

Article

CrowdHeritage: Crowdsourcing for Improving the Quality of Cultural Heritage Metadata [†]

Eirini Kaldeli *, Orfeas Menis-Mastromichalakis, Spyros Bekiaris, Maria Ralli, Vassilis Tzouvaras and Giorgos Stamou

Artificial Intelligence and Learning Systems Laboratory, School of Electrical and Computer Engineering, National Technical University of Athens, 15780 Athens, Greece; menorf@ails.ece.ntua.gr (O.M.-M.); sbekiaris@islab.ntua.gr (S.B.); mariaral@image.ece.ntua.gr (M.R.); tzouvaras@image.ntua.gr (V.T.); gstam@cs.ntua.gr (G.S.)

* Correspondence: ekaldeli@image.ece.ntua.gr

[†] This paper is an extended version of our paper published in SMAP 2020.

Abstract: The lack of granular and rich descriptive metadata highly affects the discoverability and usability of cultural heritage collections aggregated and served through digital platforms, such as Europeana, thus compromising the user experience. In this context, metadata enrichment services through automated analysis and feature extraction along with crowdsourcing annotation services can offer a great opportunity for improving the metadata quality of digital cultural content in a scalable way, while at the same time engaging different user communities and raising awareness about cultural heritage assets. To address this need, we propose the CrowdHeritage open end-to-end enrichment and crowdsourcing ecosystem, which supports an end-to-end workflow for the improvement of cultural heritage metadata by employing crowdsourcing and by combining machine and human intelligence to serve the particular requirements of the cultural heritage domain. The proposed solution repurposes, extends, and combines in an innovative way general-purpose state-of-the-art AI tools, semantic technologies, and aggregation mechanisms with a novel crowdsourcing platform, so as to support seamless enrichment workflows for improving the quality of CH metadata in a scalable, cost-effective, and amusing way.

Keywords: crowdsourcing; automatic enrichment; user validation; annotation model; cultural heritage



Citation: Kaldeli, E.; Bekiaris, S.; Menis-Mastromichalakis, O.; Ralli, M.; Tzouvaras, V.; Stamou, G. CrowdHeritage: Crowdsourcing for Improving the Quality of Cultural Heritage Metadata [†]. *Information* **2021**, *12*, 64. <https://doi.org/10.3390/info12020064>

Academic Editor: Evaggelos Spyrou
Received: 15 January 2021
Accepted: 2 February 2021
Published: 4 February 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Digital technology is transforming the way in which Cultural Heritage (CH) is produced, presented, and experienced. Accelerated digital evolution in the form of massive digitization and annotation activities along with actions towards multi-modal cultural content generation from all possible sources has resulted in vast amounts of digital content being available through a variety of cultural institutions, such as museums, libraries, archives and galleries. In addition, the evolution of web technologies has contributed to making the Web the core platform for the circulation, distribution and consumption of a broad range of cultural content. It has been estimated that 300 million objects from European heritage institutions have been digitized, representing 10% of the region's cultural heritage, out of which only about one-third is available online (Europeana Strategy 2020–2025, <https://pro.europeana.eu/page/strategy-2020-2025-summary>).

In the last two decades, a number of initiatives at organizational, regional, national and international level have focused on aggregating and facilitating access to digital cultural content, giving rise to a number of thematic, domain-based as well as cross-domain CH hubs and web platforms, such as the European Digital Library Europeana (www.europeana.eu), the Digital Public Library of America (DPLA, <https://dp.la>), and many others. These initiatives aim on one hand to streamline the aggregation process and make it easier for CH Institutions to prepare and share high-quality content and, on the

other, to engage users from different audiences—from educators and creatives to researchers and the general public—via a number of added-value services that make content more readily available for browsing, search, study, and reuse.

However, due to the complex, heterogeneous and multi-channel aggregation workflow and shortcomings in the data providing process, many of the digital resources served through the numerous web platforms suffer from poor metadata descriptions. The lack and insufficient quality of structured and rich descriptive metadata highly affects the accessibility, visibility and dissemination range of the available digital content, and limits the potential of added-value services and applications that reuse the available cultural material in innovative ways, consequently limiting the user experience provided by CH platforms.

Metadata quality improvement and enrichment is a major challenge that receives increasing attention in the digital cultural heritage domain and is among the top priorities of Europeana's strategy. CH Institutions have put significant efforts in improving the quality of their collections' metadata, however, the efficiency of such efforts is compromised by a problem of scale: improving or even adding new metadata to hundreds of thousands or even millions of records coming from different sources requires significant investment in time, effort, and resources which organizations cannot usually afford. Recent advances in the field of Artificial Intelligence (AI) and the availability of a variety of off-the-shelf AI content and semantic analysis tools offer remarkable opportunities for overcoming the bottleneck of scale, by providing capabilities for analyzing almost any amount and type of data and extracting useful metadata with minimal time and resources needed. For example, there is a plethora of methods used for Named Entity Recognition and Disambiguation [1–3], tools for extracting different kinds of features from audiovisual content (e.g., for object and location recognition [4], multimedia event recognition [5], audio analysis [6] etc.), and many others that can be used for automatic enrichment in different contexts in the CH domain and beyond.

Past and ongoing attempts to take advantage of automatic enrichment tools in the field of CH have demonstrated the great potential of AI techniques, e.g., entity linking for the enrichment of Irish Historical Archives [7], in the Apollonis (<https://apollonis-infrastructure.gr>) project for digital humanities, and the Visual Recognition for Cultural Heritage project (https://www.projectcest.be/wiki/Bestand:VR4CH_rapport_1-0.pdf). However, resorting to purely automatic methods has also revealed a number of technical and methodological limitations that deter the effectiveness, scalability, and reuse potential of automatic enrichment tools as well as the degree to which these are exploited by the CH institutions. Firstly, the accuracy of results is negatively affected by the fact that the automatic enrichment tools have been trained and tuned on corpora outside the CH sector. The availability of human-annotated data can produce a considerable improvement in accuracy, however, the acquisition of appropriate labeled data is a costly process. Moreover, parties interested in exploiting state-of-the-art AI tools for enriching their datasets, lack a streamlined and scalable process for evaluating the automating enrichments and for deciding which are acceptable for being published and presented on their platforms. Manual validation of all automatic enrichments is costly and evidently does not scale up, with existing tools lacking the necessary functionalities to support this process.

1.1. Crowdsourcing as a Means for Metadata Quality Improvement in the CH Domain

Crowdsourcing, viewed as the process of harnessing the intelligence of a large, heterogeneous group of people to solve problems at scales and rates that no single individual can, offers a remarkable opportunity to mitigate these limitations, while at the same time engaging users and raising awareness about cultural heritage assets published on CH platforms. In this context, metadata enrichment services through automated analysis and feature extraction along with crowdsourcing annotation services can be combined with human intelligence to achieve high-quality metadata at scale with minimum effort and cost. It should be noted that when we use the term “crowdsourcing”, we also consider what is often referred to as “nichesourcing”, a strategy that aims to tap into expert knowl-

edge available in niche communities rather than the general “crowd”. As a matter of fact, a “nichesourcing” approach has been identified as being particularly useful in many contexts in the CH domain [8] and such a method is adopted in many of the campaigns that required expert knowledge organized via the CrowdHeritage platform (see Section 3.1).

In this paper, we present CrowdHeritage, an ecosystem that aims to combine machine and human intelligence in order to improve the metadata quality of digital cultural heritage collections available in CH portals. The proposed ecosystem builds on the CrowdHeritage crowdsourcing platform and initial results presented in [9] to support an end-to-end workflow that exploits the power of the crowd to execute useful tasks for the enrichment and validation of selected cultural heritage metadata and utilizes AI automated methods to extract metadata from CH content. The users are engaged through crowdsourcing campaigns and are enabled to add annotations (e.g., semantic tagging, image tagging, geotagging etc.), depending on the type of content and missing metadata, and validate existing ones (e.g., by upvoting or downvoting) in user-friendly and engaging ways (e.g., through leaderboards or rewards). By interlinking the CrowdHeritage platform with existing platforms and tools that are in use by the CH community, the ecosystem is able to cover and improve the complete workflow of CH high-quality data supply, from the import of data from different sources and their mapping to a common schema for further processing to publishing the enrichments to CH repositories for consumption by end-users.

The CrowdHeritage ecosystem can be used to either enrich CH metadata with new annotations added by crowd-taskers or to validate annotations produced automatically by AI-driven tools. In the latter case, moderation by humans can overcome the errors of the machine [10] at a large scale, thus leading to the production of more reliable end results. At the same time, the manual task expected by the participant is highly facilitated, given that users are presented with automatically suggested annotations, which significantly narrow the “search space” and reduce the time spent for enrichment in comparison with creating new annotations from scratch. The resulting human-annotated or -validated enrichments can be published to digital CH platforms and directly used to improve the search and presentation services of these platforms, but also further harnessed to improve other tasks such as curation and collections creation (e.g., by exploiting metadata to make suggestions of CH items with similar characteristics). Moreover, the labeled datasets can also be exploited as good-quality CH-specific training data to improve the accuracy of AI-driven tools used in the field. This exploitation potential can prove to be particularly useful given that existing AI tools are trained and tuned on corpora outside the CH sector, such as textual data from news agencies or general-purpose image corpora, which have different characteristics from CH-related records.

Besides the benefits with regards to the improvement of CH collections quality and searchability, the CrowdHeritage ecosystem also plays a role as a means of mediating participation in collaborative and interactive projects in CH and humanities and for mobilizing crowd-taskers with different backgrounds and motivations-domain experts, culture lovers, students, the general public to contribute their knowledge to research and social causes. In this context, CrowdHeritage can thus benefit CH institutions by enabling them to engage with new audiences, researchers who can recruit willing collaborators, the educational process by engaging pupils and students in meaningful tasks, and citizens at large who expand their knowledge and exercise responsible citizenship in a fun way.

1.2. Related Work

Despite the increasing interest and involvement of museums and other CH institutions in crowdsourcing approaches, an overview of projects in the heritage field that tap into the crowd with the aim to enhance digital collections has not yet been inventoried and has only recently attracted broad attention [11]. An abundance of platforms exists for the organization of crowdsourcing projects that aim to involve contributors in performing different types of tasks [12]. Most notably, the Zooniverse [13] platform enables the organization of online citizen science project that involves the public in academic research.

The platform hosts a large number of projects on different disciplines, with projects from natural sciences representing the lion's share but with interesting examples of projects from the heritage and humanities domains. Zooniverse has been used in a number of projects in the field of CH [14].

There is also a number of crowdsourcing platforms that have been designed to serve the particular needs of the CH domain. A number of online platforms have been used for the collection of user-provided content that may refer to different formats, e.g., to public-donated photographs, which can be further analyzed for generating 3D models of monuments and sites [15–17], to recordings (<https://sounds.bl.uk/Sound-Maps/Your-Accents>) or to stories (<https://geheugenvancentrum.amsterdam>). Transcription is probably one of the most common objectives of platforms that employ crowdsourcing for improving the accessibility of digital heritage collections, such as the Libcrowd (www.libcrowds.com) platform for transcribing elements such as titles, locations etc. of the British Library's collections and the Europeana Transcribathon platform (europeana.transcribathon.eu) for transcribing handwritten materials on Europeana. Both platforms are open for contributors but do not provide a service for the organization of custom campaigns, which allows organizers to specify a setup of their design, e.g., by selecting the content of their choice, the types of tasks for the contributors etc.

One of the first crowdsourcing initiatives that invited the crowd to add tags to museum collections was the Steve.museum project [18]. This initiative was quite preliminary, without supporting the connection to semantic vocabularies and with limited user functionalities. The Waisda? platform [19] adopted a gaming approach to engage users in adding tags to audiovisual collections, thus improving their searchability [20]. CrowdHeritage shares many common concerns with the Accurator [8] nichesourcing methodology and annotation tool for CH, which uses an RDF triple store to store the descriptive metadata and annotations. The methodology and tool have been tested on three different case studies (each of which involved up to 18 contributors), however, Accurator is not available online as a platform (although it can still be deployed on a server environment through its openly available source code) and its repository has been archived.

What makes CrowdHeritage for metadata enrichment unique is that is an easy to use platform, designed to take into account the particularities and requirements set forth by the CH domain without being bound to specific types of heritage. It offers a rich variety of features and customization possibilities that make it applicable in a large variety of application use cases and enable any interested party to organize a campaign in the way they wish and. At the same time, it can be seamlessly integrated into existing workflows at a place in the CH community, through the mapping to common standards regarding the representation of descriptive metadata for collection objects, the interconnection with established CH aggregation and presentation platforms such as Europeana, the modular presentation of different types and formats of CH objects, and the support of a rich set of relevant vocabularies. The platform's interconnection with automatic enrichment tools is also a characteristic that further improves the overall annotation workflow as well as the quality of the achieved annotations and opens a large number of future possibilities. In this respect, CrowdHeritage can take inspiration from methods and tools that effectively combine AI and crowdsourcing techniques [21], such as the CrowdTruth machine-human computing framework [22] for collecting annotation data for training and evaluation of cognitive computing systems and the Figure Eight (formerly CrowdFlower) AI-assisted data annotation platform in line with the human-in-the-loop model (appen.com/figure-eight-is-now-appen).

2. Materials and Methods

Figure 1 provides an architectural overview of the main components of the overall CrowdHeritage ecosystem and their interactions, hinting also to the workflows that the ecosystem can support. At the core of the system stands the CrowdHeritage crowdsourcing platform (crowdheritage.eu), which connects the human actors and the products of

their intelligence with the software components. A detailed description of the internal architecture of the CrowdHeritage platform is provided in Section 2.1.

The CrowdHeritage platform is interconnected with all other components of the ecosystem:

1. The automatic enrichment tools, which currently include three different AI tools for extracting features from images and for named entity identification. The tools interact with the crowdsourcing platform along two directions: they can supply it with datasets that have been automatically enriched and ask the crowd to validate (down- or up-vote) the automatic annotations; and they can take datasets annotated by humans as ground-truth data. The benefits of this interchange are reciprocal. On the one hand, the human annotators are guided and facilitated in their tasks, which become more focused and less cumbersome. On the other hand, automatic tools can be trained and tweaked on good-quality labeled CH-specific corpora and thus improve their accuracy by taking into account the special characteristics of this domain. More information about the AI tools and their interaction with human intelligence is provided in Section 2.5.
2. The MINT open web-based data aggregation platform [23], which is used by more than 550 CH organizations and 8 Europeana aggregators for the aggregation and management of their metadata records and it is the main component used for ingesting datasets to Europeana. The platform offers a user and organization management system that allows the operation of different aggregation schemes (thematic or cross-domain, international, national or regional) and registered organizations can upload (via HTTP, FTP, or OAI-PMH) their metadata records in XML or csv serialization. MINT offers a visual mapping editor that enables users to map their dataset records to a desired XML target schema. The role of MINT in the CrowdHeritage ecosystem is two-fold: to feed the crowdsourcing platform with datasets of metadata records in the right format to be enriched; and, vice versa, to insert the outcomes of the crowdsourcing process, i.e., annotations (see Section 2.4), as enrichment-additions to the original metadata records and to ultimately publish the enriched datasets to the CH presentation platforms. It should be mentioned that the data loaded via MINT can be either new datasets provided by content providers or datasets sourced from the Europeana platform, thus enabling the improvement of already published metadata.
3. The CH presentation platforms, which serve the CH data to the end-user, offering a number of added-value services (e.g., search, collection views, etc.). The CrowdHeritage platform is connected to the Europeana Search and Record API and can source cultural resources stored in Europeana in order to make them available for crowdsourcing. Besides being interlinked with Europeana, interconnection with systems and platforms maintained by CH organizations and especially national and domain aggregators themselves for the custom organization and presentation of their content, such as the EFHA platform (<https://fashionheritage.eu/browse>), is also supported via MINT. Through the Validation Editor of the crowdsourcing platform (see Section 2.3), campaign organizers can review the added annotations and decide which are acceptable for being published and presented on their platforms and on Europeana. The CrowdHeritage workflow supports two different ingestion routes for publishing the enrichments to Europeana: besides publishing them as part of augmented metadata records via MINT as already mentioned, there is also the option to submit them as separate annotations by making use of the Europeana Annotation API (pro.europeana.eu/page/annotations). The later route is more appropriate in the cases where content providers prefer to keep the added tags decoupled from the original metadata records.

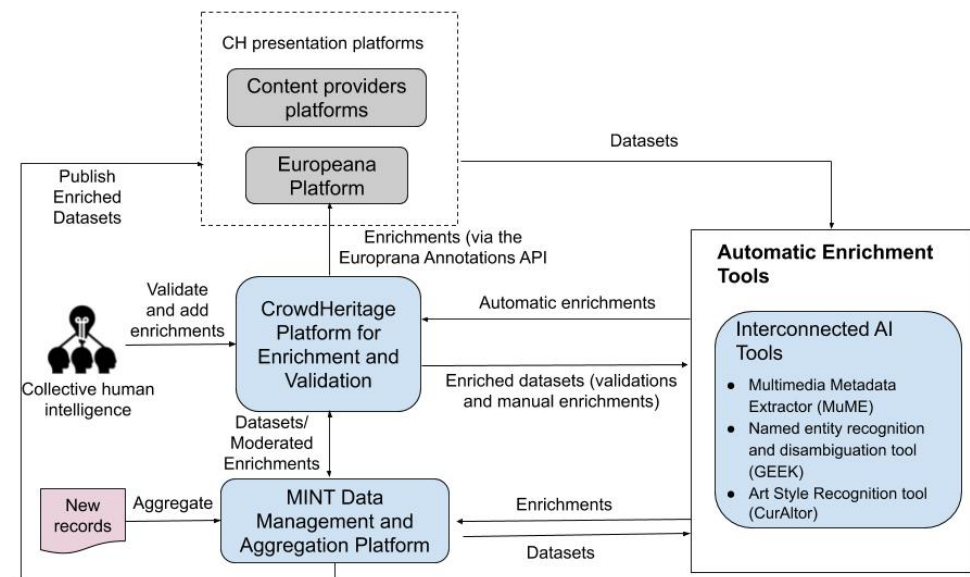


Figure 1. Architectural overview of the CrowdHeritage ecosystem.

2.1. The Main Platform Internal Components

The development of the CrowdHeritage crowdsourcing platform was conducted in collaboration with stakeholders who were involved in the definition of the user scenarios and evaluation tasks by actively participating in the necessary discussions in order to concretize the functional requirements for the technological platform, identify the content for the campaigns and carefully shape every detail regarding the campaigns' execution. Cultural Heritage Institutions (CHIs) and associations such as the ModeMuseum Antwerpen (<https://www.momu.be>), the Network of European Museum Organisations (<https://www.ne-mo.org>) and the Philharmonie de Paris, strongly influenced the objectives and outcomes of CrowdHeritage, and broader target audiences (students, teachers) provided useful feedback for the proper functioning of the CrowdHeritage platform. The development of the platform was implemented in line with the agile principles, where three versions of functional requirements were developed, each one directing the next development sprint and taking into account the evaluation results of the previous iteration. Both the frontend and the backend of the platform are deployed on servers maintained by the National Technical University of Athens, who is also responsible for the overall administration and maintenance of the CrowdHeritage platform. The source code is available on GitHub (<https://github.com/ails-lab/crowdheritage>), licensed under the Apache 2 license. For the backend, the Play Framework (<https://www.playframework.com>) is used, following the model–view–controller architectural pattern, and the code has been written in Java and Scala. The frontend uses the Aurelia (<https://aurelia.io>) JavaScript client framework, along with Node.js (<https://nodejs.org>) as a runtime environment. The languages and technologies used for the development are JavaScript, Less.js, and HTML5.

The CrowdHeritage platform consists of three basic internal components: (i) the content aggregation and collection management system; (ii) the Crowdsourcing Web Spaces, where end-users can navigate through selected collections of cultural records and add different annotation types (see Section 2.2 for more details); and (iii) the administrative user interfaces facilitating the design, customization, and validation process of the campaigns (described in Section 2.3). The internal platform architecture and the connections between its components is illustrated in Figure 2.

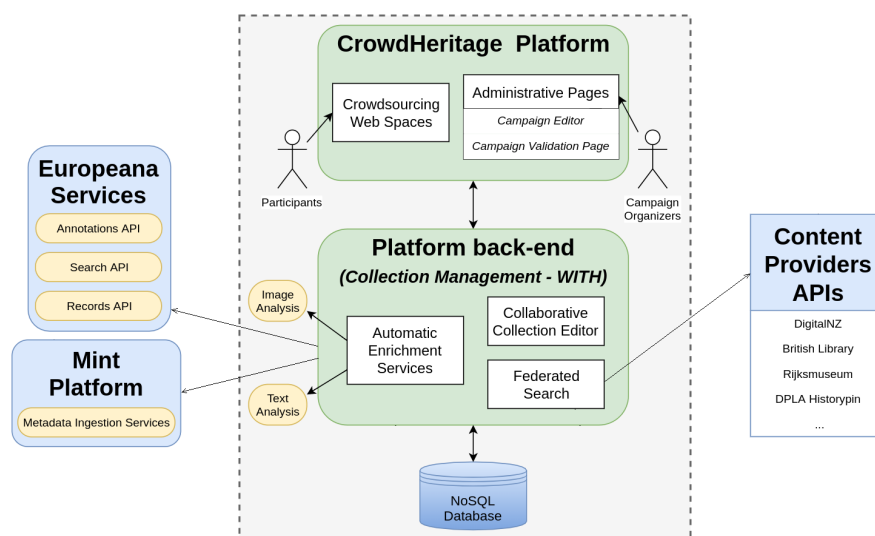


Figure 2. CrowdHeritage platform overview.

The backend layer of the platform, and particularly its content aggregation and collection management system, is built on top of the WithCulture platform (www.withculture.eu) [24,25], which provides access to digital CH resources which provides access to digital cultural heritage items from different repositories and offers a number of added-value services for the structuring and creative reuse and of that content. By making use of the WithCulture capabilities, the CrowdHeritage platform can collect CH-related data from various sources and take advantage of the federated and faceted search services that allow for the simultaneous search of multiple searchable CH repositories such as Europeana, DPLA, DigitalNZ (www.digitalnz.org), the Rijksmuseum (www.rijksmuseum.nl), and others, giving access to a huge set of heterogeneous items (images, videos, different metadata schemata etc.). The aggregated CH data is converted into a homogeneous data model compatible with the Europeana Data Model (EDM, <https://pro.europeana.eu/page/edm-documentation>), organized into thematic collections by one or more collaborators and stored in a NoSQL database. Opting for a NoSQL database enabled the maintenance of large volumes of sourced data, which are often semi-structured and follow different complex schemas depending on the content provider, while at the same allowing for more flexibility and easier updates to the target schema for representing CH records used by the platform.

The same item stored in the WithCulture database can become part of more than one thematic collection, since each collection consists of a set of references to items along with additional metadata referring to the collection itself (e.g., title, description, creator etc.). A thematic collection can then be made available for crowdsourcing by linking it to a specific campaign setup via the Campaign Editor (see Section 2.3) and for automatic enrichment with the use of appropriate AI tools as described in Section 2.5. The same collection can become an input to more than one campaign and a campaign may refer to the enrichment of more than one collections (see also Section 2.3). Each item can be associated with multiple annotations, coming from different annotators and campaigns, following the Annotation Model described in Section 2.4, which contains all the necessary information that links each item with all the collected annotations referring to it.

The CrowdHeritage platform has been designed to fully support multilingualism in a dynamic way, both with regard to the platform's interface, including campaign descriptions and instructions, as well as the with regards to the support for multilingual vocabularies and thesauri. Currently, the interface is available in English, Italian and French and can easily support any language as long as appropriate translations are provided. The annotating process also currently supplies translated tags in the three above languages and the uploading of any multilingual vocabulary can be supported, in order to present the labels and auto-complete functionalities to the user in the language they have selected.

2.2. Functionalities for End Users

Through the platform's landing page, the unregistered user can have a look at the list of ongoing, completed or upcoming campaigns and browse through each campaign via the dedicated Crowdsourcing Web Spaces. A web space consists of a set of campaign-specific pages, including a summary page that provides an overview of the campaign, presenting its goal, progress, and relevant statistics and collection pages for browsing the involved CH items in a contextual setting, where the item view only showcases annotations concerning the respective campaign along with the original metadata.

By browsing through the Crowdsourcing Web Spaces, the registered CrowdHeritage users are able to contribute to an ongoing campaign via a simple and user-friendly interface and with a quick learning curve. The content served by the platform is organized under thematic collections of cultural records enabling end-users to navigate, choose a collection, browse through the records and their metadata, and select the ones they wish to enrich, by adding annotations to them. The underlying annotation model (see Section 2.4) is expressive enough to cover a large variety of different representations, from simple textual tags to linking to Web Resources of various formats, while the User Interface (UI) is designed to serve to the end-user the different structures supported by the model in a comprehensive and functional way. In this respect, the UI currently facilitates the semantic annotation of records with terms from controlled online vocabularies and thesauri, color-tagging, and geotagging items.

The annotation process begins with the users browsing through the thematic collections of a campaign and selecting one to contribute to. By choosing the collection they wish to annotate, the users are presented with a grid consisting of the collection's items, with the option to filter out the ones they have already annotated. The annotating process begins with the user clicking on an item of the grid and work their way in the collection in a serial way. Alternatively, the participants can choose to be presented with a (currently random) selection of items for enrichment. In semantic tagging, users can tag the records by typing the desired tag into the relevant text field, which displays a list of suggested terms derived from the selected thesauri or vocabularies, supported by an auto-complete functionality. Color-tagging is accomplished by selecting the desired color from a palette of available colors and in geotagging users can pinpoint their location on a map. The users can also validate existing annotations by up-voting or down-voting them, depending on whether or not they agree with them, or even delete their own annotations. The annotation process for the end-users is illustrated in Figure 3.

The CrowdHeritage platform also provides information and statistics about each campaign, like its percentage-based progress depending on the set annotation goal, the total count of the contributors, and gamification elements such as leaderboards consisting of the most active users, with the aim to make crowdsourcing a transparent and engaging experience. The user is encouraged to add more annotations via a point and earning and reward system: usually, they gain two points for new annotations and one point for an up-vote or down-vote, pursuing the gold, silver or bronze badge which, depending on the campaign, can be accompanied with a prize. On every campaign page, CrowdHeritage provides statistics for each user regarding their contribution to the campaign, e.g., the total number of new annotations they have added, down-voted or up-voted, the number of digital cultural objects they have annotated, and their ranking in the campaign leaderboard, based on the awarded points for their contribution and determining the badge they have earned. Furthermore, the user karma points are calculated based on the percentage of upvotes versus downvotes that the tags inserted by the user have received. The karma percentage can be seen as a means of peer-reviewing that provides a quick way of identifying malicious users, who may insert quick and unrelated annotations in order to gain points.

Non logged-in users can browse the list of the available campaigns with basic information: title, description, banner and thumbnail, start date, end date, contributors, annotation target, current number of annotations, and percentage of completion and can filter the list of campaigns according to their status. By opening a campaign, more detailed information

and statistics appear as well as a grid with the collections available for crowdsourcing containing their records, visualization (photo, video, sound), metadata, and existing annotations. The campaign leaderboard is also visible, illustrating the most active users of the campaign.

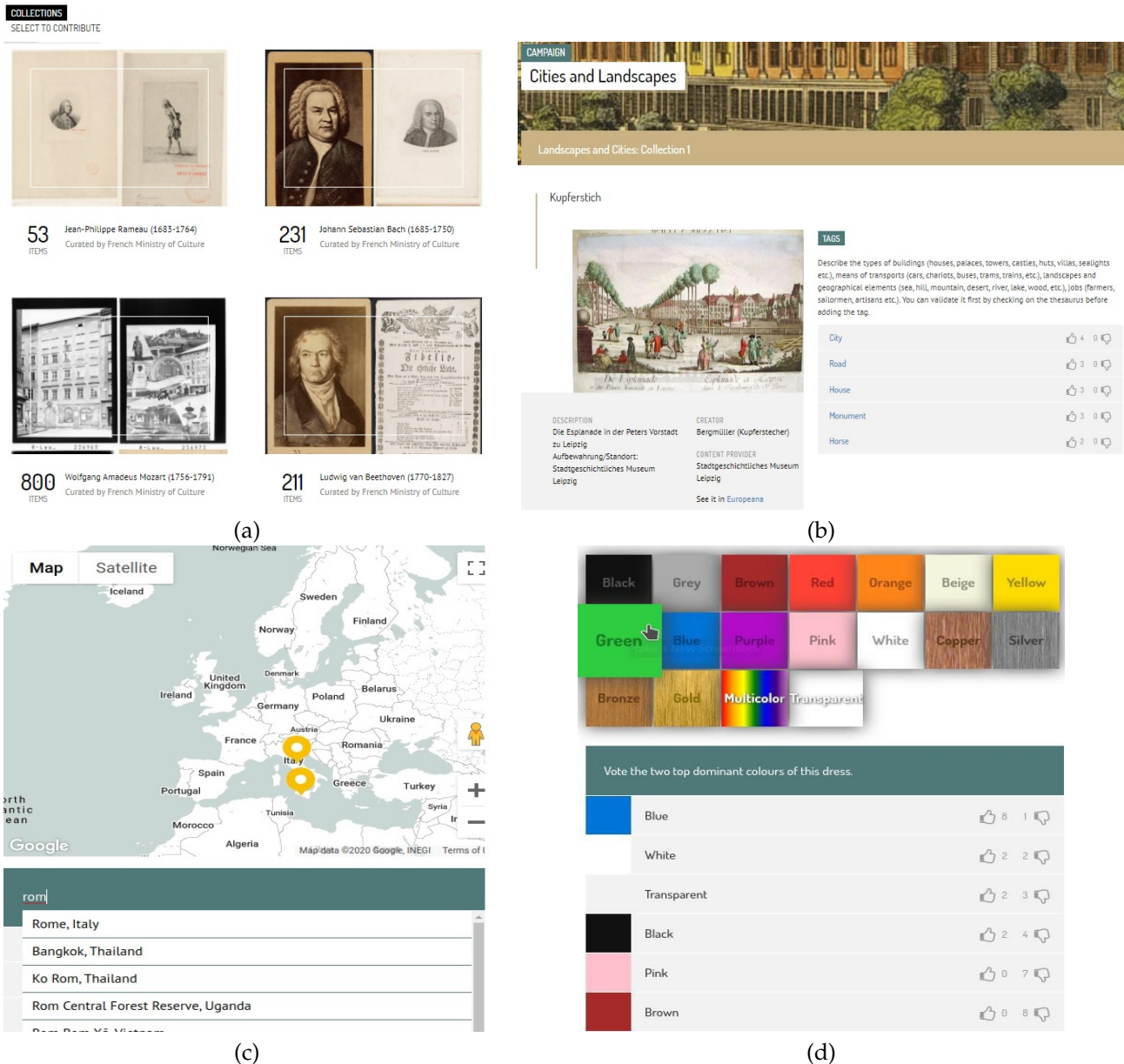


Figure 3. Annotation process: (a) collections, (b) semantic tagging, (c) geo-tagging and (d) color-tagging.

2.3. Administrative Functionalities

The platform also offers administrative functionalities for the campaign organizers, including a custom Campaign Editor and a Validation Editor. By taking advantage of these functionalities, a user who has been granted administrative permissions can launch their own customized crowdsourcing campaign in order to enrich selected digital cultural collections by deriving annotations from the public and finally moderate the campaign results by validating the produced annotations.

The Campaign Editor enables the setup, editing, deletion, and preview of custom crowdsourcing campaigns. It allows the administrative user to specify the appearance of the campaign, the content to be used, and the annotation process to be followed. The user




can define a set of features, such as the title, description and duration, choose a banner, and select the content to be used in the campaign either by directly importing a Europeana collection or by searching into Europeana and curating their own collections. Subsequently, they are able to design the desired annotation process by setting a target for the campaign, selecting from a wide variety of vocabularies and thesauri to be used, and choosing the type of desired annotations from semantic tagging, geotagging or color-tagging. They can also compile the instructions for participants and describe the prizes for the top contributors.

New campaigns are set up and launched in a fully configurable and dynamic way. Each campaign initialized through the editor interface, is stored as a separate entry in the database, which includes all the information specified by the organizer. Based on these parameters, the set of campaign-specific HTML pages that constitute the Crowdsourcing Web Space (see Section 2.2) is automatically constructed and assigned a dedicated URL based on the campaign's name, e.g., <https://crowdheritage.eu/en/garment-classification>. The implementation also enables the campaign-specific navigation of collections, so that, depending on the URL via which a CH collection is accessed (e.g., via crowdheritage.eu/en/garment-classification/collection/5daac5aa4c74793bb0b68a40), the end-user is presented only with the annotations resulting from the respective campaign. The new campaign is also added to the list of campaigns on the platform's landing page.

At the moment, the platform UI supports the use of several widely used Linked Data vocabularies, datasets, and ontologies from which campaign organizers can choose to use as the campaign's vocabulary. These include the Getty Art and Architecture Thesaurus (AAT, <https://www.getty.edu/research/tools/vocabularies/aat>), the General Multilingual Environmental Thesaurus (www.eionet.europa.eu/gemet), the Musical Instruments Museums Online (MIMO) thesaurus (www.mimo-db.eu), the Europeana Fashion thesaurus (<http://thesaurus.europeanafashion.eu>), the KULeuven Photography Vocabulary (<http://bib.arts.kuleuven.be/photoVocabulary>), Wordnet (wordnet.princeton.edu), and knowledge bases such as Wikidata (www.wikidata.org), DBpedia (wiki.dbpedia.org) and Geonames (www.geonames.org). In all the above vocabularies and datasets, each resource is always accompanied by one or more textual labels, possibly in several languages. These labels provide textual representations for the specific resource and are used for indexing the resources and facilitating term lookup.

In the platform backend, a Thesauri Manager has also been developed to support, through an offline process, the import of more Linked Data vocabularies and datasets that can be subsequently selected as a campaign's vocabulary. The Thesauri Manager converts the imported vocabularies from their source format (e.g., SKOS thesauri, OWL ontologies, N-triples datasets) to a common JSON EDM-consistent representation, stores them in the database, and indexes them to allow fast search and retrieval.

The Validation Editor provides to campaign organizers access to an interface via which they can review the annotations produced by the crowd or automatic tools (see Section 2.5) and filter them according to their own acceptance criteria. Moderation is necessary in cases where expert knowledge is required on top of the crowd contribution, ambiguity needs to be resolved (e.g., the dominant color of an outfit) or some correct, yet unhelpful information needs to be removed (e.g., records were tagged with obvious but too general annotations such as "womenswear"). Through the validation interface, the organizers can view the popular tags of campaigns, click on them, find out the records tagged with each term, and un-tag the irrelevant records assuring useful and valid annotations. A visualization of the process is depicted in Figure 4. Through the Validation Editor the campaign organizers can also view the profile of the campaign participants, including their karma points, and thus providing a means for identifying misbehaving participants.

 PUBLISH TO EUROPEANA
  EXPORT CONTRIBUTORS (CSV)
  EXPORT ANNOTATIONS (JSON)

Select a **tag** from the list and view all the campaign records that have been annotated with it.

city ship monument house building sea boat hill mountain

human settlement road Italian Republic car wood garden river man

village lake wheeled vehicle


Search for a Tag :

You have selected **2** annotations :

Cancel Selection Delete Annotations

SHIP

49 ANNOTATED ITEMS

SORT BY UPVOTES 








Figure 4. Validation page.

2.4. Annotations Metadata Model

An Annotation is the primary data structure used by the CrowdHeritage platform and refers to a piece of information associated with a CH resource, that can be added on top of the resource without modifying the resource itself. The CrowdHeritage platform adopts the W3C's Web Annotation Model (<https://www.w3.org/TR/annotation-model/>) as its underlying model for representing annotations, which provides a structured format for enabling annotations to be shared and reused across different hardware and software platforms. The CrowdHeritage annotation model does not support the full extent of the W3C model but, at the same time, introduces some dedicated terms that serve the particular needs of crowdsourcing. These include, most notably, the Annotation Score section, which helps us to establish a metric regarding the annotation's value calculated by the users as well as to keep track of the users that evaluated that annotation.

In brief, an annotation consists of an id, a motivation, a list of annotators, a body, a target, and a list of scores. An annotation may be generated either automatically by a content analysis software, a web-service etc., or manually by a human annotator. Hence, the list of annotators contains all relevant information about the origins of each annotation. The target of annotation points to the CH record which the particular annotation refers to. The core part of the annotation is its body, which links the CH object with some resource. The resource may take several formats (e.g., textual body), but usually includes a label with an IRI (Internationalized Resource Identifier) which is derived from some vocabulary or thesaurus (see also Section 2.3). Finally, the list of scores holds information about the users that have up-voted or down-voted the particular annotation. The annotation model used in the platform is depicted in the class diagram of Figure 5.

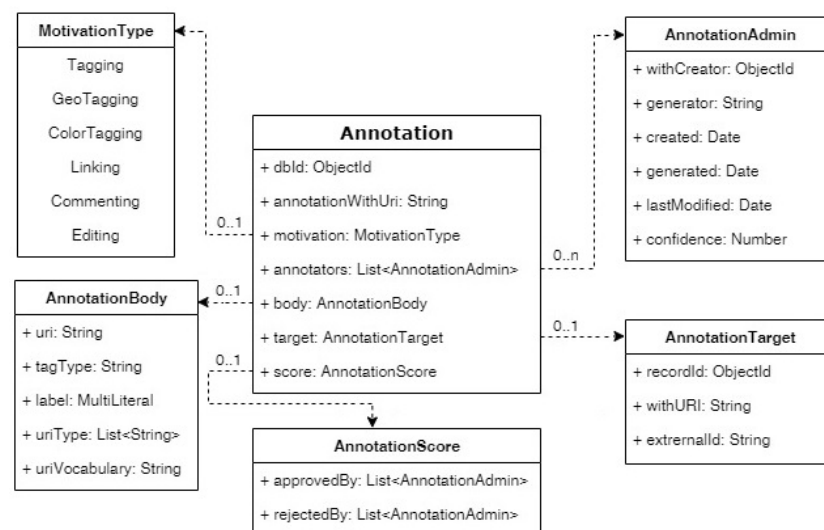


Figure 5. Annotation class diagram.

2.5. Interlinking with AI Enrichment Tools

A number of AI tools for the automatic annotation of big amounts of records have been integrated into the CrowdHeritage ecosystem as already mentioned. Though a bidirectional flow of data, the combination of human intelligence and the automatic tools can improve the quality of the end results of the enrichment process as well as the accuracy of the AI tools. AI tools can be used to produce annotations quickly and efficiently, performing challenging tasks that would conventionally require a remarkable amount of resources. However, manual validation at a large scale is required to spot and filter out inaccurate automatic results. At the same time, automatic annotations can give campaigns a significant head start because most of the required tags are already provided to the users, who only have to evaluate them, saving a remarkable amount of time and resources that are needed to create annotations from scratch.

Additionally, another direction we are following concerns the use of crowd feedback to fine-tune the AI tools. This use case paves the way for an active learning cycle to take place in the field of CH, putting crowdsourcing at the service of improving the accuracy of automatic enrichment in the field. The organization of crowdsourcing campaigns enables the creation of ground truth cultural-heritage-related data in order to retrain or fine-tune the AI systems, so that they take into account the particularities of that specific knowledge field. This human-in-the-loop direction has so far been only preliminary explored in the case of named entity recognition, as described below, with promising initial outputs, but further work is required to define a consistent methodology and results. Automatic enrichment tools can be integrated into the ecosystem as separate components through an API endpoint. So far, we have tested the use of third-party AI tools such as AIDA [26] and have also created three custom AI systems, that are particularly relevant to the cultural heritage domain. The following tools have been developed, integrated into CrowdHeritage and used in real-life cultural-related scenarios as part of the whole ecosystem:

2.5.1. Multimedia Metadata Extractor (MuME)

MuME [4] is an AI tool that processes and extracts meaningful metadata from multimedia content, that supports the recognition of faces and known persons, object detection, and landmark recognition, having a wide range of applications in the CH domain, from visual arts, fashion and design to photography and cinematography. In the context of the CrowdHeritage ecosystem, it has been mainly used to extract the dominant colors of an image or a specific object of the image, by assigning scores indicating the prevalence of the colors it identifies. In the fashion domain, color is a primary factor of garments' categorization and MuME has been employed with great success in fashion-related campaigns

(see Section 3.1) to perform visual recognition tasks like the classification of catwalk photos according to the clothing's primary color. In this campaign, the users were asked to validate the annotations produced by MuME and create new annotations in case the tool had failed to detect the garment's primary colors. The tool was used in the context of the fashion-focused campaigns to produce more 8.000 annotations for around 2.500 records, mainly representing catwalk pictures. Around 70% of the final annotations which were identified and assessed as acceptable at the end of the campaign were produced by MuME.

2.5.2. Graphical Entity Extraction Kit (GEEK)

GEEK [3] is a named entity recognition and disambiguation tool that extracts named entities in text and links them to a knowledge base using a graph-based method, taking into account measures of entity commonness, relatedness, and contextual similarity. It works in two steps: (i) it extracts text spans that refer to named entities, such as persons, locations, and organizations. (ii) 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. GEEK has been employed to extract important locations, organizations, artists, providers and many more, from the metadata of cultural heritage records. The fields of metadata (e.g., description, creator, title, etc.) were evaluated and adapted named entity recognition and disambiguation was performed for the fields that were considered adequate for each case.

GEEK was initially trained on news articles to gain general domain knowledge. We are currently in the process of tweaking the tool's parameters using as ground truth CH-related data produced through crowdsourcing campaigns that involved CH experts and were organized via the CrowdHeritage platform. Our target is to make the tool put more emphasis on the most important hyper-parameters for the CH domain, including entity commonness, relatedness, and contextual similarity. Preliminary tests with enriched datasets acquired via the CrowdHeritage platform hinted to several cases where the re-trained tool linked terms to entities which are more relevant to the CH area in comparison with the general-purpose tool (e.g., the original tool linked the term "renaissance" in an art-related text to the renaissance time period, while the re-trained tool linked it to the cultural movement of the renaissance). Further experiments need to be conducted to reach conclusions with respect to the parameter weights and the accuracy of the tools on different datasets.

2.5.3. CurAItor

CurAItor [27] is a state-of-the-art deep learning art style recognition system. The main part of the system consists of a deep ensemble network that uses Convolutional Neural Networks to process images of paintings and recognize their art style. The system was provided with thousands of images of artworks that were classified by art experts, extracting the required art-related knowledge through the training process. CurAItor is able to recognize 24 distinct art styles (Cubism, Baroque, Rococo, Pop Art, Expressionism i.a.) of various characteristics, time and art periods. Art style recognition is a challenging task even for experts, but at the same time, it's extremely important for art studying and artwork or artist categorization. Through the art style of a painting, one can deduce valid historical facts and extract useful information about artworks, artists, art movements and in general cultural movements and periods. CurAItor as part of our ecosystem is giving us access to this hard-to-find expert knowledge and can be employed to recognize automatically the art style of various paintings, while assisted by the crowd in cases that the system is not certain about the result. A preliminary test-campaign to evaluate the quality of CurAItor's results and the effect of crowd intelligence on that task has been run, where users were provided with the top-scoring options and could validate the tool's results or propose alternative labels. The test demonstrated how the use of an automatic tool can

highly facilitate a cumbersome and time-consuming manual task, with crowd involvement ultimately leading to better end-results.

3. Results and Discussion

3.1. Campaigns Setup and Implementation

So far, the CrowdHeritage platform has been used to organize twelve trans-European crowdsourcing campaigns, concerning six different themes: fashion, music, European cities and landscapes, sports, fifties in Europe, and the Chinese heritage. All campaigns' overviews and results can be found on the online platform. The main objectives of the campaigns were to demonstrate: (i) how the platform can be used to improve the quality of cultural items from different datasets suffering from poor metadata, thus facilitating the searchability, visibility and re-use of cultural material on Europeana and providers' platforms; and (ii) the user potential of the proposed tools through the engagement of different target groups (e.g., pupils, music experts, fashion researchers, broader public, etc.). Each campaign lasted three months and a final prize, offered by the partners, was awarded to the persons with the highest number of contributions in the context of each campaign. The interplay between crowdsourcing and the AI tools has been explored to a lesser extent, as part of the campaign on fashion (see also Section 2.5.1) and of two extra closed testing campaigns (see also Section 2.5.2), and remains to be further investigated as part of future work (see also Section 4).

In the fashion thematic area, two campaigns were conducted in collaboration with the European Fashion Heritage Association (<https://fashionheritage.eu>), focusing on data quality improvement of fashion-related content. The first campaign targeted a broad audience of fashion lovers and fashionistas who were invited to validate the dominant colors of fashion garments in catwalk photos from the Europeana Fashion datasets, previously-identified through automatic machine learning analysis (see Section 2.5.1). The second campaign involved mainly fashion scholars and students with the goal to improve the metadata quality of content, by providing annotations that require expert knowledge, namely adding or validating object type information related to fashion garments and linking them with terms from the Europeana Fashion Thesaurus.

In the music thematic area, two crowdsourcing campaigns were organized with the assistance of the French Ministry of Culture and in close collaboration with the Philharmonie de Paris (philharmoniedeparis.fr). The target audience involved music professionals (musicians, music teachers and scholars) as well as the broader public interested in music. The first campaign focused on musical instruments and the description and recognition of early musical instruments on medieval depictions from manuscripts, in order to enrich the Europeana datasets of musical instruments exploiting human knowledge using the MIMO vocabulary. The second music campaign was about famous composers and focused on their representation of cultural objects and images. This campaign addressed the general public and music lovers to find the right and/or the best representation of a composer and complete appropriately the corresponding metadata on Europeana.

Two campaigns were conducted in collaboration with the Michael Culture Association (MCA, <http://www.michael-culture.eu>) targeting the annotation of a variety of different themed European Collections, including Art, Maps and Geography, WWI, Photography, as well as sport-relevant content (e.g., sports event/game, competition, hobby and sport equipment) with Wikidata terms. The target audience involved both the cultural sector and teachers and pupils from schools in Italy and France. The campaign confirmed the interest of the Educational community in the CrowdHeritage platform, as a means to motivate students to learn about and contribute to various aspects of European heritage.

With respect to Europe in the 1950s, participants were requested to get involved in four thematic campaigns by adding annotations about outfits, characteristic architecture and stylish interiors, vehicles, traffic infrastructure, suitcases and bags as well as photographic qualities (light, contrast, shadows, perspective). The campaigns were promoted both

digitally and physically by the organization of events, focus groups and assignments for students of the KU Leuven university (www.kuleuven.be).

Finally, two campaigns (one of which is still open for contribution) concerning the Chinese Heritage have been organized, with the aim to demonstrate how European and Chinese Cultures are intermingled. The campaigns have been conducted in collaboration with Photoconsortium (<https://photoconsortium.net>) and have also been showcased in the context of class assignments for the students of the KU Leuven university. The first campaign is about scenes and people from China and the second one (still running) focuses on Chinese artifacts, targeting the general public. When both of these campaigns have concluded, the whole crowdsourcing effort will have accomplished the metadata improvement of at least 20,000 objects already on Europeana.

3.2. Campaigns and User Evaluation Results

For each campaign, a set of goals had been specified regarding the extent of participation, the user engagement, and the quantity and quality of achieved annotations, as assessed after the campaigns' completion. Overall, the targets of every campaign were met or exceeded. The comprehensive crowdsourcing endeavor so far has managed to aggregate annotations on more than 12,000 CH records, producing almost 40,000 new annotations and more than 85,000 validations added by more than 350 contributors. The campaigns' results-after the organizers' moderation-can be found in Table 1.

Table 1. Campaign statistics.

Campaigns Theme	Metric Type				
	Users	Items	New Tags	Upvotes	Downvotes
Fashion	87	4113	11,396 ^a	23,158	4113
Music	53	1265	560	5533	1573
Sports and Cities	67	1531	7470	23,633	1034
Fifties	126	3719	11,401	15,433	149
China	32	1487	7892	11,012	181
All	365	12,115	38,713	78,769	7050

^a1834 manual and 6549 automatic color annotations.

A user evaluation has also been performed involving both campaign organizers and contributors, amounting to 55 people in total. To evaluate the experience served by the CrowdHeritage platform and the organized campaigns, a questionnaire had been set up. A Likert scale was used to elaborate the questions and appreciate the user's level of satisfaction. The questions concern the clarity of the CrowdHeritage platform and campaigns, the platform's general performances, and its usability both as a campaign organizer and as a contributor. The online questionnaire and the feedback received demonstrated high levels of satisfaction among most of the participants with respect to the clarity, the user-friendliness, and performance of the CrowdHeritage platform.

84% of the respondents found the content of the campaign interesting and engaging (strongly agreed or agreed) and 84% felt that their contributions are valuable and were worth the effort. Both the objectives of the campaigns and the general purpose of the platform were deemed as clear and understandable by more than 90% of the respondents. More than 85% of the participants found the tagging process easy and intuitive and this was across different target groups, including CH professionals and members of the general public. 78.5% were satisfied with the performance of the platform (referring to aspects such as pages and image loading time etc.). Some performance issues were experienced during the annotation process in some of the campaigns with respect to the loading of thumbnails, which were due both to the CrowdHeritage and the Europeana platforms. Of the participants, 70.6% expressed strong interest in participating again in a campaign on the CrowdHeritage platform.

The platform's gamification set up with the leaderboard and the ranking methodology (karma points) was also assessed to be effective for encouraging contributors. During some campaigns, there have been contributors who tried to get higher scores and did not properly respect the instructions, e.g., by adding random tags. In order to prevent this kind of misbehavior by users who were motivated only to get a higher number of points, the karma percentage (see also Section 2.2) has been included in the profile pages of the user and taken into account in the awarding process.

Participants were also keen to share suggestions for improving the platform, which mainly referred to improving engagement and dissemination, such as sharing on social media, further gamification elements, and enabling interaction between users via a forum or a comment area. Moreover, contributors asked for support for more languages and the possibility to add more tags, have a larger choice of vocabularies, and be able to add free text as well. A large part of the respondent to the questionnaire mentioned that they would be happy with a mobile version of the platform. Although the platform has been designed so it can be responsive on tablets and smartphones, it seems that the participants are willing to get a dedicated app.

4. Conclusions

The CrowdHeritage ecosystem offers an end-to-end solution that couples machine-driven enrichment tools with the power of collective human intelligence, mobilized via a user-friendly crowdsourcing platform that supports the organization and launching of engaging campaigns in the CH sector. This way, it provides an efficient way for making the complete workflow of high-quality cultural data supply easier, more scalable, and cost-effective, thus streamlining and simplifying the work of aggregators and data providers. Besides providing better services for CH institutions, the CrowdHeritage ecosystem also contributes a step forward to enhancing the way digital cultural heritage is experienced by end users. The more comprehensive the metadata accompanying a digital heritage object, the more likely it is to be viewed, understood, and used by educators, creatives, culture lovers, researchers and citizen at large. By facilitating the improvement of metadata quality, the CrowdHeritage ecosystem and platform enable users to effectively discover what they are looking for, browse and go deeper into a subject, and understand its context and interconnections with other cultural heritage objects. At the same time, the CrowdHeritage platform provides the technological means to stimulate a more participatory approach to cultural heritage and engage experts as well as the general audience in its improvement.

The current implementation of the CrowdHeritage ecosystem and platform is a first important step that opens up multiple directions for future extensions, improvements, and reuse possibilities. From the implementation-technical perspective, a number of new features and improvements are planned to make the platform more functional, widen the use cases it can support, and enhance user experience. Future work towards such features include a user interface for the dynamic upload of custom vocabularies and thesauri; addition of personalization features so that so that the items and tasks suggested to the participants are selected taking into account the user's history; extensions to the Validation Editor so as to support more advanced filtering and moderation based on certain criteria to be specified by the user (e.g., based on the automatic confidence levels, the popularity of an annotation, etc.). Furthermore, more consistent efforts have to be invested into the adoption of appropriate user-moderation tools and policies. Although malicious behaviors are usually detected through our karma-points system, we also plan to provide functionality for users to directly report such behaviors. Afterwards, system moderators can manually investigate the validity of each report and act accordingly. Automatic spam and policy-violation detection methods are also being considered for campaigns in which permissible contributions by participants are not controlled via a vocabulary, i.e., allow the entry of free text.

From the more research-oriented perspective, further work and experiments need to be directed towards closing the loop of the active learning cycle [21]. To this end,

a methodology that defines a set of selection criteria for the datasets and tasks to be assigned to humans, quality thresholds etc. is yet to be specified. Elaborating on preliminary results with respect to the fine-tuning of the GEEK entity extraction tool, more extensive experiments need to be conducted by using the acquired CH-specific ground-truth datasets, in order to gain deeper insight on how the tool behaves under different parameterizations and in comparison with the general-purpose datasets. Moreover, we plan to integrate to the ecosystem and consider in our human-in-the-loop methodology more AI tools that focus on extracting different types of elements from different types of content depending on the needs of case studies from the CH domain, e.g., object extraction from image and video, speech to text, Optical Character Recognition, etc.

Considering the possibilities of impact on the CH and citizen engagement field, another interesting direction for future work concerns extending the CrowdHeritage platform to support scenarios that go beyond the enrichment of metadata towards inviting the end-user to contribute with genuine thoughts and content and associate CH collections with interpretations, emotions, and other items that can ultimately lead to richer and multi-vocal perceptions of CH collections.

Author Contributions: E.K. is mainly responsible for the conception and design of the overall framework and has written crucial parts of the current article. M.R. has written most parts of the conference paper on which the current extended article relies on, especially with respect to text describing concrete functionalities. S.B., M.R., and O.M.-M. have been the main developers of the actual tools mentioned herein and have all contributed with text describing the respective functionalities. V.T. and G.S. have been overseeing and guiding the implementation process and have provided essential feedback and made revisions to the current article. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been co-financed through (i) the Connecting Europe Facility (CEF) project CrowdHeritage: Crowdsourcing Platform for Enriching Europeana Metadata (No A2017/1564820); (ii) the Greek Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH-CREATE-INNOVATE (T1EDK-01728-ANTIKLEIA); and (iii) from the European Union's Horizon 2020 research and innovation program under grant agreement No 770158. The sole responsibility of this publication lies with the authors. The European Union is not responsible for any use that may be made of the information contained therein.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the user evaluation.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ganea, O.; Ganea, M.; Lucchi, A.; Eickhoff, C.; Hofmann, T. Probabilistic Bag-Of-Hyperlinks Model for Entity Linking. In Proceedings of the 25th International Conference on World Wide Web WWW, Montréal, QC, Canada, 11–15 April 2016; pp. 927–938.
2. Piccinno, F.; Ferragina, P. From TagME to WAT: A new entity annotator. In Proceedings of the First ACM International Workshop on Entity Recognition & Disambiguation ERD'14, Gold Coast, QLD, Australia, 11 July 2014; Carmel, D., Chang, M., Gabrilovich, E., Hsu, B.P., Wang, K., Eds.; ACM: New York, NY, USA, 2014; pp. 55–62.
3. Mandalios, A.; Tzamaloukas, K.; Chortaras, A.; Stamou, G. GEEK: Incremental Graph-based Entity Disambiguation. In Proceedings of the Linked Data on the Web Co-Located with the 2018 Web Conference-LDOW@WWW, Lyon, France, 23 April 2018.
4. Varytimidis, C.; Tsatiris, G.; Rapantzikos, K.; Kollias, S. A systemic approach to automatic metadata extraction from multimedia content. In Proceedings of the 2016 IEEE Symposium Series on Computational Intelligence-SSCI, Athens, Greece, 6–9 December 2016; pp. 1–7.
5. Amid, E.; Mesaros, A.; Palomäki, K.; Laaksonen, J.; Kurimo, M. Unsupervised feature extraction for multimedia event detection and ranking using audio content. In Proceedings of the ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing-Proceedings, Florence, Italy, 4–9 May 2014; pp. 5939–5943. [[CrossRef](#)]
6. Fillon, T.; Simonnot, J.; Mifune, M.F.; Khoury, S.; Pellerin, G.; Le Coz, M.; Amy de la Bretèque, E.; Doukhan, D.; Fourer, D.; Rouas, J.L.; et al. Telemeta: An open-source web framework for ethnomusicological audio archives management and automatic analysis. In Proceedings of the 1st Digital Libraries for Musicology workshop (DLfM 2014), London, UK, 12 September 2014.

7. Munnely, G.; Pandit, H.J.; Lawless, S. Exploring Linked Data for the Automatic Enrichment of Historical Archives. In *Lecture Notes in Computer Science, Proceedings of the Semantic Web: ESWC 2018 Satellite Events, Heraklion, Greece, 3–7 June 2018*; Springer: Berlin/Heidelberg, Germany, 2018; Volume 11155, pp. 423–433.
8. Dijkshoorn, C.; Boer, V.; Aroyo, L.; Schreiber, G. Accurator: Nichesourcing for Cultural Heritage. *Hum. Comput.* **2017**, *6*. [[CrossRef](#)]
9. Ralli, M.; Bekiaris, S.; Kaldeli, E.; Menis-Mastromichalakis, O.; Sofou, N.; Tzouvaras, V.; Stamou, G. CrowdHeritage: Improving the quality of Cultural Heritage through crowdsourcing methods. In *Proceedings of the 15th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP 2020)*, Zakynthos, Greece, 29–30 October 2020; pp. 1–6.
10. Stiller, J.; Petras, V.; Gäde, M.; Isaac, A. Automatic Enrichments with Controlled Vocabularies in Europeana: Challenges and Consequences. In *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection*; Ioannides, M., Magnenat-Thalmann, N., Fink, E., Žarnić, R., Yen, A.Y., Quak, E., Eds.; Springer International Publishing: Cham, Switzerland, 2014; pp. 238–247.
11. Navarrete, T. Crowdsourcing the digital transformation of heritage. In *Digital Transformation in the Cultural and Creative Industries*; Massi, M., Marilena Vecco, Y.L., Eds.; Taylor and Francis: Abingdon, UK, 2020; Chapter 4, pp. 99–116.
12. Mourelatos, V.; Tzagarakis, M.; Dimara, E. A review of online crowdsourcing platforms. *South-East. Eur. J. Econ.* **2016**, *14*, 59–74.
13. Simpson, R.J.; Page, K.R.; Roure, D.D. Zooniverse: Observing the world’s largest citizen science platform. In *Proceedings of the 23rd International World Wide Web Conference, WWW ’14*, Seoul, Korea, 7–11 April 2014; pp. 1049–1054.
14. Ellis, A.; Gluckman, D.; Cooper, A.; Andrew, G. Your Paintings: A nation’s Oil Paintings Go Online, Tagged by the Public. In *Museums and the Web Conference*. Available online: https://www.museumsandtheweb.com/mw2012/papers/your_paintings_a_nation_s_oil_paintings_go_onl.html (accessed on 3 February 2021).
15. Karl, R.; Roberts, J.; Wilson, A.; Möller, K.; Miles, H.; Edwards, B.; Tiddeman, B.; Labrosse, F.; La Trobe-Bateman, E. Picture This! Community-Led Production of Alternative Views of the Heritage of Gwynedd. *J. Community Archaeol. Herit.* **2014**, *1*, 23–36. [[CrossRef](#)]
16. Vincent, M.; Flores Gutiérrez, M.; Coughenour, C.; Manuel, V.; Bendicho, V.M.; Remondino, F.; Fritsch, D. Crowd-sourcing the 3D Digital Reconstructions of Lost Cultural Heritage. In *Proceedings of the 2015 Digital Heritage, Granada, Spain, 28 September–2 October 2015*; pp. 171–172.
17. Dhonju, H.; Xiao, W.; Mills, J.; Sarhosis, V. Share Our Cultural Heritage (SOCH): Worldwide 3D Heritage Reconstruction and Visualization via Web and Mobile GIS. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 360. [[CrossRef](#)]
18. Chun, S.; Cherry, R.; Hiwiller, D.; Trant, J.; Wyman, B. Steve.museum: An Ongoing Experiment in Social Tagging, Folksonomy, and Museums. In *Museums and the Web 2006: Proceedings*; Archives and Museum Informatics: Toronto, ON, Canada, 2006.
19. Oomen, J.; Aroyo, L. Crowdsourcing in the Cultural Heritage Domain: Opportunities and Challenges. In *Proceedings of the 5th International Conference on Communities and Technologies, Brisbane, Australia, 29 June–2 July 2011*; pp. 138–149.
20. Gligorov, R.; Hildebrand, M.; Van Ossenbruggen, J.; Aroyo, L.; Schreiber, G. An Evaluation of Labelling-Game Data for Video Retrieval. In *Proceedings of the 35th European Conference on Advances in Information Retrieval, Moscow, Russia, 24–27 March 2013*.
21. Gilyazev, R.; Turdakov, D. Active learning and crowdsourcing: A survey of annotation optimization methods. *Proc. Inst. Syst. Program. RAS* **2018**, *30*, 215–250. [[CrossRef](#)]
22. Inel, O.; Khamkham, K.; Cristea, T.; Dumitrache, A.; Rutjes, A.; van der Ploeg, J.; Romaszko, L.; Aroyo, L.; Sips, R. CrowdTruth: Machine-Human Computation Framework for Harnessing Disagreement in Gathering Annotated Data. In *Lecture Notes in Computer Science, Proceedings of the ISWC 2014-13th International Semantic Web Conference, Riva del Garda, Italy, 19–23 October 2014*; Springer: Berlin/Heidelberg, Germany, 2014; Volume 8797, pp. 486–504.
23. Drosopoulos, N.; Tzouvaras, V.; Simou, N.; Christaki, A.; Stabenau, A.; Pardalis, K.; Xenikoudakis, F.; Kollias, S. A Metadata Interoperability Platform. In *Museums and the Web 2012*; Springer: Berlin/Heidelberg, Germany, 2012.
24. Chortaras, A.; Christaki, A.; Drosopoulos, N.; Kaldeli, E.; Ralli, M.; Sofou, A.; Stabenau, A.; Stamou, G.; Tzouvaras, V. WITH: Human-Computer Collaboration for Data Annotation and Enrichment. In *Proceedings of the Companion of the The Web Conference 2018, Lyon, France, 23–27 April 2018*; pp. 1117–1125.
25. Giazitzoglou, M.; Tzouvaras, V.; Chortaras, A.; Alfonso, E.; Drosopoulos, N. The WITH platform: Where culture meets creativity. In *Proceedings of the Posters and Demos Track of the 13th International Conference on Semantic Systems-SEMANTiCS 2017, Amsterdam, The Netherlands, 11–14 September 2017*.
26. Hoffart, J.; Yosef, M.; Bordino, I.; Fuerstenau, H.; Pinkal, M.; Spaniol, M.; Taneva, B.; Thater, S.; Weikum, G. Robust Disambiguation of Named Entities in Text. In *Proceedings of the EMNLP 2011-Conference on Empirical Methods in Natural Language Processing, Edinburgh, Scotland, UK, 27–31 July 2011*; pp. 782–792.
27. Menis-Mastromichalakis, O.; Sofou, N.; Stamou, G. Deep Ensemble Art Style Recognition. In *Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020*; pp. 1–8. [[CrossRef](#)]