



Content Quality Plan

Main Authors:

Mina Karatza, Vassiliki Psatha, Orfeas-Menis Mastromichalakis, Kostas Konstantinidis, POSTSCRIPTUM

Erik Buelinckx, KIK-IRPA

Valentina Bachi, Sofie Taes, PHOTOCONSORTIUM

Table of content:

1. Introduction	3
2. Content quality plan from digitization to reuse	4
3. The Content Quality Scheme	13
4. Events and knowledge sharing	14
5. Conclusion	16

1. Introduction

1.1. Project and Content Quality Plan objectives

The core concept of PAGODE is to contribute to a rich user experience and heightened audience engagement with Europeana and to propose a holistic approach in the aggregation, curation and presentation of Chinese cultural heritage preserved in Europe.

PAGODE aims at promoting cultural diversity giving a wide range of audiences' access to Chinese cultural heritage, raising awareness of the richness of collections preserved in Europe, improving knowledge of cross-cultural exchanges, encouraging creative reuse of open-licensed materials, and facilitating the work of educators, researchers and curators to lever on the multi-faceted culture and history of Europe and its multiculturalism. In order to arrive at creating new cross-border perspectives, PAGODE proposes to connect existing digital historical resources and to generate new ones through the publication of additional digital collections, improving on existing collections' metadata and creating opportunities for the participation and engagement of the public.

The content quality plan aims to support these objectives and will be available for the participating content providers as a point of reference, ensuring that the **quality requirements for publication in Europeana are met** for both digital objects and metadata. Furthermore, this document is offering **guidance and support** with regards to **digitization, documentation** and interoperability standards, semantic **enrichment, aggregation** and **digital curation**. In a broader sense, this document hopes to operate as a guide for CHIs interested in contributing to the PAGODE action, or are involved in similar initiatives in the realm of digital cultural heritage.

1.2 PAGODE: a content-oriented project

As PAGODE is a project that, in the end, is directed towards user engagement, from the very early stage it became clear that the main driver for and connector between all different project aspects was content. Users are engaged, more active and involved if content is easy to retrieve, high in quality, and relatable via compelling storytelling. In all of these cases, content is a crucial factor: rich and semantically significant metadata make objects more easily findable and explorable; quality is a matter of data and metadata, therefore adopting standards for both aspects of content is important; and the selection of the right objects for storytelling and other dissemination purposes - including participatory activities - is key.

With content at the core of all activities - from semantics and curation to digitization, aggregation, ingestion, enrichment and dissemination - PAGODE established a subgroup with representatives of partners involved in the above-mentioned actions and content providing partners in particular. Together, this group worked towards reaching the qualitative and quantitative targets to be met:

- digitization (10,000 new items) and ingestion to Europeana (Activity 3)
- enrichment (20,000 items) in Europeana automatically enriched via NLP (Activity 4)
- annotation activities (2,000 items) enriched via crowdsourcing (Activity 2)

As all of these activities unfold throughout the project and involve content in different stages of development and evaluation, the workflows connected to content also span the entire project period. Moreover: they do not run in a continuous line from A to Z, but in a dynamic, constant dialogue between content providers and the project Activity leaders.

Such agility in interaction is an important aspect of content quality control, as it allows for different aspects of 'content' (records, metadata, enriched metadata, editorial output) to be tuned to each other and to the project's end goals.

2. Content quality plan from digitization to reuse

2.1 Content selection

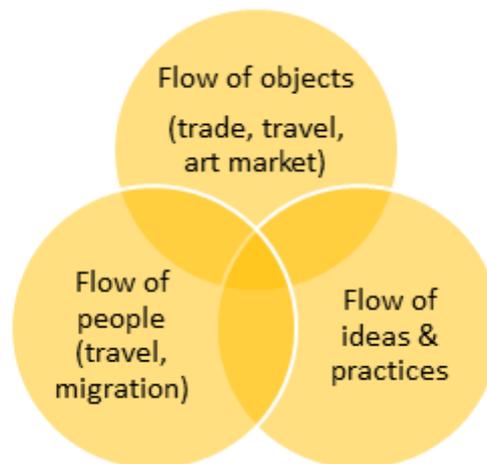
Content selection entails the coordination of the selection of content for digitization and/or publication and/or enrichment by [partners](#) and [associate partners](#). The aim is to ensure that the content - both in terms of data and metadata - offered to Europeana is interesting, original, consistent in quality, representative for the providers featured and in support of the curatorial direction set out by the semantic and editorial teams. Because of the collaborative approach, compliance criteria are followed to ensure that through the diversity of materials, topics and organizational standards, a unified contribution to Europeana can be ensured.

The selection was finalized and two streams of content were in the loop: records to be digitized and existing digital records. While the first stream entails digitization activities both streams of content entail metadata creation inspired by the semantic framework as a curatorial aspect

2.2 Development of semantic framework

The semantic background for the curation of China-related records in Europeana has been developed in several steps. First, the definition or the conceptual scheme of what constitutes Chinese CH in Europe was formulated. Secondly, based on the classification of various types of objects in that scheme, metadata specifications in the form of keyword lists were created.

The [semantic framework](#) aimed at defining what can be considered as Chinese and China-related heritage in Europe, was both inspired by and directive towards the content selection process. The semantic team at the Department of Asian Studies of Ljubljana University explored existing Europeana collections and envisaged partner contributions, before applying theoretical models from the fields of anthropology, social history, art history, linguistic and cultural studies to the materials under investigation. This resulted in a rationale describing the connections between China and Europe in terms of flows: the flow of people, of ideas and of objects. Subsequently selection activities were continued and finalized with the aim of representing the three flows.



The three flows of interactions and exchange between China and Europe

The conceptual scheme revealed the great diversity in types of objects that were included as CCH in Europe. The metadata standards for particular types of objects may vary considerably as well as the metadata practices of the various content providers. All this complicates the task of creating a standard scheme for the metadata curation of the CCH in Europeana. For these reasons it was decided to create different keyword lists for different purposes/users – a long keyword list for automatic enrichment and several shorter lists for manual annotations/mapping.

2.3 Documentation guidelines to prepare for metadata creation

Already in the content selection phase, content holders screen their raw (often analogue) archival assets to pick the most interesting items for publication online. Possibly the items come with existing documentation, or need some research work to create, or check, the descriptive information connected to each item. The documentation accompanying the cultural heritage items needs to be “translated” in meaningful and correct metadata.

In order to guarantee the high quality of metadata provided by PAGODE project, it is necessary to ensure that the above-mentioned semantic framework is reflected in both existing digital collections and the newly digitized content. The role of content providers is crucial in that phase and their content should be described with appropriate and rich keywords as defined in the semantic framework. These keywords serve as a support to content providers in the creation of the metadata for the newly digitized content; the lists are used for the annotation pilot and for the automated enrichment activity.

During the documentation of the content, it is recommended that content providers take into account the requirements of the Europeana Publishing Framework in the light of matching with the highest tier possible, ideally, meeting tier C.

As part of project’s requirement, it is mandatory that the newly digitized content is provided under a rights statement that allows for free reuse ([CC BY](#), [CC BY-SA](#), [CC0](#), [PDM](#)), meaning that the content providers should include in the documentation of their collections (items) the licenses under which the digital content is shared.

For more details about the requirements of the Europeana Publishing Framework see the document [“Meeting the Europeana Publishing Framework: summary about tier 4 and tier C”](#)

Additionally, in order to provide high quality metadata several enrichment processes were ran on the datasets of content providers, in order to embed links to authority files and vocabularies, such as AAT and Wikidata, when missing. Manual curation of the datasets took place via crowdsourcing in the PAGODE Annotation Pilot, and the collected annotations were ingested to Europeana and displayed in each record as additional information in the dedicated section “Keywords from the Community”. All these actions are described in detail in the following sections [2.5.1 Crowdsourcing annotation campaigns on records already published in Europeana via Annotation API](#) and [2.5.2 Automatic annotations on records already published in Europeana via Annotation API](#).

2.4 Digitization guidelines

According to Task 3, high quality digitization of 10,000 new objects relevant to Chinese CH hosted in collections in Europe was implemented. The new content digitized (in order to be offered for publication in Europeana) is compliant with the requirements of Europeana Publishing Framework. The Europeana Publishing Framework considers 4 levels (“tiers”) of content quality (1 to 4), to which correspond different levels of user experience in the Europeana portal. The contractual requirement in PAGODE project is that the new content digitized comply with Tier 4 of the Europeana Publishing Framework. In terms of technical requirements, Tier 4 of the Europeana Publishing Framework determines that the digital objects should be provided in as high a technical quality as possible (with a minimum of 0.95 megapixel in size). For more details see the document [“Meeting the Europeana Publishing Framework: summary about tier 4 and tier C”](#).

There are no real digitization standards yet for use in the CH field, but some best practices. The main, but not only, ones are [FADGI](#) (Federal Agencies Digitization Guidelines Initiative, USA) and the Dutch [Metamorfoze](#). Both work with a system from lower to high quality, and especially FADGI pays also attention to environment, machines, conditions.

In order to ensure the sustainability of the project outcomes, the digital content should be preserved and several actions of digital preservation should be taken by the content providers, such as protecting the integrity and identity of data and ensuring the accessibility to them (i.e., maintain the digital content available online and after the end of the funding period). The importance of being able to rely on file format standards was studied in the European project [PREFORMA](#) (PREservation FORMAts for culture information/e-archives). Some tools were created in this and other projects for checking the digitally created output. UNESCO, as an international organization working also for the safeguarding of our heritage, created some years ago the [Digital Preservation Programme](#). It’s recommended to organizations to get familiar with these existing best practices and projects to make the right decisions before starting digitizing their collections.

2.5 A short guide to EDM mandatory elements

There are mandatory elements in EDM that must be present for the metadata to be processed further for publication. The list below is an up-to-date summary of these mandatory elements. For more details

and the full guidelines, see the EDM reference [documents](#). The EDM documentation incorporates the other metadata elements recommended to create a rich and full description of your objects.

1. Each metadata record must contain either a title (dc:title) or a description (dc:description). The values in these fields need to be unique and meaningful across the submitted dataset. All metadata records in the dataset cannot be given identical titles or descriptions that do not specify the cultural heritage object sufficiently because your object would consequently not be discoverable.
2. Each metadata record describing a text object (e.g. book, manuscript, letter) must provide the language of the document in the metadata (dc:language). In other cases where languages are represented (some audio or video objects or photographs of inscriptions for example) we recommend that you include the language in the metadata. This allows us to present more material in the user's own language.
3. Each metadata record must contain the type of digital object (edm:WebResource) specified in edm:type. This metadata field can only be populated with one of the following five fixed values: TEXT, IMAGE, SOUND, VIDEO, and 3D. Many users want to search according to media and accurate completion of this field allows them to do so.
4. Each metadata record must provide some context and details about the objects described by the metadata. This additional information can be either the subject of the cultural heritage object (dc:subject), its nature or genre (dc:type), the location or place the object depicts (dcterms:spatial), or the temporal topic of the object (dcterms:temporal). The more data you can provide here, the more opportunity the users have to find it using their search terms.
5. In the case of content collected during collection days or crowdsourcing activities, each metadata record describing digital objects contributed by users (e.g. during the crowdsourcing campaigns) must show edm:ugc = true. This allows us to maintain Europeana Collections' reputation as an authoritative resource, conferred by the work of the cultural heritage institutions, separating user knowledge and curated or professional knowledge.
6. Each metadata record must contain the information of the cultural heritage institution that provides the data to an aggregator (edm:dataProvider). Users want to know where the data comes from, and we want to attribute properly.
7. Each metadata record must contain information about the data partner to Europeana (edm:provider). The value will be identical to the data provider information if the cultural heritage institution that owns the object also provides the digital representation of the object to Europeana. If the cultural heritage institution collaborates with an aggregator in order to deliver the data to Europeana, the data partner to Europeana is the aggregator.
8. Each metadata record must provide at least one link (URL) to the digital object either in the context of a cultural heritage institution or an aggregator's website or as a direct link. It is strongly recommended to provide a link to a web view of the digital object, which is a direct link to the actual file that is ready for download (e.g. jpg, mp3, pdf in edm:isShownBy). The data partner can also assign a link to the digital object in the context of the organisation's website or a link to a (book or image) viewer that shows the digital object (edm:isShownAt). We highly recommend that you provide both edm:isShownAt and edm:isShownBy along with the data, so that the user is more likely to use your data.

9. Each metadata record must contain a valid rights statement for the object using edm:rights and the corresponding URI to the rights statement. The list of valid rights statements is published on [Europeana Pro](#). A valid rights statement tells the user what they may or may not do with your digital objects. Rights statements are also machine-readable, which makes them fully searchable via the API.
10. Each resource (edm:ProvidedCHO, edm:WebResource, ore:Aggregation and each contextual entity) described in a metadata record must contain a unique and persistent identifier (e.g. rdf:about of the ProvidedCHO class in EDM and edm:aggregatedCHO) that will be used to generate the permalink to the record on Europeana Collections as well as the Europeana identifier for both Europeana Collections and the APIs. The persistence of this identifier will guarantee that the links to each object remain when the metadata record is updated, which, again, will improve user experience and likelihood of your material being used.
11. Each metadata record must be submitted in UTF-8-character encoding to ensure a legible display of the data in both Europeana Collections and the APIs. Machine-readable encodings like HTML tags are not allowed in the metadata.

2.6 Metadata Enrichment

Metadata is the information which is associated with a digital object and which provides various types of descriptions. The information is created by the curator/content holder, according to their own workflow and practices. The metadata for the items digitized in PAGODE are therefore created from the scratch, also basing on the semantic framework developed in the project.

Of course, the level of richness for the information included by the content holder may vary, and, since the early days of digitization, a growing importance has been given to the matter: nowadays it is fundamental to provide the digital object with accurate and rich information, for a more efficient retrieval of the object by user search and for establishing connections with similar objects. For this reason, enriching metadata has become a way to overcome data quality issues of records that are already published since a long time, while providing more context and multilingual information. Metadata enrichment has become part of Europeana and its data providers' strategy. Understanding the key aspects of the process has become necessary to provide best practices and guidelines.

Generally, a metadata enrichment task can be described as a process that improves metadata about an object by adding more information about the object that this metadata describes. The term 'enrichment' can be used to refer to the process - e.g., the application of an enrichment tool - or its result - the new metadata created at the end of the process. Additionally, enrichment is often supported through the implementation of thesauri or authority files that provide a trusted source for additional information. In the case of PAGODE, the keywords list produced for semantic enrichment (cfr. 2.2) are used to complement other existing vocabularies such as those about photography and about fashion, developed in previous EU-funded projects and now reused for PAGODE project.

Enrichments can be created manually, semi-automatically or automatically (e.g., by means of information extraction). Annotations can be one form of automatic enrichments.

In the context of PAGODE, enrichments are created in a threefold way:

- collected by users via the crowdsourcing campaigns,
- automatically generated with AI and NLP tools,
- and in two specific cases implemented via the National Aggregators which support metadata enrichment as a regular part of their ingestion pipeline; this applies for Museovirasto - Finnish Heritage Agency, aggregating content via FINNA, and Benaki Museum aggregating content via EKT.

2.6.1 Crowdsourcing annotation campaigns

The core concept of crowdsourcing activities in PAGODE is to engage a community of users who review selected collections from Europeana and add to each record more information selected from appropriate keywords lists. This form of user curation allows people to create new data about the objects, based on trusted sources, engaging them with digital cultural content in a participatory and compelling way.

The crowdsourcing activities are part of the PAGODE Annotation pilot, through which the campaigns are deployed. These activities make use of the WITHCrowd environment, an open-source platform that integrates and supports the Europeana APIs, and specifically the campaign is implemented in the [CrowdHeritage website](#).

The PAGODE campaign is composed of three phases, two of them with a thematic focus and a third one, compilation of collections:

- Scenes and people of China, including collections of heritage photographs (Autumn 2020)
- Chinese Artifacts (Spring 2021), including museal collections of cultural heritage objects.
- Summer Annotation sprint (Summer 2021), on China-related heritage collections recently added to Europeana

The campaign is available at the address <https://crowdheritage.eu/en/china>.

Participants look at each digital resource (photograph) and enrich the information by adding tags from a dropdown list of keywords; they can also review the tags added by others up-voting or down-voting tags, and thus advising the list of keywords associated with the cultural records. Participation is encouraged by a challenge format, with a leaderboard displaying the most active contributors.

The new tags (= additional metadata enriching the digital object from this manual curation) generated via the Annotation Pilot are checked and validated by the PAGODE Content Team to ensure their quality, relevance and accuracy, and they are then sent to Europeana via the Annotation API. Representatives of Europeana Foundation confirmed that the annotation provided via the Annotation API will be displayed in Europeana as 'Keywords from the community' in association to the source metadata originally provided by the content holder.

2.6.2 Automatic annotations on records already published in Europeana

There are various techniques employed and, in this section, only the main points are provided below.

Automated enrichment by Artificial Intelligence (AI) and Natural Language Processing (NLP) on records already published in Europeana is the third implementation scenario for metadata curation. A

combination of AI and NLP allows to recognize automatically information that is embedded in the content and in the metadata, and to add relevant terms and links to authority files in the existing metadata. This has the advantage of enabling bulk enrichment of large datasets, which would require a huge effort and extended time scale if done manually. In the scope of PAGODE, automated enrichment is performed by use of algorithms trained with the list of keywords developed in Activity 3 and described in chapter 2, running on metadata from a selected set from Europeana collections. The algorithms recognize the terms and enrich the metadata with the link to the respective entries in Getty AAT or Wikidata, and the enrichments are made available to Europeana, via the Annotation API, for display in the respective records.

Enrichment with smart algorithms for string matching as well as Named Entity Recognition and Disambiguation (NERD) were applied on the records' metadata. The NERD algorithm extracts text spans that refer to named entities:

- Persons
- Locations
- Organizations

Smart Named Entity recognition tools as well as smart string-matching tools (using the keyword list provided by the University of Ljubljana) are applied on more than 25.000 Europeana records and create enrichments with the respective entity URIs. The following NERD tools are used:

GEEK is a named entity recognition and disambiguation tool that links text to external knowledge sources, like Wikipedia. It works in two steps:

- It extracts text spans that refer to named entities, such as persons, locations, and organizations.
- It jointly disambiguates these named entities, by generating sets of candidate entities from external knowledge bases, and then iteratively eliminating the least likely ones, until we are left with the most likely mapping of textual mentions to their corresponding knowledge base entities.

In the context of PAGODE, where the content is culture-oriented, GEEK may still be used to extract such entities like curators of pieces of art, significant locations, monuments, and so forth. Considering the sentence: "Leonardo painted Gioconda in the early 1500s, and now the painting is stored in the Louvre." Leonardo needs to be linked to the polymath, and not the famous Hollywood actor. Gioconda needs to be linked to Mona Lisa, and not Ponchielli's opera. Louvre needs to be recognized as the French museum. The aim of GEEK is to eliminate such candidate entities that don't appear to fit with the context of the given text.

[AIDA](#) is a framework and online tool for entity detection and disambiguation. Given a natural-language text or a Web table, it maps mentions of ambiguous names onto canonical entities (e.g., individual people or places) registered in the [YAGO2 knowledge base](#).

Enrichment tools and enrichment workflows are validated during their development and employment to ensure they are suitably adjusted.

The enrichments as automatically generated in PAGODE are forwarded to Europeana for publication via the Annotation API.

2.6.3 Annotations via national aggregators (Finland, Greece)

Annotations via National Aggregators happen in PAGODE as far as two associate partners are concerned, Museovirasto in Finland and Benaki Museum in Greece. In both cases, the content provider sends their collections to Europeana via the National Aggregator, for which the enrichment is part of the regular ingestion pipeline.

[Finland National Aggregator \(FINNA\)](#) is integrated with several external systems which further expand the features of the search service. This integration enables the enrichment of metadata through the Finto ontology service as well as searches in indexes that are outside Finna.

[Greece's National Documentation Centre \(EKT\)](#) has developed a semantic enrichment and homogenization scheme. It is based on www.semantics.gr, a platform for publishing vocabularies and thesauri that includes a mapping tool for content enrichment and contextualization. The tool sets mapping rules from multiple metadata field values to vocabulary terms. It uses a self-improving automatic suggestion mechanism and additionally supports the curator when intervening in the enrichment process. It has also extended the aggregator infrastructure with a parametric tool for the normalization of chronological values which is based on regular expression processing. As a result the content of searchculture.gr was enriched and homogenized in respect to types, chronologies and historical periods, allowing to publish it as Linked Open Data and to enhance the portal with new multilingual search and navigation features. The semantic enrichment strategy of EKT is extensively described in the following [article](#).

2.7 Ingestion of content to Europeana and mapping process for publication into Europeana

The ingestion workflow is a procedure that, starting from the source database of a content provider, enables the metadata associated to the digital objects, which are organized according to the content provider's structure and schema, to be mapped to the Europeana Data Model, and thus collected by the Europeana engine for publication in the portal. The process therefore includes three actors: the content providers who offer their dataset, the aggregator who by their mapping infrastructure allows the dataset to be "converted" to the Europeana Data Format, and the Europeana Data Processing Team to harvest and publish the metadata in the Europeana portal. In the framework of PAGODE, it is Photoconsortium as the Europeana accredited aggregator for photography to coordinate the entire process by liaising with the content provider and with Europeana.

Different scenarios are envisaged for the aggregation and ingestion of content from PAGODE project to Europeana portal:

- In the case of PAGODE's content partners United Archives, and KIK-IRPA, the operations will be done by using Photoconsortium's aggregation tools (namely the MINT software), to perform mapping of the source metadata to EDM, and then to offer the records via OAI-PMH server for Europeana data processing team to harvest them for further publication in the Europeana portal. Also, some of the associated partners, such as e.g. The Slovene Ethnographic Museum, Promoter Digital Gallery, KADOC, ICIMSS use the same MINT software, for which Photoconsortium will provide assistance.

- In other cases, the conversion to the EDM is done through different routes and tools, according to the practice and also to the national regulations in the countries of the content providers. The Greek associate partners, for example, the Benaki Museum, maps and ingests the metadata with the National Documentation Centre (EKT), the accredited National Cultural Data Aggregator for Europeana. Museovirasto (Finland) makes use of FINNA, the Finnish National Aggregator. Leiden University Libraries and Wereldculturen Museum make use of RCE/NISV, the Dutch National Aggregator. For these three cases, the original metadata offered by the content provider are also enriched “along the way” by the National Aggregators (cfr. 2.6.3).

While Photoconsortium maintains its overall role of coordination, the involved National Aggregators provide the ingestion services to the national content providers and are responsible for the successful and timely delivery of the metadata to Europeana.

2.7.1 Quality monitoring for ingestion phase

According to the project’s timeline, the ingestion started in May 2021, corresponding to Milestone 6 “Kick-off of the ingestion stage”. Photoconsortium is in charge of supporting all content providers in order to get to the ingestion phase, and to collaborate with the Europeana Data Processing Team ensuring a successful publication in the Europeana portal. Making tests and checks on the actual metadata and the various intermediate steps of the workflow is the best way to ensure that all the quality requirements are met and that the final result in Europeana is a rich, reusable and interesting digital content.

In a linear overview for the entire process, there are phases to come first and others to follow in order for content providers to be ready to start the mapping and ingestion process, these including content selection, digitization and metadata creation, and publication of the resources in the content provider’s website.

These phases need to happen ahead the dataset is ready for initiating the mapping to Europeana, but planning and discussion was initiated in PAGODE since the very early stage of the project, also involving the Europeana Data Processing Team with tests and dedicated discussions. The planning is needed to ensure that once the mapping is done and the records are available for Europeana on a OAI-PMH server the processing at Europeana side goes smooth, with publication of the records on Europeana portal as a final outcome.

Various steps of check are foreseen at different levels, i.e. in the content provider’s website, in the aggregator’s mapping infrastructure and in Europeana’s harvesting and preview infrastructure, before publication can happen in Europeana Portal: the checks are done both in terms of technical functionality and correctness of metadata, and in terms of proper and nice visualization as preview for Europeana portal. Feedback is provided for possible corrections; after this entire process of iterative check and refinements, that implies both automated checks, manual checks and corrections if needed, the dataset is published on Europeana Portal, and finally becomes visible for the public and available for reuse in editorials.

The [full detail of the ingestion process and the different quality checks](#) performed in PAGODE is published in the project website.

2.8 Reusing records: digital curation

Europeana currently gives access to over 50 million records of any kind: it is therefore crucial to help users navigate this immense number of resources. For this reason, the concept of digital curation of the Europeana collections helps making the different collections more visible and retrievable by interested users. Being PAGODE a content and user-oriented project, the digital curation of both the new content supplied by project partners and of the existing China-related content in Europeana offers new perspectives for telling stories that are appealing to the users, by shedding a different light on digital collections in a secure, trusted and open online environment. The editorial strategy developed in the content selection, aggregation and ingestion stages was executed in the form of editorials (blogs, galleries, exhibition, featured datasets) in and beyond Europeana. In this way the thematic environment dedicated to Chinese heritage in Europeana is established in such a way that content is easily retrievable, well-organized, compelling, engaging and inclusive. Curatorial considerations for dissemination purposes mostly concerned: the adequate representation of the three flows established in the semantic framework; a good visibility for partner and associate partner content contributed in the scope of PAGODE; and the connection of that content with existing Chinese and China-related collections in Europeana. In that way, new networks of references and meaning are generated, and the inclusivity and impact of the storytelling get maximized.

3. The Content Quality Scheme

The following graph illustrates the process and the workflow connected to content in the scope of PAGODE, from selection to dissemination and reuse. It visualizes the steps and actions starting from the left pillar blocks. At this stage curation and the development of a semantic framework happened in parallel and with different iterations, at the very beginning of the project.

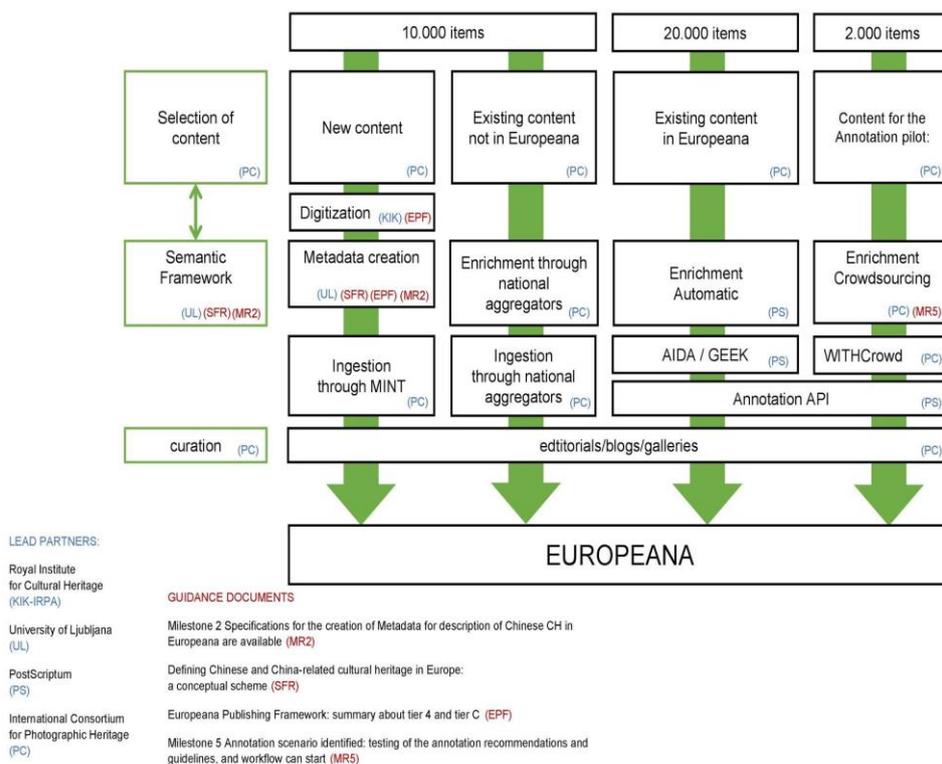
At that point, two streams of assets came into view: existing digitized collections and yet to be digitized content. In both cases, work on metadata ensued making sure that again the semantic framework was reflected.

As different aggregators were involved in working with different partners and associate partners, a central coordination of aggregation processes is of primal concern. Using the semantic framework, the curatorial/editorial strategy, as well as the Europeana and PAGODE-specific requirements as guiding principles, different workflows and standards were channelled as such that the content delivered to Europeana reflects a common approach.

In parallel, enrichments are created in a threefold way (cfr.2.6).

The editorial strategy developed in the content selection, aggregation and ingestion stages is executed in the form of editorials (blogs, galleries, exhibition, featured datasets) in and beyond Europeana.

Finally, all materials gathered in the scope of PAGODE is centralized in a dedicated thematic space within Europeana, where users can explore and directly access the content and editorials produced throughout its lifecycle.



The PAGODE Content quality scheme

4. Events and knowledge sharing

The events described below aimed at facilitating interaction and the sharing of knowledge experiences with our partners, associate partners and with the general public.

4.1 At the semantic workshop in Ljubljana (online)

Due to the Covid-19 pandemic, the semantic (metadata) workshop that was scheduled for July 2020 in Ljubljana was moved online. It was held as a two-half-day-event on 9-10 July 2020. The invitation to participate was widely circulated and a total of 58 participants registered for the workshop and participated in its activity. The semantic background developed by the PAGODE team at the University of Ljubljana was presented, discussed, and agreed upon. While presentations were only allowed for the registered speakers, discussions were open to all.

After the welcome messages of the hosting partner and of the Technical Coordinator, a group of 18 presenters included an academic keynote speaker, a member of the Collections Engagement team at Europeana, representatives of the PAGODE partners, various content owners, including associate partners and potential associate partners of PAGODE, and Chinese partner institutions.

The main outcome of the workshop was the specifications for the creation of metadata for description of Chinese CH in Europeana were made available to the content providers in the form of the conceptual scheme and the keyword lists, which was also shared widely for dissemination and uptake from other stakeholders in the Europeana family and beyond. The fine tuning of the metadata specifications for the newly digitised content continued in collaboration with partners and associate content providers, based on their specific needs.

The workshop summary with the programme and all the presentations in downloadable format can be found [here](#).

4.2 At the technical workshop in Athens (online)

Similarly, to the semantic workshop, the technical workshop in Athens, initially scheduled to take place in Athens in October 2020 was organized virtually. Having said that, the technical scope of the meeting as indicated in the Grant Agreement was expanded, in order to create an engaging capacity building seminar that included a range of interesting presentations from selected speakers and hands-on activities. The online event took place on October 26th 2020 via the ZOOM platform. The workshop was organized not only for the project partners but also for content holders outside the consortium, namely the developers of MINT and the national aggregators in Greece and Finland. A total of 50 attendees registered to the Workshop, representing governmental organizations, European museums, libraries, universities, archives, Europeana aggregators, SME's, as well as Chinese Cultural Heritage Institutions. As foreseen originally, participants had the opportunity to discuss and share knowledge on metadata curation, strategies for the use of vocabularies and thesauri, semantic enrichment by natural language processing and entity extraction methods. As special guests of the event, representatives of the National Library of China and of the Palace Museum in Beijing gave an interesting overview of the status of online access to cultural heritage digital resources in China. The workshop was also an occasion to discuss and review the technical specification of the metadata for the description of Chinese Cultural Heritage (CH) content and the overall Content Quality Plan. The workshop presenters included experts from the MINT aggregation tool (that is used by the content providers that deliver their content through Photoconsortium aggregator) and from the Greek National Aggregator (that is used by the Greek associate content provider Benaki Museum). Both interventions discussed different enrichment strategies, illustrating the use of controlled vocabularies is connected with the general scope of increasing the discoverability of Chinese CH linked contents. Particular attention was given to illustrate the Europeana requirements, with a focus on the Europeana Semantic Enrichment Framework and on illustrating the Europeana Publishing Framework to content providers who are new contributors to Europeana. The Europeana data management team was also involved in the organization of the event and the communication team at Europeana contributed to the dissemination of the event.

Additionally, the event was a perfect occasion for enabling an 'annotation sprint' to progress with the crowdsourcing annotation campaign launched in the frame of the Annotation Pilot (T2.2). This session allowed to already get nearly 4,800 annotations on the 707 records which were selected for the theme 'Scenes and People from China'. From the group of participants in the workshop, 18 people actively joined the 'annotation sprint', allowing to 100% complete the target (that was conservatively set to 4,000 expected annotations). All this was achieved in less than 2 hours, demonstrating the interest on the subject matter and the effectiveness of the chosen tool. The 'annotation sprint' for this current

campaign was therefore a big success. The Workshop was followed by a ‘private meeting’ where content providers from PAGODE (both beneficiaries and associate partners) were able to address specific questions to the experts of Europeana Data Processing team.

The workshop summary with the programme and all the presentations in downloadable format can be found [here](#).

5. Conclusion

As discussed, the content quality plan aims to support PAGODE objectives and to be a point of reference of all relevant publications and ensuring that the quality requirements for publication in Europeana are met for both digital objects and metadata. Furthermore, this document is offering guidance and support with regards to digitization, documentation and interoperability standards, semantic enrichment, aggregation and digital curation. In a broader sense, this document hopes to operate as a best practice guide for CHIs interested in contributing to the PAGODE action, or are involved in similar initiatives in the realm of digital cultural heritage.

In the scope of PAGODE the challenge was to gather all materials in a dedicated thematic space within Europeana, where users can explore and directly access the content and editorials produced throughout its lifecycle. This was not an easy task considering the diverse sources of content, spanning from the wide range of content providers (partners and associate partners) to the existing content already published in Europeana. It is acknowledged that this thematic hub will improve user engagement with Chinese heritage in Europeana from the sides of the general public, the education sector, researchers, CHIs, and the Europeana ecosystem alike.

The digital content created by digitization of physical originals (such as photographs and negatives, books and physical objects) met the so-called tier 4 of the Europeana Publishing Framework. This entails content quality criteria to be met already in the digitization phase. Quality monitoring was also met in ingestion phase and was initiated since the very early stage of the project involving the Europeana data processing team. The planning was needed to ensure that once the mapping is done and the records are available for Europeana the processing at Europeana side goes smooth, with publication of the records on Europeana portal as a final outcome.

The PAGODE workflow required to put efforts into helping the data providers and aggregators to reach a higher level of data quality. The first step is to understand what good data quality actually means, and this has been explored in this plan by examining all data standards and metadata mandatory elements. The next step was to then implement these recommendations in practice, and support PAGODE ecosystem to do so too. This is not an easy task but the consortium is well positioned to liaise with multiple parties and share different efforts happening among partners and associate partners. As Europeana constant efforts to give a better user experience and to better surface the content of partners, PAGODE consortium has worked well on curation of new and existing datasets by producing structured scientific knowledge and guidelines on what can be considered as Chinese CH in Europe. For this reason,

consulting the Semantic framework that was developed becomes a sine qua non condition in order for quality to be ensured.

In the realm of cultural heritage collections, Digital Curation provides new perspectives on, connections between and meanings of the digital materials. In that sense, curation involves actions which entail not only documentation, but also establishing the connections of digital assets with notions of time, space, provenance/ownership, materiality/format and imaging. By enhancing digital objects with such information and encapsulating that network of meanings in a narrative through engaging storytelling, Digital Curation is of vital importance when it comes to creating awareness and improving on the visibility of online heritage collections.

During the project lifecycle PAGODE guidelines and reports were produced and published [online](#). Next to this, public actions such as the Semantic Workshop in July 2020 and the Technical Workshop in October 2020, represent an occasion of knowledge transfer and capacity building, offering resources that are made available to many other stakeholders from all levels of the cultural sector.

The present Content Quality Plan is a reference document available to the content providers and it is recommended to be disseminated widely through the PAGODE communication channels, for knowledge transfer and uptake from other stakeholders in the Europeana family and beyond. The Content Quality plan is also the 'meeting point' for increasing understanding on technical issues connected with ingestion to Europeana, quality improvement and semantic enrichment.

The thematic focus on Chinese cultural heritage preserved in Europe, the semantic framework elaborated and developed and the exchange knowledge and expertise with primary Chinese institutions ensures not only the achievement of the primary scope of the project that is enriching Europeana but a broader aim is accomplished; Collaborations among European museums and archives is enabled and international relations with Chinese cultural heritage institutions are being fostered.