





Content

Content	3
Executive summary	(
Introduction	10
The purpose of the Europeana Content StrategyEuropeana Strategy and Publishing Framework	10 11
Principles of the Europeana Content Strategy	12
Quality instead of quantityDemand-driven content acquisition and publication	12 12
Priorities for content acquisition	14
 User demand for Europeana Europeana Thematic Collections	14 15
• Data quality	16
Type of content relevant for EuropeanaDate range of content relevant to Europeana	17
Geographic scope of content acquisition	19
Non-digital objects	19
Access to digital objects	20
Masterpieces and lesser known collections for Europeana	20
Crowdsourced content	2′
Metadata and content from Wikimedia Foundation projects	2
Active (pull) and passive (push) content acquisition	24
Priorities for publishing in Europeana	24
Dissemination of content	2.
Conclusion	25

Authors: Henning Scholz, Cécile Devarenne, Nuno Freire, Panagiotis Kyrou, Joris Pekel, Milena Popova, Conxa Rodà, Malgorzata Szynkielewska, Wim van Dongen, Erwin Verbruggen, Uldis Zarins

March 2017





Unless otherwise noted the contents of this document are licensed under the terms of a Creative Commons Attribution ShareAlike 4.0 licence.





Executive summary

Europeana's long-term goal is to "get the right content to the right user at the right time". Our Content Strategy drives the development and management of the content accessible via Europeana Collections and the Europeana APIs in pursuit of this goal.

Its guiding principles are:

- Quality instead of quantity
 - ° Use the <u>Europeana Publishing Framework (EPF)</u> to define high-quality content and encourage institutions to open their data
 - ° Prioritise the improvement of existing published data over adding new data
 - ° Remove existing non-compliant data as specified by the <u>Europeana</u> Publishing Guide
 - ° Showcase high-quality data in Europeana Thematic Collections
 - ° Achieve the first priority of the Europeana Strategy to improve data quality
 - ° Flip the relation between content in Tier 1 of the EPF and content in Tier 2 and higher no later than 2020
- Demand-driven content acquisition and publication
 - ° Analyse statistics of content accessed via Europeana Collections and Europeana APIs
 - ° Conduct surveys and interview users to establish their content needs
 - ° Evaluate new content against relevance to Thematic Collections and EPF compliance

Following these principles the Content Strategy concludes the following:

¹ From the Content Strategy Alliance definition of Content Strategy: http://contentstrategyalliance.com/the-beginnings/csa-charter/

Type of content

Europeana is not in a position to decide on a definition of 'cultural heritage'. Therefore, it will continue to be inclusive and in principle publish all types of content that museums, libraries, archives, and audiovisual institutions, based on their respective digital strategies, consider important enough for publication in Europeana. However, Europeana will, based on experience, discuss with data partners if publishing certain types of content is worth the effort.

Date range

Europeana is inclusive of the cultural heritage from centuries long past to modern times. Making 20th century content accessible is important to avoid limiting the value of our services; for example, when engaging with the education sector. It is at the discretion of the cultural heritage institutions to determine what is old enough to be deemed cultural heritage.

Geographic

Europeana focusses on content that is about Europe, Europeans, created in Europe, or owned by a European institution. Most of the material in Europeana will come from European institutions. But in cases where it would strengthen our themes and the sources are Tier 2 or higher, Europeana may actively pursue content from non-European sources.

Non-digital objects

Metadata records in Europeana always need to point to a freely accessible digital object. This is specified in more detail in the Europeana Publishing Guide.

Access to digital objects

Our users also need to be able to access the digital objects without authentication, passwords or filters on geographical location.

• Masterpieces and lesser known collections

Europeana is unique in offering a discovery service for lesser known collections, and therefore will not give special preference to

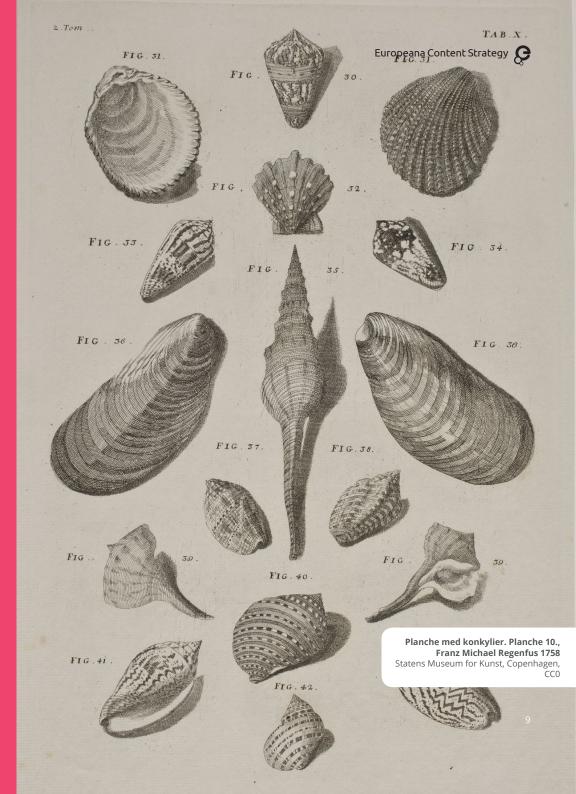
masterpieces in the content acquisition process unless there is a specific demand; for example, when specific masterpieces are important for a thematic collection.

Crowdsourced content

Some well-known and trusted digital libraries and resources for digital cultural heritage have been developed by volunteers and often have resources that are otherwise not available via data partners with whom we work. If user demand for such crowdsourced resources (user created content) exists and the quality of these resources is sufficient, then this content is relevant and can be published in Europeana. We will always mark crowdsourced content in Europeana as such making it clear it is not content provided by a Europeana data partner. In some cases, to increase user participation, Europeana will actively design Pan-European campaigns to crowdsource content.

Metadata and content from Wikimedia Foundation projects
 Integration of metadata and content from Wikimedia Foundation projects into Europeana products can help improve user experience of these products. Depending on the source of the data, Europeana will select the most appropriate integration approach.

The remainder of the Content Strategy document elaborates on the above conclusions in detail and specifies how Europeana intends to put these principles into practice. It details the rationale behind each element of the strategy and how it relates to Europeana's strategy and other relevant policies. The operational aspects of the Content Strategy as well as the data requirements to implement the Content Strategy are specified in the Europeana Publishing Guide.



Introduction

We transform the world with culture! By 'we' we mean everyone in Europeana's network across the cultural heritage and creative communities. We share the dream of a world where every citizen will have access to all cultural heritage. We believe that seamless access to culture enables a positive and necessary transformation of the world we live in. We hope our work contributes to that. With our products and services we want to contribute to a stronger feeling of European identity, to unite Europe through culture; and we want our products and services to lead to real economic change. When people - through our work - get a broader perspective on the horrors of the First World War, when they understand the relevance of migration for the development of Europe or when they express themselves through art from different countries, we have achieved our goal. If we can save cultural institutions money because they can re-use our exhibitions platform, or if a cultural entrepreneur develops a new business using our API, we have achieved our goal too. "We want to build on Europe's rich heritage and make it easier for people to use, whether for work, learning or just for fun."2

The purpose of the Europeana Content Strategy

We "dream of a world where every citizen will have access to all cultural heritage"³. If we want to make it easier for people to use cultural heritage, rich metadata are necessary so that cultural heritage objects can be identified and discovered. The content needs to be of high quality too, so it can be used for work, learning or fun. Giving access to all cultural heritage with a focus on data quality requires a balanced content acquisition approach that is not specified in the Europeana Strategy 2015-2020.

This Content Strategy helps us to plan the development and management of the content accessible via Europeana Collections and the Europeana APIs so we can "get the right content to the right user at the right time"⁴. Improving the quality of existing data while actively searching for new data from new institutions to be added to the database is key to our content development.

This Content Strategy specifies among other things the geographic scope of our content acquisition, the importance of masterpieces and lesser known collections for Europeana and how we deal with non-digital objects and crowdsourced content.

The Content Strategy is driven by user demands to tailor Europeana to user needs. This method will inform us of the themes we want to prioritise for publication, the quality aspects prioritised for improvement and the functionalities we offer to distribute the content (e.g. search improvements). By making this connection between content and user needs more explicit, we also aim to engage more with our data partners, to help them identify the right content. This way we can also show the value of content within our organisation and to our partners, in line with our brand values of 'usable, reliable and mutual'.

User demands and technologies will change and evolve over time. Therefore, our Content Strategy needs to be adapted to these changes from time to time. As the Content Strategy is built on the Europeana Strategy (see also the following chapter), strategy refinements or a new strategy will require the Content Strategy to be revised. In line with our annual cycle of business planning and prioritising our work for the following year, we will also look at the Content Strategy, review it and adapt certain aspects if necessary.

Europeana Strategy and Publishing Framework

With the Europeana Strategy we set ourselves three priorities: to improve data quality, to open the data and to create value for partners. The Europeana Content Strategy builds on the Europeana Strategy and guides the Europeana Publishing Framework (EPF). The EPF helps us to achieve the objectives set forth in the Europeana Strategy by setting out four scenarios for sharing collections via Europeana. Depending on the digital strategy of the individual cultural institution and the quality and copyright status of content they make accessible, we are able to provide different types of services to the institutions in order to support them in fulfilling their respective missions. In clarifying the benefits and added value of open and high-quality content, the EPF helps us encourage institutions to open their data and improve the quality of what they make accessible. For the different types of content, the **Europeana Publishing Guide** specifies the quality requirements to comply with one of the four scenarios or tiers. If institutions want to make sure that a collection is used in all contexts and across all applications, they will need to make it available in high quality and under a licence that allows free reuse.

http://contentstrategyalliance.com/the-beginnings/csa-charter/

² Our mission: http://www.europeana.eu/

³ http://strategy2020.europeana.eu/

From the Content Strategy Alliance definition of Content Strategy:



Principles of the Europeana Content Strategy

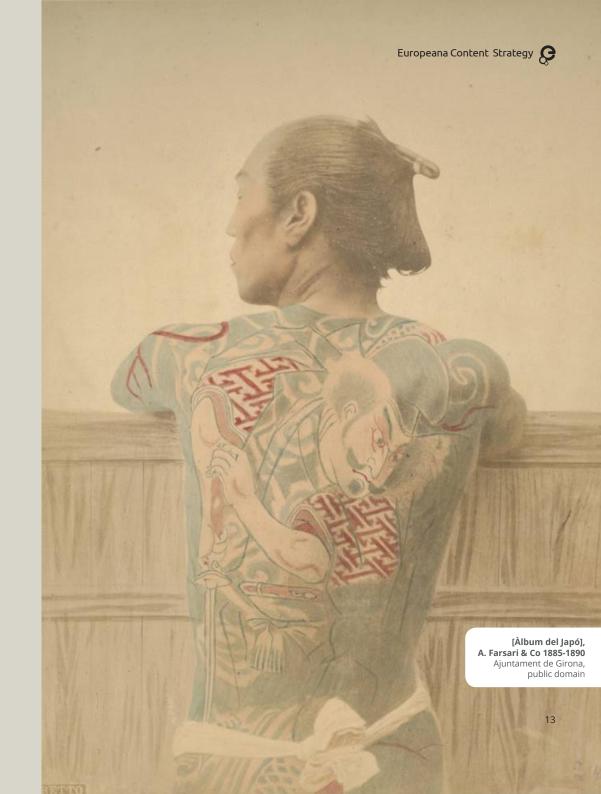
Quality instead of quantity

Improving metadata and content quality is the number one priority in achieving our strategy's goals. Now that we have made a critical mass of content available, publishing with the goal to increase the quantity of content only is no longer a priority for us. We even encourage data partners to invest in collections that are compliant with Tier 4 of the EPF while keeping in mind that such high-quality collections may be very small and may only encompass a few hundred items. We would even accept to decrease the total amount of content that we make accessible if it improves the overall quality of the content.

Following the logic of a user-driven Content Strategy, we acknowledge that data quality can be defined differently depending on the audiences. Keeping this in mind for an overall quality improvement strategy, we have defined good metadata quality as resulting from a series of trusted processes: findable, readable, standardised, meaningful to audiences, clear on re-use, and visible. During 2017, we will also introduce measurable metadata quality values that are automatically created when content is first published or updated in Europeana.

Demand-driven content acquisition and publication

There is a lot more content in museums, libraries, archives, and audiovisual institutions than we can publish in Europeana in a meaningful way. It is acceptable to leave gaps in our collection, also because not all of the content is equally interesting for our users. Therefore, we listen to our users and their needs for cultural heritage content to prioritise content acquisition and publication (see section below). User demands are also important when it comes to access: access is the ability to view and download a digital object or metadata that is available online, which only makes sense if it is understood in connection with discoverability and usability.



Priorities for content acquisition

User demand for Europeana

User demands are key to develop the content available via Europeana Collections and APIs. The scope of the demand-driven Content Strategy should not be limited by the personas created for individual services or other audiences like aggregators/experts/network; it should be universal. A methodology will be developed that will assist us to identify and navigate from and towards user needs and/or new audiences. This tool will be flexible enough to give us direction on different levels and can be adapted to the needs of each specific content sourcing project.

The baseline for developing this methodology is a segmentation of all the possible sources of demand into four categories.

1. Audience we serve now

To identify the needs of both visitors and audiences we already serve we analyse our statistics (most viewed items, most visited exhibitions, etc) and/or perform a direct user research (like the bi-annual user research Europeana runs continually). The qualitative user research will help us to develop a deeper understanding of the drivers behind the statistics.

2. Audience we want to serve

These are audiences we want to connect with for strategic reasons (big opportunities, business needs, partners needs, etc). First, we need to understand why we want to reach these audiences and what the benefits are of reaching out to them. Then we need to be able to empathise with the needs of such audiences. A market research and validation process should take place to develop a deeper understanding of what those audiences are looking for.

3. Audience we might want to serve

These are the audiences that we potentially can serve, but no leads have been generated yet. Most important here is to start generating leads, for example, by creating channels and points that will attract the attention of such audiences. In addition to that, we should focus on

starting small and validate our insights quickly. Here we should focus on market research (similar to the above), user validation (instant contact with users to validate our insights) and affiliated audiences (if it is of interest to group A it might also be of interest to groups B and C).

4. Trending

Here we position timed or random opportunities that might appear and will allow us to generate further leads. Examples of such opportunities are Christmas, the Olympic Games, anniversaries, etc. This category is based on foresight and planning. With some planning and a bit of effort we can generate leads for these random or timed opportunities by already including relevant content and reach out to different audiences based on the event.

After having completed the above ideation, validation and understanding of the user demand we need to take action. Based on the specific demand's use case, the quality and quantity of content will then respectively improve and increase.

Europeana Thematic Collections

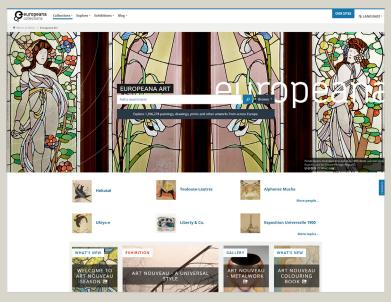
Europeana will introduce a new approach to reach end-users. The Europeana Thematic Collections are a key element of this approach. They show a filtered view of the Europeana database based on broad themes⁵ to encourage a deeper engagement with the Collections. The decision to develop a new thematic collection depends on four factors: content base, proven demand, community of partners for curating and sustaining the collection, and a strong concept.6

Information for data partners to ensure their collection records are featured in a thematic collection in Europeana can be found in the Europeana Publishing Guide. For developing virtual exhibitions and specific campaigns, more content typically needs to be actively sourced and/or existing content improved; for example, to get the content into Tiers 3 or 4 of the Europeana Publishing Framework. User demand will inform us what type of content will be actively sourced.

⁵ This can also include resource types like newspapers for example.

⁶ See the Policy for Europeana Thematic Collections: http://pro.europeana.eu/files/ Europeana Professional/Europeana Network/europeana-policy-on-thematic-collections.pdf

The concrete offer to make content more visible in Europeana or feature it in exhibitions or social media campaigns accompanying the thematic collections will be an incentive for cultural institutions to join us. The community of partners will have an important role in reaching out to institutions and support the institutions in publishing their content in Europeana. However, the scope of a Thematic Collection should always be subject related and not just cover the reach of the partners' community.



Homepage of the Europeana Art Collections.

Data quality

A digital cultural heritage collection's data quality determines whether or not a collection is accepted for publication. If the metadata are not compliant with the stipulations as set out in the Europeana Publishing Guide, a collection will not be published in Europeana.

While the <u>Europeana Publishing Guide</u> and Tier 1 of the EPF defines the minimum requirements to publish in Europeana, our ambition is to not stop

at the minimum level but to publish the highest possible quality of cultural heritage objects online. We are not aiming to increase the number of objects compliant with Tier 1 of the EPF, but to significantly increase the number of objects compliant with Tier 2, 3 and 4, to the detriment of Tier 1 compliant objects. That also means that we make every effort to improve data published in the past, and that are not compliant with Tier 1, to be compliant with Tier 2 and higher.

As technology and demands change over time, quality criteria are also changing over time. This may also have implications for the content published many years ago in Europeana. We will work with the partner that published the data to either improve them or remove them from Europeana. For up to date acceptance criteria of data to be published in Europeana and for process and implementation details for removing data, see the Europeana Publishing Guide. Our EDM documentation will also be regularly reviewed and updated to adopt stricter criteria.

To ensure continuous data quality improvements we stop publishing in Europeana for three months per year: January, April, July-August. In these months we identify quality issues (e.g. broken links), work with data partners to fix them, clean up the database and apply larger scale improvements, but we do not publish new data.

The <u>Data Quality Committee</u> (<u>DQC</u>) was established in late 2015 as a standing committee that will work on various aspects of the data quality challenge and with a particular focus on the discovery of cultural heritage objects. Formally defined as a Europeana Network and Europeana Tech working group, the DQC will work on recommendations for quality improvements connected to usage scenarios. All these recommendations will be considered for an annual data quality plan to set clear targets for data quality improvements across the Europeana database.

Type of content relevant for Europeana

Europeana is a Pan-European and cross-domain aggregator that provides seamless access to culture. Culture can be defined as "a set of artistic, philosophical, technological, and scientific works and ideas that might be seen as forming a European cultural heritage and are testimony to a

common European civilisation"⁷. But other experts in the field have established <u>much broader definitions of culture</u>, making culture a very vague concept with no clear criteria regarding what is and is not cultural heritage. Cultural heritage therefore encompasses a wide variety of themes and potential content types to be published in Europeana.

As the offer is very diverse, users take advantage of this offer and explore the diversity of what is published in Europeana. Looking at statistics for top viewed items in Europeana, famous paintings like Rembrandt's Nightwatch are equally popular for users as Dublin street directories or watermarked pictures of German football players with their family. Being inclusive consequently has both value for cultural heritage institutions and for users. Therefore, Europeana will continue to be inclusive and in principle publish all types of content that museums, libraries, archives, and audiovisual institutions based on their digital strategy consider to be important enough for publication in Europeana, while also actively looking for content that may need to be included in Thematic Collections and other channels. Also, Europeana will, based on experience, discuss with data providers if publishing certain types of content in Europeana is worth the effort. Other initiatives or platforms may have more suitable publishing channels to reach the institution's target audience. With regard to other types of content, we will work with an editorial board that we will set up to help with content that is deemed unsuitable for publication in Europeana by certain stakeholders.

Date range of content relevant to Europeana

Europeana is inclusive of the cultural heritage from centuries long past to modern times. Acquiring content from the 20th century is complicated by the fact that most works, in particular works from the latter part of the 20th century, are still protected by copyright. This leads to the so-called '20th century black hole', which we also <u>find in Europeana</u>. It is still a priority for Europeana to close this gap and make as much 20th century content accessible as possible to avoid limiting the value of our services, for example, when engaging with the education sector.

Providing free and unrestricted access to in-copyright content created within the last 70 years is a big challenge for cultural institutions. We

therefore designed the entry level to publish in Europeana (i.e. Tier 1 of the EPF) in a way that allows cultural institutions to at least make their incopyright collections discoverable in Europeana Collections.

Geographic scope of content acquisition

Our mission statement makes it clear that we want to build on Europe's rich heritage. This includes heritage that was created, located or preserved in Europe. It also includes heritage created by Europeans, heritage about Europe, Europeans or subjects of a European dimension. It is common that collections of such heritage are hosted by non-European cultural institutions, such as the Getty Open Content Images or Yale Center for British Art. Europeana has already published such collections from non-European institutions in the past and also made connections with more than 800,000 non-European sources as part of Europeana 1914-1918 (DPLA, DigitalNZ, Trove, Canadiana).

We will continue this approach and provide users with the possibility to discover Pan-European cultural heritage. However, we do require that non-European sources have a fit with one of the themes we are working on and comply with at least Tier 2 of the EPF. The Collections Manager and Advisory Board of a Thematic Collection are expected to review non-European sources and recommend whether or not it should be published in Europeana. The Business and Product Owner of Europeana Collections together with the Data Partner Services Manager will then decide upon the publication of a collection from a non-European institution.

Non-digital objects

In the context of the Europeana Content Strategy, non-digital objects are any objects for which a digital representation is not available. A discovery service of non-digital objects lies outside of Europeana's mission, whose focus is on digital objects. However, we acknowledge there are cases where non-digital objects play an important role in Europeana (e.g. in usability) due to the informative potential that non-digital objects provide to other digital objects. These cases are specified in more detail in the Europeana Publishing Guide. It is important to keep in mind that these cases are exceptional and that non-digital object aggregation will otherwise not be implemented in Europeana. Bibliographic records without digital surrogates - for example, with no explicit hasPart/isPartOf relations with other digital objects - will not be considered for publication in Europeana.

⁷ Tagiuri 157-158, Giacomo Tagiuri, 2014, Forging Identity: The EU and European Culture, Survival, vol. 56, no. 1, pp. 157-178

Access to digital objects

<u>URIs</u> are the means to access digital objects in Europeana Collections and the Europeana APIs and are included in the metadata of the cultural heritage objects. These links can point directly to the digital object or to a website or viewer where this digital object is shown in context. Embeddable links to some media types and the International Image Interoperability Framework (IIIF) are also supported by Europeana to allow access to digital objects. In any case, we only publish collections where our users get access to the digital objects without authentication, passwords or filters on geographical location.



Page 27 of Heidelberger Turnierbuch (Jost Pirckhammer 1886, Universitätsbibliothek Heidelberg, CC BY-SA) in the Europeana IIIF viewer

Masterpieces and lesser known collections for Europeana

From a user perspective, it is important that masterpieces and lesser known collections are available in Europeana. Users expect to find famous works to be in a database that aims to give access to all of Europe's culture. Cultural institutions value their masterpieces being accessible in Europeana, for example, to contribute to virtual exhibitions. Further, masterpieces help attract new institutions that can have their content appear next to world-famous cultural heritage works. However, usage

statistics show that even if masterpieces are included in digital collections they do not always, or consistently, turn out to be the most popular items of the institution. That reflects the unique selling proposition of Europeana: users are using Europeana largely as a discovery service for the long tail of cultural heritage, to find niche items that cannot be found anywhere else online. Therefore, Europeana will not give special preference to masterpieces in the content acquisition process unless there is a specific demand, for example, when specific masterpieces are important for a Thematic Collection as was the case with Europeana 280.

Crowdsourced content

Crowdsourcing is "the practice of obtaining needed services, ideas, or content by soliciting contributions from a large group of people and especially from the online community rather than from traditional employees or suppliers"8. Some well-known and trusted digital libraries and resources for digital cultural heritage have been developed by volunteers and are available via the Internet Archive, Project Gutenberg and others. They often have resources that are otherwise not available via the data partners with whom we work. For these collections to be published in Europeana, the same principles apply than for any other content: if a user demand for specific crowdsourced resources exists and the quality of these resources is sufficient, then this content is relevant for Europeana and can be published. We follow the same approach as suggested above for the publication of non-European sources: The Collections Manager and Advisory Board of a Thematic Collection are expected to review crowdsourced content and recommend whether or not it should be published in Europeana. The Business and Product Owner of Europeana Collections together with the Data Partner Services Manager will then decide upon the publication of a collection of crowdsourced content.

⁸ http://www.merriam-webster.com/dictionary/crowdsourcing



Metadata and content from Wikimedia Foundation projects

Integration of metadata and content from Wikimedia Foundation projects into Europeana products can help improve user experience of these products. Europeana can use a source like Wikidata to annotate existing Europeana objects. These annotations will be clearly marked as coming from Wikidata. Europeana will not publish and display, but may link to, digital files from Wikimedia Commons or Wikisource that have been uploaded from cultural heritage institutions' websites (or scanned from publications) by individual Wikimedians without consent from the institutions in question. Europeana can publish digital files from Wikimedia Commons media that have been crowdsourced in initiatives like Wiki Loves Monuments. This will be done with the support of the Wikimedia Commons community.

Active (pull) and passive (push) content acquisition

A lot of content currently published in Europeana is selected based on what is available in cultural heritage institutions and pushed to Europeana for publication. This is an important element of the content acquisition process, but we will, in future, more actively influence what content gets published in Europeana. In addition to the push approach, we will adopt a more targeted content acquisition process to pull in new collections. This active content acquisition process is driven by user demands, which helps when engaging with our data partners to identify the right content for publishing in Europeana. In these cases Europeana will approach potential data partners either directly or through the appropriate aggregator to explore how to make this content available through Europeana.

Priorities for publishing in Europeana

Looking at what is published in Europeana to date, what is digitised but not published in Europeana and what can be digitised in the coming years (see Enumerate Core Survey 3), it is clear that Europeana cannot deal with the amount of data in a meaningful way and publish everything quickly and in high quality. Instead of becoming exclusive and reject collections based on the type of content, it is more important to focus on quality, user demands and a thematic fit to prioritise collections during the ingestion and publication process. Metadata and content quality, but also user demands and business priorities, will define which collections receive more attention during the publication process (see Europeana Publishing Guide for more details).

Dissemination of content

Europeana provides access to content to fulfill user demands. It is continuously investing in dissemination and distribution of content to make sure that users know which content is available and to stimulate more users to engage with the content. The Europeana Strategy and the strategic refinements following the Council Conclusions are clear on our direction to distribute content. Focussing on specific markets and audiences with specific products and services will help us make Europeana better known. We also offer trainings and workshops for our audiences to enable them to work with our products more efficiently and get access to the content they want and need. The needs of our audiences will also be considered in our product design and development work, to make our content work for users who work on the technical platforms. In addition to the work on the front end, the backend also needs to allow users to find the content they expect to find. We will respond to that with an improved search architecture.

Conclusion

Implementing the Europeana Content Strategy should lead to a significant improvement of the Europeana database in the coming years. By the end of 2016, about 83% of the content in Europeana complies with Tier 1 of the EPF, and approximately 17% only complies with Tier 2 and higher. This makes it very difficult to fulfil user demands and provide a compelling user experience. Following user demands and setting up more thematic collections should help improve content accordingly. The more thematic collections will be realised, the more high-quality content will be needed. Therefore, we expect the relation between content in Tier 1 of the EPF and content in Tier 2 and higher to flip by 2020; so by 2020 we aim to have only 20% Tier 1 content and 80% Tier 2+ content.

24





Europeana DSI is co-financed by the European Union's Connecting Europe Facility

The sole responsibility of this publication lies with the author. The European Union is not responsible for any use that may be made of the information contained therein.