Human Computation (2024) 11:1:47-70

ISSN: 2330-8001, DOI: 10.15346/hc.v11i1.2

Crowdsourcing as a Pedagogical Tool in Computer Science Higher Education: a Case Study

VASSILIS LYBERATOS, NATIONAL TECHNICAL UNIVERSITY OF ATHENS

SPYRIDON KANTARELIS, NATIONAL TECHNICAL UNIVERSITY OF ATHENS

EIRINI KALDELI, NATIONAL TECHNICAL UNIVERSITY OF ATHENS

SPYROS BEKIARIS, NATIONAL TECHNICAL UNIVERSITY OF ATHENS

PANAGIOTIS TZORTZIS, NATIONAL TECHNICAL UNIVERSITY OF ATHENS

ORFEAS MENIS-MASTROMICHALAKIS, NATIONAL TECHNICAL UNIVERSITY OF ATHENS

GIORGOS STAMOU, NATIONAL TECHNICAL UNIVERSITY OF ATHENS

ABSTRACT

The approach used and the insights gained from employing crowdsourcing techniques in a computer science homework assignment for higher education students are described in this paper. Making use of a platform that supports crowdsourcing in the cultural heritage domain, students were solicited to enrich the metadata associated with a selection of music tracks. The results of the campaign were further analyzed and exploited by students through the use of semantic web technologies, with the aim to construct/lead to the construction of a music knowledge that was utilised for designing a recommendation system. In total, 98 students participated in the campaign, contributing more than 6400 annotations concerning 854 tracks. The process also led to the creation of an openly available annotated dataset, which can be useful for machine learning models for music auto-tagging. The outcomes of the campaign and the feedback collected via an online questionnaire allow us to draw useful insights about the challenges and benefits of incorporating crowdsourcing techniques into computer science curricula and the educational gains for students.

1. INTRODUCTION

In recent years, we are witnessing an increasing number of studies that explore the use of crowdsourcing in education, embracing various disciplines, topics, and educational levels (Jiang et al., 2018). Most of the existing work looks into crowdsourcing as a means of creating or assessing educational material and collecting feedback from students (Alenezi and Faisal, 2020; Zahirović Suhonjić et al., 2019; Wang, 2016). Fewer studies tap into how involvement in a crowdsourcing experiment can serve an educational goal in itself (Mikhailova et al., 2022; Khan et al., 2020). We argue that incorporating crowdsourcing in the form of real-world and discipline-appropriate exercises in education curricula can bring about multiple benefits for students. First, students can learn about the potential as well as the challenges (technological and methodological) associated with the planning and execution of a crowdsourcing process and how this can be useful within the context of their discipline. Additionally, the design of appropriate crowdsourcing tasks can help students familiarize themselves with important domain-specific concepts in a hands-on way and understand how the data collected through the process can be further exploited and enable new possibilities. Lastly, the participatory nature of crowdsourcing can stimulate learning and engagement (Hills, 2015) and add a collaborative and creative touch to more traditional teaching procedures.

In tandem with its educational potential, crowdsourcing has been extensively used and studied in the context of citizen science, by providing the methodology and tools that enable the engagement of individuals who voluntarily contribute to knowledge production with a scholarly focus (Shanley et al., 2019). Higher Education Institutions (HEI), including universities as well as vocational schools and colleges, can play a significant role in this context, by providing technical, material, and human resources, reinforce open science policies, stimulate cross-disciplinary collaborations, and hone the competences of new generations of scientists, researchers, and innovators as already mentioned. In such a setting, crowdsourcing can be seen as a driver of both citizen-enhanced open science and educational learning. Similarly, students are invited to play a dual role: act as citizens/contributors and as scientists/researchers.

The role that crowdsourcing can play in the Computer Science (CS) programs of HEIs in particular remains quite unexplored. Despite its close interrelations with IT practice and research, considering both the use of digital technology as a facilitator of crowdsourcing and, vice versa, the extensive application of crowdsourcing techniques in the IT domain (e.g. active learning, dataset construction for Machine Learning (ML)), crowdsourcing is not among the subjects which usually make up a CS curriculum. In fact, crowdsourcing as a practice and technology remains peripheral to CS-relevant HE programs, with references to it and hands-on experiments being incidental (note that the crowdsourcing initiatives in education reviewed in (Jiang et al., 2018) include only one case study from CS). Our main objective in this paper is to alleviate this gap by reporting on a case study that investigates how crowdsourcing can be incorporated into a CS curriculum as a component and facilitator of a mini-project assignment that can teach students useful lessons. As part of the case study, students were invited to participate in an online campaign with the aim to enrich the metadata of a music tracks collection (curated by the authors' team; see Section 3.1). Students were then instructed to analyze the enriched dataset and apply semantic web technologies to construct a knowledge base and use it to extract useful information from it.

In this context, our research set out to explore the impact crowdsourcing had along two main perspectives: learning outcomes achieved in terms of new knowledge and skills acquired, especially within the scope and objectives of the CS curriculum (e.g. to what extent did students feel that they improved relevant skills? what did they learn?); and how the participation experience was perceived by students (e.g. did they feel engaged? did they enjoy the process?). Concerning the educational gains, we were particularly concerned with the extent to which crowdsourcing assists students to gain deeper insights into the structure and shortcomings of data, into the processes and technological infrastructures that can be used to acquire richer and higher-quality data, as well as into how the enhanced data can be further utilized. Regarding the extent to which students felt engaged, we were mainly interested in affective characteristics, relating to feelings and attitudes. Another direction we explored regards the role of technological tools, considering the requirements for certain features as well as the way in which the digital platform that was used influenced the experience of the participants (which features were the most appreciated? what issues were identified?). By putting themselves in the position of the contributor and platform end-user and, in parallel, by drawing on their capacity as CS students and technical experts, participants were able to provide insightful perspectives along this strand of inquiry.

Overall, the case study exemplifies how crowdsourcing can fit in a CS course and serve its intended didactic objectives. By describing in detail the methodology that was followed, from the data curation and the campaign setup to the exploitation of results and the evaluation approach, the technological tools that were used, the challenges that were encountered and the way in which these were overcome, the current paper points both to the benefits as well as the limitations of the approach and can thus serve as a paradigm and as a source of inspiration for incorporating crowdsourcing in CS curricula and beyond. As confirmed in the survey performed in (Vance-Chalcraft et al., 2022), the availability of such detailed descriptions and accompanying open resources are highly appreciated by educators in tertiary education to facilitate the incorporation of citizen science projects into post-secondary courses (Vance-Chalcraft et al., 2022). Besides the educational benefits, in line with the principles of open and citizen science, the data generated by the case study is further processed and made openly available, thus contributing to ongoing developments in research on music tagging research.

Finally, our case study adds one more dimension with a strong interdisciplinary orientation, that of digital humanities. The curated dataset given (constructed by the authors' team; see Section 3.1) acting as the baseline for the crowdsourcing campaign as well as the design of the associated enrichment tasks were driven by established practices in the cultural heritage (CH) domain. Music was selected as a type of heritage which is quite popular among students and can be enjoyed and appreciated without any special requirements for expert knowledge. Crowdsourcing has been employed quite broadly in multiple settings within these fields - in the context of projects led either by museums' departments or by universities and research institutes - mainly as a process that invites members of the public "to tag and classify, transcribe, organize, and otherwise add value to digital CH collection content" (Owens, 2013). Such approaches customarily operate within the traditional functioning of digital humanities, looking into how computational tools can be harnessed to support the humanities researcher and CH professional. Although the current study provides useful insights

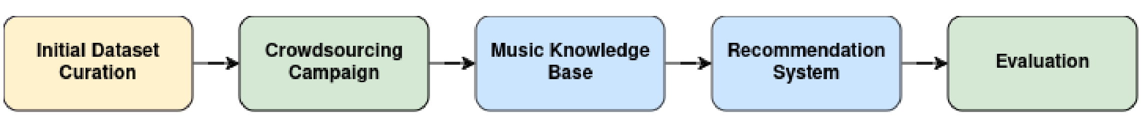


***Figure 1.* The steps followed in our case study are as follows: the first box indicates a step implemented by our team; the second and fifth boxes represent steps handled in collaboration with the students; and the third and fourth boxes denote steps completed by the students.**

along this perspective, as a demonstrator of how digital technology can be employed for the enrichment of CH collections, it places its main focus on the opposite direction, which so far has received much less attention: How can digital heritage collections, and music collections in particular, be utilized in a CS context? What are the potential benefits for CS students and IT research? And, ultimately, by connecting to the strands of research discussed earlier: how can concepts, processes, and tools used in cultural heritage, computer science, and citizen science be meaningfully combined within a higher education context and what kind of conclusions can be drawn from this interplay?

The article is organized as follows: Section 2 reviews related work on crowdsourcing in education, citizen science, digital humanities, Music Information Retrieval (MIR) tasks, and crowdsourcing platforms. Section 3 describes the methodology we followed to prepare our case study and Section 5 provides a detailed account of the students' assignment. The results of the crowdsourcing campaign and the students' evaluations for the case study are presented in Section 5. In Section 6, we draw conclusions from the case study and provide recommendations for educators and campaign designers interested in integrating crowdsourcing into their curricula. Figure 1 illustrates the specific steps taken to implement our case study.

2. RELATED WORK

According to the typology suggested in (Jiang et al., 2018), crowdsourcing is used in educational activities to serve four main objectives: create educational content; collect feedback from learners; exchange complementary knowledge by resorting to external crowds; and by providing practical experience. Its most prevalent uses concern educational material generation and assessment (Alenezi and Faisal, 2020; Zahirović Suhonjić et al., 2019) and this is also true for CS-related curricula in higher education. For example, (Pirttinen and Leinonen, 2021) describes a tool that can be used as a means to support teachers and students to create and review programming assignments. The main motivation behind such initiatives is premised on the potential benefits of crowdsourcing concerning optimizing the lecturing process and stimulating student involvement through knowledge co-creation and sharing, in line with contemporary learner-centered approaches to education (Lambert and McCombs, 1998). 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

The suggested paper expands upon the case study presented in (Lyberatos et al., 2023b), offering a more comprehensive examination and analysis of the preliminary findings. The case study employs crowdsourcing in a project-based setting (Blumenfeld et al., 1991), inviting students to grapple with a real-world problem — that is creating a music knowledge base and a recommendation system. Tapping into the multiple prospective benefits of project-based learning, in (Khan et al., 2020), the potential of resorting to crowdsourcing platforms for sourcing realistic tasks that can replace traditional assignments addressed to students of industrial design is discussed. The current case study aims to further investigate and leverage the employment of crowdsourcing in such a context, aiming at similar educational gains but from a different perspective: it approaches crowdsourcing not merely as a pool of possible pre-designed tasks, but rather as a methodology and technique that can be adapted to the specific course objectives and that is worth learning in its own right.

Applications in CS-related higher education curricula (informatics, computer engineering, etc) that adopt crowdsourcing as a means to provide practical experience in a setting relevant to the students' discipline are few. In (Kasumba, 2022), a crowdsourcing experiment conducted as part of a research project, involving data science students in rating homework reviews, had the unplanned effect of serving as a learning opportunity for students. In (Chen and Luo, 2014), students of software engineering were assigned the task to test commercial software and through this process achieved industrial-strength training. The most common practice, which is also followed by our work, is that instructors assume the role of the requester and students that of crowd-workers. An interesting exception to this is (Guo et al., 2018), where graduate and undergraduate computer science students were asked to design and deploy their own crowdsourcing projects. The current case study adds to this line of work, by placing the focus on the challenges and possibilities of crowdsourced-enabled data enrichment in serving CS-relevant learning objectives and by contributing novel evidence and multi-dimensional insights grounded on an extensive analysis of feedback collected from students about multiple aspects (skills, engagement, usability, etc).

Within the last years, there is an increasing number of initiatives that apply crowdsourcing in citizen science-oriented settings within formal and informal learning environments (schools and universities) (Kloetzer et al., 2021). Most such initiatives involve children and adolescents at the primary and secondary levels (Roche et al., 2020), while citizen science projects in tertiary education remain fewer (Vance-Chalcraft et al., 2022). In an application of citizen science in an undergraduate environmental studies course (Heigl and Zaller, 2014), students were engaged in reporting roadkilled animals, thus gaining a deeper understanding of ecological problems and their solutions. Another case study (Mikhailova et al., 2022), involving students from biology and environmental studies in field data collection, concludes that students enjoyed the learning process and improved their understanding of the domain as well as of crowdsourcing as a method for data collection. The crowdsourcing task selected for the current case study involves data enrichment of a music collection (Gomes et al., 2012).

In this respect, the current paper contributes to ongoing efforts (Zhang et al., 2018; Bogdanov et al., 2019b; Laurier et al., 2009; Humphrey et al., 2018; Aljanaki et al., 2016), many of which resort to crowdsourcing methods, to increase the availability and quality of annotated datasets that can be useful for prototyping systems for MIR tasks (Choi et al., 2016; Lyberatos et al., 2023a) and particularly tasks concerning genre (Costa et al., 2011), instrument (Shreevathsa et al., 2020), and emotion recognition (Liu et al., 2017). One of the shortcomings of such annotated datasets is that most of the music tracks are released under licenses that do not permit their publication. Due to this limitation, it is common that published datasets only contain features that are derived from audio analysis, without including the raw audio data (Bogdanov et al., 2019a), or that they only publish short samples of the music tracks (Panda et al., 2018). With respect to datasets annotated with emotion labels entered by human subjects, in particular, the subjectivity associated with the task makes it especially time-consuming, labor-intensive, and prone to errors, resulting in limited availability of such datasets (Soleymani et al., 2013). In this context, by making openly available a carefully moderated collection of the annotations gathered via the crowdsourcing campaign, the case study also plays a part in developments in music auto-tagging.

A broad range of crowdsourcing platforms have been proposed and tested in education. For example, (Farasat et al., 2017) describes a platform for the collaborative creation and refinement of large "banks" of subject matter problems in higher STEM education. For CS-related curricula, in particular, the open-source platform CrowdSorcerer supports novice programmers in creating and evaluating programming assignments (Pirttinen et al., 2018). Multiple platforms are used in experiments that bring citizen science in education, with the selection depending on the particular circumstances and the task at hand (e.g. platforms for collecting geospatial data, for software testing, etc). For the collection of data that are of interest to the CH domain, technology usage ranges from general-purpose platforms such as Zooniverse (Simpson et al., 2014) to tools tailored to the needs of CH, such as the Transcribathon (Europeana Transcribathon platform, 2019) and the CrowdHeritage (Kaldeli et al., 2021a,b) platforms. In this work we extended the utilities of the CrowdHeritage platform, creating a more comprehensive environment for crowdsourcing (see Section3.3).

3. METHODOLOGY FOR PREPARING THE CASE STUDY

The case study was conducted as part of an assignment involving fourth-year undergraduate informatics students of the National Technical University of Athens who attended the course "Knowledge Systems and Technologies" in the spring semester of 2022. The main objective of the course is to introduce students to the fundamentals of description logics, methodologies for object-oriented knowledge representation, management, evolution, automated reasoning, and semantic data integration algorithms. Specific emphasis is given to the analysis of W3C standards for semantic data and knowledge representation (XML, RDF, OWL, etc), ontology engineering and applications of knowledge-based systems and intelligent web services (Stamou and Chortaras, 2017). The course includes a semester-long multi-step assignment that aims to familiarize students with the abovementioned concepts and associated tools via hands-on tasks. In line with these educational objectives, the case study set out to introduce concepts from digital CH as well as crowdsourcing to this purely CS-oriented curriculum and broaden the scope of the assignment towards an interdisciplinary direction.

In its first step, the assignment focused on familiarizing students with the curated dataset and on engaging them as annotators in a crowdsourcing campaign as a means to enrich the dataset with additional useful knowledge. The main objective in this respect was for students to understand the shortcomings of real-world datasets and how raw, inadequate, or inconsistent forms of data can be transformed into well-structured, normalized, and inter-linked formats. Next, students were asked to transform the enriched data structure into a knowledge graph containing RDF triples and build an ontology that describes the data by constructing concepts, roles, axioms, and instances syntactically and semantically correct. Finally, students were solicited to use various methods to infer extra information and exploit it to make meaningful recommendations on music.

This section outlines the step-by-step methodology we used to set up the case study. It includes our preparations for curating the initial dataset to be enriched (Section 3.1) and for organizing the crowdsourcing campaign, including the definition of the enrichment tasks (see Sections 3.2 and 3.3).

3.1. Dataset curation

We decided to use the Europeana digital library [1] to source the data that constituted the starting point of the crowdsourcing campaign and the subsequent assignment steps. Europeana currently aggregates more than 58 million records coming from CH Institutions (CHI) across Europe, including

[1] https://www.europeana.eu/

a diverse set of audio files on the theme of Europe's Music Heritage. CH items on the Europeana platform are described via a well-defined established metadata structure, the Europeana Data Model (EDM) (Europeana Data Model, 2017), which conveys important information about the items, such as their title, free text description, creator, etc. These metadata fields are essential for the accessibility and discoverability of the rich and disparate collections made available through the Europeana platform, helping users to find and understand the objects they are interested in. It should be noted that all metadata published on the Europeana platform are licensed under a CCO license (Creative Commons Zero Universal Public Domain Dedication).

The first step towards the preparation of the case study concerned the curation of the dataset that would constitute the starting point of the crowdsourcing campaign and the subsequent assignment steps. We started by scouting the music content available on the Europeana platform through the Europeana Search API, which provides a way to search for metadata records and media on the Europeana repository and supports advanced queries and filtering.

The following selection criteria were used to guide the curation process:

- Quality of metadata that accompanied the music tracks. Metadata records on the Europeana platform often suffer from poor metadata, either due to many empty properties or inconsistent values (e.g. the EDM property "`dc:contributor`" sometimes includes composers and sometimes interpreters). In order to build an initial knowledge base that can act as a sufficiently expressive starting point for further enrichment, we filtered out metadata records that lacked information considered essential for building an initial knowledge base (e.g. information about the creator, the year of publication, etc).
- Quality and length of audio files. Audio files longer than 6 minutes were discarded, to filter out files that represented more than one music tracks (e.g. recordings of a whole concert or album) as well as to avoid assigning overly time-consuming tasks to students. The sound quality was also evaluated on sample files, which were considered indicative of the overall sound quality provided by a provider.
- Genre coverage. In order to serve the needs of the assignment and facilitate meaningful recommendations, the selection process aimed to cover a wide coverage of music genres (from classical and folk to rock and rap).

By performing a series of API queries reflecting the criteria described above and evaluating a sample of the results, we ended up using data from the following CHIs: the "Internet Archive", the "Internet Culturale / Biblioteca Nazionale Braidense - Milano", and the "Fondazione Biblioteca Europea di Informazione e Cultura (BEIC)". Eventually, 854 songs were collected. As far as the licenses of the music tracks themselves are concerned, these vary depending on the provider: the Internet Archive has collected recordings from musicians under a trade-friendly statement [2], which allows for the non-commercial exchange of the recordings; the tracks provided by the "Biblioteca Nazionale Braidense - Milano" are not restricted by copyright [3]; and the license associated with the tracks aggregated by BEIC is stated as "Preview Only", with no further information.

A post-filtering procedure on the curated metadata records was necessary since not all metadata

[2] wiki.etree.org/index.php?page=TradeFriendly

[3] rightsstatements.org/page/NoC-OKLR/1.0/?language=en

Table 1. Metadata specifications

Property name	Correspondence to EDM property	Specification
EuropeanaID	rdf:about	the music track's unique Europeana record ID
Title	dc:title	the music track's title
Year	dc:date	the year when the performance was recorded
Duration	ebucore:duration	the duration of the track in milliseconds
Composer	dc:creator	the composer of the music track
DateOfBirth	rdaGr2:dateOfBirth	the date of birth of the music track's composer
DateOfDeath	rdaGr2:dateOfDeath	the date of death of the music track's composer
Biography	rdaGr2:biographicalInformation	the biography of the music track's composer
Publisher	dc:publisher	the publisher of the music track
Place	skos:prefLabel	the place where the performance was recorded

fields included in the returned records are characterized by consistent values. Either some metadata fields were missing from the majority of records or contained values that were inconsistent with respect to the intended semantics (e.g. in some cases "`dc:description`" included information about the album of the track and in others about the location of the concert). Moreover, fields that do not contribute information that can be helpful in the framework of a knowledge system for music, such as the name of the data aggregator, have been discarded. The post-filtering process resulted in the metadata properties shown in Table 1. We used the User Gallery tool on the Europeana platform to organize the items and the Europeana Record API to retrieve the metadata records in JSON format, which we processed to create a CSV file with the filtered metadata.

In order to organize the items, we first used the User Gallery tool provided to create 3 galleries, i.e. collections of items selected by the user, on the Europeana platform, one per institution. We used Python scripts to issue requests to the Europeana Record API [4] in order to retrieve the metadata records in JSON format. Finally, we processed the returned JSON structures returned by the calls, applying the criteria described above, in order to create a CSV file with the filtered metadata. The file was used as the basic data for the purposes of the assignment.

3.2. Definition of the enrichment goals

The information conveyed by the original metadata properties sourced from the Europeana platform is quite limited, allowing only for quite basic queries and restricting the potential for their meaningful further exploitation. In order to enable higher flexibility, richer ontology structures, and

[4]https://pro.europeana.eu/page/record

more reliable recommendations, the curated dataset has to be enriched with more information which can be exploited by the later stages of the assignment (see Section 4). At the same time, the more extensive and specialized information is added, the more expert knowledge, effort, and time is required. For example, retrieving detailed information about the performance and featured artists (e.g. singers, musicians) requires dedicated research. It should also be noted that performing raw audio analysis for extracting sonic characteristics, such as "instrumentalness" or "danceability", used by established music recommendation systems (Spotify recommendation system, 2022) is beyond the scope of the specific CS course. Similarly, taking into consideration the size of the class and time constraints, the analysis of user taste profiles as a means to inform recommendations on music was not considered as part of the assignment.

Weighing in the above considerations and in order to achieve a middleground between desired richness and feasibility, the manual enrichment process focused on collecting data along the following three aspects: "Emotion", "Genre" and "Instruments". These enrichment goals were formulated as crowdsourcing tasks to be carried out by students via their participation in an appropriately designed campaign (see Section 3.3). The terms for all metadata fields and respective type of tasks correspond to Wikidata URIs and were selected based on specific criteria, as explained below. The use of Wikidata allows students to more effectively exploit the collected information in order to build more complex concepts and queries by consulting additional knowledge that can be derived from the semantic web (see Section 4).

Emotion reflects how the audience feels when listening to a music track. The use of emotion in creating music playlists is a frequently occurring concept with users and can be exploited for making meaningful music recommendations. Obviously, this is partly a subjective issue - every person perceives a music piece on their own way, although a majority of people would usually agree whether a song is melancholic or joyful. The subjective dimension of emotion is an additional reason why an aggregated opinion by the crowd can help us derive an "average" metric about what kind of emotion a song gives rise to. In order to represent emotion within music, we based on the circumplex model developed by James Russell (Russell, 1980). The model is oriented around two dimensions: arousal represents the vertical axis and valence represents the horizontal axis. The emotion values-tags that we used included: *Arousal, Joy, Pleasure, Calmness, Boredom, Sadness, Anxiety and Fear*. Their place on Russell's model is shown on Fig. 2. The main advantage of using this model is its simplicity, which entails that users being asked to express emotional ratings should find it fast and easy to engage with (Griffiths et al., 2021).

Genre is a characteristic that is commonly used to organise music tracks and is exploited by music recommendation systems, sometimes in combination with emotion (Zhu et al., 2006). Information about the genre of a music track may be included in the original Europeana metadata records under the property "`dc:subject`". However, the inspection of Europeana data led us to the conclusion that (i) this information is missing from a large number of items; (ii) that its values are frequently inaccurate; and (iii) that they do not follow a common and consistent classification system. In fact, the property "`dc:subject`" takes free text values, i.e. does not enforce selection from a controlled list of terms. As a result, genre values in Europeana metadata records are highly heterogeneous (for example, in some cases, such as "Art", they are too coarse and in others, such as "Roots Rock", too granular), making it difficult to computationally assess the correlation between different tracks. In order to derive consistent information about music genres, we decided to use a predefined terminol-

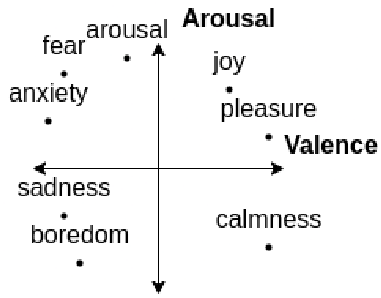


***Figure 2.* Emotion tags in circumplex model**

ogy from which students should select the most appropriate term during the campaign. Taking into account the fact that our annotators are not music experts, we selected tags that represent typical music categories (e.g. *Rock*) instead of tags with highly specific usage (e.g. *Alternative Rock*). Considering also the coverage of the curated dataset, the following controlled list of terms was used: *Pop, Rock, Country, Classical, Opera, Instrumental, Funk, Hip-hop, Reggae, Jazz*, and *Traditional Folk*.

The instruments used in a track is another important musical characteristic. Compared to the other categories mentioned above, it is the characteristic that requires the most familiarity with music. For the musical instrument annotation task, we used 12 different terms. Similarly to our approach on representing genre, we selected rather high-level terms with a broad application (e.g. *Brass* instead of *Trumpet, Trombone etc*) as well as basic instruments (such as *Piano*), whose sound can be adequately recognized even by non-experts. Considering also the most common instruments appearing in the curated dataset, we ended up with the following instrument tags: *Piano, Electric Guitar, Acoustic Guitar, Drums, Synthesizer, Violin, Harmonica, Banjo, Bass, Woodwind, Brass*.

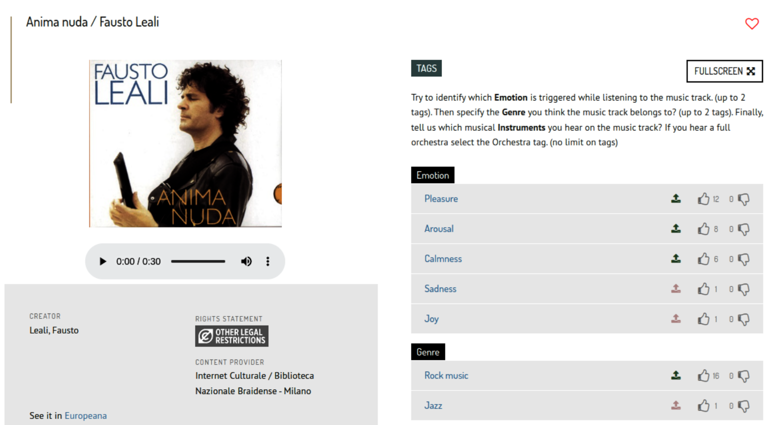


***Figure 3.* Campaign snapshot**

Orchestra was added to the list given the fact that many music tracks of our dataset are performed by a symphony orchestra and it would be difficult and time-consuming for students to discern individual instruments.

In addition to the three musical characteristics represented via controlled lists of Wikidata terms as mentioned above, students were also given the possibility to add free text comments about the music track they listened to. These comments could be exploited in later stages of the assignment, complementary to the "Emotion" tag, in order to further enrich the dataset using Natural Language Processing (NLP) techniques for sentiment analysis (see Section 4). Additionally, we gave the annotators the choice to mark music tracks they liked as "favorites", using a dedicated button on the crowdsourcing platform. This information was used to create users' profiles which helped the students to experiment with recommendations (see Section 4).

3.3. Setup of the campaign by using and extending the CrowdHeritage platform

After preparing the data and defining the enrichment objectives, the next step was to set up and run the crowdsourcing campaign that would allow students to perform the actual enrichment tasks in a well-defined and collaborative way. More specifically, students were invited to listen to the music tracks, recognize their musical characteristics, and tag the items appropriately by selecting values from the controlled lists of terms defined above. The open-source CrowdHeritage platform (Kaldeli et al., 2021a) was used to this end. The platform supports the organization of online crowdsourcing campaigns for the enrichment and validation of CH metadata. Through a user friendly interface which supports playful features such as leaderboards and rewards, users are invited to add new annotations or validate (via crowdvoting) existing ones produced either automatically by AI tools or added by other users of the platform. The platform has been used so far for the organization of multiple crowdsourcing campaigns in the CH domain, engaging different audiences (CH professionals, educators and students, CH enthusiasts, citizens etc) who have conducted various types of enrichment and validation tasks.

CrowdHeritage can parse data in the EDM format and is connected to the Europeana Search and Record APIs, thus facilitating the import of resources from the Europeana platform. Among its enrichment capabilities (e.g. color-tagging, geo-tagging), it supports the semantic annotation of records with terms from controlled vocabularies: users can add tags by typing in a dedicated text field and select from a list of suggested terms derived from a selected vocabulary, supported by an auto-complete functionality. This feature is particularly fit for performing the enrichment tasks described in section 3.2. Another particularly useful functionality refers to the validation mechanism supported by the platform, which can be seen as a means of peer-reviewing. Users can up- or down-vote existing annotations, depending on whether or not they agree with them. This validation input is further analyzed to identify questionable annotations and users with malicious or unreliable behaviour. A validation editor allows campaign organizers to review the produced annotations and to post-edit or filter them according to their criteria (e.g. based on the popularity of an annotation). These peer-validation and moderation mechanisms proved helpful for assessing the crowdsourced enrichments and for maintaining reliable end results, as explained in Section 5.1.

In order to serve the specific requirements that emerged from the needs of the case study, CrowdHeritage was supplemented with some critical novel features, which are also useful for future applications of the platform since they streamline and expand its capabilities. To begin with, the data import capability was extended to support the direct import of a custom collection curated via the Europeana portal, thus allowing us to readily retrieve the curated data described in Section 3.1 and make them available for crowdsourcing. An important limitation of the previous version of the CrowdHeritage platform referred to the expressive power of the annotation model it uses, which builds on the W3C Web Annotation Model (W3C Annotation Model, 2017). So far, annotations referred to an item as a whole and it was not possible to distinguish between tags targeting different attributes of the metadata record (e.g. emotion or genre in our case). Moreover, it was not possible to assign to users more than one annotation task grouped under the same item. To overcome these limitations, both the backend and the user interface of the platform were extended so as to enable campaign organizers to create multiple tagging modules within the scope of the same item and assign different custom terminologies/vocabularies to each module. The annotation model was also extended to support the representation of free text comments by the campaign participants. Another improvement regarded the annotation user interface, so that during semantic tagging the user is presented with the dropdown list of terms upon clicking on the textbox. In order to facilitate the building of a recommendation system, a "favorites" functionality was added, so that the user can select their preferred items (see Section 3.2). Fig. 3 provides an example of how a CH item and associated tasks are presented to the user. Lastly, the vocabulary ingestion pipeline was streamlined, so as to support the seamless upload and parsing of CSV files with terms, resulting in vocabularies that are decoupled from specific campaigns and can be reused for different purposes across the platform.

By making use of the platform's administrative functionalities, a campaign with concrete instructions was set up, which run for 18 days. The campaign setup included the import of the curated dataset; the definition of the annotation tasks by making use of the vocabularies/lists of controlled terms as defined in Section 3.2; and the specification of the campaign's overall objective, associated instructions, duration etc. The curated dataset was divided into eight sets of items with respective micro-tasks, in order to ensure balanced contributions by participants across the data. Students were advised to semantically annotate about 80 music tracks each and were encouraged to also add comments expressing additional information and their thoughts in free text. The completed campaign can be accessed here: Campaign.

4. USING THE ENRICHED DATASET TO BUILD AND QUERY A MUSIC KNOWLEDGE BASE

At first, the annotations collected from the campaign underwent a review and filtering procedure (see Section 5.1) and were then parsed and embedded as new properties to the EDM metadata records. The resulting enriched dataset was moderated (see Section 5.1) and provided to the students as a CSV file. Students were advised to transform the tabular data to a knowledge graph (Hogan et al., 2022).

The next step was to build an ontology linked with the graph using the ProtÃl'gÃl' editor (Musen, 2015). The objective of this step was to teach the students how to structure the conceptual knowledge that can be inferred from the individual track instances into a generalized semantic model (as captured by the ontology) expressed in the form of concepts and properties. For example, the concepts "Song" and "Composer" can be used to represent the set of all items-tracks and composers respectively, while the property "hasComposer" can be used to connect a song with its composer.

The transformation of the dataset into a graph associated with an accompanying ontology opened the possibility for further automatic enrichment of the data using semantic techniques and enabled the support for advanced queries. The main techniques for further automatic data enrichment introduced to the students included: (i) accessing additional knowledge from external Linked Open Data resources; (ii) applying NLP on the free text comments; and (iii) extending the ontology by creating new concepts through axioms. Regarding (i), the students were advised to exploit the Wikidata URIs included in the metadata records and use the Wikidata SPARQL endpoint [5] in order to retrieve additional information and link it to the knowledge graph's entities. For example, using the composer's name, the students could construct a SPARQL query that fetches the artistic movements that characterize this composer or the location that the composer was born. Regarding (ii), students were solicited to apply a sentiment intensity analysis model (Hutto and Gilbert, 2014) to analyze the free text comments added by students through the campaign and extract additional sentiment metadata features. Specifically, the model predicts how positive, negative or neutral a comment is. This additionally retrieved information was incorporated in the knowledge graph and used as an extra characteristic (referring to likeability) for identifying tracks that may be relevant for the user. As for (iii), the students were instructed to create novel concepts in the ontology, in order to support more expressive queries by combining existing information. For example, the concept *CalmJazzSong* can be defined via an appropriate axiom that groups together music tracks that have *Jazz* as its genre and *Calmness* as a relevant emotion; while the concept *NineteenthCenturyComposer* can be used for representing composers who were born in the nineteenth century.

At the point that the students had created a music knowledge base by linking their extended ontology with the enriched knowledge graph, they were able to apply SPARQL queries to it with the aim to identify tracks similar to a given track based on multiple criteria and thus make recommendations. The enriched information added via the crowdsourcing campaign was extensively exploited by the students and allowed them to construct complex and smarter concepts as well as SPARQL queries that can take into consideration multiple aspects that define a music track. Students experimented with different combinations of properties such as *hasGenre, hasInstrument, hasEmotion* to fetch tracks with certain characteristics. An example of a SPARQL query that returns all music tracks that have *Rock* as genre and *Joy* as emotion is shown in Fig. 4, where `ns` is the namespace for concepts and `prop` is the namespace for properties. Compound concepts of the extended ontology were used by students as parameters in their SPARQL queries in order to make them more concise. For example, the concept *JoyfulRockSong* can be defined through an axiom that groups together songs which are of genre *Rock* and have emotion *Joy*.

In order to support students in experimenting with and evaluating recommendations, some favorites lists of music tracks were created by considering some artificial users and track selections as well as the favorite lists created by students during the crowdsourcing campaign through the use of the respective functionality of the CrowdHeritage platform (see Section 3.2). By making use of these lists as a reference, students were able to make recommendations by applying similarity SPARQL queries based on the metadata features of the first song in each favorite list (a decision made to simplify the task of finding similar tracks). In this way, students were able to experiment with SPARQL queries that combined different criteria (e.g. common composer, emotion, genre), compare the re-

[5] https://query.wikidata.org/bigdata/namespace/wdq/sparql

```
select ?x
where {
?x rdf:type ns:Song.
?x prop:hasGenre ns:Rock.
?x prop:hasEmotion ns:Joy.}
```

***Figure 4.* Example of a SPARQL query.**

sults returned against the favorite lists, and get familiarized with the concepts of evaluation metrics, such as precision and recall.

5. RESULTS AND EVALUATION

Overall, the crowdsourcing campaign involved 98 participants, 68 males, and 30 females, all of whom were students of the course "Knowledge Systems and Technologies" of age 21-23 years old. Below, we provide an overview of the annotations contributed during the campaign. We then discuss the information collected via an online survey that was conducted after the completion of the campaign.

5.1. Campaign results

The campaign led to the addition of 8399 annotation tags in total, while there have been 49351 up-votes and 495 down-votes of annotations added by other users. A moderation process was necessary to review and filter out the results which were considered of questionable validity. The number of up- versus down-votes received by an annotation was used as the main criterion to assess its reliability and resolve issues of ambiguity, subjectivity, malicious or irresponsible behavior via a majority vote. The annotations' moderation took place by making use of the validation editor provided by the CrowdHeritage platform (Section 3.3), which allows campaign organizers to review the annotations produced during a campaign and filter them according to their own acceptance criteria. During the moderation process, only the two top-ranked annotations per **Emotion** and **Genre** were kept and only if these had an up- versus down-votes difference of at least two. For the **Instruments** property in particular, only values with a votes difference above five were kept. This rather strict pruning criterion was decided because many students mentioned in their feedback (see Section 5.3) that they did not have the necessary expertise to recognize musical instruments. In addition to this filtering process, the annotations of a random sample of 80 music tracks were reviewed by two music experts, who concluded that the enriched metadata were of high quality.

As a result of the post-filtering process, 5147 annotations were kept: 1248 of them refer to genre tags, 1643 to emotion tags, 1422 to instrument tags and 834 represent free text comments. In Fig. 5 the statistics of annotation tags per metadata property are presented. For the instrument annotation task, we notice that tracks with knowable and distinguishable sounds such as *Drums* and *Orchestra* are the most annotated ones. Furthermore, the distribution of the Genre tags demonstrates that *Instrumental, Rock, Classical* and *Pop* are the most dominant tags. As for the emotion property, we observe that positive emotions are the most common, a finding that confirms the bias towards positive emotions in music datasets discussed in previous work (Cano, 2017).

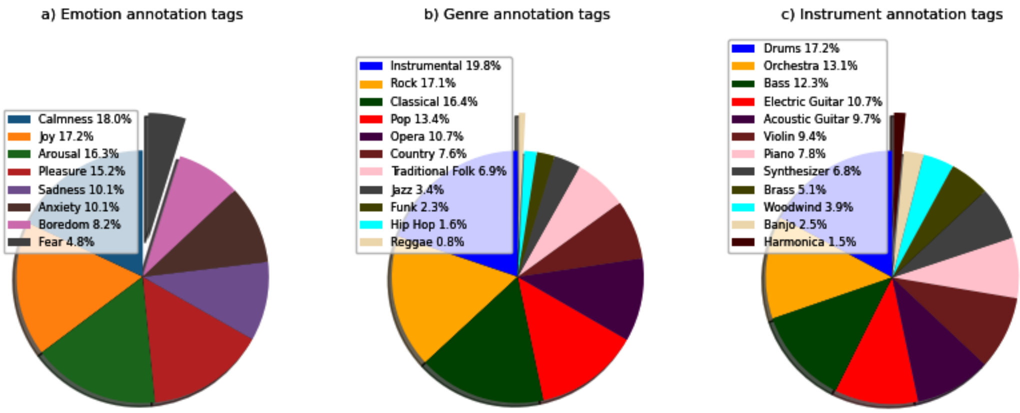


***Figure 5.* Annotation tags per category from the campaign**

In order to further assess the quality of the collected metadata, we analyzed them by using association rules. We applied the Apriori algorithm (Agrawal et al., 1994) on the total of the tag element sets in order to study their correlation. We took into account the support metric in order to observe the most frequently observed pair tags. The support metric is calculated by counting the occurrences of both tags appearing in the same set and dividing them by the total number of set. Table 2 shows which paired tags appear more frequently. We observe that the most common paired tags reflect intuitive knowledge about music (e.g. Rock-Drums, Rock-Electric Guitar, Classical-Orchestra), while paired tags connecting emotion and genre (e.g. Calmness-Instrumental) are also in accordance with prior findings (Worlu, 2017).

***Table 2.* Most frequent paired tags**

Support	Pair of Tags
0.201	Joy, Drums
0.192	Rock, Drums
0.185	Drums, Arousal
0.181	Electric Guitar, Drums
0.162	Electric Guitar, Rock
0.147	Classical, Instrumental
0.147	Classical, Orchestra
0.145	Calmness, Instrumental
0.144	Joy, Arousal
0.138	Bass, Drums
0.134	Instrumental, Orchestra

5.2. Open annotated dataset

Although the use of ML was beyond the scope of the CS course in which the case study took place, the annotated dataset that resulted from the crowdsourcing campaign and the respective post-filtering process (see Section 5.1) can be valuable for the prototyping and evaluation of MIR systems. To this end, the collected music tracks, metadata, and moderated enrichments were made openly available, so that they can be freely reused as data amenable for computational purposes (Lyberatos et al., 2023b). All three categories of enrichments-annotations (genre, instruments, emotion), as well as the properties retrieved from the EDM (see Section 3.1), are included in a single dataset. All metadata with their enrichments are made available under a CCO license.

All tracks are annotated with respect to genre, emotion, and identified instruments using the value lists described in Section 3.1. It should be noted that the rather strict filtering criteria already mentioned ensure that only annotations for which there is high certainty for their validity are maintained. The filtering based on the up-/down-voting of annotations in particular compensates for factors commonly identified as leading to poor annotations, such as inattentive labeling, listener fatigue, or other errors (Soleymani et al., 2013).

5.3. Evaluation by participants

The online survey addressed to students consisted of a combination of closed and open questions. First, we aimed to understand how the students experienced the crowdsourcing process as a part of their mini-project assignment. Relevant questions investigated: the degree to which the crowdsourcing objectives were lucid; what students identified as the main benefits of introducing a crowdsourcing campaign in the assignment; the degree and ways in which the process improved or extended students' knowledge and skills; the kind of feelings their participation gave rise to (e.g. boring, joyful, interesting etc); and the types of difficulties they experienced when performing their tasks (e.g. lack of skills, fatigue). Secondly, we aimed to collect feedback about the CrowdHeritage platform as a tool for contributing to crowdsourcing campaigns. Questions in this track focused on the overall usability of the platform; the usefulness and efficiency of different sub-components/functionalities (e.g. item view, annotations views, profile and contributions view, favorites etc); and on identifying certain shortcomings and collecting recommendations for further improvements.

35 students provided answers to the online questionnaire (5 females and 30 males). This low participation in the survey in comparison with the number of students who contributed to the campaign (36% of the campaign participants) is mainly attributed to the fact that answering the questionnaire was not seen as an integral/necessary step of the course assignment. It should be noted, however, that many students opted to use the free commenting functionality of the CrowdHeritage platform as a means to express their perceptions and provide feedback.

The objective of the campaign as well as of the overall assignment was well-understood by the students (97% described the objectives as "very clear/clear" and 3% as "clear enough"). 52% of the students described their participation experience as interesting or very interesting , 37% as neutral (neither boring nor very interesting) and 11% as boring. All students expressed that they had some knowledge gains: 77% of the students declared that their knowledge and skills were improved and expanded to a very large or large degree and 33% to a sufficient degree. 88% of the students stated that they enhanced their practical and technical skills, e.g. learned how to use certain technolog-

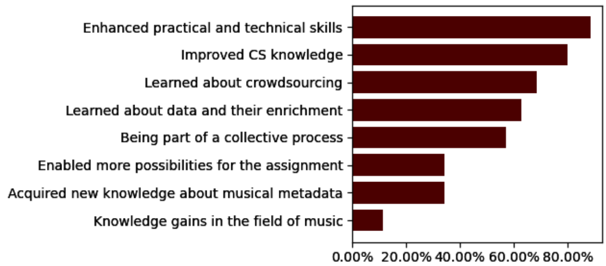


***Figure 6.* Most appreciated benefits gained by students**

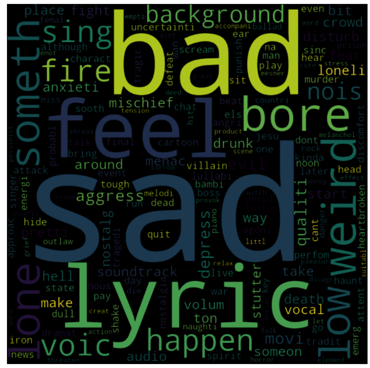
ical frameworks, and 80% that they improved their CS scientific knowledge, e.g. with respect to semantic web principles, (see Fig. 6). Highly appreciated benefits also included: learning about the potential of crowdsourcing and how this is conducted (69%); gaining a deeper understanding of the data, their shortcomings, and the value of their enrichment (63%); the participatory elements that crowdsourcing added to the assignment (56%); and facilitating, through the data enrichment, more interesting things in later stages of the assignment (34%). Only 34% stated that they acquired new knowledge about cultural and musical metadata (e.g. their structure, properties). Knowledge gains in the field of music (e.g. learning about new songs, genres, to identify instruments) were mentioned only by 11% of the students. All students declared that they would consider employing crowdsourcing as a means for data enrichment in the future.

The most commonly mentioned factor that hampered students' degree of engagement concerned the fact that certain tasks, and particularly the identification of genres and instruments, required a degree of music sophistication which many students did not possess (46% of the students encountered this difficulty). This finding is aligned with the observations made in previous work discussing music content annotation campaigns (Samiotis et al., 2021), which express the concern that crowd workers are often expected to annotate complicated music artefacts that demand certain skills that participants may lack. Moreover, the completion of all the annotations tasks expected by a user (each user was encouraged to annotate 80 items) was perceived as too time-consuming (35% of students pointed to this issue). Some students mentioned that, in some cases, the available choices from the controlled vocabulary lists were not sufficient to convey what they would like to express, while others identified some music tracks as being of poor quality.

Overall, 94% of the students agreed or strongly agreed with the fact that the CrowdHeritage platform was very usable and user-friendly. As described in the open answers, students particularly appreciated the structure of the annotations view, the presentation of available choices,and the ease of adding new tags and of up-/down-voting other users' annotations. Most criticism referred to the navigation between the different items and the need for more detailed views that allow users to inspect their contributions and progress. Some students reported that they experienced responsiveness issues, with either the sliding between the different items or the loading of certain tracks taking too long. Some of them also mentioned that it would have been useful if more tag values were made available.

Their involvement in the crowdsourcing process, which put them in the position of the contributor/end user of the CrowdHeritage platform and offered them the possibility to mobilize their

(a) Words in positive comments



(b) Words in negative comments

Figure 7. ***Wordclouds about the free text comments added by students during the crowdsourcing campaign.***

expertise in the design and use of digital technologies, allowed students to make some useful recommendations for the further improvement of the platform. Many students made various suggestions for improving and expanding the gamification elements of the platform and its incentives for engagement. They also suggested adding views which would allow users to continuousy monitor their progress. Moreover, they proposed functionalities that would enable users to interact with each other, e.g. in case of conflicting opinions on an annotation, there should be a flag and a chat window for discussing the argument.

Besides the questionnaire, the free text commenting functionality supported by the CrowdHeritage platform can also be seen as indicator of the engagement of the participants. Overall, students added 834 comments under different tasks-tracks. The high number of free comments demonstrates that students felt the need to express their thoughts besides the framework of the strictly defined enrichment tasks and thus reflects a genuine sense of involvement. Although the commenting functionality was initially intended as an extra means to collect additional information about music tracks (see Section 3.2), it was also used by students as a means to express their feelings about the music tracks and their overall experience. For example, indicative comments that students wrote include "something bad is happening", "made me uncomfortable", "inner peace", "nice vibes", "party time", "happy attitude". Fig. 7 presents in the form of wordclouds the frequency of words in the comments classified as positive and negative by the sentiment analysis model.

6. CONCLUSIONS AND DISCUSSION

The current work exemplifies crowdsourcing as a promising practice in CS-related curricula of higher education, illustrating how it can be embedded in a homework mini-project and incur benefits both for students and research. The methodology followed, the tools used, and the overall experience accrued can pave the way for embracing crowdsourcing in other frameworks within the scope of CS curricula. For example, crowdsourcing could be used to in combination with ML tasks, to enrich an ontology and its relations, or in a course on human-computer interaction, with an emphasis on the UX features that should characterize platforms used for conducting crowdsourcing tasks. The insights and recommendations for improving the CrowdHeritage platform collected by students in the current case study already point to interesting ideas towards this direction. Parts of the methodology that was followed can also be useful for educational purposes in other disciplines, besides CS, such as digital humanities. An interesting direction that can be explored in various disciplines is ways to engage students as requesters in the preparation phases of the crowdsourcing lifecycle and ask them to design their own crowdsourcing projects in order to solve a specific problem.

Revisiting the two main strands of inquiry we set out to investigate, concerning the educational and the engagement implications of crowdsourcing, the assessment of the results and the feedback received from the students point us to some interesting conclusions. Transparency and clarity about the objectives of the crowdsourcing process and its functioning in the overall flow of the CS assignment was considered crucial by the students so as to understand the relevance of the project and how their contributions would be used and their skills would improve, thus attaining their interest and investment in the project.

The knowledge gains from the crowdsourcing enrichment process are evidenced by the deep understanding which students acquired about the metadata structure and its characteristics as well as the gradual process they followed to construct a knowledge graph and an ontology of increasing richness and expressiveness. The multiple and genuine ways in which students exploited the enriched data to develop complex concepts and queries and build added-value features also attest to the conclusion that the assignment served its educational purpose. As manifested by the students' responses, what was mostly appreciated concerned competences which advanced their CS expertise. Students also got acquainted with the practical technical challenges behind crowdsourcing, especially concerning the UX features that make a platform successful. This is reflected in the apt feedback and recommendations students provided about the CrowdHeritage platform. Knowledge benefits from performing the music annotation tasks themselves, in their role as crowd workers, were much less acknowledged.

Concerning the engagement dimension, feelings appeared to be mixed. Although almost all students liked the incorporation of the crowdsourcing campaign in the assignment and perceived the platform as user-friendly, almost half of them described their participation experience as neutral or even boring. Crowdsourcing was mostly appreciated in a rather instrumental way, as a practical means to achieve an interesting end. This can be partly attributed to the quite demanding goal that was set (asking students to annotate 80 tracks each) and the fact that many students felt that certain tasks required quite advanced music sophistication that they lacked.

Even so, we cannot ignore the fact that the commonly praised participatory and affective benefits of citizen science and crowdsourcing were not the most cherished ones by students. This resonates with recent criticisms on the way in which crowdsourced citizen science is touted as an enjoyable and participatory experience, while at the same time its labor ramifications and the repetitive or mundane nature of the crowdsourced tasks are understated (Kidd, 2018; Del Savio et al., 2016). Further experiments and more in-depth evaluation in higher education settings is required to shed more light on this aspect.

Although the current study lays its primary focus on the role and impact of crowdsourcing within the CS higher education community, the publication of the carefully filtered annotated dataset is also an important outcome that can prove helpful for the research and ML communities. An inspection of the annotations' characteristics allowed us to draw some useful insights concerning the human subjects' behavior, the correlation between tags, and the overall annotations' quality. Further work is required in order to yield the dataset readily amenable for the development and evaluation of MIR models. An expansion of the dataset (e.g. more music tracks covering different genres) would enhance and widen its usefulness for computational models. Further data reliability analysis (e.g. to extract diversity measures, agreement likelihood etc) and experimentation is required to demonstrate the dataset's validity and possible usages and to establish a benchmark for the MIR community.

Building on the practical experience we gained and the feedback we collected from the students, the current case study allows us to draw some recommendations, which can be proved useful for campaign designers who seek to engage students and educators who wish to incorporate crowdsourcing as part of their curricula, in CS and beyond:

– *The crowdsourcing setup should fit naturally in the objectives of the course and clearly explained to the students.* The foremost motivation of students attending a course and taking over a project assignment is to improve skills relevant to the course's stated objectives. This consideration should be the primary guiding principle for designing the crowdsourcing process and task, so as to maintain constructive participation on behalf of the students. The tasks should be carefully designed so that they are meaningful within the scope of the particular course and the role of crowdsourcing should be well-defined and well-explained to the students. Connecting to a real-world problem that relates to informatics students is important for making the overall project/assignment purposeful.
– *The crowdsourcing tasks should not be too cumbersome.* Granted that the more effort is invested in a crowdsourcing task, the better and more utilizable the end result is. However, the fact that crowdsourcing entails (often repetitive) labor should not be overlooked. Therefore, educators should maintain an appropriate balance between achieving an end result that is of sufficient quality and quantity and avoiding frustration on behalf of the students. Crowdsourcing with its collaborative and playful elements can stimulate additional engagement and motivation, yet, if the tasks are too time-consuming or difficult for students to complete, students' interest and engagement can be compromised. This is particularly so given that in educational settings, intrinsic motivation is the principal driving force for participation. Choosing tasks that are close to students' interests and assessing the required time for completing certain tasks in advance can help keep up engagement.
– *Emphasis should be given to crowdsourcing as a process and not just as a task in itself.* The learning benefits for students lie less on completing some specific crowdsourcing task (e.g. tagging, validating) and more on familiarising themselves with crowdsourcing as a methodology and understanding its possible uses and associated challenges. Engaging students in conversations about aspects pertaining to this methodology, from the steps it encompasses to the design of the digital tools that are used, can lead to fruitful learning outcomes. Depending on the course's topic, the focus of such discussions can vary. Asking students in particular to make specific suggestions for improving the digital tools can open up many directions for follow-up discussions.
– *Particular attention should be paid to data curation and preparation.* The effort for selecting, collecting, and cleaning the data to be used should not be underestimated, since it has to satisfy multiple criteria, which depend on the project's focus. Criteria to consider include quality (which

may refer to different aspects depending on the application, e.g. may refer to image quality; metadata etc); quantity (which should be sufficient within the project's scope to draw useful conclusions); format (amenable for computational purposes, that is be in appropriate formats to allow its parsing and analysis by students); licenses. Depending on the course's objectives, the tasks of data sourcing, curation and processing could be designed as an assignment for students with potentially useful learning outcomes (e.g. interconnect with APIs, familiarisation with data formats and processing etc)

– *It is of added-value if the results of the crowdsourcing are made openly available and have an impact beyond the course.* Depending on the crowdsourcing objectives, the results can be useful for researchers, business or end users of certain platforms. In particular, crowdsourcing provides an excellent opportunity for demonstrating how educational institutions and students can contribute to open science. By knowing that the crowdsourcing process will lead to reusable results that can be helpful for the research community or a particular sector (e.g. as training data, as data published in a digital library etc), students and educators feel that their efforts have a value that goes beyond the narrow scope of the particular course and acquire extra motivation. Moreover, such an approach sets the floor for teaching to students the importance of open data and compliance with FAIR (Findable, Accessible, Interoperable, Reproductible) principles in a hands-on way. Some additional work may be required to this end (e.g. appropriate post-processing and packaging of results, communication with interested stakeholders).

ACKNOWLEDGMENTS

The work is co-funded by the Erasmus+ Program of the European Union, project number 2020-1-BE02-KA203-074727. We thank the students of the course "Knowledge Systems and Technologies" for their contributions to the crowdsourcing campaign and valuable feedback. Part of this article was first published in Lecture Notes on Data Engineering and Communications Technologies (Vol. 190, pp. 224-240, 2024, Springer Nature). The extended content features additional material in the introduction as well as related work in digital humanities, MIR, and crowdsourcing. It also offers a more detailed account of the dataset curation, the CrowdHeritage platform enhancements, and the student recommendation system, along with extra figures.

7. REFERENCES

Agrawal, R, Srikant, R, and others, . (1994). Fast algorithms for mining association rules. In *Proc. 20th int. conf. very large data bases, VLDB*, Vol. 1215. Santiago, Chile, 487–499.

Alenezi, H and Faisal, H. (2020). Utilizing crowdsourcing and machine learning in education: Literature review. *Education and Information Technologies* 25 (07 2020), 1–16. DOI:http://dx.doi.org/10.1007/s10639-020-10102-w

Aljanaki, A, Wiering, F, and Veltkamp, R. C. (2016). Studying emotion induced by music through a crowdsourcing game. *Information Processing & Management* 52, 1 (2016), 115–128.

Blumenfeld, P. C, Soloway, E, Marx, R. W, Krajcik, J. S, Guzdial, M, and Palincsar, A. (1991). Motivating project-based learning: Sustaining the doing, supporting the learning. *Educational psychologist* 26, 3-4 (1991), 369–398.

Bogdanov, D, Porter, A, Schreiber, H, Urbano, J, and Oramas, S. (2019)a. The acousticbrainz genre dataset: Multi-source, multi-level, multi-label, and large-scale. In *Proceedings of the 20th Conference of the International Society for Music Information Retrieval (ISMIR 2019): 2019 Nov 4-8; Delft, The Netherlands.[Canada]: ISMIR; 2019.* International Society for Music Information Retrieval (ISMIR).

Bogdanov, D, Won, M, Tovstogan, P, Porter, A, and Serra, X. (2019)b. The MTG-Jamendo Dataset for Automatic Music Tagging. In *Machine Learning for Music Discovery Workshop, International Conference on Machine Learning (ICML 2019)*. Long Beach, CA, United States. http://hdl.handle.net/10230/42015

Chen, Z and Luo, B. (2014). Quasi-crowdsourcing testing for educational projects. In *Companion Proceedings of the 36th International Conference on Software Engineering*. 272–275.

Choi, K, Fazekas, G, and Sandler, M. (2016). Automatic tagging using deep convolutional neural networks. *arXiv preprint arXiv:1606.00298* (2016).

Costa, Y. M, Oliveira, L. S, Koericb, A. L, and Gouyon, F. (2011). Music genre recognition using spectrograms. In *2011 18th International conference on systems, signals and image processing*. IEEE, 1–4.

Del Savio, L, Prainsack, B, and Buyx, A. (2016). Crowdsourcing the human gut: Is crowdsourcing also 'citizen science'? *Journal of Science Communication* 15, 3 (2016), A03.

Europeana Data Model (2017). https://pro.europeana.eu/page/edm-documentation. (2017). Accessed: 2022-06-30.

Europeana Transcribathon platform (2019). europeana.transcribathon.eu. (2019). Accessed: 2022-06-30.

Farasat, A, Nikolaev, A, Miller, S, and Gopalsamy, R. (2017). CROWDLEARNING: Towards Collaborative Problem-Posing at Scale. In *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale (L@S '17)*. Association for Computing Machinery, New York, NY, USA, 221âĂŞ224. DOI:http://dx.doi.org/10.1145/3051457.3053990

Gomes, C, Schneider, D, Moraes, K, and de Souza, J. (2012). Crowdsourcing for music: Survey and taxonomy. In *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. 832–839. DOI:http://dx.doi.org/10.1109/ICSMC.2012.6377831

Griffiths, D, Cunningham, S, Weinel, J, and Picking, R. (2021). A multi-genre model for music emotion recognition using linear regressors. *Journal of New Music Research* 50, 4 (2021), 355–372. DOI:http://dx.doi.org/10.1080/09298215.2021.1977336

Guo, H, Ajmeri, N, and Singh, M. P. (2018). Teaching crowdsourcing: an experience report. *IEEE Internet Computing* 22, 6 (2018), 44–52.

Heigl, F and Zaller, J. G. (2014). Using a citizen science approach in higher education: a case study reporting roadkills in Austria. *Human Computation* 1, 2 (2014).

Hills, T. T. (2015). Crowdsourcing content creation in the classroom. *Journal of Computing in Higher Education* 27, 1 (2015), 47–67.

Hogan, A, Blomqvist, E, Cochez, M, D'amato, C, Melo, G. D, Gutierrez, C, Kirrane, S, Gayo, J. E. L, Navigli, R, Neumaier, S, Ngomo, A.-C. N, Polleres, A, Rashid, S. M, Rula, A, Schmelzeisen, L, Sequeda, J, Staab, S, and Zimmermann, A. (2022). Knowledge Graphs. *Comput. Surveys* 54, 4 (may 2022), 1–37. DOI:http://dx.doi.org/10.1145/3447772

Humphrey, E, Durand, S, and McFee, B. (2018). OpenMIC-2018: An Open Data-set for Multiple Instrument Recognition.. In *ISMIR*. 438–444.

Hutto, C and Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, Vol. 8. 216–225.

Jiang, Y, Schlagwein, D, and Benatallah, B. (2018). A Review on Crowdsourcing for Education: State of the Art of Literature and Practice. *PACIS* (2018), 180.

Kaldeli, E, Menis-Mastromichalakis, O, Bekiaris, S, Ralli, M, Tzouvaras, V, and Stamou, G. (2021)a. CrowdHeritage: Crowdsourcing for Improving the Quality of Cultural Heritage Metadata. *Information* 12, 2 (2021). DOI:http://dx.doi.org/10.3390/info12020064

Kaldeli, E, Menis-Mastromichalakis, O, Bekiaris, S, Ralli, M, Tzouvaras, V, and Stamou, G. (2021)b. CrowdHeritage: Crowdsourcing for Improving the Quality of Cultural Heritage Metadata. *Information* 12, 2 (2021). DOI:http://dx.doi.org/10.3390/info12020064

Kasumba, R. (2022). *Application of Crowdsourcing and Machine Learning to Predict Sentiments in Textual Student Feedback in Large Computer Science Classes*. Master's thesis. Washington University in St. Louis.

Khan, V.-J, Papangelis, K, and Markopoulos, P. (2020). Completing a Crowdsourcing Task Instead of an Assignment; What do University Students Think?. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–8.

Kidd, J. (2018). Public heritage and the promise of the digital. *The Oxford Handbook of Public Heritage Theory and Practice* (2018), 197–208.

Kloetzer, L, Lorke, J, Roche, J, Golumbic, Y, Winter, S, and Jõgeva, A. (2021). Learning in citizen science. *The science of citizen science* 283 (2021).

Lambert, N. M and McCombs, B. L. (1998). *How students learn: Reforming schools through learner-centered education.* American Psychological Association.

Laurier, C, Meyers, O, SerrÃă, J, Blech, M, and Herrera, P. (2009). Music Mood Annotator Design and Integration. *Content-Based Multimedia Indexing, International Workshop on* 0 (06 2009), 156–161. DOI:http://dx.doi.org/10.1109/CBMI.2009.45

Liu, X, Chen, Q, Wu, X, Liu, Y, and Liu, Y. (2017). CNN based music emotion classification. *arXiv preprint arXiv:1704.05665* (2017).

Lyberatos, V, Kantarelis, S, Dervakos, E, and Stamou, G. (2023)a. Perceptual Musical Features for Interpretable Audio Tagging. *arXiv preprint arXiv:2312.11234* (2023).

Lyberatos, V, Kantarelis, S, Kaldeli, E, Bekiaris, S, Tzortzis, P, Mastromichalakis, O. M, and Stamou, G. (2023)b. Employing Crowdsourcing for Enriching a Music Knowledge Base in Higher Education. (2023).

Mikhailova, E. A, Post, C. J, Zurqani, H. A, and Younts, G. L. (2022). Teaching Field Data Crowdsourcing Using a GPS-Enabled Cellphone Application: Soil Erosion by Water as a Case Study. *Education Sciences* 12, 3 (2022). DOI:http://dx.doi.org/10.3390/educsci12030151

Musen, M. A. (2015). The protégé project: a look back and a look forward. *AI Matters* 1, 4 (2015), 4–12. DOI:http://dx.doi.org/10.1145/2757001.2757003

Owens, T. (2013). Digital cultural heritage and the crowd. *Curator: The Museum Journal* 56, 1 (2013), 121–130.

Panda, R, Malheiro, R, and Paiva, R. P. (2018). Musical texture and expressivity features for music emotion recognition. In *19th International Society for Music Information Retrieval Conference (ISMIR 2018*. 383–391.

Pirttinen, N, Kangas, V, Nikkarinen, I, Nygren, H, Leinonen, J, and Hellas, A. (2018). Crowdsourcing Programming Assignments with CrowdSorcerer. In *Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE 2018)*. Association for Computing Machinery, New York, NY, USA, 326âĂŞ331. DOI:http://dx.doi.org/10.1145/3197091.3197117

Pirttinen, N and Leinonen, J. (2021). Exploring the Complexity of Crowdsourced Programming Assignments. In *Seventh SPLICE Workshop at SIGCSE 2021 âĂIJCS Education Infrastructure for All III: From Ideas to PracticeâĂİ*.

Roche, J, Bell, L, Galvão, C, Golumbic, Y. N, Kloetzer, L, Knoben, N, Laakso, M, Lorke, J, Mannion, G, Massetti, L, and others, . (2020). Citizen science, education, and learning: challenges and opportunities. *Frontiers in Sociology* 5 (2020), 613814.

Russell, J. (1980). A Circumplex Model of Affect. *Journal of Personality and Social Psychology* 39 (12 1980), 1161–1178. DOI:http://dx.doi.org/10.1037/h0077714

Samiotis, I. P, Qiu, S, Lofi, C, Yang, J, Gadiraju, U, and Bozzon, A. (2021). Exploring the Music Perception Skills of Crowd Workers. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* 9, 1 (Oct. 2021), 108–119. https://ojs.aaai.org/index.php/HCOMP/article/view/18944

Shanley, L, Parker, A, Schade, S, and Bonn, A. (2019). Policy Perspectives on Citizen Science and Crowdsourcing. *Citizen Science: Theory and Practice* 4 (12 2019). DOI:http://dx.doi.org/10.5334/cstp.293

Shreevathsa, P, Harshith, M, M., A. R, and Ashwini, . (2020). Music Instrument Recognition using Machine Learning Algorithms. In *2020 International Conference on Computation, Automation and Knowledge Management (ICCAKM)*. 161–166. DOI:http://dx.doi.org/10.1109/ICCAKM46823.2020.9051514

Simpson, R, Page, K. R, and De Roure, D. (2014). Zooniverse: observing the world's largest citizen science platform. In *Proceedings of the 23rd international conference on world wide web*. 1049–1054.

Soleymani, M, Caro, M. N, Schmidt, E. M, Sha, C.-Y, and Yang, Y.-H. (2013). 1000 songs for emotional analysis of music. In *Proceedings of the 2nd ACM international workshop on Crowdsourcing for multimedia*. 1–6.

Spotify recommendation system (2022). www.music-tomorrow.com/blog/how-spotify-recommendation-system-works-a-complete-guide-2022. (2022). Accessed: 2022-06-30.

Stamou, G and Chortaras, A. (2017). Ontological query answering over semantic data. In *Reasoning Web International Summer School*. Springer, 29–63.

Vance-Chalcraft, H. D, Hurlbert, A. H, Styrsky, J. N, Gates, T. A, Bowser, G, Hitchcock, C. B, Reyes, M. A, and Cooper, C. B. (2022). Citizen Science in Postsecondary Education: Current Practices and Knowledge Gaps. *BioScience* 72, 3 (2022), 276–288.

W3C Annotation Model (2017). www.w3.org/TR/annotation-model. (2017). Accessed: 2022-06-30.

Wang, L. (2016). Employing Wikibook project in a linguistics course to promote peer teaching and learning. *Education and Information Technologies* 21, 2 (2016), 453–470.

Worlu, C. (2017). Predicting listener's mood based on music genre: an adapted reproduced model of Russell and Thayer. *Journal of Technology Management and Business* 4, 1 (2017).

Zahirović Suhonjić, A, Despotović-Zrakić, M, Labus, A, Bogdanović, Z, and Barać, D. (2019). Fostering studentsâĂŹ participation in creating educational content through crowdsourcing. *Interactive Learning Environments* 27, 1 (2019), 72–85.

Zhang, K, Zhang, H, Li, S, Yang, C, and Sun, L. (2018). The PMEmo Dataset for Music Emotion Recognition. In *Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval (ICMR '18)*. Association for Computing Machinery, New York, NY, USA, 135âĂŞ142. DOI:http://dx.doi.org/10.1145/3206025.3206037

Zhu, X, Shi, Y.-Y, Kim, H.-G, and Eom, K.-W. (2006). An integrated music recommendation system. *IEEE Transactions on Consumer Electronics* 52, 3 (2006), 917–925. DOI:http://dx.doi.org/10.1109/TCE.2006.1706489

Cano, E. (2017). Crowdsourcing Emotions in Music Domain. *International Journal of Artificial Intelligence & Applications* 8 (07 2017), 25–40. DOI:http://dx.doi.org/10.5121/ijaia.2017.8403