European Commission
Innovation and Networks Executive Agency (INEA) CEF Telecom

Europeana Common Culture
Landscape of national aggregation in Europe and establishment of emerging national aggregators

Information on the Action	
Grant Agreement Nº	INEA/CEF/ICT/A2018/1633581
Action Title (Art. 1 of G.A.)	Europeana Common Culture
Action Nº (Art. 1 of G.A.)	2018-EU-IA-0015

Milestone contributors	
Author(s)	Tamara Butigan, National Library of Serbia Kathryn Cassidy, Trinity College Dublin Natalia Jeszke, Poznan Supercomputing and Networking Center Peeter Kondratjev, Consortium of Estonian Libraries Network Aleksandra Nowak, Poznan Supercomputing and Networking Center Tomasz Parkoła, Poznan Supercomputing and Networking Center Frederik Truyen, KU Leuven Artūrs Žogla, National Library of Latvia
Contributor(s)	Rob Davies, Cyprus University of Technology Lore Felmayer, Kulturpool Henning Scholz, Europeana Foundation
Reviewer(s)	Valentine Charles, Europeana Foundation

Table of Contents

Table of Contents 2

Document history 5

List of abbreviations 7

Executive summary 9

Introduction 10

Methodology of work 12

Survey on aggregation landscape 14

Purpose and scope 14

Methodology 15

Analysis of responses 16

General information / aggregation DNA 16

Aggregator's mission 17

Scope of aggregation 17

When did it all start? 19

Authorities support 20

Partners/Relations 21

Types of CHIs served by aggregators 22

Type of aggregation 22

Communication 23

What are you communicating to your partners and data providers? 23

Types of collaborating institutions 24

Partnering with other aggregators to deliver content to Europeana 25

Aggregation route from CHI to an aggregator 25

Awareness of the current (national) aggregation landscape 26

Availability of an aggregator to collaborate with 27

Aggregation activities and government policies 28

Services 28

Provided services 28

Gaps in provided services 30

Staff / capacity / expertise 31

Expertise of NAs staff 31

Support given to CHIs 33

Staff capacity 34

Competencies of staff 35

Skills lacking in the staff 36

Data 36

Size of dataset 36
Metadata formats/models 37
Linked Open Data support 39
Content aggregation 41
Technical infrastructure 41
Functional modules 42
Search portal 43
Data aggregation techniques 45
Public API 46
Harvesting workflow 47
Data ingestion and storage 47
Data processing 49
Cooperation with Europeana 50
Mission 50
Delivery of data to Europeana 51
Presentation of data in Europeana Collections 54
What are the main weaknesses of the current aggregator model (Europeana partners with aggregators which aggregate from cultural heritage institutions)? 55
Overall experience/lessons learned from setting up your aggregation services 56
Sustainability 57
Summary of the survey results 60
Establishment of emerging national aggregators 64
Consultancy service 65
Establishment of aggregators 65
Estonia 65
Introduction 65
Methodology of work 66
Future plans 67
Lessons learned 67
Serbia 68
Introduction 68
Methodology of work 69
Future plans 70
Lessons learned 70
Latvia 71
Introduction 71
Methodology of work 71
Future plans 72
Lessons learned 73
Ireland 73

Introduction 73
Methodology of work 74
Future plans 76
Lessons learned 76
Summary 77
Considerations on national aggregation policies 77
Recommendations on national aggregations practices & policies 80
Recommendations on national aggregation practices 80
Recommendations to the member states on National Aggregation Policies 85
National Digital Transformation Strategies 85
Frameworks and standards 86
Capacity building 87
Stakeholder engagement 88
Collaboration at the Europeana Aggregated level 89
Suggested actions 89
Conclusions 90
Annex 1. Survey questionnaire 92

Document history

Version	Author	Date	Changes
0.1	Aleksandra Nowak, Natalia Jeszke, Tomasz Parkoła, PSNC	29.05.2019	Initial version of the document.
0.2	Tomasz Parkoła, PSNC	29.10.2019	Improvements in the context of the analysis of survey results
0.3	Tomasz Parkoła, PSNC	13.02.2020	Sections on summary and recommendations added. Improvements.
0.4	Valentine Charles, Europeana Foundation	30.03.2020	Review of the document.
0.5	Tomasz Parkoła, PSNC	09.04.2020	Adaptations and changes based on the review.
0.6	Frederik Truyen, KU Leuven	26.04.2020	Recommendations on policy added.
0.7	Tomasz Parkoła, PSNC	12.05.2020	Introduction modified (objectives from policy recommendations integrated).
0.8	Frederik Truyen, KU Leuven	13.05.2020	Graphs relevant for Activity 4 added.
0.9	Tomasz Parkoła, PSNC	27.07.2020	Updates based on inputs from Frederik Truyen and alignment with overall document structure.
1.0	Peeter Kondratjev, ELNET Consortium of ESTONIA; Kathryn	28.07.2020	Emerging aggregators section added.

	Cassidy, TCD; Tamara Butigan; NBS; Artūrs Žogla, LNB; Tomasz Parkoła, PSNC		
1.1	Frederik Truyen, KU Leuven; Tomasz Parkoła, PSNC	27.08.2020	Updated sections relevant to Activity 4. Small changes and improvements
1.2	Rob Davies, Cyprus University of Technology; Lore Felmayer, Kulturpool; Valentine Charles, Europeana Foundation; Tomasz Parkoła, PSNC	08.09.2020	Editorial work.
1.3	Tomasz Parkoła, PSNC	02.11.2020	Cover page, small improvements.

List of abbreviations

Abbreviation	Description / explanation
CHI	Cultural Heritage Institution, includes key actors in the cultural heritage sector, such as libraries, museums, archives or galleries.
CHIs	Cultural Heritage Institutions, plural of the Cultural Heritage Institution.
DCHE	Expert Group on Digital Cultural Heritage and Europeana, details can be found on https://ec.europa.eu/digital-single-market/en/expert-group-digital-cultural-heritage-and-europeana-dche.
EAF	Europeana Aggregators Forum, a cooperation space for NAs to exchange the knowledge and best practice that supports their work. Details can be found on https://pro.europeana.eu/page/aggregators.
EDM	Europeana Data Model, mode details can be found on https://pro.europeana.eu/page/edm-documentation.
ICT	Information and Communication Technology
IPR	Intellectual Property Rights
IT	Information Technology
LOD	Linked Open Data
NA	National Aggregator, an entity that works with cultural heritage institutions to gather authentic and trustworthy data and make it accessible through Europeana or other dissemination channels. National aggregators define their scope by specific country and they work with contributors situated within that country.
NAs	National Aggregators, plural of the National Aggregator.
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting, more details can be found on https://www.openarchives.org/pmh/.
PSNC	Poznan Supercomputing and Networking Center, R&D centre affiliated to the Institute of Bioorganic Chemistry of the Polish Academy of Sciences. PSNC is a partner in the Europeana Common Culture project, leading Activity 2.
R&D	Research and Development

SaaS	Software as a Service, more details can be found on https://en.wikipedia.org/wiki/Software_as_a_service.
URL	Uniform Resource Locator
XSLT	Extensible Stylesheet Language Transformations (XSL Transformations)

Executive summary

This report provides a synthetic view of the situation of national aggregation in Europe, complemented by a detailed exposition of relevant recommendations and best practices. All the analysis and conclusions presented in this report are based on a series of workshops and a survey conducted among representatives of 22 National Aggregators.

Various aspects of the national aggregation landscape in Europe are investigated in this report. However, it mainly focuses on organisational and technical matters. The scope of the investigation covers:

- Mission, legal mandate and capacity building.
- Aggregation workflow with data models, quality standards and harvesting pipelines.
- Services offered by NAs to a broader community.
- Cooperation with external entities and community building practices.

The key outcomes of this report are summarized in the Recommendations on national aggregation practices & policies and Summary of the survey results sections.

Introduction

This report summarizes the efforts and developments undertaken in Activity 2 (especially task 2.1) and Activity 4 in the framework of Europeana Common Culture project.

The goal of the task 2.1 in Activity 2 was to understand the state of the art of aggregation in Europe, especially in the context of the technical aspects of aggregation, including solutions and approaches that are used in existing and emerging National Aggregators (NAs, entities that work with CHIs on a national level to gather cultural heritage resources and make them accessible through Europeana and/or other dissemination channels). It was also important to collect lessons learned, i.e. good and bad experiences that NAs had through the years.

Activity 4 refers to the whole ecosystem of content aggregation to Europeana, including both the NAs participating in Europeana Common Culture and the other domain and thematic aggregators that operate in Europe. The main objective of Activity 4 was to develop and agree a common policy for aggregation so that each cultural heritage institution in the participating Member States will have a specified route to aggregate content to Europeana (or others).

The rapidly evolving digital transformation makes this task even more important: a common policy for aggregation, including a roadmap to capacity building, offers a much needed instrument to further the digital turn across European Heritage Institutions. The goal of the NA cannot be limited to the aggregation of content into Europeana, it implies exerting leadership at the national or regional level towards the implementation of digital workflows and the sharing of digital expertise.

In order to understand the aggregation landscape in Europe and develop a common policy for aggregation, a series of actions were undertaken, including planning and organisation of three workshops with NAs, an online survey to collect appropriate data and dedicated events to determine aggregation policy. In the course of the work more than 89 attendees from the majority of European NA were involved and provided valuable input to the structure of the activities carried out as well as to the final shape of this deliverable.

Strategic discussions and key guidance of the work took place during the three workshops with NAs organised in Riga, The Hague and Frankfurt am Main as well as two policy related

- during Europeana event under the Romanian Presidency[1] and the 2019 fall edition of the Europeana Aggregators Forum. In addition, in order to collect detailed data about the landscape of aggregation, it was decided to run an online survey and ask all European NAs to contribute.

In the next section of this document methodology of work in the scope of task 2.1 and Activity 4 is described. It is followed by a section describing the survey conducted to understand and investigate the aggregation landscape in Europe. The next section provides recommendations on national aggregation practices and policies. These recommendations have been determined based on the survey analysis exercise as well as conducted workshops and meetings. The last section provides a summary.

[1] https://pro.europeana.eu/post/europeana-event-under-the-romanian-presidency-supporting-strong-national-infrastructures

Methodology of work

In order to achieve the goal of task 2.1 and Activity 4 it was decided to run three workshops with NAs, organise dedicated policy events and conduct a survey. The workshops with NAs were organised in the following locations:

- Riga on 11th January 2019
- The Hague on 5th April 2019
- Frankfurt am Main on 13th June 2019

The first workshop was organised during the kick-off meeting. The aim of this workshop was to agree on a general plan of the work, including the scope and time-frame of consecutive actions that had to be undertaken. The key output of this workshop was a decision that an online survey would be used to gather information on the state of the art related to each NA. The main areas of interest in the context of aggregation landscape were also identified. To achieve synergy across different activities of the project, it was agreed that the survey should be conducted together with the survey on aggregation scenario planned in task 4.1 (Survey about the aggregation scenarios active in each Member State). An initial list of questions related to task 2.1 was created by PSNC and consulted with project partners. The final version of the survey of task 2.1 and task 4.1 was sent out to targeted respondents by the end of February 2019.

The second workshop was organised on 5th April 2019 in The Hague, The Netherlands as part of the EAF. During the second workshop, the initial survey results were presented. Based on these results and insights from the workshop attendees, a draft version of the report ‘Landscape of national aggregation in Europe’ was prepared. The draft version of the report was discussed during the third and final workshop organised on 13th June 2019 in Frankfurt am Main, Germany. Based on the discussions a final version of the deliverable was prepared.

The workshops were initially planned to be held in the Netherlands, Germany and Poland. In order to initiate the work as soon as possible and to take advantage of having the majority of targeted participants already gathered together at the kick-off event, it was decided to conduct the first workshop in Latvia (Riga) instead of Poland. The total number of workshops participants was above the assumed 30 (kick-off meeting: 41; second workshop: 33; third workshop: 15).

In April 2019 a policy maker workshop was organized in Iasi (Romania) on the occasion of the launch of the Romanian presidency of the European Union. In the workshop, representatives of DG CONNECT, Europeana Foundation, the Europeana Network Association, the Romanian Ministry of Culture and representatives from Romanian Cultural Heritage Institutions held a two-day meeting to discuss the requirements and feasibility of setting up a Romanian NA. A report on this workshop[2] was published in September 2019 and communicated on Europeana Pro in October.

On May 7-8 2019 Fred Truyen presented the preliminary results of the survey and workshops[3] to the DCHE in Luxembourg, where representatives of the Member States could hear the first findings and intended recommendations. This was very positively received and it was decided that the recommendations would be presented to the EAF and then would go back to the DCHE, so that the member states could be approached.

At the Aggregators' Forum meeting in Stockholm in October 2019, Task Leader KU Leuven led a workshop on "Empowering Collaborations" between Aggregators, National as well as Domain and Thematic. The results of this workshop have been gathered in a report[4] (a summary is also available in the slide presentation[5]).

A follow-up discussion has been held at the online EAF meeting on May 6th 2020. EAF members discussed 5 themes distilled from the gathered policy recommendations:

- National Digital Transformation Strategies (includes National/Regional Aggregation)
- Frameworks and Standards
- Capacity Building
- Stakeholder engagement
- Collaboration between aggregators

Over the Summer 2020, a dedicated Task Force worked to elaborate these themes further.

[2] https://pro.europeana.eu/post/final-recommendations-under-romanian-presidency
[3] https://drive.google.com/file/d/16JIvnf-FoWMV9g8RS2-cmDNHXurhMWMt/view?usp=sharing
[4] https://drive.google.com/file/d/1GD0t9ceERQTgUCznd1z2jz3WXXMAyS5y/view
[5] https://docs.google.com/presentation/d/1nPi5NR4kgmO8qSjAneDNlySrHcFRlUkZOmvt2Mxftsw/edit?usp=sharing

Survey on aggregation landscape

In order to collect information on aggregation practices from aggregators in Europe a survey has been prepared and conducted in the framework of task 2.1. The following subsections present the scope and purpose of the survey, methodology as well as survey results. It is important to note that the survey was sent out to targeted respondents together with the survey prepared in task 4.1.

Purpose and scope

The main purpose of the survey was to identify and learn about the landscape of aggregation across Europe in the context of technical characteristics like software platform in use, standards applied or workflows followed. It also asked a set of questions relating to policy aspects, looking into relations with funding authorities, Europeana and other aggregators. The targeted respondents of the survey were active players in the aggregation activities, i.e. National Aggregators.

The final survey was composed of several dozens of questions which were divided into the following sections:

- *General information / aggregation DNA* covering name and other basic characteristics.
- *Partnerships / relations* covering cooperation practices of aggregators with external entities.
- *Services* that asked about the services provided by aggregators to cultural heritage institutions / targeted users.
- *Staff / capacity / expertise* investigating the organisation of work and responsibilities within the aggregators.
- *Data* covering questions related to metadata and data aggregated and processed.
- *Technical infrastructure* covering questions related to the set up of hardware and software environment.
- *Functional modules* asking about features offered by aggregators to their users.
- *Harvesting workflow* investigating details related to metadata and content processing.
- *Cooperation with Europeana* investigating how relations with Europeana are handled by aggregators.

- *Overall experience/lessons learned from setting up your aggregation services* investigating experiences related to the whole process of setting up an aggregator.

The survey questionnaire is available in Annex 1 and the online version is available at https://goo.gl/ryVbY8.

Methodology

The survey was prepared in the form of an online questionnaire using Google Forms[6]. The final shape of the questionnaire itself including the questions was agreed during the consultations with project partners, especially in the scope of the first workshop organised in Riga. There were several types of questions available in the survey:

- Open questions - free text questions allowing respondents to provide their individual answer.
- Choice questions - single or multiple choice questions offered a predefined list of answers and allowed users to select:
    - only one answer (single choice question)
    - one or more answers (multiple choice questions).
- Matrix questions - were used when several characteristics of a single aspect needed to be gathered. These were organised as a table where each row represented a single choice question or multiple choice question. Columns gave the specific options for answers.

To achieve synergy across different activities within the project it was agreed that the survey will be conducted together with task 4.1 (survey on aggregation scenarios). Activity 2 and activity 4 leaders jointly proposed a set of questions within the aforementioned sections.

An online survey, containing 99 questions divided into 11 sections was created. Some of the questions had additional sub questions where respondents were asked for more details based on participant answers. The survey was published online at the end of February 2019 and was open for respondents until the end of March 2019. Appropriate dissemination channels and methods were used to reach targeted respondents (e-mail, basecamp messages, reminders).

[6] https://docs.google.com/forms/u/0/

Detailed description of question groups will be presented in the following chapters together with the analysis of participants' responses.

Analysis of responses

The survey was open to all aggregators active in the current European landscape. It means that various entities that currently work with cultural heritage institutions to gather data and make it accessible through Europeana, were invited to take part in the survey. These invited institutions (aggregators) and the actions they undertake constitute the European landscape of aggregation. The aggregators themselves can be domain, thematic, regional and national[7]. Domain & thematic aggregators define their scope by cultural sector (such as museum, archive or library) or by topic and theme (such as fashion) and they work with contributors from various countries. National or regional aggregators define their scope by specific country or region and they work with contributors situated within that country/region.

Altogether the survey collected responses from 37 respondents, including some domain and thematic aggregators in addition to 22 NAs. Some respondents indicated to be both domain as well as national aggregators. It is important to note that not all domain and thematic aggregators responded to the survey, as they were not the main target.

The following subsections contain analysis of specific answers related to survey sections. Each section is briefly described in order to give an idea of what information it was intended to collect within a section. As some of the questions can not be interpreted statistically and were asked to gain overall context, they are skipped in the analysis.

General information / aggregation DNA

This set of questions was intended to gather the basic information needed to understand who the aggregator is and what the aggregator is working with. Questions covered for instance the name, mission, geographical localisation or domain of interest of the aggregator. The following subsections discuss specific questions.

[7] https://pro.europeana.eu/page/aggregators

Aggregator's mission

The most important part of this section is related to the mission statement of the aggregator. It was crucial to understand why various organizations start their aggregation process and see whether the reasons were the same or similar in the whole aggregators community. The responses indicate that there is a common view on the mission of NA, i.e. many respondents provided answers that were aligned with each other. The most important aspects of the mission mentioned in the responses are:

- **to give access to cultural heritage objects**
    - free access is important
    - resources should be of high quality
    - single point of access to the resources
- **to promote resources and cultural heritage of its country**
- **to set up quality standards and create high quality data and metadata**

It is interesting to note that the aggregation activities themselves are not identified as the core mission of the NAs. The reason is that the aggregation activities are in fact a means to achieve the stated mission of NAs.

Scope of aggregation

The scope of aggregator services is in the majority of cases cross-domain (see chart 1). There is only one case where a NA indicated domain focus (on libraries and museums).

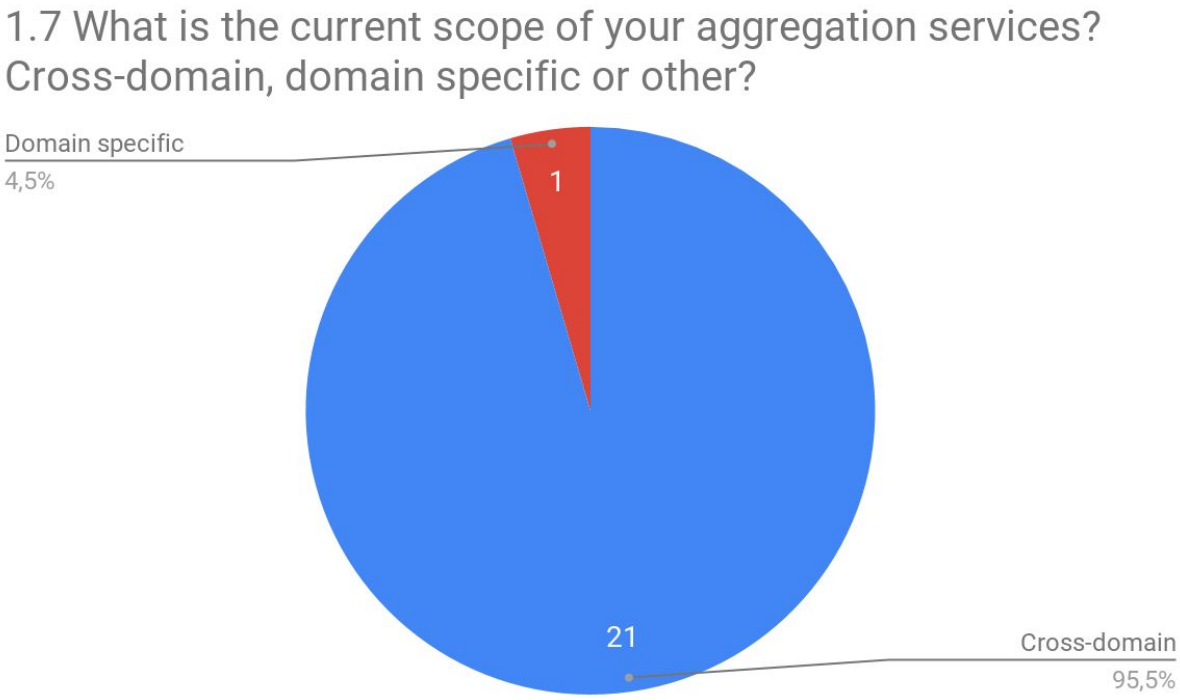


Chart 1 Scope of aggregation services in the context of domains covered.

In the follow-up question (investigating the domains of interest of the NA) it is visible that libraries, museums and archives play an important role in the aggregation activities (see chart 2). The most common pattern is an aggregation of data from both libraries and museums. Usually NA cooperate with institutions from various domains to support all the aspects of their country's cultural heritage.

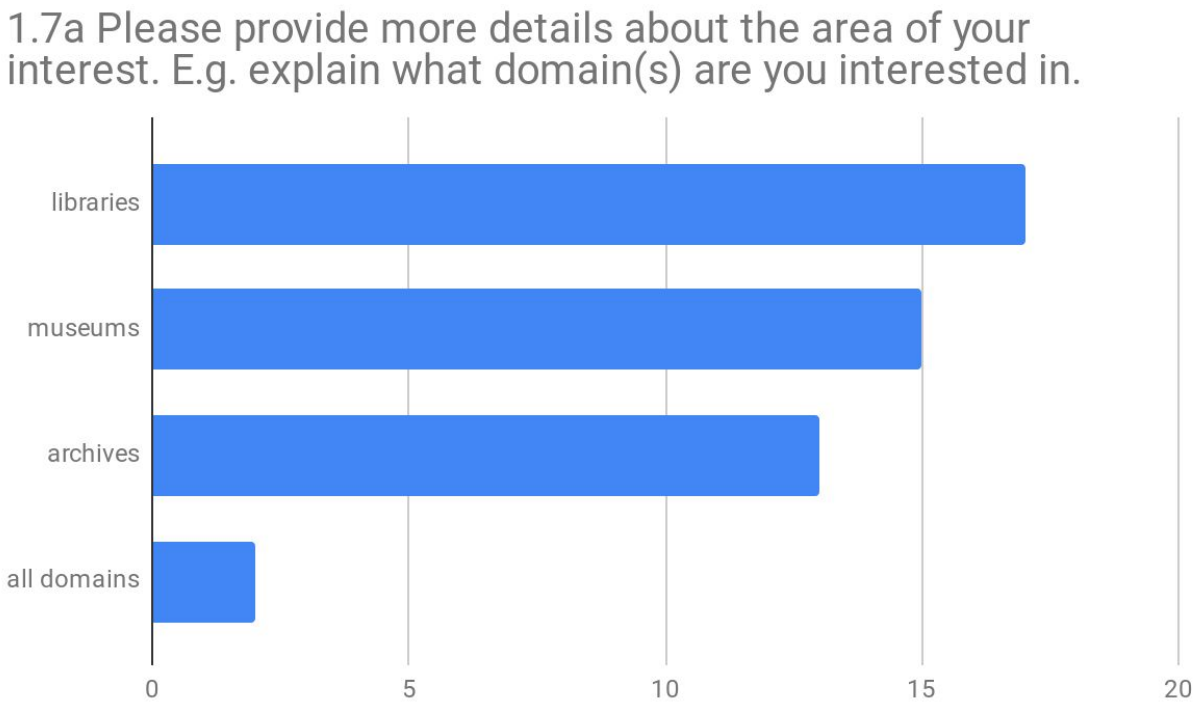


Chart 2 Area of interest (domains) of NA.

When did it all start?

One aggregator (Poland) initiated its activities in 2007, which was even before Europeana was established. The next 4 aggregators were established in 2008, which is the same year when Europeana was initiated and the Europeana Local project[8] was started. See Chart 3 for more details.

[8] https://pro.europeana.eu/project/europeanalocal

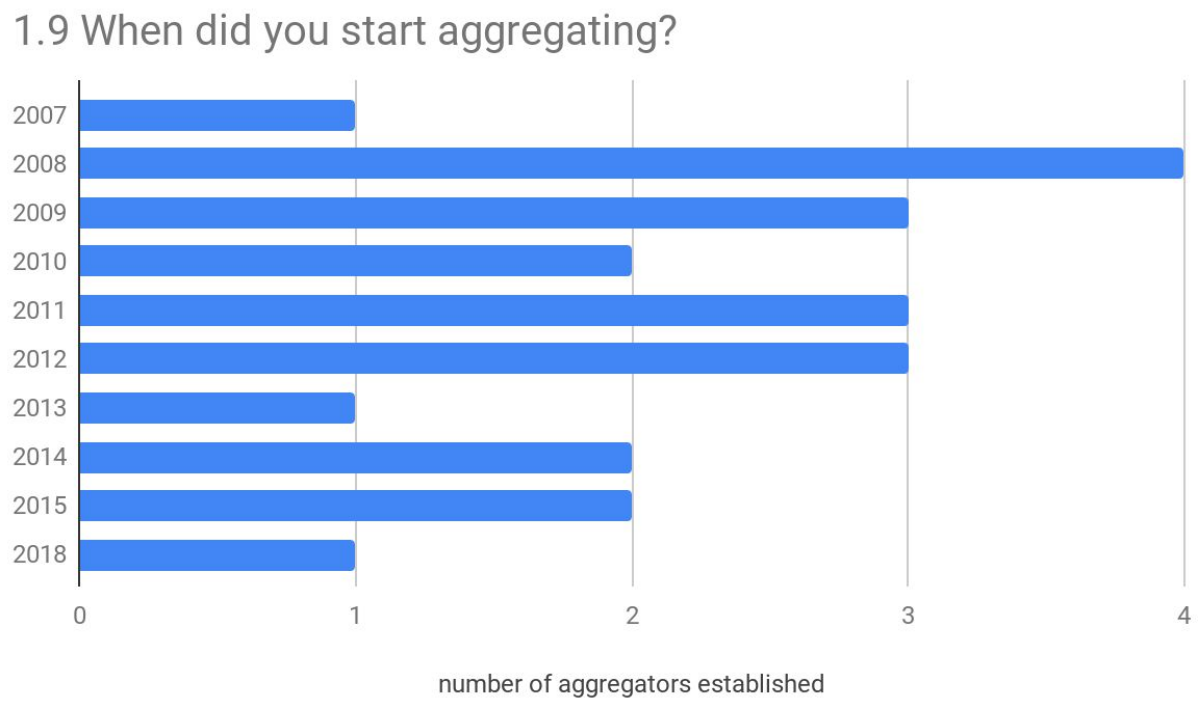


Chart 3 Number of NAs established each year.

Authorities support

The majority of NAs have a legal mandate from their government authority for their activities. Moreover, some NAs indicated that their activities were started on the authority's initiative.

The relationship of the NAs with the authorities of the respective country is a complex and individual matter. However, having a good established relationship or a legal mandate is the case for the majority of the NAs. Official support of authorities often helps NAs to have sustainable funding and organizational support.

The experiences of Aggregators (e.g. Poland, Bulgaria) that have no official mandate and at the same time are struggling to secure sufficient and sustainable funding and well deserved recognition, lead us to form a recommendation for new Aggregators: it is important to get a legal mandate from national government authorities. During the follow-up discussions with NAs (e.g. in the scope of the third workshop organised by the project) it was visible that a legal mandate is a good practice. Activities related to the recognition via the accreditation framework of NAs by the Europeana community help to tackle this issue in a bottom-up

manner, i.e. they indicate NAs that are important in the context of European aggregation landscape. These concerns should be a subject of discussions among Member State countries, e.g. in the context of the DCHE sub group or similar.

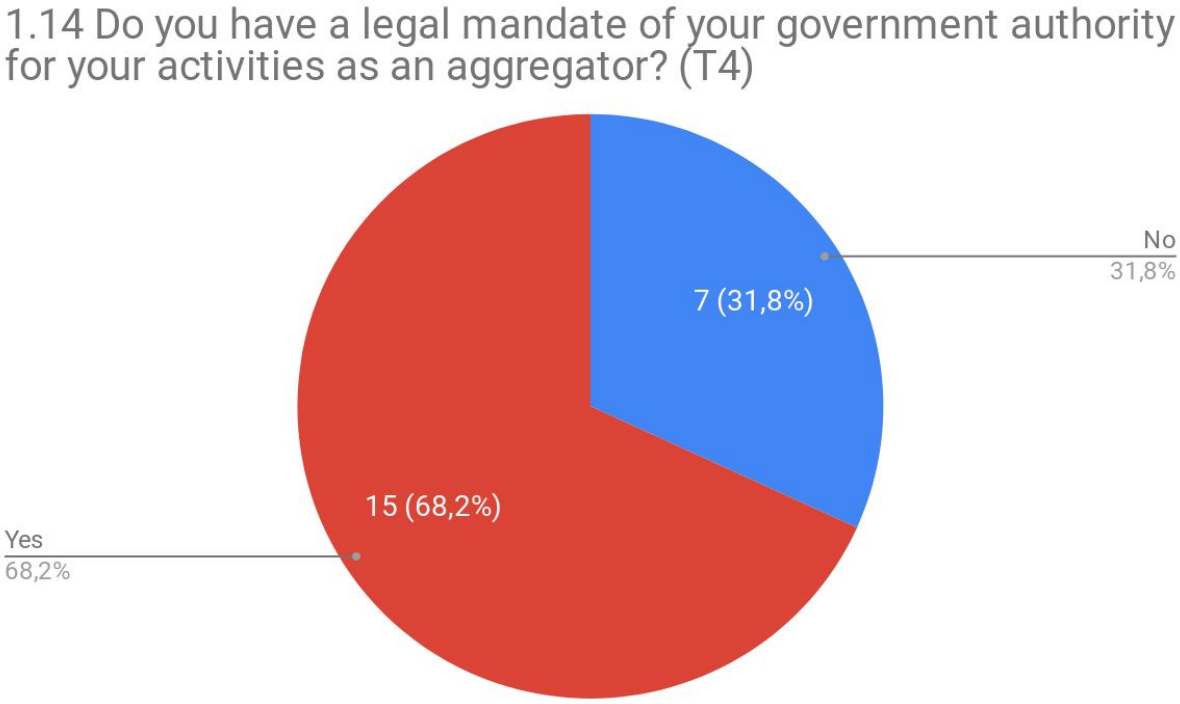


Chart 4 Number of NAs with and without a legal mandate in their country.

Partners/Relations

The section is focused on relationships that Aggregators build with their data providers. It also captures the relationship that Aggregators have or do not have with their countries' governments or other authorities and how it affects their ongoing activities.

From the technical perspective of the landscape it is interesting that the amount of data provided varies very much from country to country. There are cases where aggregators deal with several data providers, but there are also cases where they deal with hundreds. This obviously depends on the internal structure of the network of cultural heritage institutions in the country and the approach to statistics calculation, e.g. one may count consortium as a single entity, while others can count all of the institutions in the consortium separately.

Types of CHIs served by aggregators

Chart 5 presents types of CHIs served by the survey respondents. The responses indicate that many of the aggregators serve multiple types of CHIs and the majority serves at least 4 types.

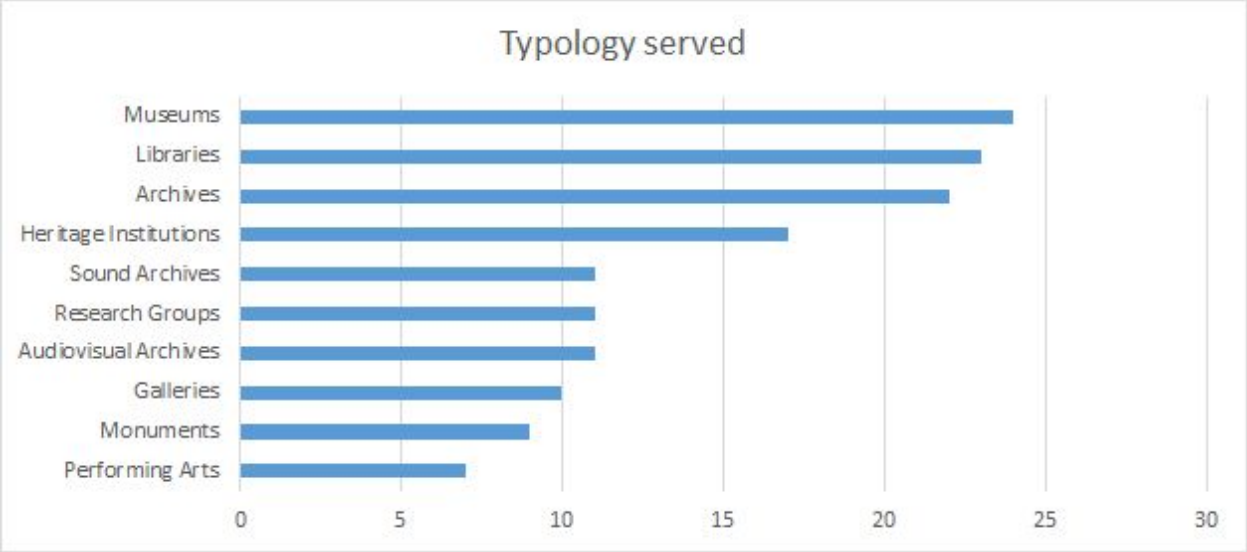


Chart 5 Types of CHIs served by NAs.

Type of aggregation

Another question related to partners and relations with them was about the aggregation type (see Chart 6). Three options were provided:

- Centralised - meaning that various institutions ingest data into one portal, according to the rules set by the portal.
- Distributed - meaning that the aggregator harvests metadata from external existing online systems and has to unify them in order to present them on the web or provide them to Europeana.
- Both - hybrid approach - where both centralised and distributed aggregation are in place.

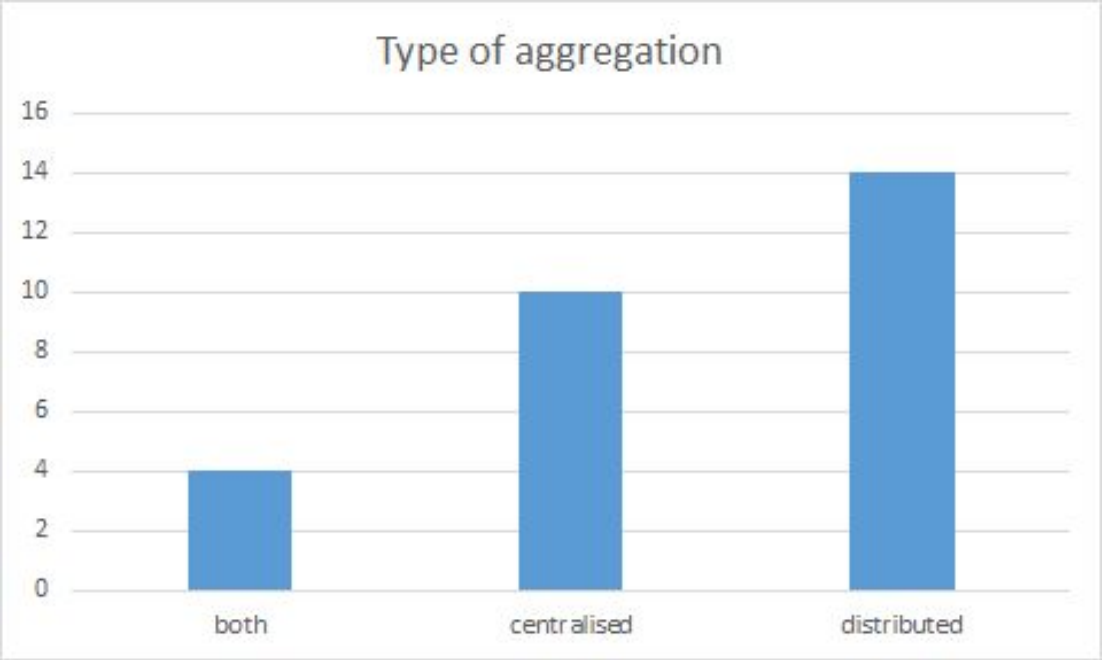


Chart 6 Type of aggregation conducted by aggregators.

Communication

This section investigated communication practices of aggregators with their data providers and as well as with aggregators from other countries and other institutions.

What are you communicating to your partners and data providers?

There are multiple topics that are disseminated via aggregators networks. However, the most common include guidelines, standards and recommendation as well as feedback related to data quality and aggregation process itself.

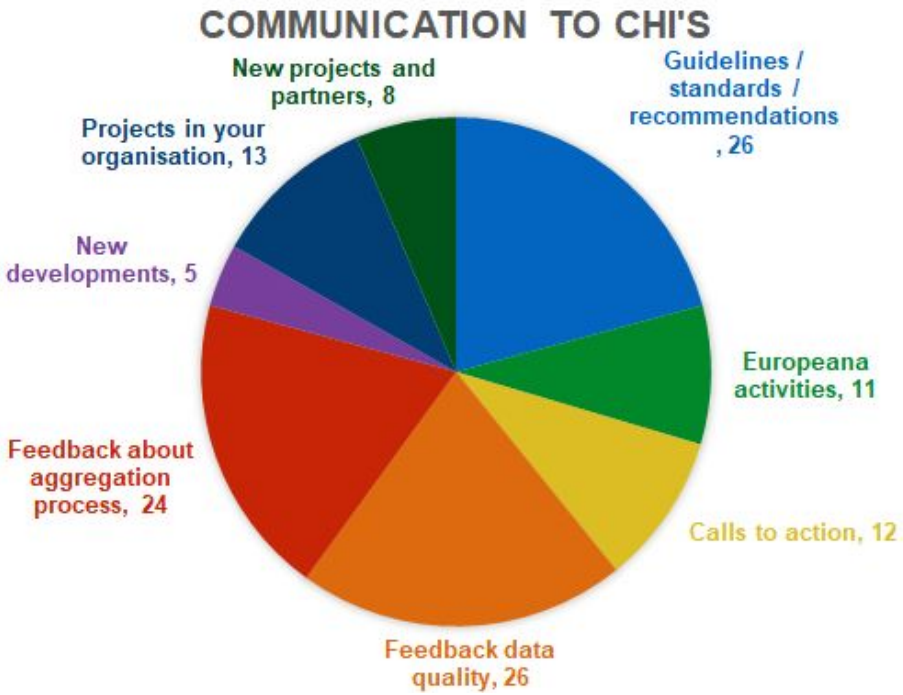


Chart 7 Topics communicated to CHIs.

Types of collaborating institutions

Aggregators collaborate with various institutions (see summary in Chart 8). However, they collaborate mostly with government bodies and other aggregators.

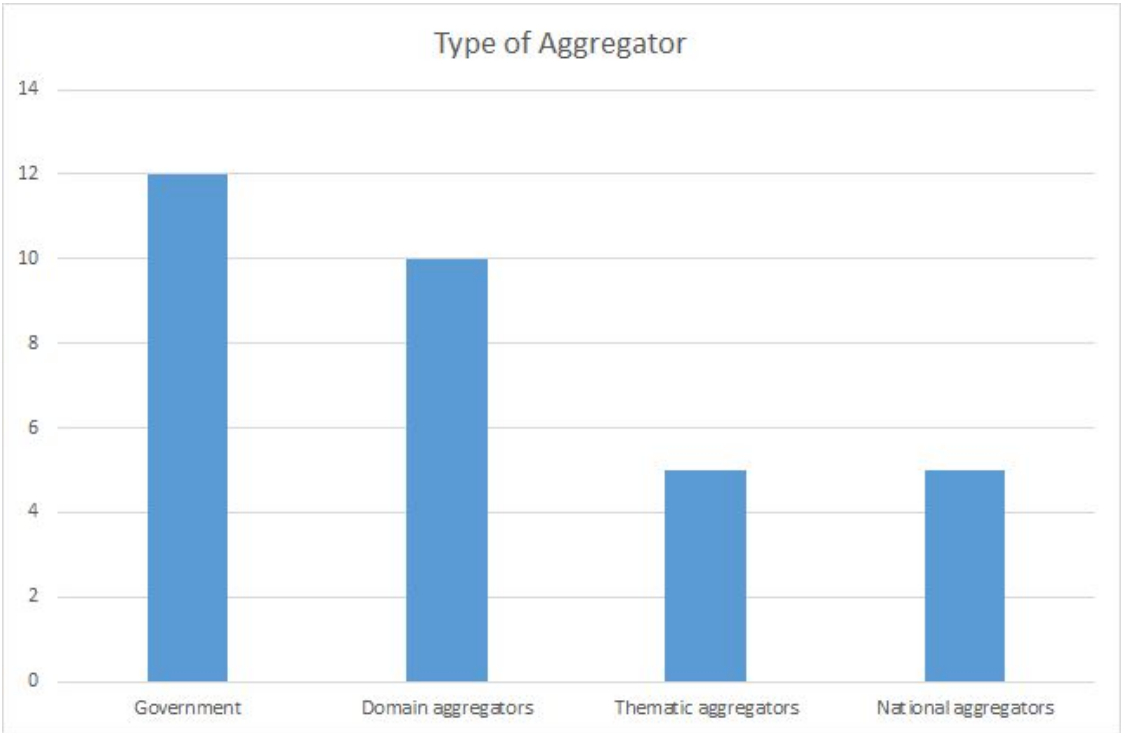


Chart 8 Topics communicated to CHIs.

Partnering with other aggregators to deliver content to Europeana

Respondents indicated that aggregators do collaborate to provide content to Europeana. The chart below shows the National Aggregators who have such partnerships. Most partnerships are with a domain or thematic aggregator.

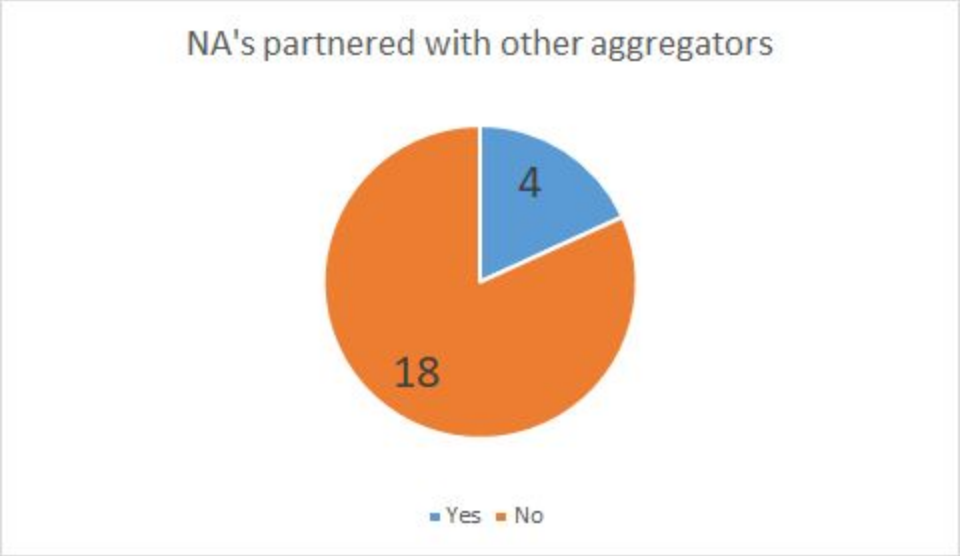


Chart 9 Collaboration with other aggregators in delivering data to Europeana.

Aggregation route from CHI to an aggregator

Most aggregators do believe that there is a clear aggregation route from CHIs to themselves. However 6 of them claim that there is no clear route available. See Chart 10 for results.

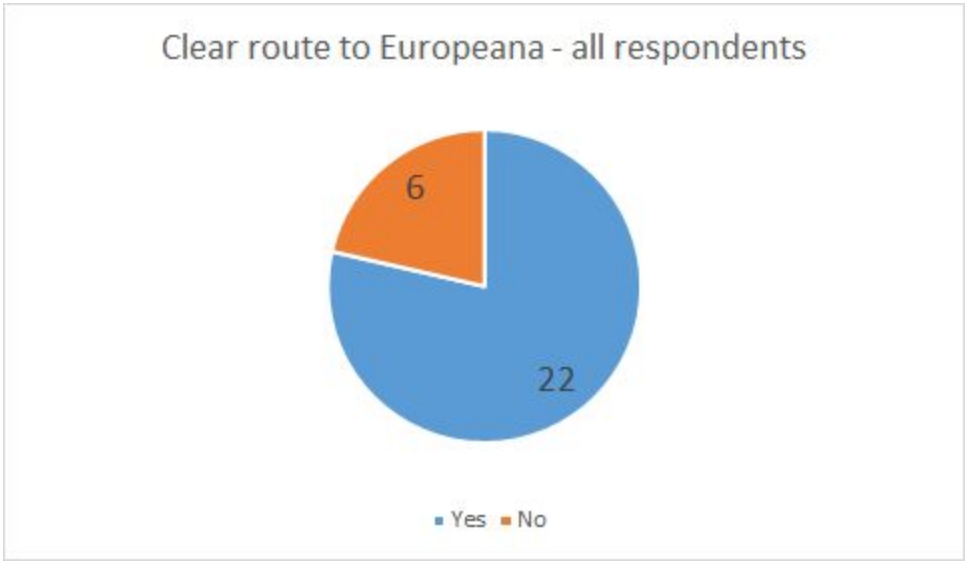


Chart 10 Opinion on clearness of an aggregation route from CHI to an aggregator.

Awareness of the current (national) aggregation landscape

Most of the aggregators claim that they have a clear view on the national aggregation landscape.

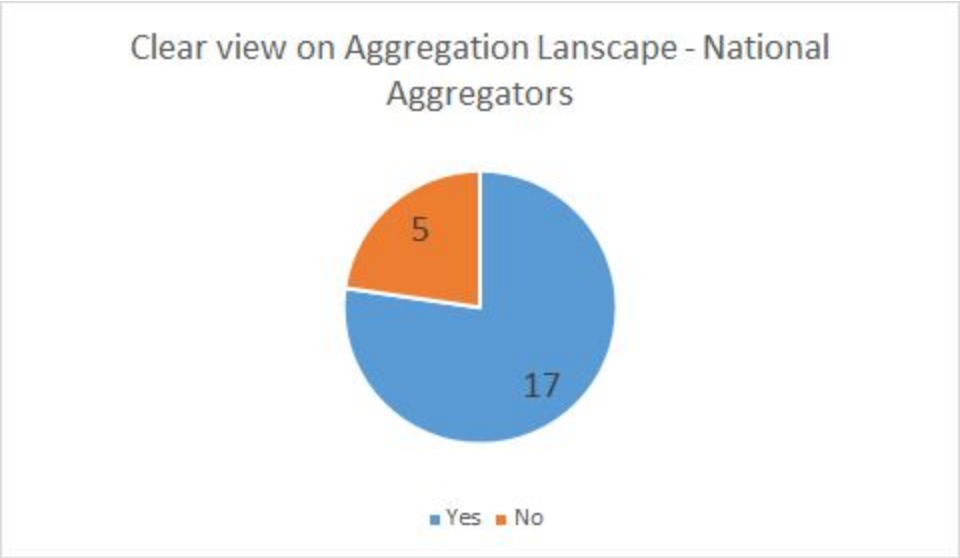


Chart 11 Awareness of the current (national) aggregation landscape.

Availability of an aggregator to collaborate with

Most aggregators claim that in every geographical location and content scope they operate there is an aggregator that can serve an interested CHI (see Chart 12 for details). However, there are still missing spots that were expressed by 3 National aggregators. There is still no national aggregator for Ireland, Belgium, and some indicate regions that have no access, without further specification.

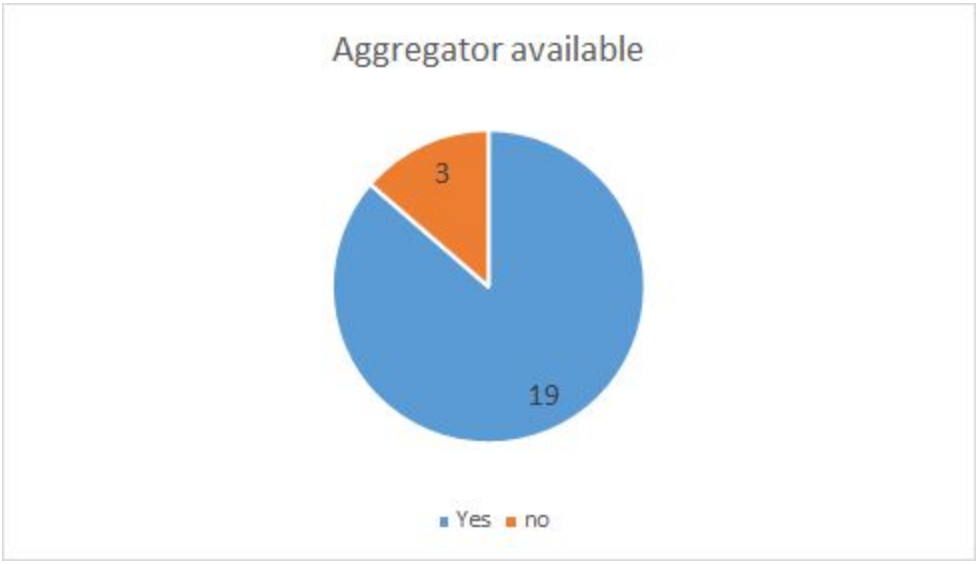


Chart 12 Awareness of the current (national) aggregation landscape.

Aggregation activities and government policies

Most of the surveyed aggregators state that government policies do not hamper aggregation activities (see Chart 13 that depicts results). However there are 5 respondents which claim that there are hampering policies issued by the government.

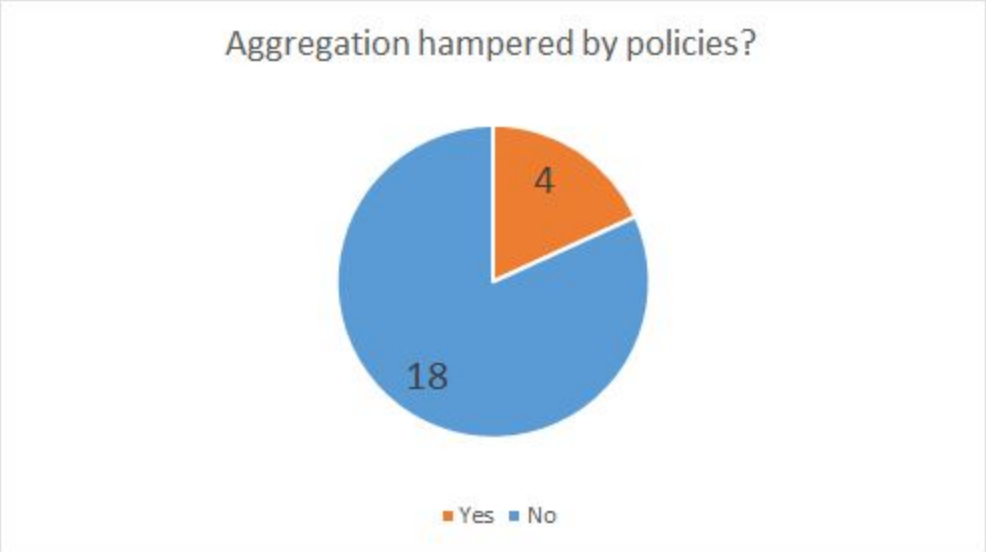


Chart 13 Awareness of the current (national) aggregation landscape.

Services

This section investigates what are the services that are offered by the aggregator to their partners and what are the gaps in this regard.

Provided services

Based on the provided responses (see Chart 14) it is clear that NAs offer their CHI's the following key services:

- Metadata mapping
- Workflow guidelines
- Technical support
- Consultancy

Metadata mapping is the most common service offered by NAs. Transformation of data between different formats and schemas is usually the main part in the aggregation workflow, therefore this result is natural and expected .

In addition, NAs also have extensive knowledge and experience in the cultural heritage domain. This is why they often provide support and consultancy for their partners in an operational manner, e.g technical support, consultancy and workflow guidelines. These activities lead to a common understanding of the aggregation pipelines and help improve efficiency and quality of results.

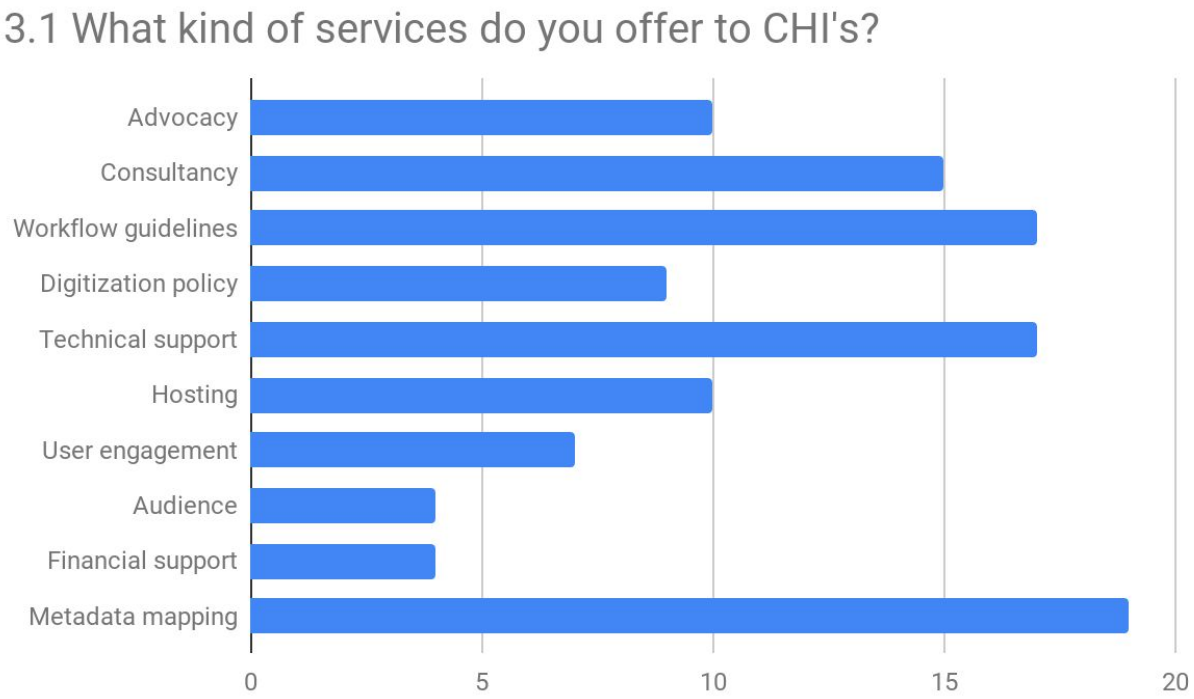


Chart 14 Services offered by NAs to cultural heritage institutions.

Correct usage of IPR is a necessity for all cultural heritage institutions. Therefore the survey investigates separately this topic. Results confirm that IPR guidelines and training is an important area of consultancies - offered by over half of the NAs (see chart 15).

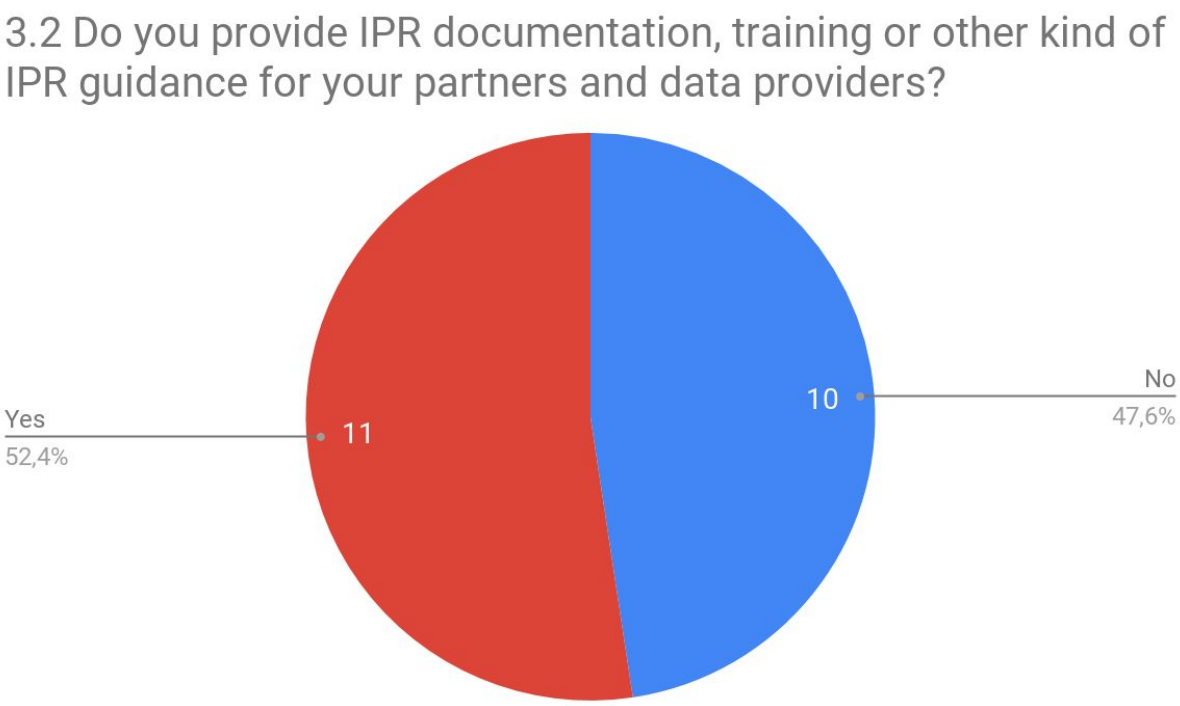


Chart 15 Number of NAs in the context of provided IPR support.

Gaps in provided services

Almost 30% of NAs were able to identify gaps in their services (see Chart 16). The most important gaps that they identified are as follows:

- IPR training
- Metadata mapping
- Archiving
- Digital exhibitions
- 3D and video objects

It means that although the majority of NAs feel good with IPR and metadata mapping, some of the NAs (less than 30%) are still struggling with these core topics. Archiving and digital exhibitions are not part of the usual core service provided by NAs, therefore it is natural that some of them are not able to handle these kinds of requests. 3D and video objects are, it seems, still not fully embraced as media types by NAs.

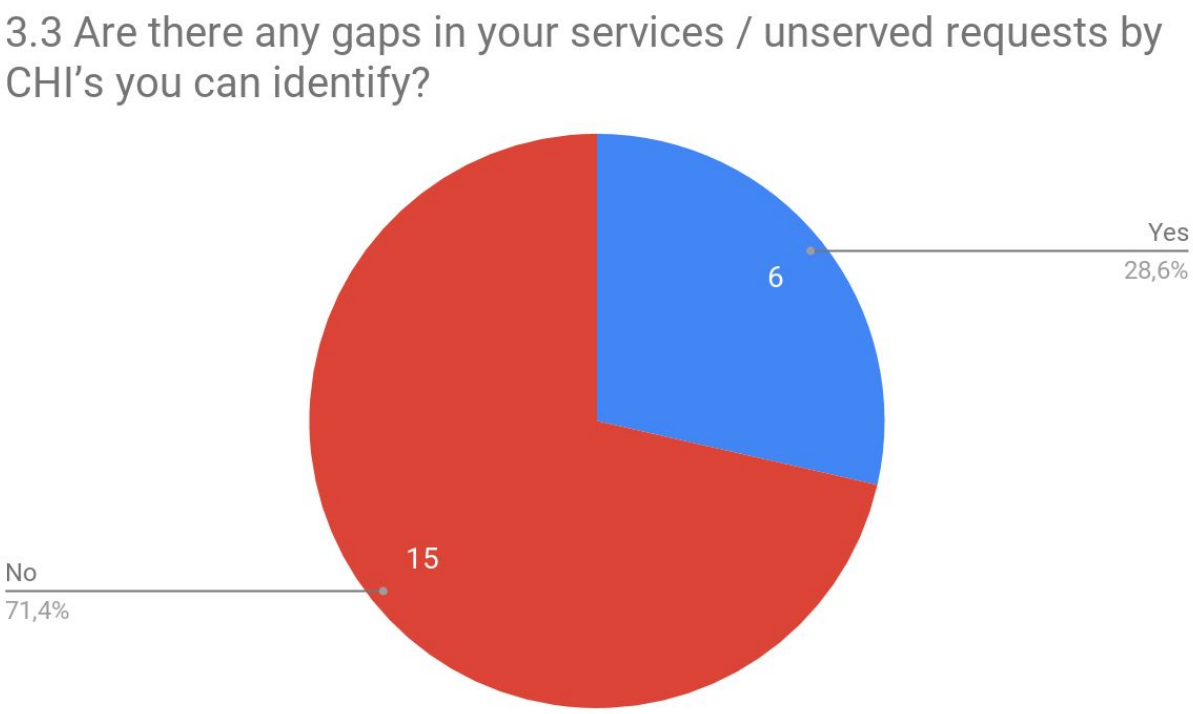


Chart 16 Number of NAs in the context of identified gaps in provided services.

Staff / capacity / expertise

This section covered questions related to staff that are employed to work on aggregation. The intention was to learn if all employees are employed directly by the aggregator or whether the staff is outsourced. It was considered important to know whether NAs believe that their staff has all the competences they need.

Expertise of NAs staff

There are two key groups of employees hired to make aggregator services run: IT developers/operations; as well as metadata/information specialists (see Chart 17). On one hand NAs work on a daily basis with data and metadata, therefore it is important for its staff to have proper competences - this is why hiring information and metadata specialists is so popular. On the other hand, aggregation services have software components, therefore it is essential to have employees that will make this software operational.

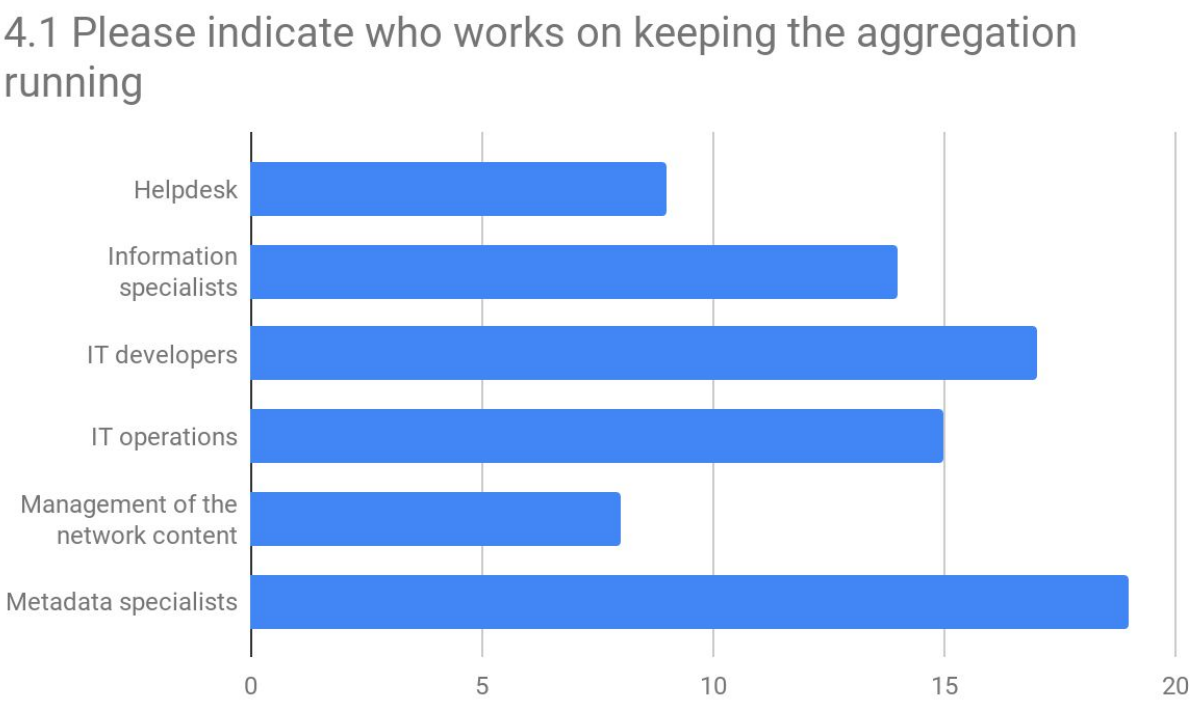


Chart 17 Staff involved in keeping the aggregation activities running.

It is interesting to note that 55% of IT developers are outsourced (see Chart 18). The reason is that NAs are usually organizations based in the cultural heritage domain, where technical staff is not always hired on an in-house basis. However, 45% of NAs hire IT developers and have them in their teams on a daily basis. There are also NAs (e.g. Poland, Cyprus) that originate from the ICT R&D sector and are actively engaged in the aggregation activities, so this approach is more natural for them. Over half of the responses indicate that external companies are hired to handle technical issues. Sometimes this is outsourced only for some of the components (e.g. hardware issues are delegated to the hosting company).

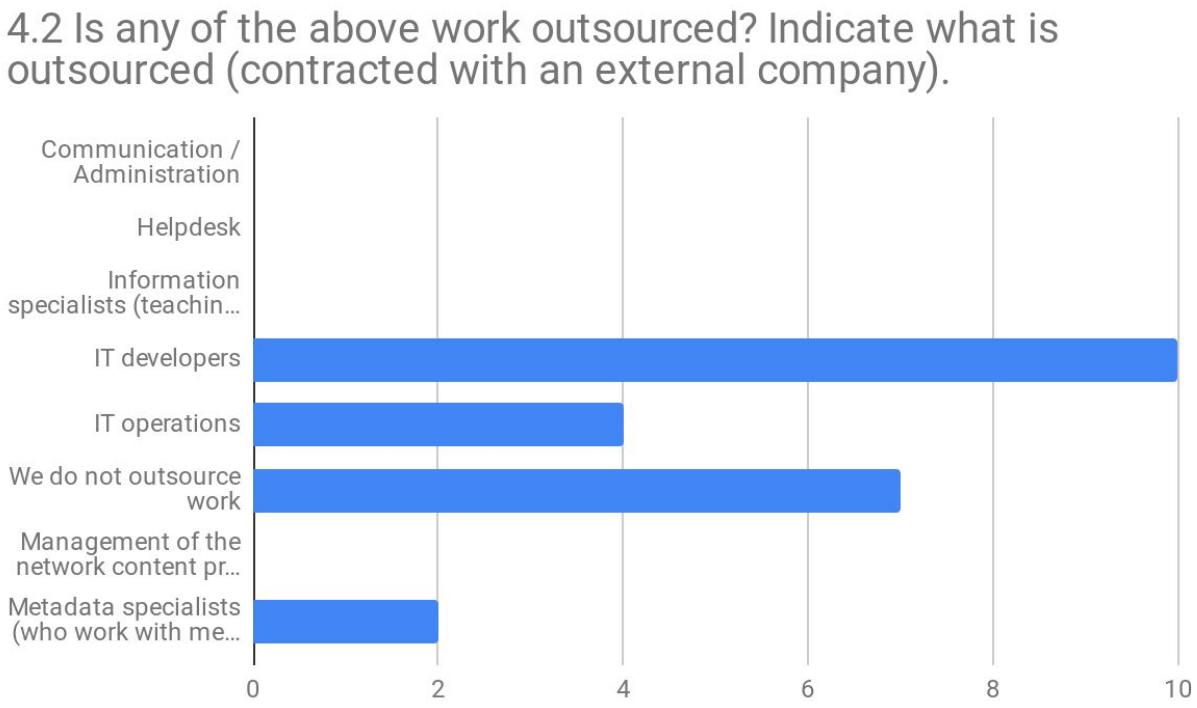


Chart 18 Activities that are outsourced by NAs.

Support given to CHIs

All of the NAs stated that they offer some kind of support for their data providers. In most cases NAs provide support related to quality and modeling of data (including mapping, metadata quality) as well as general technical support (e.g. safe storage, infrastructure). There were also a few responses about Intellectual Property Rights.

NAs can be contacted by CHIs by email or telephone/Skype calls (see Chart 19). Personal meetings or ticketing systems are rarely used as a communication tool. Although ticketing or Customer Relationship Management tools are not widely used in the community, they seem to have a set of features that would be beneficial to build an efficient communication with partners: e.g. history of communication, search options, bulk change. During the second workshop in the Hague, the idea of using specialized software was discussed. Aggregators stressed that ticketing systems could be an option depending on the number of CHIs that need to be served.

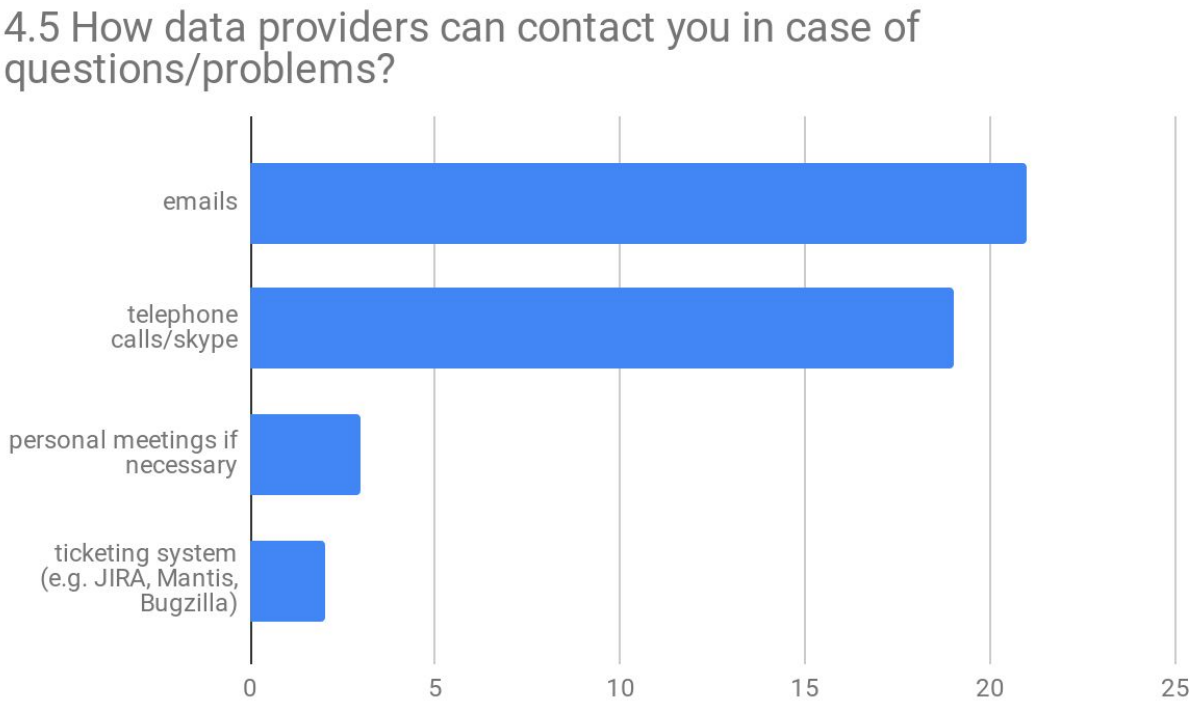


Chart 19 NAs communication channels for solving problems or answering questions from CHIs.

Staff capacity

Chart 20 presents answers related to the question “Do you have sufficient staff capacity?”. Unfortunately, only 10 aggregators claim they have sufficient staff capacity, while others do not.

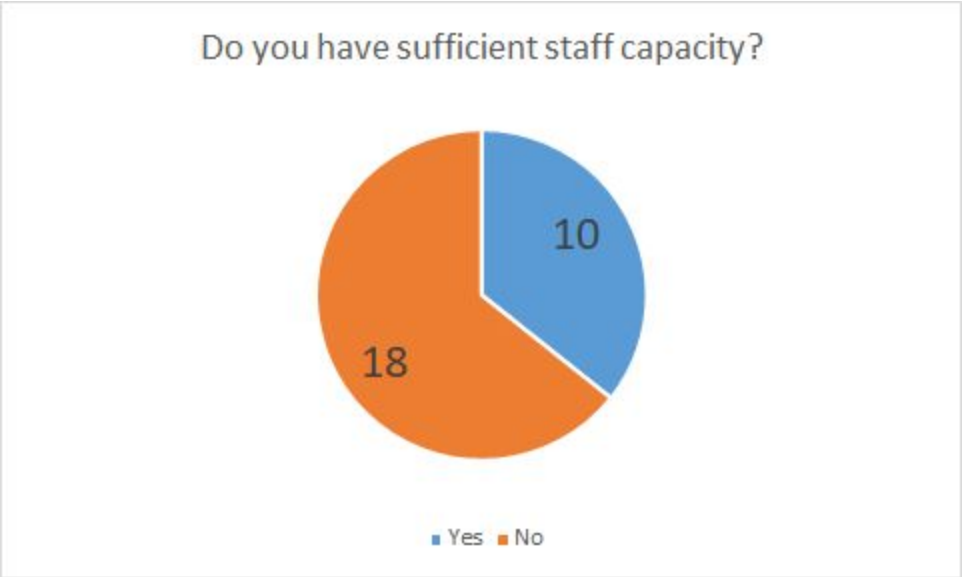


Chart 20 Opinion of all responders (national, domain and thematic) aggregators on their staff capacity.

Competencies of staff

Chart 21 depicts opinions of aggregators on the staff competencies. Most aggregators believe they have proper competencies in their staff. However, taking into consideration answers from Chart 20 the capacity is not sufficient.

Chart 21 Opinion of aggregators on their staff competencies.

Skills lacking in the staff

From the 8 respondents that claim not having required competencies in their staff, the following skills are mentioned:

- Data modelling and management
- User research
- UI and graphic design
- Data curation and metadata transformation skills
- Technical skills
- Digital preservation skills
- Communication skills
- legal skills, IPR knowledge
- XSL transformation
- Openrefine

Data

This section covers questions related to data and metadata. Collected information includes the size of the dataset that NAs are working with and its growth rate, the type of harvested content and its storage method as well as metadata models used by NAs and the level of implementation of LOD principles.

Size of dataset

The number of objects aggregated by NAs ranges from a very small number (dozens) in the case of emerging aggregators, to a large dataset (millions) in the case of well established aggregators. Chart 22 presents statistics in this regard. The NAs have been divided into three more or less equal groups and based on that, three sections have been derived. It is visible that there are three distinct ranges, i.e. 0-100 000 objects, 100 000 - 1 000 000 and above 1 000 000. The number of aggregated objects does not indicate advancement of NAs, their competence or efficiency. It is mainly for statistical purposes.

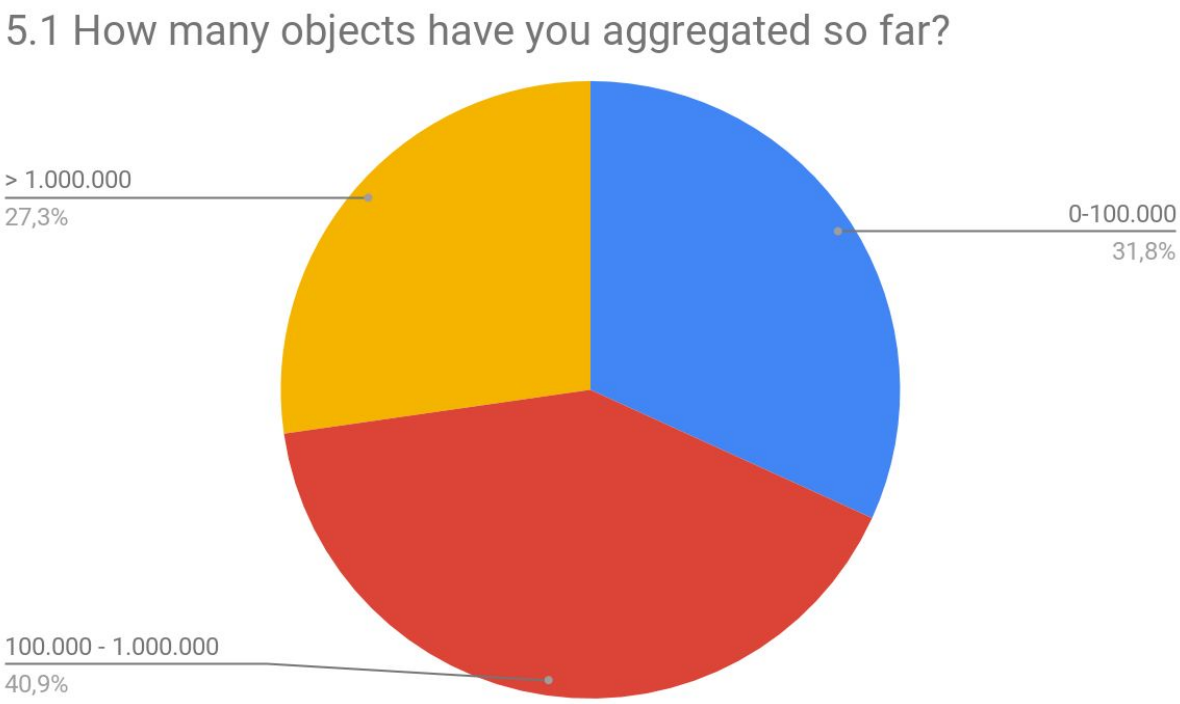


Chart 22 Number of objects aggregated by NAs

Metadata formats/models

Responses to the survey (see Chart 23 and 24) clearly indicate that there are multiple metadata models supported by NAs with the Dublin Core metadata[9] schema being the most popular one. The majority of NAs support more than 1 metadata schema and they perform metadata mappings between the schemas. It is important to note that during the second workshop some of the NAs stated that flexibility in regard to accepted metadata schemas assures better quality, i.e. CHIs provide metadata in the original format, that is best known by them and appropriately used. If the mapping is done by CHIs without thorough investigation and training there is a high risk of misunderstandings and inappropriate results. Therefore the majority of NAs accept various metadata formats and do metadata mapping on their own, knowing exactly what the expected outcomes of the mapping activity are and having direct contact with CHIs. However, there is one NA, which expects that the CHIs provide metadata in only one specific format accepted by them. In order to reach that goal, the activities of the NA are heavily focused on appropriate training of cooperating CHIs.

[9] https://dublincore.org/

In addition, the second workshop delivered a very interesting insight into the landscape of aggregation, i.e. lack of standardisation in collection management tools is one of the main concerns when it comes to high quality and consistent metadata. Collection management tools are often not aligned with NAs expectations related to data delivery methods. Therefore either CHIs or NAs struggle to derive appropriate information from existing data. For instance, it is visible in the context of LOD, where collection management tools usually do not have direct support for ingestion and/or export of this kind of data. However, Europeana is requiring more and more the use of LOD and encouraging NAs to find solutions in order to provide such information.

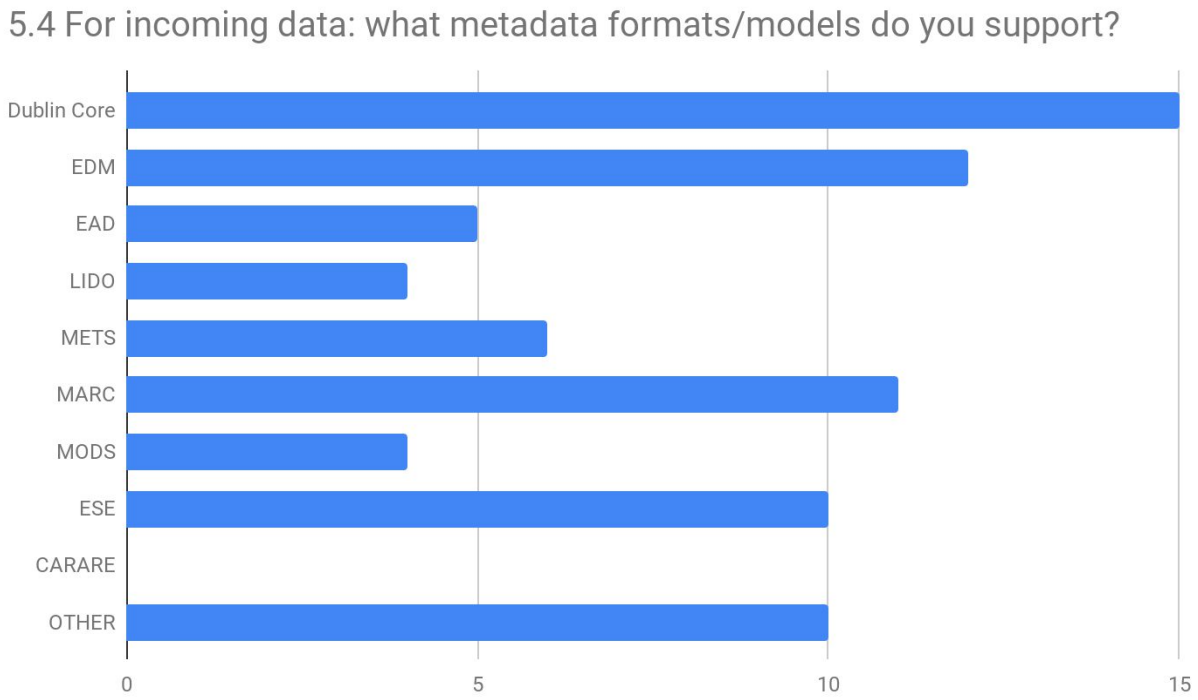


Chart 23 Metadata formats/models accepted by NAs when aggregating records.

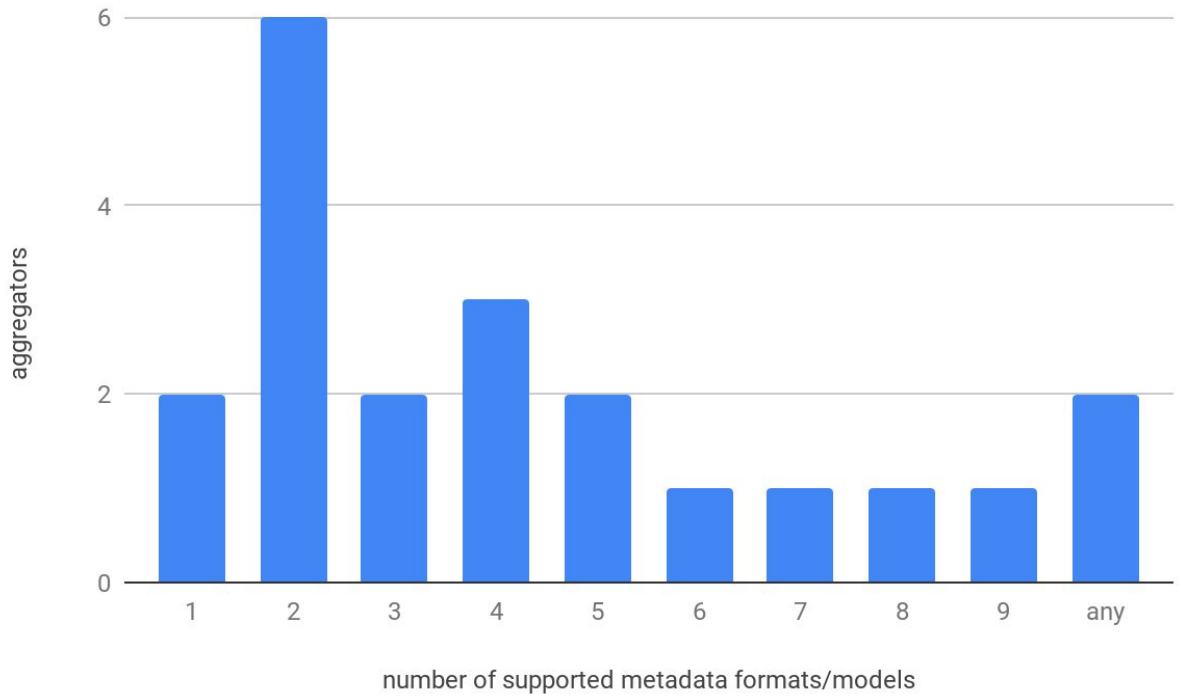


Chart 24 Number of supported metadata formats by NAs.

Linked Open Data support

Almost half of NAs use linked data vocabularies (see Chart 25). Geonames[10] is the most popular service in this regard (7 NAs use this vocabulary - see Chart 26). Other data vocabularies are less popular, as they are used by maximum 3 NAs.

In general, discussions during the workshops indicate that NAs are interested in using LOD to increase data quality. This is especially important to increase metadata tier of their datasets in the context of Europeana Publishing Framework[11]. There are also different approaches to this topic. Some NAs don't use LOD themselves, but their sub aggregators or data providers do so. Other NAs do it themselves either manually or automatically. There are also additional problems in this regard, e.g. vocabulary sets don't have translations (full or at all) in some of the languages or, as already mentioned, a lack of support for LOD in collection management systems.

[10] https://www.geonames.org/
[11] https://pro.europeana.eu/post/publishing-framework

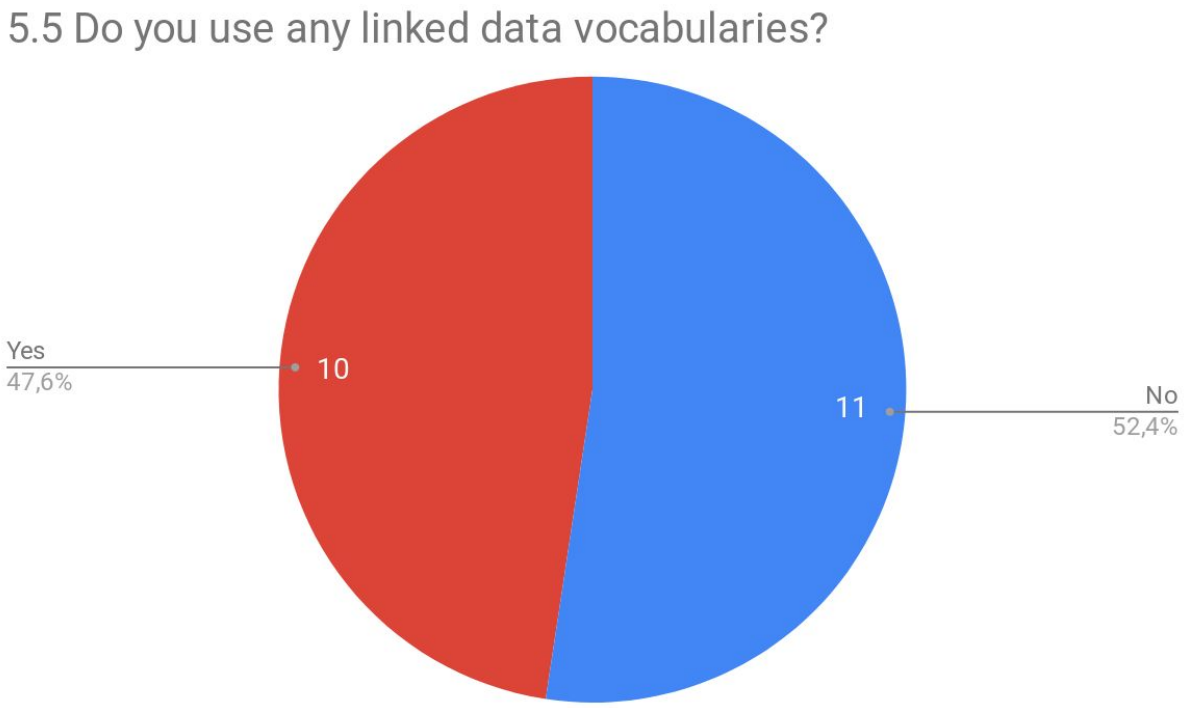


Chart 25 Number of NAs that use linked data vocabularies.

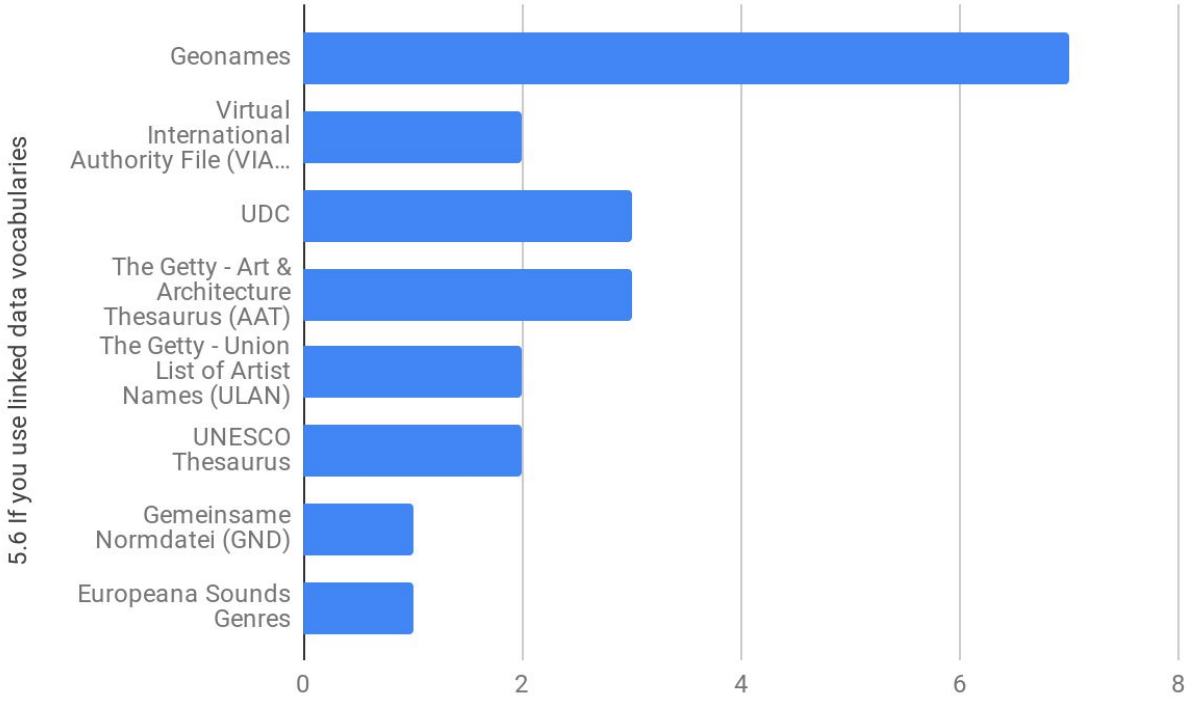


Chart 26 Types of linked data vocabularies that are used by NAs.

Content aggregation

11 out of 21 NAs aggregate content and store digital files in their infrastructure (see Chart 27). The reason for that is usually related to the fact that content providers do not have their own website or technical infrastructure, therefore it is not possible to link to their objects - the objects simply need to be stored in the aggregator's infrastructure.

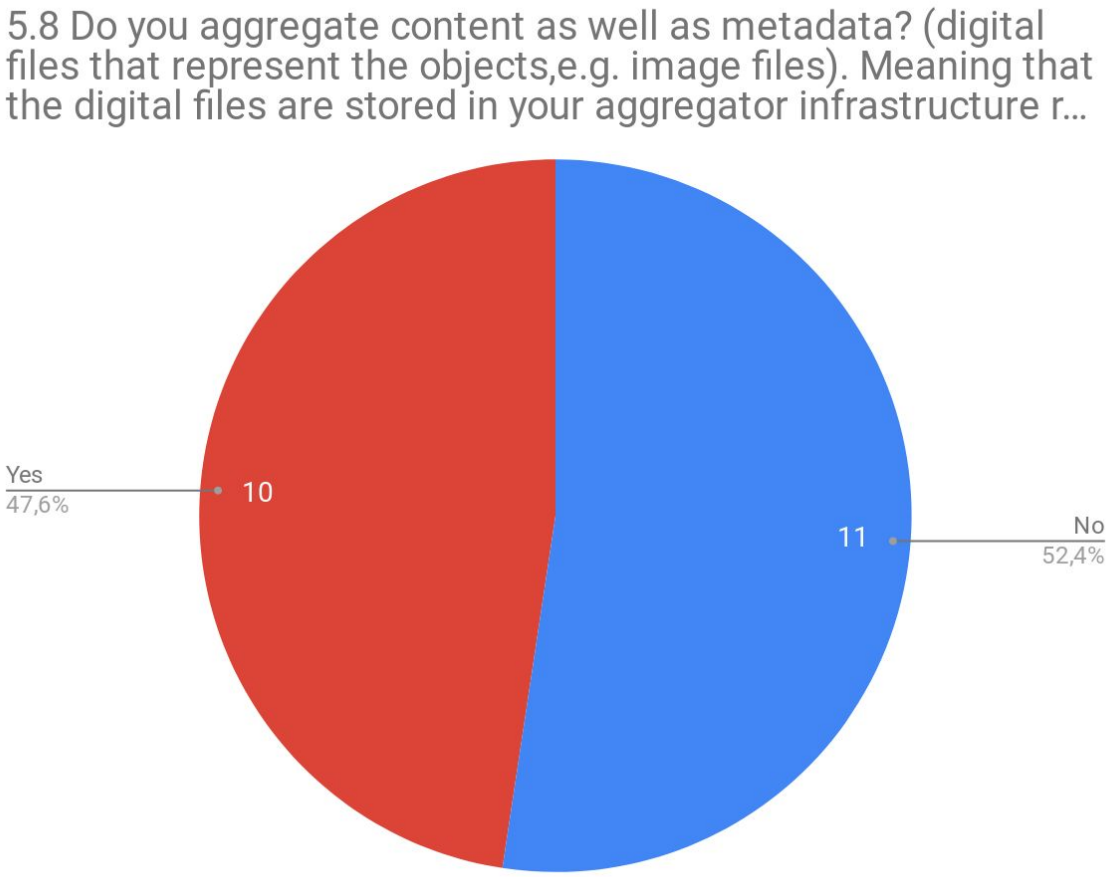


Chart 27 Number of NAs that aggregate content together with metadata.

Technical infrastructure

This section focuses on the technical infrastructure of the aggregator. It investigates the main software modules and components as well as hardware infrastructure.

There are multiple software components used by NAs, the most used ones are mapping and processing. Harvesting, storage systems and external API are the second most indicated components in the technical infrastructure. The most common attempt to

develop these modules is by an in-house and/or open-source approach. Aggregators usually operate the system in their own data centre. There are several NAs that use the tools in a SaaS model.

There are no software vendors which provide ready to use tools that meet all of the needs. This is why aggregators usually have either custom systems (developed in-house or ordered from external companies) or existing solutions tailored to their needs. It is important to note that more and more of this software is created as open-source. Because the aggregation market is currently quite small it is not highly probable that commercial companies will be interested in producing a special product to sell to all the aggregators. It seems therefore that one of the options to create a common technical backend solution would be for all of the NAs to work together.

In conclusion, it is valuable to mention that new aggregators should be aware that before starting their core activities they will have to build complex toolsets. Most probably, at present, they will need to build the system from scratch, by either their own development team or external contractor. An essential prerequisite in this context is to analyse what are the needs of all the parties that will use the technical solution. It needs to cover both the end users who will search and view the data as well as the aggregator's staff that will deliver the data to the system. Aggregators have to collect their requirements, assess the priorities and address them with their technical skills and budget.

Hardware varies very much between aggregators and it is usually adjusted size-wise to how big the aggregator is and how much data needs to be processed. What is interesting is that some of the aggregators use advanced setup with modern technologies like Ceph[12], Hadoop[13] or MongoDb[14] clusters.

Functional modules

This section covers functional modules offered by the NAs for the users. It also focused on the data delivery method (from CHI to aggregator's system).

[12] https://ceph.io/
[13] https://hadoop.apache.org/
[14] https://www.mongodb.com/

Search portal

Aggregators are usually focused on the presentation of digital objects to the end users as 86% of them have a search portal (see Chart 28). The core functionalities of the search portal are basic search, advanced search and filtering of the results (see Chart 29). As can be seen in Chart 30, creating news and articles related to aggregators activities is also quite popular. In addition, 9 aggregators create featured collections where they can promote selected objects to the users. User accounts are a feature provided by a couple of aggregator services. Moreover, comments given during the workshops suggest that this feature is not commonly used by the end users. Less popular functionalities are IIIF streaming for content, own-developed API or dedicated blog or information page.

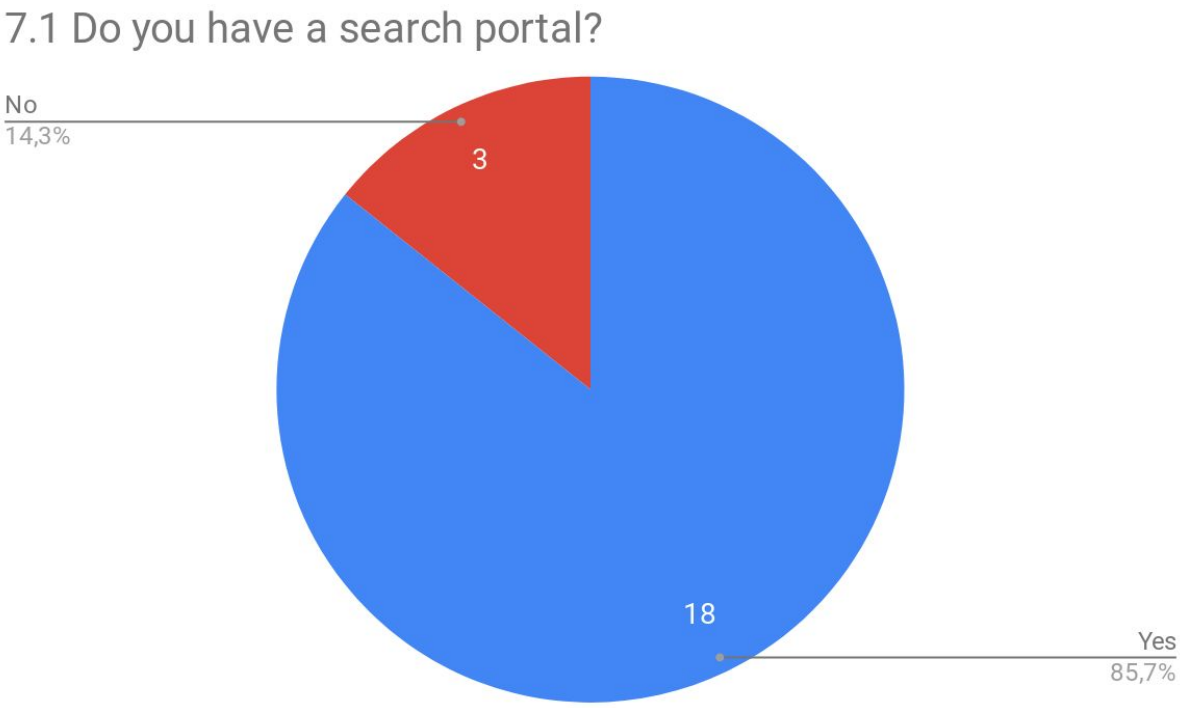


Chart 28 Number of NAs that provide an Internet search portal.

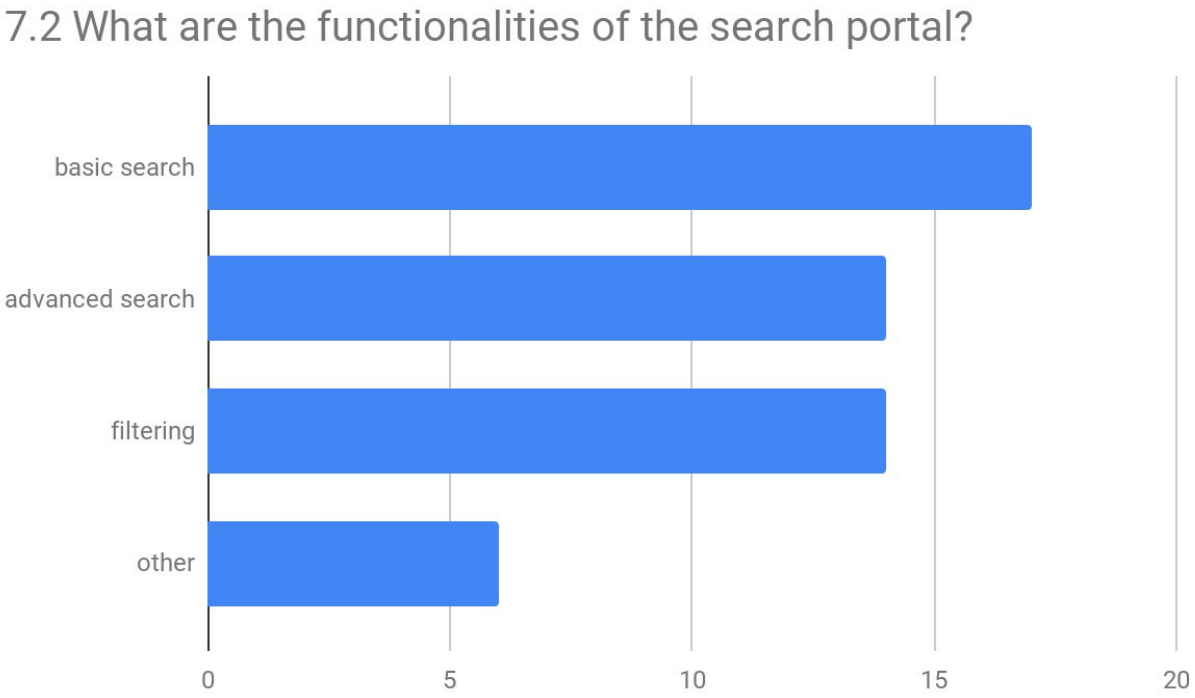


Chart 29 Main functionalities of the existing search portals provided by NAs.

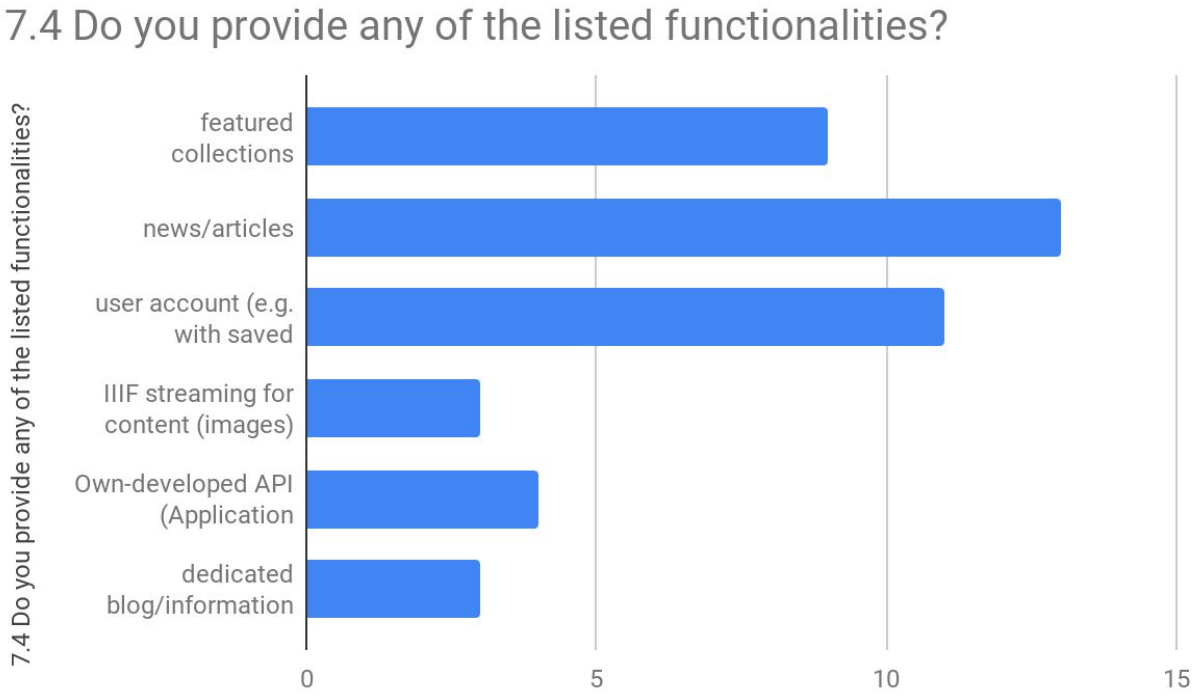


Chart 30 Statistics related to additional functionalities apart from the search portal.

The technical solution that is the most commonly used by aggregators to implement search functions in the portal is using search engines (see Chart 31). The majority of aggregators use Lucene based solutions, with Apache Solr[15] being the most popular. In a couple of cases a mixed approach is used (Search engine + database). Only 3 NAs base their search portal on database only.

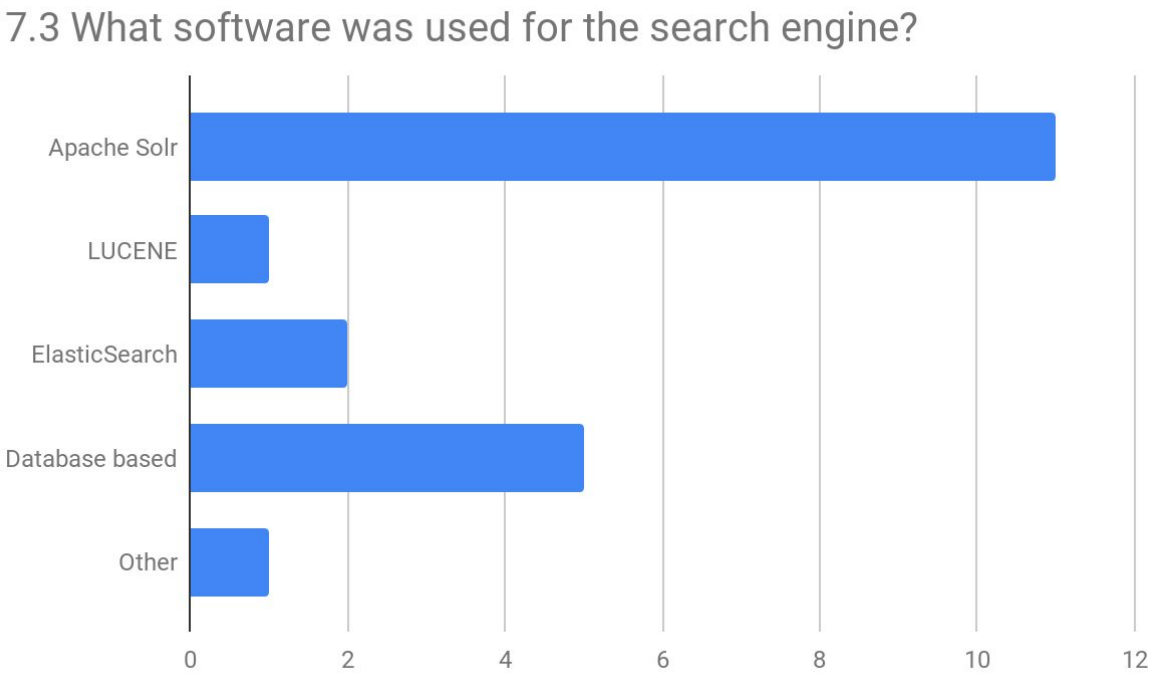


Chart 31 Software used by NAs to provide search features.

Data aggregation techniques

The technical solution used for data aggregation is mostly OAI-PMH, as a dedicated protocol for metadata harvesting and it appears to be one of the most adopted options for NAs to collect data (see Chart 32. This technique is used by the majority of NAs but it is not the only one. Other aggregators have opted for alternatives as they couldn't (fully) implement the protocol. Finally, in practice, technical solutions for data aggregation must be chosen based on the landscape of technical solutions used by data providers in the country. Based on the current practices it is reasonable for NAs to recommend OAI-PMH to their data providers (especially new ones, that are just building their technical solutions) but should be prepared to also support other methods of data delivery (for exceptional

[15] https://lucene.apache.org/solr/

cases, cf. “Enabling better aggregation and discovery of cultural heritage content for Europeana and its partner institutions” master thesis by Julien A. Raemy[16]).

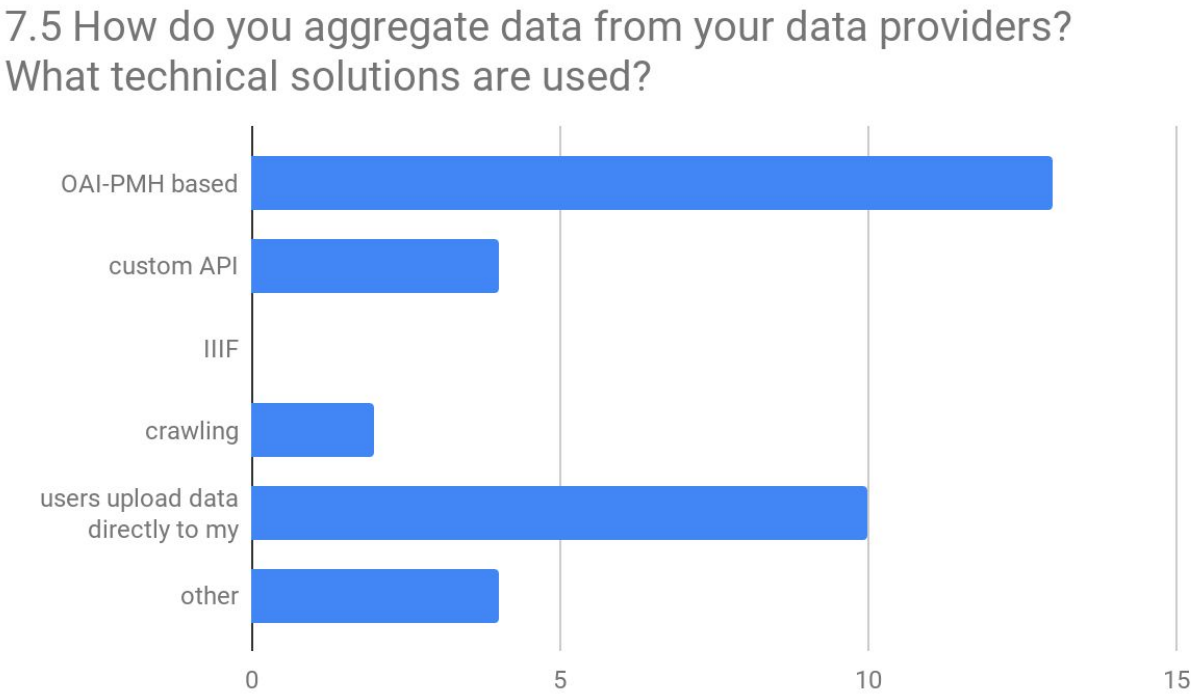


Chart 32 Types of technical solutions used to aggregate data by NAs.

Public API

A public API is exposed by almost half of NAs (see Chart 33). They usually expose an API related to search and view functionalities. Having them in an API helps to integrate with external portals or applications and therefore helps to promote digital heritage objects collected by NAs.

[16] https://julsraemy.github.io/assets/doc/Mastersthesis_europeana_raemyjulien_FV.pdf

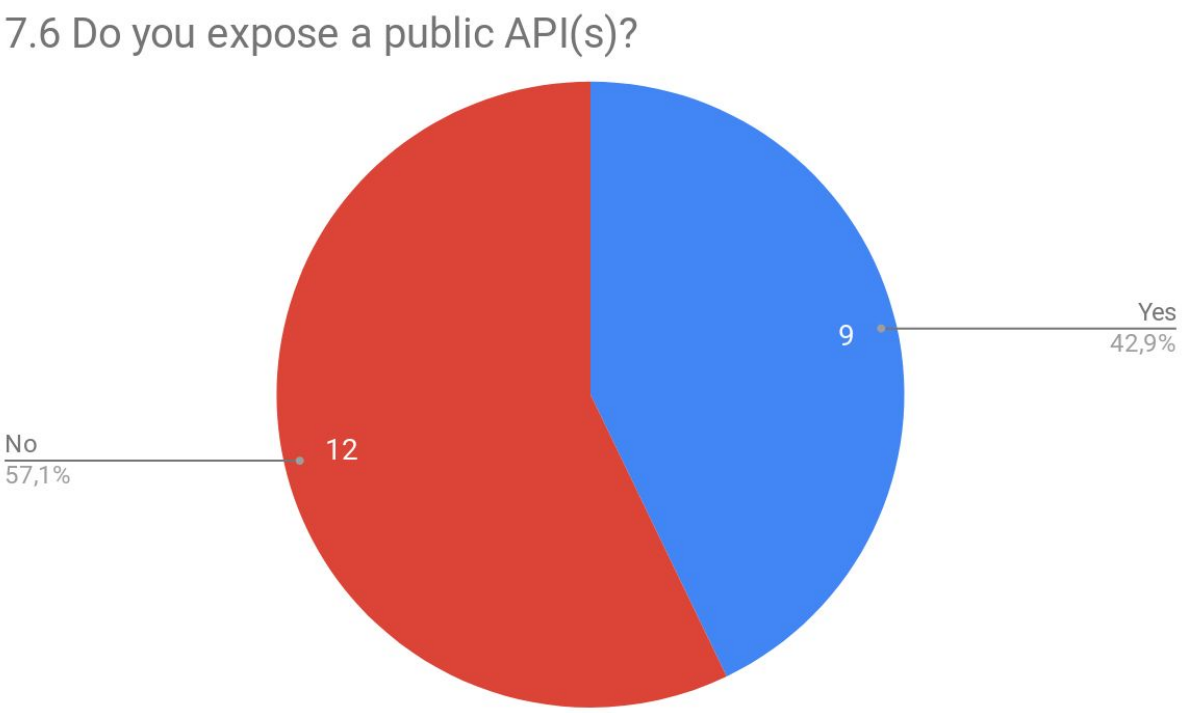


Chart 33 Number of NAs that expose public API.

Harvesting workflow

This section gathered information about the process of data delivery to an aggregation system and further data processing in the system itself. The questions covered investigated storage methods as well as processing approaches such as transformation, enrichment, validation or normalisation.

Data ingestion and storage

Ingestion plans are not a practice used by the majority of aggregators (see Chart 34). Automation or a lack of automation is not an indicator on how well an aggregator is achieving its mission. It's an individual characteristic of every aggregator that depends on the resources, technical capabilities, maturity of technical solutions used by data providers and how aggregators developed their relationships with data providers.

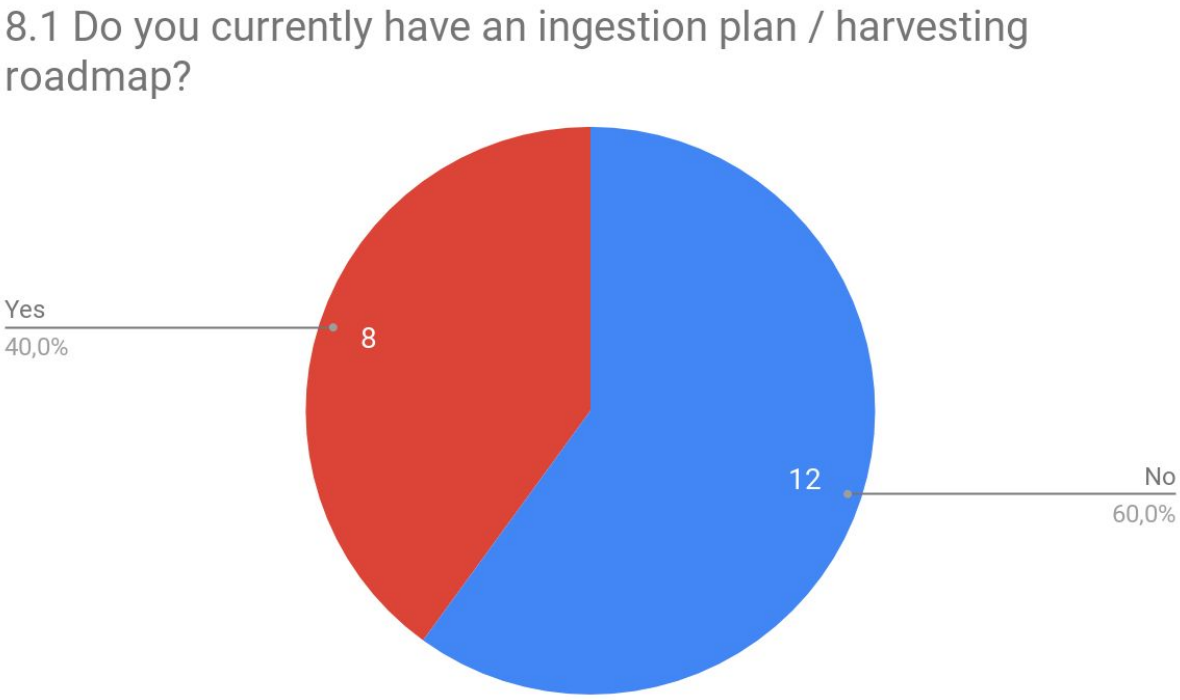


Chart 34 Number of NAs that have an ingestion plan or harvesting roadmap.

NAs often try to automate their processes or some parts of it but it's not the case for everyone (see Chart 35). The answers visible on Chart 35 are also aligned with question 7.5 (Data aggregation techniques in Functional modules), i.e. in question 7.5 10 respondents state that they need to deal with the data manually and it is similar in question 8.2 (see answers on Chart 35) - 10 respondents state that they deal with harvesting in a manual way.

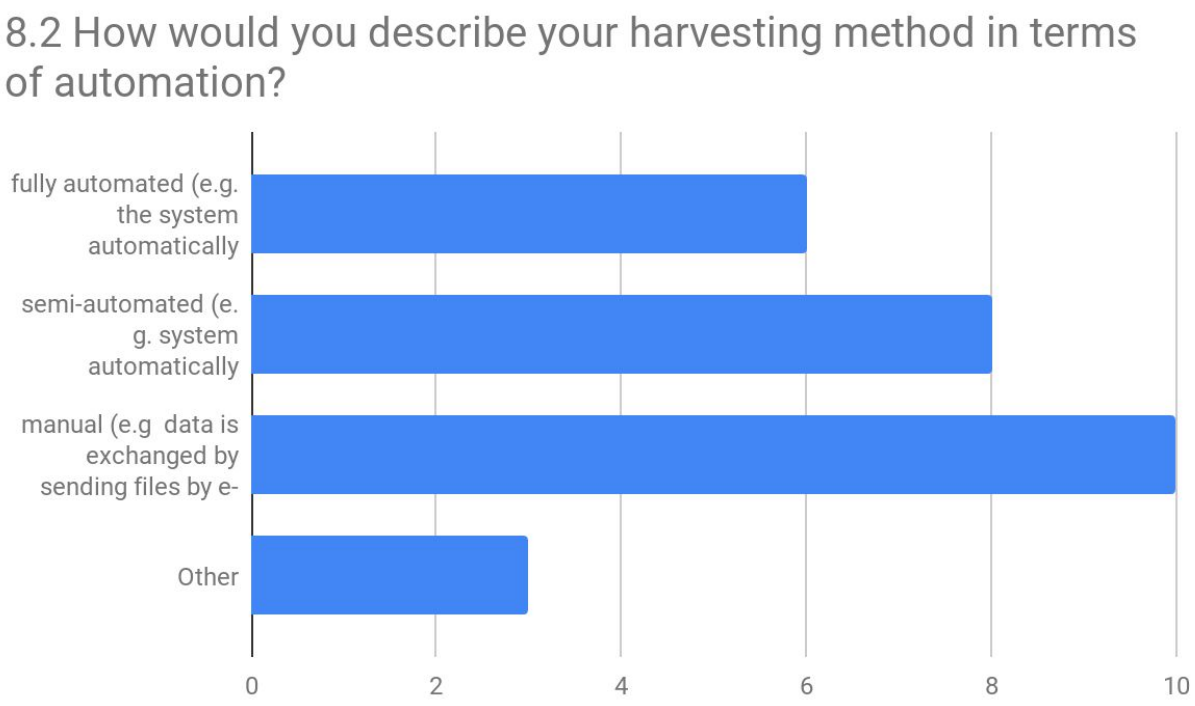


Chart 35 Harvesting methods in terms of automation.

Question 8.3 investigated data storage and delivery methods. It was an open question and therefore the responses had different levels of details, e.g. in some cases no storage details were given. However, based on the answers it was possible to conclude that there are 3 main ways of delivering data to an aggregator system:

- via protocol (mostly OAI-PMH),
- direct upload by data providers,
- direct upload by aggregators (because data was sent via email or file transfer).

The data is usually stored as xml files in a database (relational or not) or directly in the search engine.

Data processing

NAs process their data using: metadata enrichment, normalisation and validation. Each option is done by almost half of NAs and they usually execute at least two of the options. The discussions among NAs indicate that they should consider introducing quality rules for their data partners. Setting a quality standard and validating data against it is a way of ensuring good data quality. Validation might check presence of mandatory fields or

correctness of data (xml) format and structure and what is important - it can be easily automated, e.g. as a step run just after data harvesting. Quality can be further improved by applying normalization. NAs usually clean aggregated metadata, e.g. normalize URL or use a strict list of values in certain fields. This helps to make harvested datasets unified and creates a better user experience for end users of the search portal. Enrichment seems to be the most sophisticated processing step performed on the data level. Based on the information that is given in the metadata fields, more comprehensive information is added to the document either manually or (semi-)automatically, e.g. links to LOD vocabularies are added. A manual process is more precise but challenging and time-consuming when dealing with big data sets. An automatic process on the other hand is faster, but can be more error prone. Although this step is a challenging one, it should be considered by the NAs, especially if they want their data to be compliant with higher tiers as defined by the Europeana Publishing Framework.

Tools used to normalise, validate or enrich data are usually self-developed as their functionalities strongly depend on the data model used in the aggregation system. There are also other steps run by different aggregators like review/acceptance by data providers in a preview portal, creation of Digital Object Identifiers[17] or creation of backups.

Cooperation with Europeana

The section focused on whether and how NAs are cooperating with Europeana. What steps are undertaken to deliver data to Europeana and how often it is done. Also if the NAs are active in promoting Europeana.

Mission

All the respondents claim that Europeana is a part of their mission in one or more of the following ways:

- promotion and dissemination of national resources
- network building and exchange of experiences
- help with content quality (standards, frameworks, etc.)
- NA was designated by authorities to cooperate with Europeana
- European integration

[17] https://www.doi.org/

Delivery of data to Europeana

The majority of respondents provide their data to Europeana (see Chart 36). Only two (emerging) NAs were not providing data to Europeana at the time of completing the survey. In addition, one NA did not respond to the question.

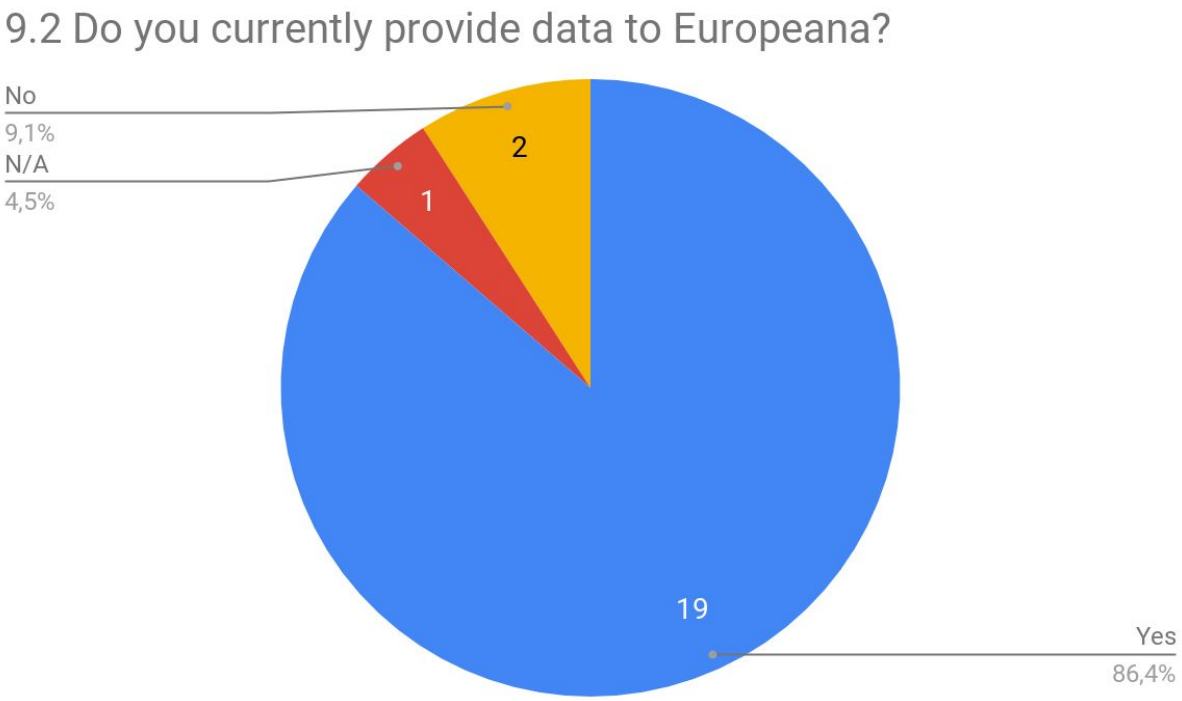


Chart 36 Number of NAs that provide data to Euroepana.

The next question is related to accreditation status of NAs. In 2019 Europeana Aggregators Forum set up an accreditation scheme to identify active aggregators and acknowledge their role as active and trusted partners for cultural heritage institutions to work with, but also for other aggregators to share knowledge and best practices with[18]. Accredited aggregators are featured on a dedicated page on Europeana Pro[19]. Responses to the question about accreditation status (see Chart 37) indicate that 15 NAs are already Europeana accredited aggregators, 3 of them are not accredited aggregators. 4 did not respond to that question.

[18] https://pro.europeana.eu/post/breathing-new-life-into-the-europeana-aggregators-forum
[19] https://pro.europeana.eu/page/aggregators

Although the accreditation process has just started, it seems that NAs are eager to get the certification and be part of the professional data delivery process to Europeana[20].

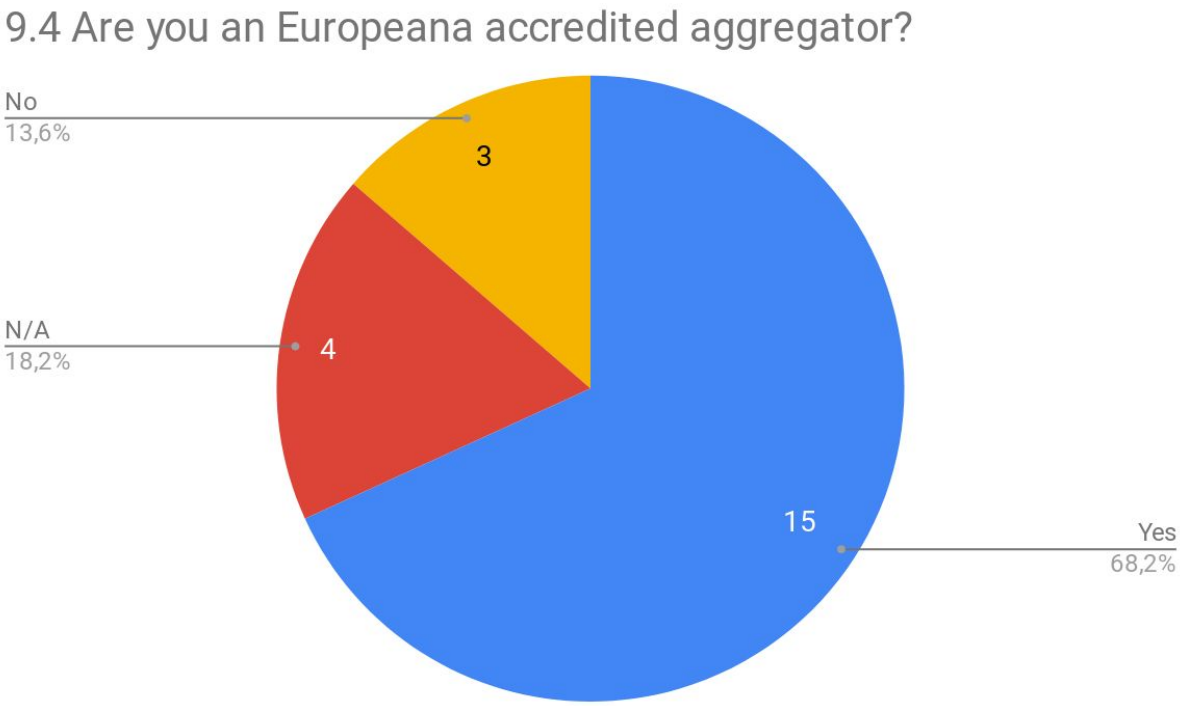


Chart 37 Number of NAs that are Europeana accredited aggregators.

The next questions in the survey investigated how often aggregators send their data to Europeana and how big are the updates. Based on the answers it is possible to state that data is delivered on a regular basis, but the time frame between updates varies from 1 month to even a year. It is an individual decision of each NA. For instance, the responses show that 8 aggregators last delivered data to Europeana in early 2019 and 6 of them before 2019. Other NAs did not state their last ingestion date. It is therefore clear that the updates are regular, but not very frequent.

Based on answers to question 9.11 (see Chart 38) it is visible that the majority of aggregators (13) do not send all of their records to Europeana, 4 send all of their records and 3 didn't give an answer.

[20] At the time of writing the report, there are already 37 accredited aggregators.

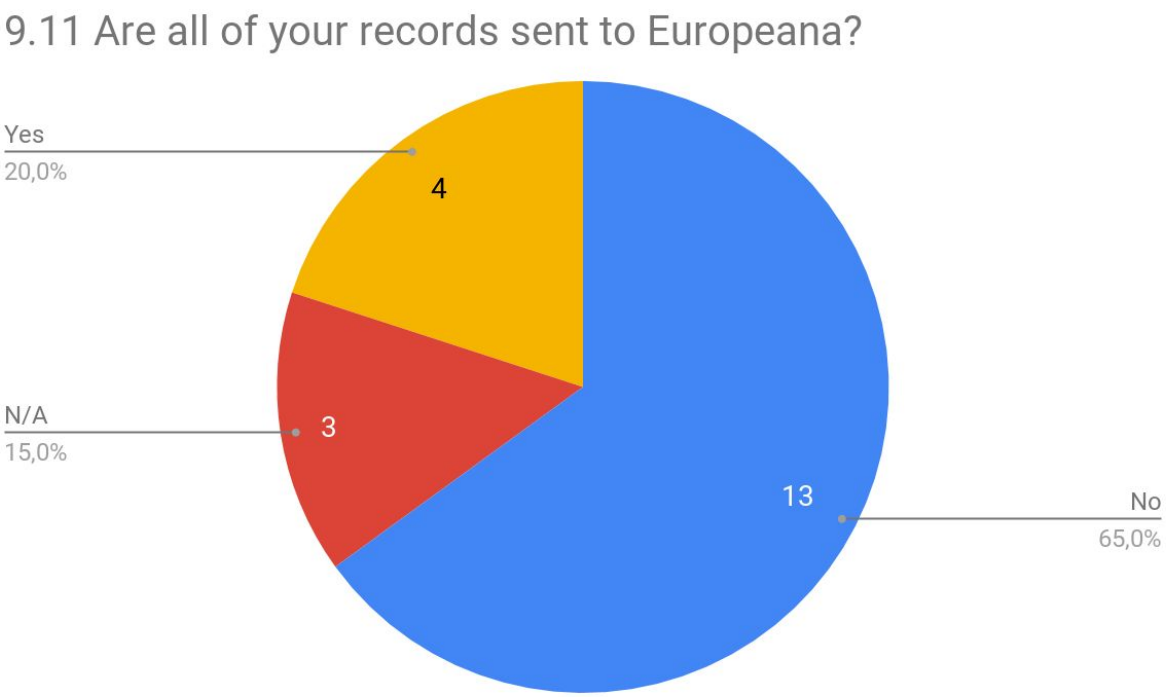


Chart 38 Number of NAs that send all their records to Europeana.

The majority (13) of NAs use the OAI-PMH protocol to deliver data to Europeana (see Chart 39). 1 is using a mix of OAI-PMH and other methods, 3 send data in ZIP packages, 3 did not answer this question.

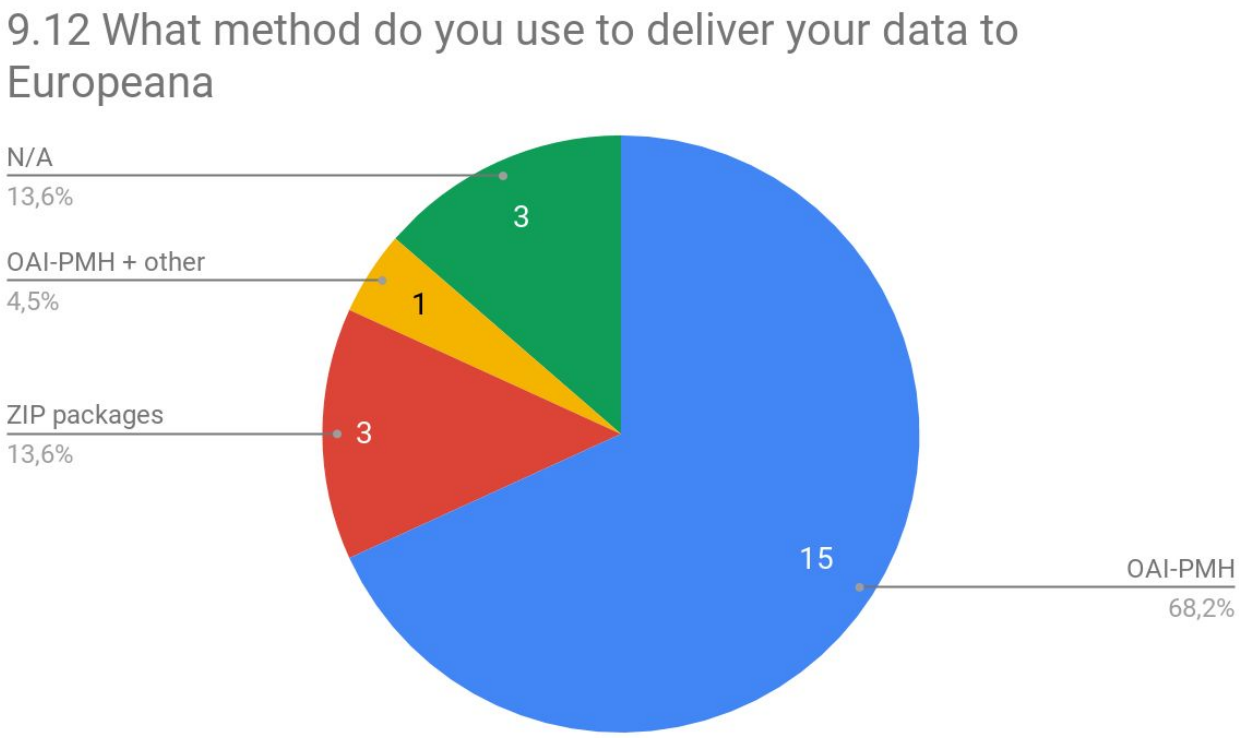


Chart 39 Data delivery method used by NAs.

Presentation of data in Europeana Collections

The Europeana Collections portal presents content files coming either from NAs or from data providers. Respondents indicated (see Chart 40) that in most cases content files are coming from data providers services (9 respondents). 6 NAs provide content files from their infrastructure or service and 3 NAs use a mixed approach. 4 NAs did not respond to the question.

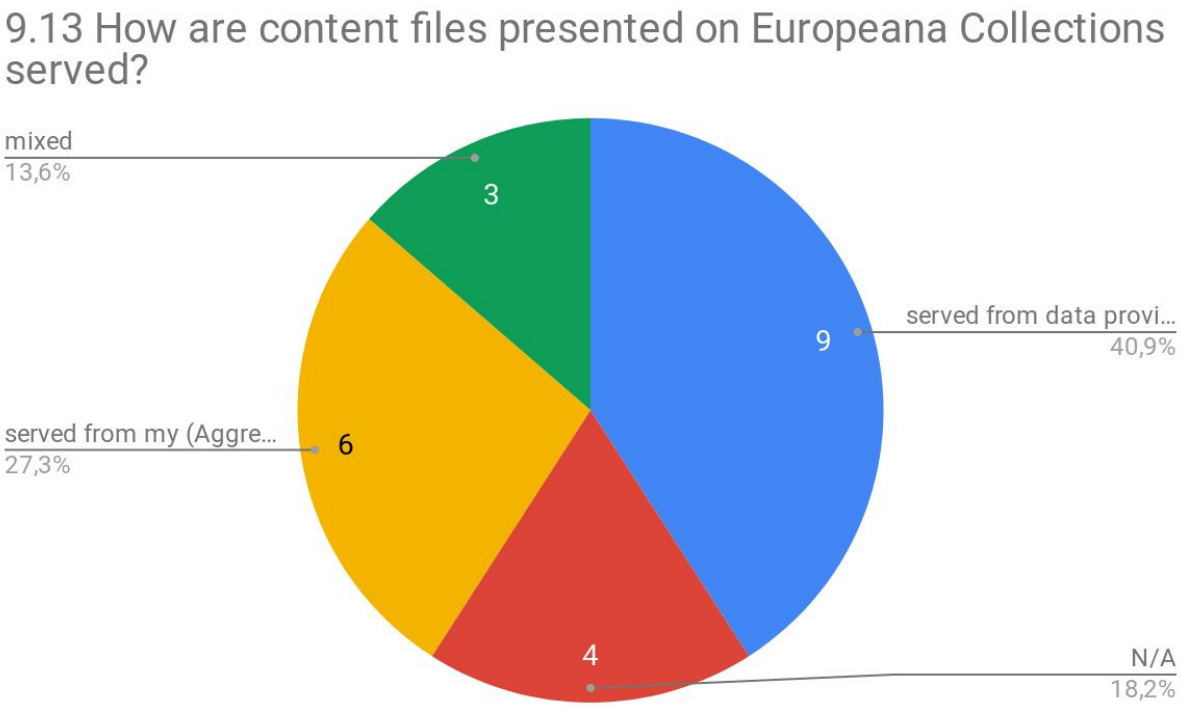


Chart 40 Method to serve content files that are presented in Europeana Collections portal.

The next question investigated the method of transformation of NA records to the EDM. The most popular answer by respondents was that the process is automated by XSLT. The other answers included “in-house developed tool” or “custom code.”. Some of the aggregators, despite having automatic processes, have to in some cases perform manual conversions: “In some cases we transform manually from ESE and MARC21”.

Question 9.15 “If you aren’t cooperating with Europeana now, have you done it in the past?” was a badly-worded question because there were multiple “yes” answers from NAs who are actively cooperating with Europeana at the moment (based on answers from other questions). Therefore analysis of this question was not performed.

What are the main weaknesses of the current aggregator model (Europeana partners with aggregators which aggregate from cultural heritage institutions)?

This was a free form question (9.17), and we got a variety of answers. Main points mentioned included:

- The main weakness of the model is that the same data could be sent via domain aggregators or present/past projects (Athena, Partage) and national aggregators, creating duplicate records in Europeana.
- If there is a strong national aggregator in the country, which ingests the content from all the regional digitization centres, this national aggregator should be appointed as the main Europeana aggregator and all the content from the partners should be redirected to the Europeana via the national aggregator.
- Whole model and underlying infrastructure is still based on classic aggregation (with copying and losing of data), still partly manual, and takes still a long time to get something published. Not scalable for the future.
- Incomplete: not all potential data providers can find a route to Europeana.
- The process is quite labor-intensive and not fully automated.
- Lack of clear strategy at national level: National Agregators under-resourced

Overall experience/lessons learned from setting up your aggregation services

In the last section NAs were asked to provide information about their lessons learned from their process of becoming an aggregator. Answers from this section are important because they show the overall experience of NAs in the context of their previous and current activities. In fact, this kind of information is a great input for emerging NAs and guidelines on how to approach the process of setting up a NA.

Among the good experiences that NAs had, the two most important factors are related to:

- recognition and support of the respective governmental bodies (Ministry of Culture or alike)
- good collaboration with data providers and CHI that are responsible for delivering data.

Appropriate organisation of the work within the NAs was also indicated as one of the success factors. Good technical staff and help from Europeana was also appreciated.

When it comes to bad experiences the NAs indicated the following factors:

- Lack of official support/recognition and sustainable funding of respective governmental bodies (e.g. Ministry of Culture).

- Cooperation with stakeholders did not go well and therefore NAs were struggling with their activities. More information on guidelines, tools, metadata mapping, cooperation with data providers and Europeana is needed.

NAs were also asked what they would have done differently (what would they improve) if they had a second chance to set up NA from scratch. The most common answers included:

- Work on toolset that is used in the aggregation activities (automate, introduce validation, work on quality from the beginning)
- Better focus on cooperation with data providers and metadata mapping that is crucial in this context.

Sustainability

This section investigated how the aggregator activities are funded and how it affects their daily activities. It investigated how aggregators planned to sustain their efforts, whether they planned future collaboration with Europeana, the future trends they identified as having a potential impact on their activities as well as perceived threats that could jeopardize future activities.

11.1 Is there a dedicated partner agreement in place?

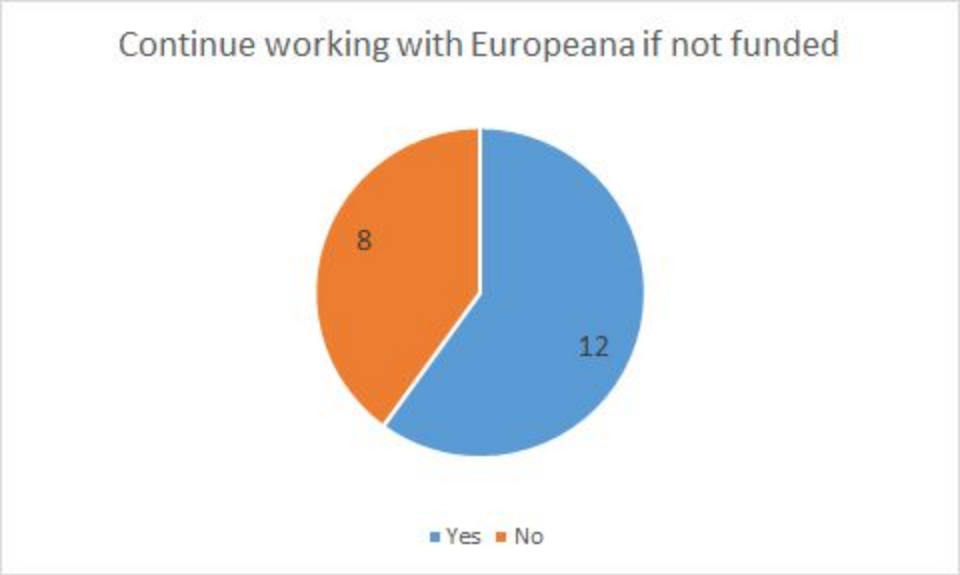
11.2 How are your aggregation services funded?

This was an open question, and not all respondents (and not all NAs) responded to it. The majority indicated that it is funded by the government, however it is not always indicated whether it is taken from the annual budget of the institution or directly provided through a specific budget for the task.

11.3 Until when is funding provided?

Here we got only 18 answers, the majority indicating that there is no time limit, and that the budget is sometimes reviewed through annual budget or as part of multi-year plans.

11.4 Will you continue working with Europeana after your current funding ends/gets limited?

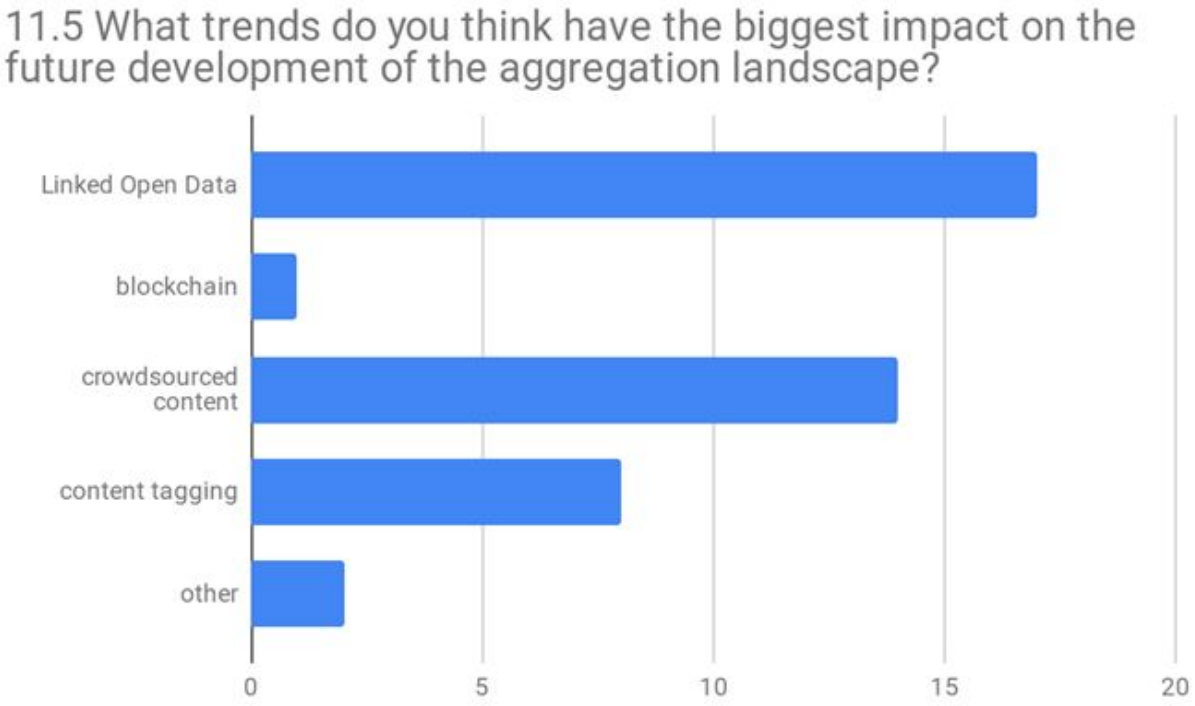


This highlights that aggregation can only take place if funding is secured.

11.4a How will you fund these cooperative activities?

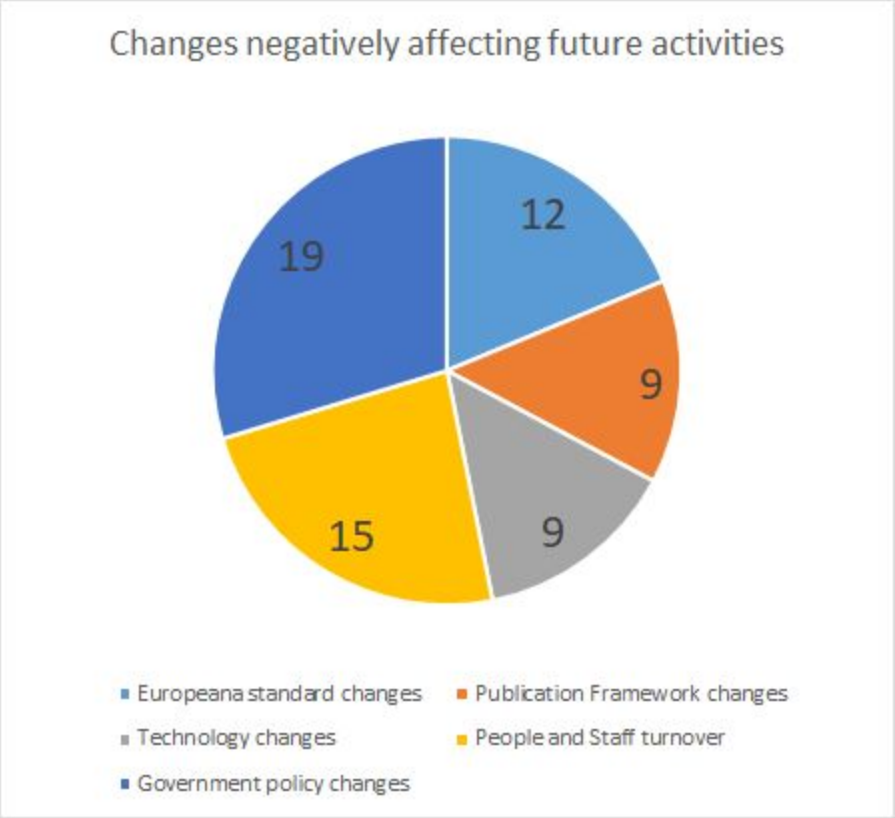
It is clear from the type of funding indicated that this will mostly be through the institutional funding by the government.

11.5 What trends do you think have the biggest impact on the future development of the aggregation landscape?



No surprises here: Linked open Data and Crowdsourcing top the list. The work in Activity 6 of the Europeana Common Culture Project addresses crowdsourcing.

11.6 Which type of changes (might) negatively affect your activities ?



The biggest concern is changes in government policy, followed by people and staff turnover.

Summary of the survey results

The conducted survey investigated various aspects of data aggregation, including capacity, services, data, technical infrastructure as well as cooperation with external entities, partners/relationships and sustainability. This section summarizes the most important conclusions derived from responses and defines what a NA is, its key role, duties and relevant activities.

NAs have a **common mission to give access to and promote cultural heritage**. Respondents indicated that providing good quality (meta)data is a priority in these activities. All NAs except one are cross-domain, which clearly indicates that the activities related to aggregation are broad and target various types of resources. A good practice identified within NA is to have a legal mandate from national authorities. It helps in smooth set up of an NA and sustainable operations.

NAs provide various services to CHIs, but the most important one is related to **metadata mapping**. For most of the NAs this is the core activity that brings consistency and a common view on aggregated resources. Although metadata mapping is core activity of the NAs there are still difficulties in this context (30% of NAs indicate this as a gap in their offering). There are different approaches to metadata aggregation. New aggregators might choose to support only one format or to support multiple formats and perform metadata mappings. Decisions should be based on a deep analysis of partners that will provide the data, their competencies in providing good quality metadata and variety of software solutions that exist in the national market. However, the overwhelming majority of NAs support multiple formats and focus on metadata mapping. If an aggregator decides to support only one (specific) schema it should be thoroughly investigated whether it can efficiently be imposed on collaborating data providers. An important conclusion from the survey and related discussions is that **software vendors should be advised to introduce standardisation** in their products, especially in the context of aggregation needs. It mostly relates to the systems that deliver digital resources in an online manner. These systems usually have an import module from the library catalogue systems, which are not perfectly aligned with demands coming from the online world, especially in the context of Linked Open Data. Improvements in this context could directly bring benefits to the aggregation activities on a national and European level.

Each NA implements relevant aggregation workflow(s). These workflows are complicated processes, composed of many steps. The majority of NAs run transformation on their data. This is connected with the fact that many of them support multiple input formats. Data is usually transformed to an internal aggregator schema so all the records are unified and standardised. Transformation is usually performed using custom software, often based on XSLT technology. The NAs try to automate the process, however almost half of them admit that they execute the harvesting workflow manually, and only 6 of them do it automatically (others indicate a semi-automated approach). NAs process their data using metadata enrichment, normalisation and validation. Each option is done by almost half of NAs and they usually execute at least two of the options. The discussions among NAs indicate that they should consider introducing quality rules for their data partners. **Setting a quality standard and validating aggregated data against it is a way of ensuring good data quality**.

NAs involve relevant staff that is composed of two main groups: **metadata/information specialists and IT specialists**. These two groups build NAs and deliver most of its capacity and operational quality. While metadata and information specialists are usually employed on a permanent basis, IT specialists are mostly outsourced (e.g. via external companies).

NAs provide support and consultancy in the context of digitisation related issues and aggregation workflow pipeline. In case CHIs want to communicate with NAs they usually take advantage of emails or telephone/online meeting tools. Discussions during workshops with NAs indicate that tools such as ticketing systems could be employed to make the communication more efficient and trustworthy. A special focus in the survey was given to IPR-related support and consultancy. From the responses it is visible that almost half of respondents provide documentation and support in this context. Still, the majority of the NAs do not tackle this issue, forcing CHIs to investigate and decide on their individual approach. This is clearly visible in the gap analysis in the services provided by NAs, i.e. the top indicated gap is IPR training. From the responses it is also visible that audiovisual and 3D content is not entirely supported.

Part of the core activities of all the NAs is providing data to Europeana. NAs are eager to deliver data in appropriate format and to be involved in ongoing activities, e.g. the accreditation process. Although some of the NAs were created especially to provide data to Europeana (e.g. during the Europeana Local project) they now state that is not a primary reason for their existence. Only one aggregator claims that delivering data to Europeana is the main rationale for its existence. Europeana is for sure an excellent way to increase the visibility of the national cultural heritage. Europeana and Europeana Network is also a source of good practices and quality standards. Becoming a Europeana Accredited Partner is a way of acknowledging the role of aggregators as active and trusted partners for cultural heritage institutions to work with on the national level. It also might be helpful when applying for various national and European projects. The majority of aggregators promote Europeana by: various publications, workshops (for both end users and Cultural Heritage Institutions), presentations at conferences, social media, links and logos on their websites. By promoting Europeana they promote European Cultural Heritage and join efforts of the members of the Europeana Network Association[21]. Aggregators deliver data on a regular basis but the time frame is different for different NAs. In principle, frequent and regular

[21] https://pro.europeana.eu/network-association/sign-up

deliveries assure that the Europeana Collections portal presents the most actual versions of objects and it should be the objective for all the NAs. However, arranging ingestion schedules of Europeana and all NAs is challenging, therefore special focus should be given to this aspect. Aggregators don't have to send all their data to Europeana. Usually some of the data is skipped mainly because of not sufficient or not available copyright statements. Other reasons include: data do not meet the minimum requirements of Europeana Publishing Framework, data of some providers is delivered using different channels (e.g. thematic provider), data providers haven't signed Data Exchange Agreement[22]. Content of the digital objects presented on Europeana Collections can be served either from data providers services/websites or NA services.

Software and hardware infrastructure varies among NAs. In the context of aggregation software, there are individual technical solutions, often with no common components shared between NAs. It is very interesting when confronting this with the fact that NAs share a common mission, provide similar services and have similar staff. Although, from this perspective, it would be natural to share common technical solutions, especially in the context of metadata aggregation for Europeana, it is not not visible in the responses. One of the reasons could be the fact that NAs emerged more or less during the same period and none of them considered shared solutions as an option, taking into consideration that forming aggregation infrastructure on a national level is already a huge challenge.

NAs have a lot in common when it comes to functional modules. **A search portal seems to be the most practical and commonly used way of presenting cultural heritage objects** to end users and can be implemented by NAs if they want to promote their resources through a national-level portal. The most common technology used to implement a search feature in the portal is a search engine, such as Apache Solr[23]. However, there are already existing and successful aggregators that don't have search portals and their data is available only in Europeana. Public API is an additional way of promotion and dissemination of aggregated data for almost half of the NAs. The API is usually related to the search and view features of the search portal. In order to aggregate data, NAs use the OAI-PMH protocol in most cases. This is a clear message for the data providers as well as NAs to promote and support this way of interaction, as it is the most popular method for

[22] https://pro.europeana.eu/page/the-data-exchange-agreement
[23] https://lucene.apache.org/solr/

aggregation, which seems to be the de-facto standard. However, there are special cases, where aggregation is done in a different way.

More than a half of NAs store data (digital content) in their infrastructure. There can be various reasons to do so, e.g. lack of appropriate infrastructure at the CHI level or assuring reliable access to digital objects, without relying on data providers systems (e.g. addressing the dead links problem). However, such an approach requires NAs to have appropriate technical solutions and storage, which can be especially challenging for large datasets. This way of serving content might also be seen as a problem by data providers themselves, if they observe a decreased number of visitors on their own websites.

NAs are in general satisfied with what they have accomplished throughout the years of their existence. Based on the responses it is visible that one additional statement should be included in NAs overall mission, i.e. **building good relationships with cultural heritage institutions, national authorities and Europeana**. NAs have taken care of these relationships, e.g. with CHIs by providing support and expertise (training, feedback, promotion), with Europeana by active participation in the Europeana Network and delivery of good quality data, with the authorities by seeking their official support and recognition. NAs are aware of their problems, such as: lack of proper legal status, sustainable funding, cooperation with CHIs and their experts, cooperation with Europeana, quantity over quality. As insufficient funding might be a reason for NAs to stop their work, they should actively seek financial support, e.g. with their national or local authorities.

Aggregators have tried to fulfill their mission. At the same time they built a unique set of competences and a huge knowledge base. Naturally they became experts in the digital heritage field and a perfect source of information and expertise for their data partners and other NAs. New aggregators should use the lessons learned for guidance and thoroughly investigate how the aggregation workflow should look like. The design process should be preceded by a thorough analysis, prioritising quality of data over quantity, communication with partners and clear definition of responsibilities.

Establishment of emerging national aggregators

In the framework of Europeana Common Culture project four emerging National Aggregators have been established in Estonia, Ireland, Latvia and Serbia. In order to support their actions a consultancy service has been provided and proper support has

been offered. This section details the process of setting up these four NAs together with a description of the consultancy service as well as lessons learned from the experiences.

Consultancy service

In order to ensure proper development of emerging aggregators a consultancy service has been created with the intention to answer pressing or difficult questions that could appear in the process of setting up an aggregator. In practice the consultancy service was provided via a JIRA ServiceDesk feature and in the initial months of the project the consultancy was also provided via e-mail or video conference channels. Additional support was also provided during the workshops and meetings organised as part of the project or EAF.

Establishment of aggregators

As mentioned earlier, four emerging NAs have been established within the scope of the Europeana Common Culture project. The following subsections detail the approach, operations, future plans and lessons learned for each new NA developed in the project.

Estonia

Introduction

The aim of the e-repository portal is to bring together Estonian digital cultural and scientific information, which means that it is possible to search the e-repository portal for archive records, books, pictures, sculptures, sounds, videos and digital scientific information. The e-repository portal creates a single gateway that connects different areas, both between research institutions and memory institutions.

Researchers, historians, journalists, students, lecturers, ordinary users, whether they are doing scientific research or just doing research and just want to search for structured information out of personal curiosity then the e-repository portal is created for all those users.

Estonian E-varamu portal was created within the project “Estonian e-repository and conservation of collections (first stage)”, which ended in December 2015, as an object of the Estonian research infrastructure roadmap. On behalf of Estonia, the University of Tartu, Tallinn University, Consortium of Estonian Libraries Network, Estonian Literary Museum,

Estonian National Library and National Archives participated in the creation of the portal. After the end of the development project, the portal was managed by Consortium of Estonian Libraries Network. All the future financing is coming from integrated partners as a participation fee and from the budget of Estonia Ministry of Education and Research.

More than 17 million objects are currently indexed on the E-varamu portal and this number is constantly growing. The number of data sources is also growing, as preparations are underway for the interface between the repository of the Estonian University of Life Sciences and National Library newspaper portal.

It is clear from the above that the E-repository portal is a large and complex information system that must provide a universal search service not only to the affiliated institutions, but to the entire public. There is a great demand and high expectations for the high-quality operation of the e-repository portal. For several years now, the public sector has been looking for universal consolidated solutions in the provision of public and general interest services, and the E-varamu portal has a significant role to play in making Estonian intellectual property available.

Methodology of work

The created portal is unique in Estonian conditions and enables to search simultaneously in all libraries, museums, archives and collections of other institutions that have joined the portal. In other words, it is a solution that integrates different databases and is based on a modern search engine, the ambition of which is to make all the intellectual property that is collected and stored in Estonia digitally searchable and discoverable in one place. For today over 12 different Estonian data sources have joined the E-repository: National Archives, Shared catalogue of the largest libraries of Estonia ESTER, The database of Estonian articles, Index Scriptorium Estonia (ISE), Ministry of Culture, Estonian Literary Museum, Estonian National Library, University of Tartu Library, Tallinn University Academic Library, Tallinn University of Technology Library, Ministry of Education and Research, Estonian Language School, Estonian Language School, Library of Estonian University of Life Sciences.

The e-repository portal as a software itself does not preserve digital objects but consolidates and creates an indexed base for digital information existing in special fields and institutions, highlighting the descriptions of digital material and the direction to the raw data.

From the beginning of the whole project, there has been a plan to offer the data of the e-repository portal as open as possible to users and publish it also in Europeana. In 2015, a data exchange agreement was signed with Europeana and the e-repository portal is Estonian input for the European Digital Library Europeana, created at the initiative of the European Commission. The existence of the portal thus enables both the memory institutions that have joined the portal and the Estonian state as a whole to fulfill its obligations in making the digital collections of libraries, museums and archives available in the European Union. The Europeana Common Culture project provided the means to include relevant digital objects in the Europeana Collections portal, therefore stimulating and finalising the process of setting up an aggregator for Europeana in Estonia.

The portal software is based on the software of the German Digital Library and was developed by the Fraunhofer Institute for Intelligent Analysis and Information Systems (IAIS), part of the Fraunhofer Institute in Germany, which also develops and manages the corresponding digital library in Germany.

The e-repository portal makes its data available through modern technologies using OAI-PMH and API protocols, and the data is made available in EDM format.

Future plans

In the future, it is planned to integrate and make available all data sources located in Estonia as much as possible, so that critical quality would be created through quantity. As the E-repository portal reflects interfaced databases, strong cooperation is carried out with partners to improve data quality.

Estonian E-varamu portal was created within the project “Estonian e-repository and conservation of collections (first stage)”. This “first stage” is set up knowingly because in the first stage finances was only half of the road plan and in 2020 there is a new project financing from the Estonia Ministry of Education and Research to further develop the E-varamu portal in terms of content and technology.

Lessons learned

Data quality doesn’t only rely on the quality of metadata and digital data, but also close cooperation with our partners, who need not only a technical approach but also cooperation between different institutions. This, in turn, means legal work.

In addition, great emphasis has been placed on technology and process automation, but despite this it is clear that more manpower is needed, and this is in a situation where the E-varamu portal is growing every year.

Serbia

Introduction

The idea of setting up the National Aggregator for Europeana at the National Library of Serbia existed many years before 2018 – the year the service was officially launched. It was based on the leading position the National Library had in the digitisation of cultural heritage in Serbia in the last 15 years, as well as its involvement in TEL from 2005. The decision made at the CENL level to close the TEL service (Europeana Aggregator for European National Libraries) in 2016 accelerated the efforts of the National Library of Serbia to set up the aggregation infrastructure. Legally, the National Aggregator for Europeana in Serbia is a part of the Digital Library Department at the National Library of Serbia.

Strong support was provided by the Delegation of the EU in Serbia through the EU Info Center, in recognition of the European integrations in the digitisation of cultural heritage in the efforts to set up the National Aggregator for Europeana. They helped a lot in the promotion of Europeana projects, which used to be the aggregators for Serbian digitised CH on Europeana and where 5 Serbian CHI have been partners. Finally, the Delegation gave the financial support to the project of building the aggregation infrastructure at the National Library. The Ministry of Culture and Information of Serbia also supported this project, recognising it as a part of a broader, national digital infrastructure for culture and arts - DARIAH Serbia. Also, the Ministry is supporting the annual maintenance and updates of the aggregation infrastructure, developed by Semantika, an IT company from Slovenia, a technical partner of the National Library of Serbia.

A staff of 4 librarians worked part-time on the NA in 2018, when the aggregation service started, but since April 2019, 3 persons have been working on the NA, so in-house capacities are greatly deficient. These 3 librarians are working on the Europeana Common Culture project, with the help of the technical partner in Slovenia. The aggregation workflow includes working with library datasets, but it could be adjusted to accommodate museum datasets too. There are a few museums interested in contributing to Europeana, as well as some libraries. We are still testing this kind of cooperation, because of limited resources in the first place.

The National Aggregator for Europeana in Serbia achieved the accredited aggregator status at the Spring 2019 Europeana Aggregators' Forum. However, as an emerging aggregator, it has first provided the datasets from the National Library of Serbia, in 2018, as a result of two Generic Services projects: Rise of Literacy and Migration in the Arts and Sciences.

Methodology of work

Currently, the National Library of Serbia is working with data exported from the COBISS cataloguing system, which includes the majority of libraries in Serbia. The exported data is in the COMARC xml format. The data is then transformed via an in-house tool for COMARC to EDM conversion and enriched with links leading to the digital objects at the NBS digital portal (digitalna.nb.rs) and links to Creative Commons Licenses. The obtained EDM xml dataset is then uploaded to the Aggregator for Europeana for further processing, resulting in an EDM External dataset. The link to the dataset is then sent to Europeana.

Recently, Europeana introduced a new system for submitting ingestion requests of new or updated datasets via the Jira platform, which is quite useful as it facilitates the exchange of information per dataset and tracking the progress of dataset processing. Prior to the publication of the processed dataset, Europeana sends us a preview of the dataset so that we can check for possible errors on either side. So far, the cooperation with Europeana has been very fruitful as we have attended meetings and workshops to guide us in data transformation to EDM and their experts are always available for questions and guidance. We have also been able to spot certain deficiencies in their tier level determination system, so together, we are working towards constantly improving both our systems.

Currently, the majority of the datasets come from the National Library of Serbia. As part of the Common Culture project, the National Library of Serbia is working on updating old datasets that were not submitted via the Aggregator, in order to enrich the metadata and improve the tier levels. The National Library of Serbia has also submitted new datasets as part of the project and two blogs. Also, some of the material submitted as part of the project has been used in 3 galleries.

When it comes to cooperation with data providers, so far it has been possible to include datasets from the Public Library in Bor, as part of the Common Culture project, and the Metropolitan University in Belgrade, as part of the Migration in the Arts and Sciences project. Currently, work is proceeding on including the material from the University of Arts

in Belgrade, Faculty of Music, and possibly two museums: the Museum of African Art and the Museum of Yugoslavia.

The major challenge in obtaining new data providers lies in the fact that many libraries in Serbia still do not have their own digital portal to host their digital material. So far, we have used the NBS digital portal for some of their collections, but that is not a viable long-term solution. The libraries in the COBISS system can export their own datasets in the COMARC format and the datasets are then sent to the NBS experts for further processing into EDM and EDM External. However, in the future, especially after we establish cooperation with many data providers, they could be trained to process their own data into EDM and then submit such datasets for conversion into EDM External via the Aggregator.

Future plans

The experts from the National Library of Serbia working on the Aggregator for Europeana will take part in library conferences in an effort to promote the Aggregator for Europeana and the benefits of cooperation with Europeana for cultural heritage promotion and dissemination. They are also available for questions and inquiries and ready to provide guidance for data providers interested in submitting their data via the Aggregator and for developing tools for data conversion. It is also intended to work on including not only libraries but also museums and other cultural heritage institutions and adjusting tools to accommodate their data as well. The goal is to improve the datasets to meet the highest tier levels feasible, so the partners at Semantika are currently working on enabling conversion from Dublin Core to EDM to accommodate certain datasets, and they are also working on improving the Aggregator so as to enable the use of Wikidata in elements dc:creator and dc:contributor. It is also planned to include elements such as edm:Agent, skos:note (to add the type of script of the Serbian language), dcterms:spatial and dcterms:temporal. There is a dedication to keeping up with the latest Europeana standard, and contact is maintained with their experts for any advice needed. Participation in the Aggregator-in-Residence programme would be appreciated.

Lessons learned

Working on the Aggregator presents a challenge, due to constantly changing circumstances, which forces the NA to push its own boundaries. Unfortunately, it could not update certain collections to fit the latest content and metadata tier levels, so they had to be depublished. From now on, submitted collections will be updated to the latest standards. It has also been learnt that in order to make the Aggregator a part of the national digital infrastructure for cultural heritage in Serbia, stronger support is needed from the Ministry of Culture in the future.

Latvia

Introduction

The National Library of Latvia (NLL) has been aggregating content since almost the very beginning of establishment of Europeana and for several years the main metadata format used was ESE, which was later replaced by EDM. For the most part National Library of Latvia has provided its own datasets to Europeana with the exception of a few smaller partner institutions that lacked the capacity to send content to Europeana directly by themselves.

Although the National Library of Latvia has always been an active participant in Europeana Aggregators community activities, it only achieved accredited aggregator status in Autumn, 2019, transforming itself from emerging aggregator to established one. This means that NLL will be the main (and only) gateway to Europeana for most CHIs in Latvia, but it is still unclear whether the National Archive of Latvia will keep a separate data channel to Europeana through their specific aggregators.

There is no specific financing for the National Aggregator in Latvia per se, but the aggregator is a part of IT infrastructure of NLL that has an annual funding within the budget of NLL.

The National Aggregator runs as a collection of services within the digital object management (DOM) system of NLL. DOM is one of the main systems of NLL and has technical support from NLL's IT staff: both for hardware and software parts of the system.

As DOM has been developed specifically for NLL by an external IT company in Latvia, NLL has a dedicated annual budget for paid system upgrades, which lately have been used to make upgrades necessary for aggregator functionality.

There is currently one dedicated NLL staff member for all technical tasks with preparing EDM data for the National Aggregator. Another two staff members are involved in resolving technical issues with receiving data from partners and there is also a metadata editor responsible for overall data quality.

The National Aggregator in Latvia does not have a specific target audience as it collects a wide variety of digital objects: both in terms of format and content.

Methodology of work

National aggregator functionality is part of the Digital Object Management (DOM) system, which is used by NLL as the main storage of all digital objects created by NLL. It now also holds more and more partner CHI objects specifically for providing them to Europeana.

DOM has an EDM-based (although not exactly EDM) internal metadata format, which can be translated into EDM fairly easily. As part of improving metadata quality in the Common Culture project, many metadata fields have now been updated with a language parameter. Until recently data providers could upload their data to NLL in various different formats (XML, Excel, even Word) and NLL technical staff then converted data into DOM internal data model and ingested transformed data into the system. However, in the future NLL plans to provide a data ingest service to all data providers, who will be responsible for preparing data according to NLL specifications.

Currently there is no generic agreement used by NLL for cooperating with data providers and each potential partner is treated on a case-by-case basis. This might change in the future as more CHIs might express interest in aggregating their content to Europeana, in which case a standard agreement might be created.

NLL has used two methods for sending content to Europeana: a.) a single XML file for the whole dataset or b.) OAI-PMH service. Recently, OAI-PMH has been the preferred method for NLL when uploading data to Europeana. The communication with Europeana has also recently changed from emails to the new Jira system Europeana now uses for communication with aggregators.

To help disseminate and promote content in Europeana, NLL has participated in most collection days organized by Europeana (Europeana 1989, Europeana 1914-1918, Europeana Migration, etc.). NLL has also written posts for the Europeana blog, one of the most recent being a collaborative blog post with Latvian museum of Energy for CC project.

Future plans

One of the main challenges for NLL when aggregating content from various CHIs with different digital object management systems, is the different controlled vocabularies, classifiers and thesauri used. A typical example would be lists of persons, which almost every data provider has. However, each data provider has different methods of identifying persons and in some cases the values in person lists are not controlled at all.

To deal with different controlled vocabularies, NLL and partners are working on a centralised system that should be used and referenced by all data providers involved. Still, it is expected that, at least initially, a lot of manual mapping will be required.

In case archives will agree to go through a National Aggregator workflow, NLL has to decide how to best represent their records as archives use a hierarchical metadata structure - not typical either for NLL systems or the EDM metadata model.

The Common Culture project has been an inspiration for NLL to review data quality requirements not only for the content that is aggregated for Europeana, but for local digital objects as well. In all current and future digitization projects the data quality standards are set to be at least Tier 4 compliant (where it is possible due to copyright restrictions) and metadata quality for the newer digital objects also tends to be at least Tier B according to Europeana recommendations.

Lessons learned

While gathering data form data providers a conclusion has been reached that it is important to enforce data quality requirements as early as possible in order to not cause a lot of error corrections in later stages of data workflow. In some cases only the original data provider is capable of providing correct information on the original object, for example, identity of creators, places, events, etc.

Data providers should also be responsible for choosing the correct copyright restrictions of their digital objects as National Aggregator can only apply default copyright settings (usually more restricting than those that data holders could provide).

Many of our data providers wish to see digital object view stats in Europeana, which is also their main motivation for providing data to Europeana in the first place. It is suggested that Europeana should provide some kind of analytics tool for digital object access/view statistics for each National Aggregator.

Ireland

Introduction

Prior to the Europeana Common Culture project there had been no active National Aggregator in Ireland for several years. Although many items were previously aggregated via a former National Aggregator and via Thematic Aggregators, many of these did not meet the new Europeana Publishing Framework Content and Metadata Tier requirements.

As a National infrastructure for Ireland's Social and Cultural Heritage, the Digital Repository of Ireland (DRI) was keen to reestablish a National Aggregation service. It was particularly conscious of the risk that Irish materials would become unavailable on Europeana without a National Aggregator who could promote Europeana to Irish Institutions, and educate them about the new Content and Metadata Tier requirements.

DRI achieved accredited aggregator status at the Fall 2019 Europeana Aggregators' Forum. Official recognition by the Irish Department of Culture, Heritage and the Gaeltacht is currently being sought.

The DRI activities are funded by the Higher Education Authority and the Irish Research Council. DRI also operates a paid membership model. While the development of the Europeana Aggregation service was funded by the Common Culture project, ongoing support, maintenance and operation of the service will be funded under the DRI's Core funding.

DRI has a staff of 16 which includes software developers and archivists/librarians. They work with member organisations providing support and training in the preparation of collections. The technical team maintains the repository and continues to develop new functionality. The repository has over 35,000 published digital objects, the majority of which are suitable for aggregation to Europeana. A number of additional collections are due to be published and aggregated in the coming months.

The designated communities of the DRI include GLAM professionals, researchers, educators, genealogists and members of the public. The repository includes both digitised Cultural Heritage objects from the GLAM sector and research data from the Humanities and Social Science domains.

Methodology of work

DRI is a member-based organisation providing a range of services to member organisations. Membership is offered to any organisation with content related to Ireland's contemporary or historical social and cultural heritage. Members have access to training, education and networking events. Regular members are provided with training in online, face-to-face group and one-on-one formats. A Digital Archivist and Content and Engagement Coordinator work directly with members to help them to prepare their collections for ingest and aggregation.

DRI provides sustained access, storage and long-term digital preservation in addition to aggregation services.

DRI is actively aggregating collections by working closely with the Europeana Ingest Team. Since the beginning of the Common Culture project DRI have co-authored two blog posts for Europeana Collections, one for LGBT + History Month and one regarding our pilot aggregation, the Jacob's Biscuit Factory Photographic Collection. DRI also contributed a blog about Europeana Developments in Ireland for the Europeana Pro blog. In addition, DRI has contributed to the Seasonal Traditions in Europe Exhibition and the Industrial Heritage Thematic Collection. An article has recently been submitted on its 3D activities for an upcoming issue of Europeana Tech Insight.

In 2019 DRI was awarded a Europeana Research Grant to run an event entitled 'FAIR and Open Data: Opportunities for the Galleries, Libraries, Archives and Museums sectors', to take place in 2020.

The aggregation service is implemented as part of the DRI preservation repository platform. This is an open source platform based on Ruby on Rails and Samvera. The code is available on Github (https://github.com/Digital-Repository-of-Ireland/dri-app).

Collections in the Repository must be catalogued to a range of popular metadata standards in use in the GLAM sector. These are Dublin Core and Qualified Dublin Core, MARCXML, MODS and EAD. A number of tools are also provided to allow metadata to be generated directly from spreadsheets.

These metadata standards are all mapped to a consistent internal data model and indexed in a Solr Search Engine. An OAI-PMH EDM feed is automatically generated from the data in the Solr Index. Contextual classes for date ranges and locations are included where this data is provided by the content provider. In some cases where this is not provided, DRI staff perform metadata enrichment prior to aggregation.

DRI has so far run one dedicated Europeana outreach event which took place in October 2019 (https://www.dri.ie/engaging-with-europeana). Similar events are planned for the future, although due to the current COVID-19 pandemic, these will now likely be postponed until after the end of the Common Culture project. Virtual training and outreach activities are being considered in the meantime.

The DRI Metadata guidelines (https://www.dri.ie/publications#guidelines) are being updated to include information on Europeana Aggregation via DRI. In November 2019 DRI published a Factsheet on DRI and Europeana (https://doi.org/10.7486/DRI.v9807p04r).

DRI have made contact with all of the organisations who had aggregated collections to Europeana via the former National Aggregator, providing guidance on how they can update their collections. We are also heavily promoting Europeana Aggregation to our members.

Future plans

Support for 3D content is currently being developed and it is intended to add IIIF aggregation in the near future. We will furthermore be adding support for additional Linked Open Data sources and mapping these to contextual classes in EDM. In particular, this will include increased support for the creation of skos:concepts from LOD subject terms.

DRI are also planning to participate in the upcoming Europeana editorial season on Sport by project managing an online series of collection days titled 'The Sporting Irish'. Bringing in partners from the Europeana Network Association, Universities and Cultural Heritage Institutions in Ireland. This online event will look for contributions from a range of sports played in Ireland, at a time when all professional and amateur games have been cancelled due to the Covid-19 outbreak.

DRI are currently in discussions on using the newly updated Enrich Europeana Transcribathon platform to run a series of crowd-sourced metadata enrichment campaigns with our Data Provider Dublin City Library and Archives.

Lessons learned

As DRI is in a position where it is reestablishing a dormant National Aggregator service, and has not been able to guarantee the updating of existing Irish records, which do not meet the minimum quality requirements of the Europeana Publishing Framework. Not all content providers who had previously aggregated material are in a position to work with DRI going forward. This means that some records have had to be depublished instead of being updated.

The opportunity to be part of a Europeana blog, gallery or exhibition provided an incentive for new data providers to engage with DRI and to have their collections aggregated. Editorials also provided an additional incentive for enriching metadata or applying a licence that allowed reuse, so the collection could be included in editorials and promoted online through social media.

There is an appetite in Irish cultural institutions to engage with Europeana and a general response from institutions we reached out to was that it was always their intention to find a way to contribute a collection to Europeana. However there had been a lack of knowledge around how to engage with Europeana in the absence of a national aggregator. To help spread awareness of the Europeana aggregation process DRI wrote blog posts, organised events and published a factsheet. The primary aim of these activities was to demystify the aggregation process and explain the technical jargon around preparing your collection for aggregation to a non-technical audience in Irish cultural heritage institutions.

Summary

In the framework of European Common Culture project, the four aggregators have been working to become established aggregation entities for CHIs in their countries. Each aggregator has already developed aggregation routines and has a plan for future actions to assure proper visibility, dissemination and accessibility of respective cultural heritage resources. In 2019 three out of four aggregators already achieved Europeana Accredited Aggregator status, which means that the set up of emerging aggregators in the project activities is well progressing. There are still challenges and issues to overcome in the coming years, however the basic infrastructure and aggregation routines are in place and allow the ingestion or update of records in Europeana.

Considerations on national aggregation policies

In the fall of 2019 a workshop on “Empowering Collaborations” between aggregators was organised by KU Leuven. The results of this workshop have been gathered in a report[24]. At the end of the workshop, participants “voted” on specific collaboration questions. Some of the results are listed below:

[24] https://drive.google.com/file/d/1GD0t9ceERQTgUCznd1z2jz3WXXMAyS5y/view

Chart 41. Expected level of formality between aggregators.

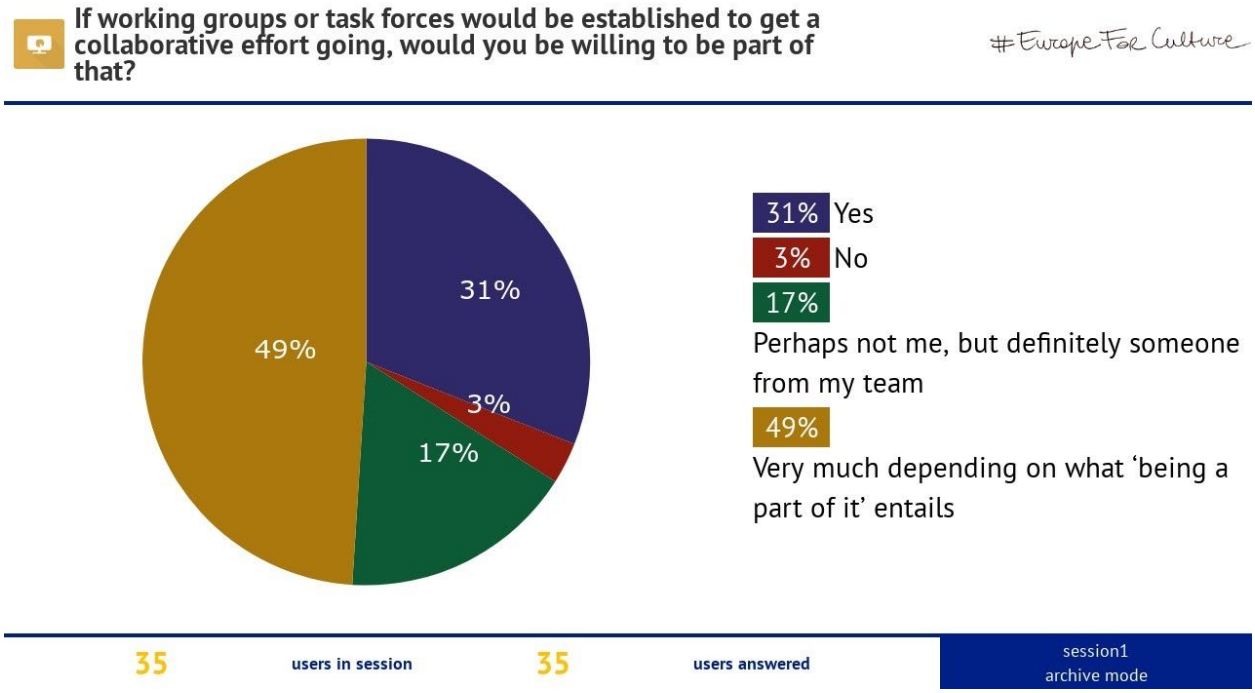


Chart 42. Willingness to participate in working groups and task forces focused on collaborative efforts of aggregators.

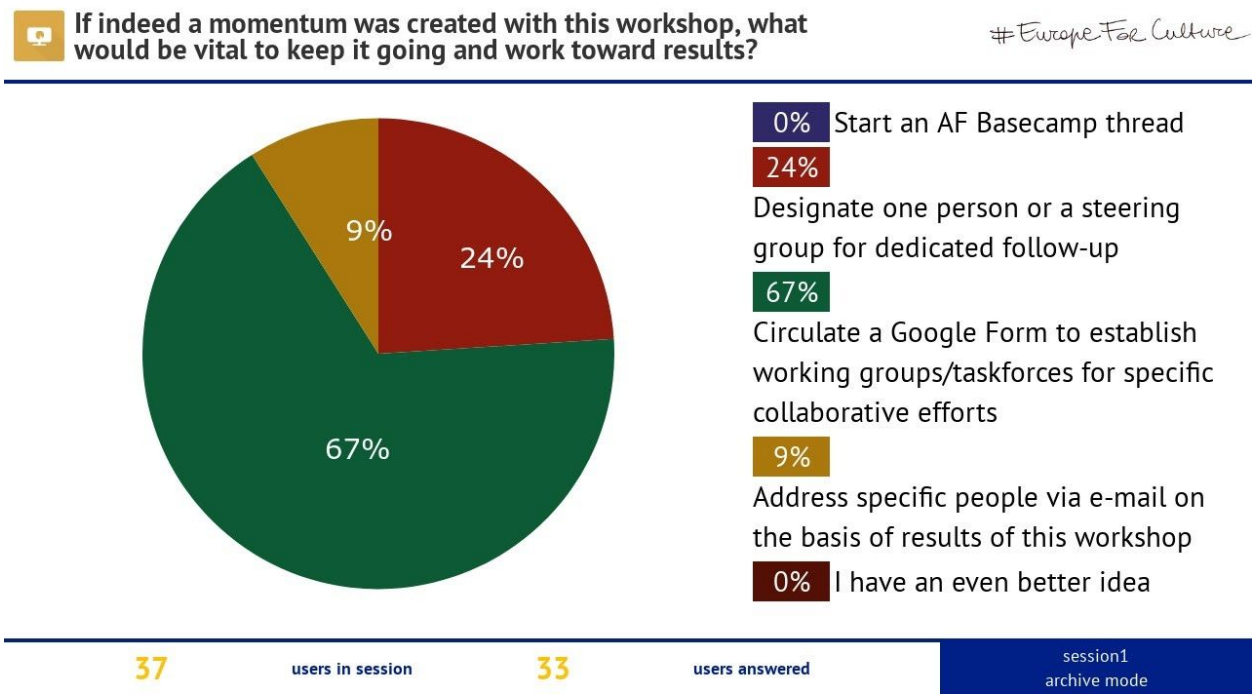


Chart 43. Indication on future steps related to collaboration between aggregators.

Several activities organised by the the EAF and CC Activity 4 team followed-up the meeting:

- A workshop report including slides with results from the questions asked during the session (using QANDR tool).
- Alert participants and distribute the report via Basecamp communication tool.
- Follow-up on working group / task force idea.
- Concertate with EAF in the next 6 months on Europeana Common Culture policy report development.
- Update on progress both on the Europeana Common Culture policy report and on the Empowering collaborations effort during the EAF of spring 2020.
- A Task force has been set up in the EAF to develop a joint policy brief that will be presented to the DCHE meeting in November 2020. This policy brief is in its final stages.

These actions are supposed to strengthen the outcomes of Europeana Common Culture project as well as broaden the scope of investigation and involve all possible stakeholders.

Recommendations on national aggregations practices & policies

Recommendations on national aggregation practices

This report as a whole provides a European landscape of national aggregation practices. Especially in the survey analysis part, there are many points highlighting that NAs have a common approach to specific challenges and duties. These approaches are in our opinion good practices related to data aggregation as a whole and present a state of the art in this field. Therefore, in this section, a list of relevant recommendations has been formulated and listed in Table 5.

These recommendations highlight key best practices and guidelines on how to operate, develop, improve or set up NA. They are intended both for emerging and established NAs. While the emerging aggregators can use the recommendations to inform future decisions on the key aspects related to their daily operations, the established aggregators can use them to revise or improve their existing approaches.

The recommendations are listed in Table 5. and are prioritised according to the MoSCoW[25] method into the following categories:

- **Must Have (M)**: requirements are critical for setting up a NA. “Must have” should be agreed with all relevant stakeholders.
- **Should Have (S)**: requirements are important, but not necessary for setting up a NA. “Should have” are to be considered only if time and resources allow it.
- **Could Have (C)**: requirement is desirable but not necessary.

We do not use the “Won’t have” (W) category, as the intention of this summary is to have a list of best practices that are important to have a state of the art aggregation.

The prioritisation has been determined in the context of workshops and survey analysis conducted and it expresses the opinion of the authors of this document.

Recommendation	Priority

[25] https://en.wikipedia.org/wiki/MoSCoW_method (11.06.2019)

Formulate your goals and your mission. The most frequently mentioned goals/mission statemebts by existing aggregators are: • to give access to cultural heritage objects • to promote resources and cultural heritage of its country • to set up quality standards and create high quality data and metadata More details can be found in the Aggregators mission section.	M
Involve relevant specialists to build capacity. Engage specialists needed to develop and operate the national aggregator. The critical roles are metadata specialists, information specialists and technical specialists. More details can be found in the Staff / capacity / expertise section.	M
Determine metadata formats and models that you will accept from the data providers. Decision should be based on deep analysis of partners that will provide the data, their competencies in providing good quality metadata and a variety of software solutions that exist in the national market. The overwhelming majority of NAs support multiple formats and focus on metadata mapping. In such a case NA needs to conduct thorough analysis in the context of transformation and mapping method. If an aggregator decides to support only one (specific) schema it should be thoroughly investigated whether it can be efficiently imposed on collaborating data providers. More details can be found in the Metadata formats/models section.	M

Apply for a legal mandate to run your National Aggregator. The majority of NAs have a legal mandate of their government authority for their activities. Moreover, some NAs indicated that their activities started because of the authorities initiative. In addition, official support of the authorities often helps NAs to have sustainable funding/operations. More details can be found in the Authorities support section.	S
Offer support to data partners. It is especially important in the context of IPR guidelines, metadata (mapping) and relevant training. More details can be found in the Services section.	S
Advise software vendors to introduce a more standardised approach in the context of aggregation needs. It mostly relates to the systems that deliver digital resources in an online manner (digital repositories, libraries, museums, archives, galleries) . These systems usually have an import module from the library catalogue systems, which are not perfectly aligned with demands coming from the online world, especially in the context of Linked Open Data. Improvements in this context could directly bring benefits to the aggregation activities on a national and European level. More details can be found in Metadata formats/models and Linked Open Data support sections.	S
Advocate and use Linked Open Data. Use LOD vocabularies, recommend usage of LOD vocabularies to data providers and advocate to support LOD vocabularies in collection management systems.	S

More details can be found in the Linked Open Data support section.	
Consider building a search portal to present aggregated resources. Search portal seems to be the most practical and commonly used way of presenting cultural heritage objects. However, not all NAs have it. Therefore conscious decisions in this regard should be undertaken. More details can be found in the Functional modules section.	S
Use standardardized technologies to harvest and expose data. There are various ways to aggregate data and expose them to external entities. However, in general, it is always good to conform with well-established standards, protocols and formats. OAI-PMH and public REST API are examples of such approaches. More details can be found in Data aggregation techniques and Public API sections.	S
Introduce content and metadata quality standards to your data partners. Setting a quality standard and validating data against it is a way of assuring good data quality. Validation might check presence of mandatory fields, correctness of data (xml) format and structure or any criteria set to support some data features (such as search and multilinguality). It can easily be run automatically, e.g. as a step run just after data harvesting. More details can be found in the Data processing section.	S
Provide your data to Europeana to increase their visibility in Europe. Providing data to Europeana is part of the core activities of all the NAs, they are eager to deliver data in appropriate format and to be involved in	S

ongoing activities, e.g. accreditation process or Europeana Network activities. More details can be found in the Cooperation with Europeana section.	
Automate your workflow or parts of it. It's an individual characteristic of every aggregator that depends on available resources, technical capabilities, maturity of technical solutions or relationships with data providers. However, aggregators often try to automate their processes or some parts of it. More details can be found in the Harvesting workflow section.	C
Implement advanced data processing. There are various techniques to improve and consolidate aggregated data. Examples include metadata enrichment, normalisation or validation. More details can be found in the Data processing section.	C

Table 5 Recommendations on national aggregation practices

Recommendations to the member states on National Aggregation Policies

National Digital Transformation Strategies

A **published national digital transformation strategy** establishes the roadmap for the Member State, and it is a driver of change to transform the sector.

A national strategy should:

- Articulate the impact of its implementation
- Be supported by an enabling policy environment and funding
- Be transparent and measurable
- Be open to public debate

A strategy delivers a roadmap identifying the scope, parameters and priorities for the digital transformation of the cultural heritage sector. It considers the impact - the changes in people or groups within society - it is aiming to achieve, and sets out the path to achieve them. This roadmap enables the Member State to achieve economies through focusing knowledge and available resources, scaled towards a common EU-wide goal.

A funded policy environment which is designed to enable the desired changes identified in a national strategy will support its implementation across the Member State, and cohesions with EU-wide development. Working cross-sectorally to jointly programme and implement change and innovation projects saves money, effectively utilises developing knowledge and skills, and produces results that are suitable for cross-sector application.

Examples:

- Copyright rules that support the digitisation efforts of cultural heritage institutions
- Mandates for the building of centres of expertise and infrastructures
- Competence centres for digitisation, such as https://www.digitisation.eu/ (see also the call DT TRansformations 20 on a European Competence centre.

A strategy that is shared openly and accompanied by clear measurements can motivate and focus the sector. This enables the sector's members to learn from each other, as well as advocate the benefits of the elements within a strategy.
Examples of Member States with national strategies that enact these components are:

- The Netherlands' Netwerk Digitaal Erfgoed
- Germany's Competence Network of the Deutschen Digitale Bibliothek

- The Digital Cultural Heritage Initiative[26] of the Finnish Ministry of Education & Culture
- The Welsh Digital Action Plan 2017-2020

An important part of a national digital transformation strategy is the aggregation of digitized (and born digital) cultural content. When we speak about a "National Aggregator" in this document, it is primarily meant as an *accredited aggregator for Europeana, with the ambition to cover the content ingestion flow from an entire region or country, rather than specializing in a specific domain or theme*.

It is Europeana's ambition to have national aggregators in each member state of the European Union, to enable a representative content representation. For participating member states it is important to be correctly represented in the Europeana collections. This can also be considered a cultural right of its citizens. Working with national aggregators strikes the right balance between the investments of the European Commission in the supporting platform and core service through the CEF and future DEP, and the contributions of content by member states (or their regions).

Therefore, it would be an important step forward if "national aggregators" somehow are also "accredited" by their own government, or at least recognized and supported in their important role. We are aware that this is up to the member states themselves to decide and that this could require legal instruments. But we think there is a similar case in the so-called "legal depot" that all countries uphold regarding publications. Of course, this goes beyond the scope of this project, but we recommend to monitor and share evolutions in this regard in the member states. Almost 70% of the surveyed National Aggregators have a legal mandate.

Anyway, a *Europeana* National Aggregator should clearly specify its relation and role in relation to the member state or region it claims to represent. And thus it should be part of an overarching national aggregation policy, where member states organize the digital collection of their heritage as part of a digital transformation strategy.

Frameworks and standards

The best way to make the playing field more equal and more representative throughout Europe is by promoting **common standards** for aggregation. Part of the innovation role of Europeana in the field of Cultural Heritage is precisely to give guidance to professionalisation and enhanced interoperability. Adopting common standards to describe and maintain metadata descriptions can facilitate trans-european exchange and support cross-border education and research. But also the licensing framework, and professional standards for procedures, and guidelines for inclusivity are part of this.

Adoption of international standards requires local environments that create opportunities to learn, share and access resources and expertise.

[26] https://minedu.fi/en/digital-cultural-heritage

Adoption of international standards should be supported by:

- Local translation of standards and supporting resources
- Pan-sector events to facilitate knowledge-sharing.

Local translation of resources and documentation drives adoption of standards. Language barriers should be minimised through active support for the translation of essential documentation such as standards and frameworks, making them more useful and more likely to be adopted. Workshops, meetings and events enable communities to access expertise, share knowledge and learn from each other. Organisations across the sector face common issues and should be provided opportunities to constructively share these issues, co-create solutions and share their own practices. This is achieved by utilising existing networks, and providing funding and support where gaps in knowledge are identified.

Example: The new EU copyright laws harmonise the rules but are also adopted at a local level which leaves open the possibility of differing implementations. Member States should identify how they can support the laws and fund access to expertise according to how they have implemented the directive.

Examples of Member States that have organized the support of standards are:

- Italy's translation of the Europeana Publishing Framework by CulturaItalia
- Poland's implementation of the Polish language Impact Playbook, supported by a series of workshops
- Belgium's Cultureel Erfgoed Standaarden Toolbox
- The Netherlands' De Basis curriculum delivered by Digitale Erfgoed Nederlands
- Rights Statements translations to 8 European languages with 7 additional translations underway, see https://rightsstatements.org/en/documentation/translations.html

Capacity building

In the first part of this Landscape document, the technical requirements and best practices for a successful National Aggregation workflow have been documented. But capacity building goes beyond this. It first and foremost means having the right people and management capacity to deploy national aggregation.

In this sense it is important to set up **"train the trainer"** initiatives. The Europeana Initiative can better take up its role as sector innovator for the Digital Transformation if it can rely on a National Aggregator Network. A good example is the **Impact Playbook**. The development of well-documented **teaching and training tools**, which can be deployed through the Network, both the Network Association as well as the EAF, can be an instrumental contribution by the Europeana Initiative in capacity building.

Accredited aggregators is a first step in a concept that would see **Europeana Certified Professionals** bringing much-needed competencies to the field. Training can be provided for specific skills, such as **running a crowdsourcing workshop**, or technical knowledge such as **using OpenRefine[27] to improve datasets**, or **implementing the Europeana Frameworks**.

With a network of National Aggregators, the Europeana Initiative will have an instrument to give the necessary outreach and impact to its actions. It becomes more **measurable**: when we deploy it through a National Aggregation Network, we can more easily indicate our European-wide impact.

Stakeholder engagement

While National Aggregators should focus primarily on capacity building, making sure they have the right expertise and infrastructure to deploy and support a nation-wide aggregation workflow, this can only succeed if the national stakeholders at different levels are on board.

It makes sense that the National aggregator for Europeana also aggregates contents on behalf of the national (or regional) level, regardless of ingestion into Europeana. In this case the National Aggregator will usually have its own harvesting infrastructure and portal. This is the best guarantee to make sure the democratic control over the representativity of the aggregated content mirrors the democratic structures of society at large.

Giving incentives to regions, provinces and cities to engage with and contribute to Europeana content is an important mission for a National Aggregator, making sure that every citizen in Europe feels “at home” in Europeana and feels represented. Currently it can be quite frustrating for a Europeana user to find out there is little to no content about their hometown while there could be abundant content from neighboring places. A more place-based approach would have additional implications for improving description and organisation, beyond the more representative availability of local content. It would give the European context and its regional diversity more depth and meaning.

Stakeholders are in the first place the Cultural Heritage Institutions falling within the scope of the National Aggregator. This sometimes poses a challenge, as the National Aggregator might be a Library, while important stakeholders could be a National or Regional Museum or Archive. In that case it is vital to build trust and active networks that facilitate collaboration.

The Europeana Initiative could provide models for this kind of national inter-institutional networks, building on Best Practice Examples already existing in the member states.

But stakeholder engagement goes further than establishing solid collaborations with national institutions and authorities. **It is also about spreading the Europeana message**

[27] https://openrefine.org/

and reaching out to citizens. Hosting Europeana collection days is one example of an activity that perfectly fits a National Aggregator in collaboration with local institutions. In collaboration with a CHI, Europeana invites citizens on location to bring an object that has special meaning to them and tell their story. These stories are recorded and published on Europeana.

Setting up a National Aggregation Network is also a good occasion to highlight Europeana as a People Network. It is important, for researchers, educators, amateurs as well as the general public to discover through Europeana, not only the wealth of valuable heritage objects and collections, but also **to learn to know the institutions and professionals who maintain and care for this heritage**. This can for example facilitate the work of researchers and educators who want to do in-depth work with the Europeana content and need more background info. But it can also bring professionals in the sector closer to their stakeholders and end-users.

In this way, the National Aggregators can stimulate the growth, not only of the Europeana Aggregator Forums, but also of the Europeana Network Association, by bringing people across Europe who care for shared values together.

Collaboration at the Europeana Aggregated level

Setting up the National Aggregation Framework also implies revisiting the way in which Aggregators work together, at the technical level, but also at the content and collections level.
It is in Europeana that the diaspora of Heritage Objects comes together, and can be united in new, European-Wide Collections. This has been worked on in several CEF Generic Services projects, and offers a unique opportunity to complement and enrich metadata of collections at the aggregated level. This way, collections which are de facto spread over Europe can be brought and displayed together - which is in particular important for European Countries and communities who have a large diaspora of citizens and heritage throughout Europe.

Besides making more complete and better described virtual European-wide collections, there is also an opportunity for National Aggregators to share expertise and work more closely together, within the EAF, with the Domain and Thematic Aggregators. Ideally, **joined curation and enrichment tools** should be available on the Europeana Platform itself, as they are currently in development for Europeana XX.

Suggested actions

Based on considerations in previous sections it is recommended to investigate and conduct the following actions related to concrete areas of aggregation policy:

- National Digital Transformation Strategies
    - In each member state at least one Europeana Accredited National Aggregator who has an officially recognized aggregation role by its governing authority.

- Frameworks and standards
    - Further elaboration and implementation of Europeana Frameworks as a common reference for Aggregation.
- Capacity building
    - Through “train the trainer” actions, the impact playbook, accreditation and benchmarking with measurable indicators capacity can be strengthened at the national/regional level.
- Stakeholders
    - Building a Stakeholder Community through visible workshops, training and certification, developed jointly at the Europeana level and distributed through the National Aggregators and the Europeana Network Association.
- Collaborations
    - Finding new, platform-supported collaborations between Aggregators (National, Domain and Thematic) at the common Aggregation level.

Conclusions

The landscape of national aggregation in Europe is composed of many actors, including NAs, domain and thematic aggregators, Europeana as well as data providers. All of the activities conducted by these actors influence and direct the way European cultural heritage is exposed/visible to the interested parties on a global scale. The analysis conducted among NAs in Europe provides important inputs on how the NAs are organised. It also shows that NAs share similar values, characteristics or approaches.

In principle, NAs are initiated with the mission to promote resources and cultural heritage of respective countries by providing access to free and high quality cultural heritage resources. This core mission is realised via aggregation workflows that take place in each NA. The common approach is to harvest data from heterogeneous data sources (data providers) and then apply relevant data transformation techniques to achieve a standardised set of data that can be consolidated on a national and international level. However, there are rare approaches to impose requirements on data providers and ensure that the standardised set of data is already available on the level of data source (data provider). Data harvesting and processing techniques vary from country to country, e.g. some NAs aggregate data automatically, while others do it manually. There are also similarities like relevant protocols of formats that are commonly used. Main channels of dissemination include national data access portals and Europeana. National portals are focused on search & browse features to help users explore cultural heritage on a national level. Cooperation with Europeana provides means to open up and present cultural

heritage in European and global context. Additional activities are undertaken in this regard, including data transformation (to conform with Europeana requirements) and following relevant ingestion pipeline.

Given various characteristics of the NAs in Europe, gathered in the conducted survey, it is possible to identify common elements relevant for a typical/model NA. These elements/characteristics can be understood as best practices or common approaches important for setting up or operating NA, and therefore they have been gathered and summarized in the form of recommendations. The recommendations are dedicated to all NAs, regardless of the fact whether they are established or emerging. Each NA should consider all recommendations and undertake informed decisions in the context of their local conditions.

At the policy level, key policies that should be addressed at the national (or regional) level have been identified: national digital transformation strategies, frameworks and standards, capacity building, stakeholder engagement, collaboration at the aggregated level.

Annex 1. Survey questionnaire

Attached as a separate PDF file generated from Google Forms used to gather the responses.