielle Fragen an einige Einrichtungen

4. Analyse der Antworten auf speziellen Fragen

5. Überarbeitung des Gesamtfragebogens

Das Vorgehen … *Bildung von Themenkomplexen*

Vier Themenkomplexe gebildet

1. Kosten

2. Umfang

3. Zugang

4. Erhaltung

Das Vorgehen …   *Expertengespräche zu den Themenkomplexen*

1. Kosten (03/2012)

2. Umfang (01/2012)

3. Zugang (06/2012)

4. Erhaltung (10/2012)

Das Vorgehen …   *Darauf aufbauend spezielle Fragen an einige Einrichtungen*

(Research Design:)
Es wäre unmöglich gewesen alle Fragen in allen Sprachen an alle zu richten. Es wurde entschieden Einrichtungen aus verschiedenen Staaten den verschiedenen Themenbereichen zuzuordnen.

Das Expertentreffen zu „Erhaltung" ergab:

„A separate high level questionnaire on digital preservation as part of the Thematic Survey was considered to be too complex, too premature and as a result ineffective"

➔ Erstmal ausgeklammert

Beteiligt:

Ungarn
Spanien
Portugal
Slowakei
Schweden

Insgesamt: 24 Einrichtungen (Große und kleine Einrichtungen)

(Um das Verfahren/Vorgehen zu zeigen wird der Themenbereich Zugang etwas ausführlicher dargestellt)

Die Fragen in Core Survey 1 lauteten:

[19] Does your organisation have an explicit (written) policy regarding the use of your digital collections?

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

[21] If yes, how?

[22] Please indicate estimated percentage of all the digital objects you have that are and/or will be accessible through the access options

Beobachtung aus den Antworten aus Core Survey 1:

The question about written policy documents showed that in a sample of 1488 institutions 31% have an explicit policy regarding the use of the digital collections. The second question in this section produced an average of 42% of institutions (n=1495) measuring in some way the frequency of accessing their digital collections. The answers to questions like these can become more useful if they are supplemented with information about the intended users, the conditions of use, and specific monitoring practices.

31% haben eine schriftlich niedergelegte Richtlinie für die Gewährung von Zugang

42% messen die Häufigkeit des Zugriffs auf ihre digitalen Sammlungen

31% haben eine schriftlich niedergelegte Richtlinie für die Gewährung von Zugang

42% messen die Häufigkeit des Zugriffs auf ihre digitalen Sammlungen

Kritik / Ideen aus Kommentaren und Expertengesprächen:

"In the reviewing of Core Survey 1 several comments related to the ambiguity that may exist in answers about such things as the "(written) policy" and "measuring the number of times digital metadata and/or digital objects are being accessed". The general criticism concerned the high variability that is possible in the underlying reality: a written policy could be a frequently used and updated document of a hundred pages giving exact regulations and guidelines for offering access to an institution's digital collections; it could just as well be a one page only text, written a few years ago and lost in a drawer somewhere.

Something similar could be the case in monitoring the number of times a digital collection is accessed. How valid is the monitoring methodology? Is it done in all of the institution's publicly available collections? What is publicly available? Is measuring organised at regular time intervals? How often then? What is being done with the results of measuring?"

… „Schriftlich niedergelegte Richtlinie" – darunter kann man viel verstehen …

… 42% messen Zugriffshäufigkeit, was bedeutet das? Wie verlässlich sind die Messungen …

„Schriftlich niedergelegte Richtlinie" … kann man hier genauer fragen?

Der Spezialfragebogen „Zugang" war deshalb umfangreicher und ergab:

There are 11 of the 24 institutes that have an explicit (written) policy regarding the use of their digital heritage collections. In 7 of these 11 cases this written policy document mentions *and* elaborates upon specific types of use of the digital heritage collections.

Für welche Nutzung ist Zugang (aus Sicht der Einrichtung) wichtig (10=sehr wichtig)

Type of use	Average
Educational purposes	9,0
Personal (cultural) development	8,8
Enjoyment	6,4
Creative reuse/Remix	4,6
Religious and commemorative use	3,7
Commercial trading	2,5

Die Frage nach der (schriftlich fixiert) intendierten Zielgruppe kann sinnvollerweise gestellt werden)

Kann man auch danach fragen, welche Kanäle des Zugangs intendiert sind?

Table 3.4 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 (currently and in 2 years from now) (n=21)

% of digital objects	Curr. Avail. (a)	Within 2 years (b)	Diff.	Core Survey 1 (a)	Core Survey 1 (b)	Diff.
Institutional website	52%	58%	6%	31%	47%	16%
Offline	48%	48%	0%	49%	61%	12%
National aggregator	27%	39%	12%	22%	37%	15%
Europeana	25%	38%	13%	15%	31%	16%
Thematic aggregator	11%	13%	2%	14%	24%	10%
Institutional API	6%	11%	5%	12%	23%	11%
3rd party API	6%	7%	1%	6%	13%	7%
Social media platforms like Flickr or Facebook	6%	8%	2%	-	-	-
Wikipedia	3%	4%	1%	3%	7%	4%
Something else	4%	9%	5%	16%	28%	12%

The platform that is mentioned the most for the presentation of the digital collections is the own institutional website, with an expected improvement of 6% in the next two years. A large improvement is expected for Europeana (from 25% now to 38% within two years).

Die Frage war leicht anders formuliert als in der Ursprungsbefragung, erbrachte jedoch ähnliche Ergebnisse. Ergebnis: Hiernach darf in bisheriger oder leicht veränderter Weise gefragt werden.

Zugriff: Wie sehen Antworten auf die Frage „Wie oft wird gemessen?" aus

Wie oft wird Zugang/Zugriff gemessen?

	%	N
On a weekly basis	6%	1
On a monthly basis	39%	7
On a yearly basis	39%	7
Other (please specify)	17%	3
Total	100%	18

Die Frage ist für die Institutionen zu beantworten. Es ist jedoch schwer, die Antworten zu interpretieren

Zugriff: Wie sehen Antworten auf die Frage „Womit wird gemessen?" aus

Womit wird Zugang/Zugriff gemessen?

Google analytics (13x)
Webaudit (1x)
database and in-house developed tools (1x)
Advanced Web Statistics 7.0 (1x)
the ILS own statistical module (1x)
AWStats (1x)
We use proprietary tools (statistics) of our integrated library system (DIGIBIB) (1x)
Logs (1x)
Database reports (1x)

Die Frage ist für die Institutionen beantwortbar. Es ist jedoch nicht leicht, die Antworten zu interpretieren

Zugriff: Wie sehen Antworten auf die Frage nach Messindikatoren aus

Total number of visitors
total sum / number of visitors / users (4)
total sum / number of unique visitors / users (3x).
Total sum of visited websites and how many of these are visited from abroad
Total number of views
Duration of session, traffic (downloads)

Access
Number of times the site was accessed
Number of times the digital objects were accessed
Access Digital records available
Access to digital objects
Access to viewed pages.

Die Frage ist für die Institutionen zu beantworten. Die Antworten sind jedoch recht vielfältig

Anmerkung der Einrichtungen zur Beantwortung der Fragen (Zusammengefasst)

On average it took the respondents 60 minutes to fill out the questionnaire. (minimum: 15 minutes, maximum: 3 hours). There were two respondents who said that it took them two days to fill out the questionnaire. Almost all respondents needed to look up things before filling out certain questions.

Wenn allein die Beantwortung dieser Fragen soviel Zeit in Anspruch nimmt, dann ist dies im Gesamtfragebogen zu bedenken …

Beteiligt:

Österreich
Estland
Finnland
Deutschland
Irland
Luxembourg
Slowenien

Insgesamt: 55 Einrichtungen (vornehmlich große Einrichtungen)

Die **Fragen nach Umfang** in Core Survey 1:

[13] Estimate the percentage of your heritage collections that has already been digitally reproduced

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

Empfehlungen der Expertenrunde zu Umfang

Umfang zu messen ist ein schwieriges Unterfangen
1. Die Komplexität der Erscheinungsformen digitaler Entitäten ist groß. Wenn die Fragen zu allgemein sind, sind die Antworten wenig aussagekräftig. Sind die Fragen zu speziell dauert das Ausfüllen zu lange (man bekommt weniger Antworten)
2. Die Umfangsbestimmung von Born-Digitals ist ein eigenes Problem (war in früheren Umfragen oft nicht separat behandelt worden)

… deshalb Empfehlung:
Nach Quelle der Angabe (Woher beziehen sie Ihre Antworten) fragen

Im Thematic Survey wurde in Core Survey 1 eine vorgegebene Liste von Digitalisat-Typen getestet. Die Befragten sollten Aussagen zur Anwendbarkeit machen. Vor allem aber wurde versucht herauszufinden woher die Einrichtungen die Angaben (**wieviel von was** ist digitalisiert?) beziehen können.

- Einige wurden nach Ihrem Sammlungsmanagement-Programm(en) (CRS) gefragt und sollten Angaben daraus beziehen.

- Einige wurden gebeten eine einzelne digitalisierte Teilsammlung detailliert zu quantifizieren

- Einige wurden gebeten nach einer überarbeiteten Liste von vorgegebenen Digitalisat-Typen zu quantifizieren.

Wenn man die Frage nach Umfang an der Quelle der Information ausrichtet ….

On average the third approach proved to be the most time consuming, with an average of 80 minutes for filling out the questionnaire. The least time consuming approach is the CRS approach with on average half an hour for filling out the questionnaire.

Table 4.7A How much time was needed to fill out the questions for the different approaches?

	Number of MinutesMean	Percentile 25	Percentile 75
CRS approach	31	10	45
Detailed subcollection approach	52	15	60
Improved Enumerate approach	80	15	60

"The results show that, whatever approach will be chosen for ENUMERATE Core Survey 2, we should indicate explicitly that just this one topic of the questionnaire will take on average about 1 hour to answer."

Ergebnis: Es kann sehr lange dauern die Anzahl von Digitalisaten nach Umfang per Arten von Digitalisat zu ermitteln. Das muss in Fragebogen berücksichtigt werden.

Beteiligt:

Belgien
Griechenland
Litauen
Niederlande
Polen
Schweiz
Tschechien
Vereinigtes Königreich

Insgesamt: 35 Einrichtungen (vornehmlich große Einrichtungen)

ENUMERATE

Die Fragen in Core Survey 1:

In the 2012 Core Survey one of the sections was about 'Digitisation Expenditure'. Essentially three kinds of information were collected: the total annual expenditures on *digitisation*, the number of staff (paid and volunteers) involved, and the sources of funding. **In the TS the aim was to widen the scope of the questions from 'digitisation' only to the total costs of creating or acquiring, maintaining and preserving digital heritage collections**. To get more specific information per cost item the TS experimented with questions in which respondents could quantify expenditures per activity. It was clear from the beginning that the challenge would be in posing questions that would relate as much as possible to the financial systems as used by heritage institutions.

Frage für Thematic Survey war, wie die Frage nach Digitalisierungskosten erweitert werden kann um alle diesbezüglichen Kosten zu ermitteln (z.B. für Pflege der Digitalisate)

Experten-Empfehlungen und Schlüsse aus Core Survey 1:

Frage für Thematic Survey war, wie die Frage nach Digitalisierungskosten erweitert werden kann um alle digitalisierungsbezogenen Kosten zu ermitteln (z.B. für Pflege der Digitalisate)

Das Expertentreffen erbrachte zudem die Anregung nach Gesamtkosten der Digitalisierung einer einzelnen spezifischen Sammlung zu fragen, um so zu ermitteln ob man Antworten erhält (oder: welche Antworten …).

Generelle Frage in Thematic Survey :

Does the most recent annual report of your institution explicitly mentions costs related to your institution's digital collections? [Yes / No / Do not know].

Antwort zeigt …
Only 10 (28.6%) out of 35 respondents (n=35) answered Yes to the question.
This may be taken as an indication that it may be difficult to separate the costs of digital collections from other reportable costs.

Generell: Nur wenige Einrichtungen weisen digitalisierungsbezogene Kosten in ihren Haushalten gesondert aus

Frage in Thematic Survey danach, wie schwer es ist, die Kosten der Digitalsierung einer spezifischen Sammlung zu bestimmen („incidental costs"):

(Je weiter unten, desto schwerer zu ermitteln:)

1. Project management

2. Acquisition of digital born material

3. Analogue-Digital conversion

4. Copyright clearance

5. Metadata creation and/or enhancement

6. Web design and/or software development

7. Logistics (shipment of collection for digitisation, etc)

8. Collection of user generated content (UGC)

9. Selection of material for digitisation

Frage in Thematic Survey danach, wie schwer es ist, die laufenden und folgenden Kosten der Digitalsierung einer spezifischen Sammlung zu bestimmen („structural costs"):

(Je weiter unten, desto schwerer zu ermitteln:)

1. Archiving (storage, including backups)

2. Licensing

3. Costs related to giving access to the digital collection

4. Project management

5. User outreach and support (I)

6. Editorial (including content selection and updating)

7. Usage analysis (including surveys, interviews, and other activities)

8. Activities concerning the (long-term) preservation of the digital collection.

Folgerungen aus Thematic Survey „Kosten":

The exercise to specify the costs of workflow activities ('activity based costing') made clear that some of the institutions were able to fill in concrete quantities. For others the exercise was too difficult or did not relate to their daily practice and local circumstances.

…

In all, due to the limited response and the lack of a harmonised terminology, we will have to be cautious when interpreting the results presented here.

Die Frage der Digitalisierungskosten - gestellt auf Grundlage von einzelnen Digitalisierungsaktivitäten - kann nur von einigen Einrichtungen beantwortet werden.

Der Plan … (um zu einer möglichst verläßlichen Statistik zu kommen)

1. Aufbauen auf Vorarbeiten (insbesondere Numeric)

2. Fragenkatalog entwickeln

3. Fragen stellen

4. Antworten analysieren

5. **Fragenkatalog verbessern**

6. Erneut fragen

3. Kritik der Fragen …

Relevant outcomes for the ACCESS questionnaire:

Some interesting qualitative outcomes of additional questions that did come out of the Thematic Survey are:
- The question on 'type of use'
- The question 'How often do you analyse the results of your access monitoring activities'

Possible changes for the Core Survey 2 based on the outcomes of the ACCESS questionnaire are:
- For the question '% of digital objects accessible' a new answering category can be added in Core Survey 2, namely: 'social media platforms like Flickr or Facebook'.
- For the question 'methods of use to measure the number of times digital metadata and/or digital objects are being accessed' an additional question can be added in the Core Survey 2: When using web statistics, which are used? (google analytics or other?)

Relevant outcomes for the SIZE questionnaire:

Some interesting outcomes of this part of the survey:
- Even though the sample is quite small, the dataset gathered seems to confirm most of the outcomes from the ENUMERATE Core Survey 1, both about the size of the digital collections and the popularity of specific collections types for digitisation.
- We have also made some progress in better understanding the composition of born digital heritage collections across Europe.

Possible changes for the Core Survey 2 based on the outcomes of the ACCESS questionnaire are:
- The survey proved that the CRS approach was the most time effective for the respondents. However, the output from this approach turned out to be not as 'rich' as in the other two approaches. This needs further discussion in the ENUMERATE team.
- There was little discussion about the classification that ENUMERATE created to designate digital heritage collection types, so this improved classification can be used in future surveys.
- Use 'descriptive metadata' instead of 'metadata'.

Relevant outcomes for the COST questionnaire:

Some interesting outcomes of this part of the survey:

- A minority of institutions report separately on their digital activities in their annual reports. If this doesn't change, it seems unlikely that future surveys will be able to collect really big datasets on costs of digitisation.
- Although the sample is small, there seems to be a trend that the cultural heritage institutions in Europe tend to spend more incidental costs on oursourcing activities than structural costs. Long term management of digital collections is very much an in-house activity.
- There is no univocal way institutions register their costs. Some institutes do actually use activity based costing, where others use another method.
- Incidental cost could be identified, structural cost were much more difficult to identify.
- Institutes use a large variety of different categories of costs. Some new categories mentioned are activity based while others are not activity based.

We understand better now which cost types can be more easily assessed by the institutes than others.

Overall conclusions

As the Thematic Survey was intended to test and improve the methodology with a smaller sample of institutions, there are few general conclusions that can be drawn about the status of digital cultural heritage in Europe. Most findings will feed into the preparations of the 2nd ENUMERATE Core Survey. The most important conclusions for this are:

Although the questionnaires were rather complex and time consuming they give us useful input for Corey Survey 2 and beyond: on average they took a lot of time to fill out, even from institutes that were very motivated to participate in the Thematic Survey. There were some institutes that didn't fill out the questionnaire in the end although they had said to do so. So careful consideration is needed to find the balance between on the one hand a questionnaire that can provide rich data to better understand the progress of digital cultural heritage in Europe and on the other hand an approach that doesn't scare away institutions from participating because of complexity or because participation is time-consuming.

Was kommen wird …

… ist schon da …

Core-Survey 2 seit letzter Woche online

www.enumerate.onderzoek.nl

Mitmachen !

Wo Zahlen zählen … muss nach Fragen gefragt werden !

Danke

Stefan Rohde-Enslin, Institut für Museumsforschung (SMB-PK)
In der Halde 1
14195 Berlin
s.rohde-enslin@smb.spk-berlin.de