.

Developing a Comprehensive Dataset for Enhancing Social Inclusion and Cohesion through Citizen Curation in Cultural Heritage

Manuel Striani[1,2,*,†], Belen Diaz-Agudo[5,†], Guillermo Jimenez-Diaz[5,†], Antonio Lieto[3,4,†] and Rossana Damiano[1,†]

1Computer Science Department, University of Turin, Turin 10149, Italy

2Computer Science Instituite DiSIT - University of Piemonte Orientale, Alessandria 15121, Italy

3Department of Political Social and Communication Sciences, CIIT Lab, University of Salerno, Fisciano 84084, Italy

4ICAR-CNR, Palermo, Italy

5Facultad de Informática, Universidad Complutense de Madrid, Spain

Abstract

This article presents a dataset developed under the EU H2020 SPICE (Social Cohesion, Participation, and Inclusion through Cultural Engagement) project. The dataset serves as a tool for identifying and examining communities that emerge from citizens' interactions with cultural heritage, capturing key representations of individuals and groups to reveal unexpected connections. An example from a SPICE case study at GAM (Galleria Civica d'Arte Moderna e Contemporanea) in Turin, Italy, illustrates the dataset's structure, focusing on the interpretation of artworks, with particular attention to the deaf community's emotional responses. The dataset primarily organizes diverse perspectives, values, and emotions expressed through non-transitive relationships. Initially designed for analyzing narrative identities within museum audiences and communities, the dataset has potential applications in education, social work, and community building.

Keywords

Cultural Heritage, Dataset for CH, Citizen Curation, Community Modelling, Museum Interaction, Social Cohesion

1. Introduction and motivations for the data collection

The SPICE project (Social Cohesion, Participation, and Inclusion through Cultural Engagement) seeks to develop tools and methods for "citizen curation", a process that empowers individuals to create and share personal interpretations and reflections on heritage objects. This is made possible through an integrated technological infrastructure [1], enabling citizens to apply curatorial practices to cultural heritage items across various institutions [2, 3]. SPICE's ultimate goal is to enhance social cohesion among diverse groups by offering citizen curation activities that encourage individuals to express their interpretations of cultural heritage while engaging at the same time with the perspectives of others. Such exchanges foster a deeper understanding and connection among different communities. Recognizing that museums operate within broader socio-cultural systems [4], the project aspires to bridge the gap between the "cultural heritage space" and the "citizen space".

The methodological framework developed within SPICE is aimed at stimulating the processes of interpretation and reflection in the cultural heritage domain. In the SPICE project, the processes of inter-

pretation and reflection are closely interconnected: the project introduces a continuous Interpretation Reflection Loop (IRL), which serves as a model for linking interpretation and reflection activities across different phases and components of the SPICE digital platform. The SPICE platform provides users with various citizen curation activities, such as selecting artifacts, tagging, and sharing personal stories and opinions. Through user-friendly interfaces, citizens are encouraged to contribute and share their rich interpretations of the cultural heritage artifacts they encounter. These contributions are then analyzed by the system to promote reflection. The platform may suggest new activities, offer recommendations, and present users with alternative perspectives, enabling them to explore and reflect on both their own contributions and those of others. This creates an inclusive and participatory interpretation reflection loop.

In this paper, we describe a dataset collected during the conduction of the SPICE case study carried out it Turin from May 2020 to April 2023 in cooperation with Galleria Civica d'Arte Moderna e Contemporanea, GAM [1]), in which we developed a web app for citizen curation called GAMGame. The GAMGame, inspired by theories of "cultural narrative identity" [5, 6], allowed the users to create and share stories using the artworks in the GAM collections, as described extensively in [7, 8]. In the context of the IRL, narrative identity explains how different emerging groups, which may share explicit attributes, overlap in complex and non-linear ways (heterarchically), representing more nuanced and emerging narrative identities.

2. Interpretation and Reflection processes in SPICE

Using the GAMGame, users can create and share stories from the museum artworks in a simple intuitive. To create a story, the user select and added the artworks from the artwork gallery; when an artwork is added, the applications requires the user to annotate it with "something personal" using hashtags, emojis and short text templates. When using the app, the users can receive affect-based recommendations about artworks and stories which are aimed at expanding diversity in story creation and exploration.

The dataset collected during the testing of the GAMGame has been integrated into the VISIR software tool, which helps identify and explore the communities emerging from citizens' interactions with cultural heritage through a visual interface [9]. The VISIR visualization tool [9] is crucial for supporting the exchange of different viewpoints within and across citizen groups. These tools help identify and highlight unexpected new groups that emerge from interactions with cultural heritage, showcasing their similarities, differences, and relationships. For curators, this involves recognizing and interpreting these emerging groups in ways that can be effectively communicated to various audiences.

The dataset includes the following attributes:

- **Citizen demographic attributes**: for citizens, 5 different attributes were collected from the users of the GAMGame through online forms: *relationshsip_with_arts* encodes the citizen's interest in art; *relationship_with_museums* encodes the frequency with which the citizen visit museums; the binary attribute *interest_in_LIS* encodes the interest in contents in Italian Sign Language (Lingua Italiana dei Segni, Italian Sign Language) (LIS); *gender* (male, female and non-binary) and *age range* are standard demographic variables.
- **Artworks attributes**: this set of attributes includes both standard museum catalogue metadata, such as *title*, *technique*, *dimension* (*size_height*, *size_width*, *size_depth*), *collection*, *year* (including *artwork_start_date* and *artwork_end_date*) and *type*, and a set of attributes added by the museum staff to investigate the visitors' relationship with the artworks. In particular, based on the experience gathered during museum labs with schools and communities, the latter include information on the artwork's *subject*, *materials*, and artistic movement (*artwork_artistic_movement*), since the visitors tend to engage with the subject depicted by the artwork, the artistic movement it belongs to, and to the materials that compose it; since visitors are sometimes aware of the artist's main biographic data, specific attributes for the artist's nationality (*artist_country*), birth and death date
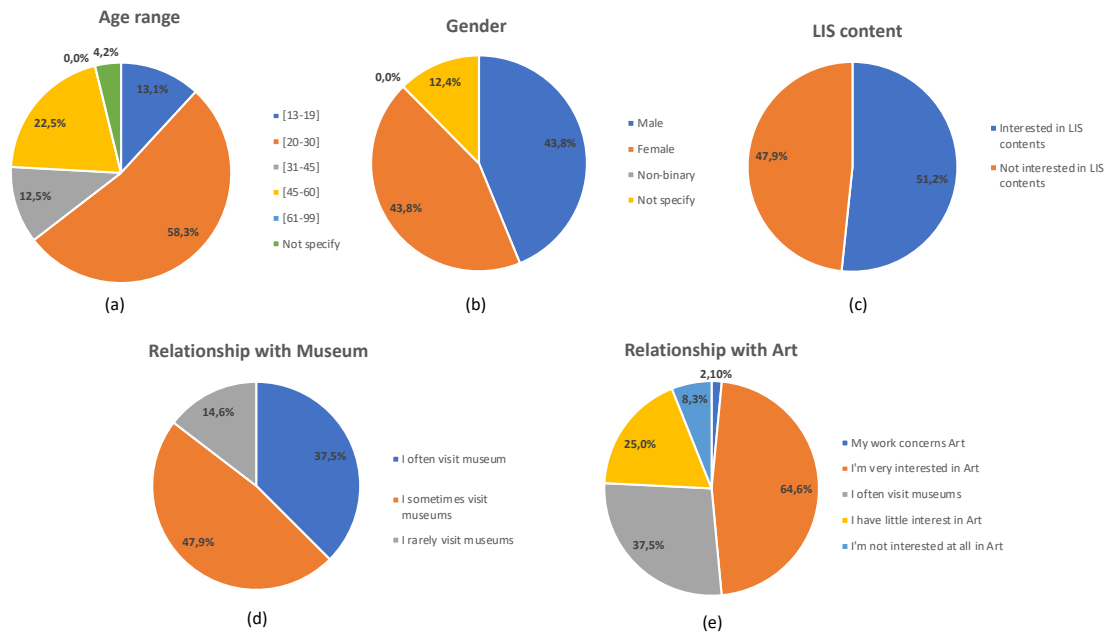

Figure 1: Statistics on user attributes in GAMGame dataset

(*artist_birth_date*, *artist_death_date*), and *gender* were also included. To avoid arbitrariness in identifying the subject values, we resorted to an authoritative, standard resource, the Iconclass[2] classification of iconographic subjects. Iconographic subjects are organized into 9 main subject types (Abstract, Non-representational Art; Religion and Magic; Nature; Human Being, Man in General; Society, Civilization, Culture; Abstract Ideas and Concepts; History; Bible; Literature; Classical Mythology and Ancient History), further subdivided into more specific types of varying specificity (e.g., "adolescent, young woman, maiden" or "sea (seascape)", recurring subjects in the GAM collection).

- **Interpretation attributes**: for each artwork, a list of emotion labels was extracted from the curatorial notes, and from the stories generated by the users through the GAMGame. Curatorial notes and user-generated text in English and Italian (comments and tags) were fed into the semantic analysis and reasoning pipeline of SPICE, which relies on Plutchik's theory of emotions [10], as described in [11]. The pipeline encompasses DEGARI 2.0 reasoning system [7, 8] to identify fine-grained, complex emotions from can use basic emotions. An example of this JSON-LDH which represents the user-interaction with GAM's collections is described in Figure 2.

The enrichment of the catalogue records with the attributes described above (Artwork attributes) was conducted on a set 56 items included in the GAMGame installation created for the pilot. The rational behind the selection consisted in presenting the users with a variety of artworks by author, type, style and time period, so as to avoid any biases towards specific artwork types. This choice is reflected by the value distributions of the artwork attributes, which encompass 12 different types of material, 26 artistic movements, 21 techniques, 4 artwork types and artists from 10 different countries.

During the European Researchers' Night at University of Torino in 2022, SPICE researchers collected 149 stories from 49 casual users who volunteered to use the GAMGame app. Figure 1 shows the pie chart statistics on the user attributes: users are evenly distributed according to gender (about 43.8% of male and female)(Figure 1 (b)); most of them are aged between 20 and 30 years old (about 58.3%) (Figure 1 (a)); they have a strong interest in receiving contents in Italian Sign Language (Lingua Italiana dei Segni, Italian Sign Language (LIS)) (52.1%) (Figure 1 (c)) and in art (64.6%) (Figure 1 (e)). Finally,

---

37.7% users stated that they often visit museums and art galleries (Figure 1 (d)). Concerning the stories, they contained 56 distinct artworks, i.e., all artworks have been selected by the users at least once. The stories contained overall 404 artworks, which yields an average story length of 2.7 artworks.

3. Description of format and schema

The dataset has been designed and implemented to be managed by a NoSQL Database Management System (DBMS), stored and distributed across multiple servers within a Linked Data Hub (LDH) system. This approach ensures a more flexible data model compared to relational databases, allowing for greater scalability in service utilization, higher availability and fault tolerance. The dataset consists of three JSON files:

- **artworks** - This dataset contains detailed records of the artworks and includes metadata such as titles, authors, creation dates, dimensions, materials, techniques, and collections they belong to. Each entry also includes descriptions, artistic movements, associated emotions, and other relevant classifications.
- **ugcUsers** - This dataset contains information about users who contribute content related to the artworks. This may include their profiles, activities, contributions, and their interactions with the artworks in the database. It can be used to track the engagement of different users with the cultural heritage content.
- **ugcContributions** - This dataset contains the contributions made by GAMGame users. These include comments, emotions, discussions, and other forms of user-generated content associated with the artworks. This file links the users (from ugcUsers.json) with their contributions, providing insights into how communities interact with and contribute to the database. An example of this entry is described in Figure 2.

Together, these datasets provide a snapshot of the interaction between users and cultural heritage in citizen curation, allowing for a deeper understanding of the engagement with the artworks through cultural narratives.

The first dataset in JSON format named "artworks.json" contains detailed records of artworks, primarily from a cultural heritage of GAM's collection. Each entry in this dataset includes various attributes related to individual artworks. Here's a breakdown of the key components found in the file:

1. **Basic Information**:
   - ID: A unique identifier for each artwork.
   - Title: The title of the artwork in GAM's Museum.
   - Inventory Number: A unique code identifying the artwork within the collection.
   - Collection: The specific collection or exhibition the artwork is part of.

2. **Artwork Details**:
   - Author: The artist who created the artwork.
   - Year: The year the artwork was created.
   - Material and Technique: The materials used and the technique applied to create the artwork.
   - Dimensions: The physical dimensions of the artwork, including height, width, and depth.
   - Acquisition: Information on how and when the artwork was acquired by the collection.

3. **Descriptive Information**:
   - Description: A narrative description providing context, historical significance, or a detailed analysis of the artwork.
   - Image URL: A link to an image of the artwork.
   - Artistic Movement: The art movement or style to which the artwork belongs.

4. **Artist Information**:
   - Birth and Death Dates: The lifespan of the artist.
   - Country: The nationality of the artist.

- Gender: The gender of the artist.

5. **Emotions and Iconography**:

   - Extracted Emotions: Emotions associated with or evoked by the artwork, as identified by curators or through our system DEGARI 2.0 for emotions extraction.
   - Iconclass Subjects: Iconographic subjects and themes depicted in the artwork, classified according to the Iconclass system (ICONCLASS API[3] for artworks)

6. **Additional Classifications**:

   - Artwork Type: The type of artwork, such as painting, sculpture, etc.
   - Technique: Specific techniques used in the creation of the artwork.
   - Materials: The materials used in the artwork.

The second JSON dataset ('ugcUsers.json') contains user demographic data related to their interactions with a cultural heritage or art-related application. Here's a summary of the key elements found in the dataset:

1. **User Identifiers**:

   - Each user in the dataset is represented by a unique 'userid'. This identifier is used to group the different pieces of demographic data related to a specific user.

2. **Demographic Information**: The dataset includes several demographic attributes for each user, such as:

   - Gender: Recorded as values like "Male "Female" or "Not specified".
   - Age: Age ranges such as "20-30 age", "31-45 age", "45-60 age", etc.
   - Relationship With Art: Indicates the user's interest in art, with values like "I have a strong interest in art", "Art interests me little", or "I have no interest in art",
   - Relationship With Museum: Describes the user's frequency of museum visits, with values like "I often visit museums", "I visit museums and exhibitions from time to time", or "I rarely visit museums and exhibitions.
   - Content in LIS (Italian Sign Language): Indicates the user's interest in content provided in Sign Language, such as "I'm not interested in sign language content" or "I would like to see content in Sign Language (LIS)".
   - Deaf: A binary value ("Yes" or "No") indicating whether the user is deaf.

3. **Context and Category**:

   - Each demographic entry is associated with a 'context', typically set to "application", which likely refers to the context in which this data was collected.
   - The 'category' for all entries is "demographics," highlighting that this dataset focuses on user demographic information.

4. **Data Structure**:

   - The data is structured with each 'userid' being a key that maps to an array of demographic entries. Each entry in this array includes fields like 'id', 'source_id', 'source', 'pname' (parameter name), 'pvalue' (parameter value), 'context', and 'datapoints'.

This dataset can be used to analyze user demographics and their relationship with art and museums, as well as their preferences for accessible content like LIS. This information could be valuable for tailoring cultural heritage experiences or exhibitions to different audience segments.

---

[3]https://iconclass.org/help/api

Table 1

Description of the JSON Dataset "artworks.json".

Field Name	Type	Description
id	String	A unique identifier for each artwork entry in the GAM's collection.
DateOfLastModify	String (DateTime)	The date and time when the record was last modified.
title	String	The title of the artwork.
Inventory	String	The inventory number of the artwork within the GAM's collection.
Collection	String	The specific collection or exhibition the artwork is part of.
Sumbject	String	The subject of the artwork (often left blank or empty).
author	String	The name of the artist who created the artwork.
year	String (Year)	The year when the artwork was created.
Material_and_echnique	String	The materials and techniques used in the creation of the artwork.
Dimension	String	The physical dimensions of the artwork, including height, width, and depth.
Definizione	String	Additional definitions or classifications of the artwork (often left blank).
Acquisizione	String	Information on how and when the artwork was acquired by the collection.
image	String (URL)	A URL linking to an image of the artwork.
description	String	A narrative description providing context, historical significance, or a detailed analysis of the artwork.
degari extracted emotions	Array of Strings	Emotions associated with or evoked by the artwork, as identified by curators or through analysis.
Artwork start date	String (Date)	The start date of the artwork's creation.
Artwork end date	String (Date)	The end date of the artwork's creation (often left blank).
Artist birth date	String (Date)	The birth date of the artist.
Artist death date	String (Date)	The death date of the artist (if applicable).
Gender	String	The gender of the artist.
Artist country	String	The country of origin or nationality of the artist.
Artist secondary country	String	A secondary country associated with the artist (if applicable)
Artwork Artistic Movement	String	The primary artistic movement or style to which the artwork belongs.
Secondary Artwork Artistic Movement	String	A secondary artistic movement or style associated with the artwork (if applicable).
Technique	String	Specific techniques used in the creation of the artwork.
Artwork type	String	The type of artwork, such as painting, sculpture, etc.
Size unity	String	The unit of measurement for the artwork's dimensions (e.g., cm).
Size height	Number	The height of the artwork.
Size width	Number	The width of the artwork.
Size depth	Number	The depth of the artwork (if applicable).
Materials	Array of Strings	A list of materials used in the artwork.
Iconclass subjects curators	Array of Strings	Iconographic subjects and themes depicted in the artwork, classified according to the Iconclass system.
iconclassIDString	String	A string representing the Iconclass IDs associated with the artwork's themes and subjects.
iconclassArrayIDs	Array of Strings	An array of Iconclass IDs linked to the artwork.
ApproxYear	Number	The approximate year of the artwork's creation.
Decade	String	The decade during which the artwork was created.

```json
1   {
2       "itMakesMeThinkAbout.emotions":{
3           "0bAAwK14":[
4               {
5                   "id":"63313e7848ebba445a3b3cee",
6                   "userid":"0bAAwK14",
7                   "origin":"35450",
8                   "source_id":"fake90e6d701748f08514b01",
9                   "pname":"itMakesMeThinkAbout.emotions",
10                  "pvalue":{
11                      "sadness":0.401,
12                      "interest":1.0,
13                      "joy":0.823
14                  },
15                  "context":"application",
16                  "category":"interest"
17              }
18          ]
19      },
20      "itMakesMeFeel.emotions":{
21          "BTKF72et":[
22              {
23                  "id":"632dab3fcfb08e51f124bcef",
24                  "userid":"BTKF72et",
25                  "origin":"39347",
26                  "source_id":"fake90e6d701748f08514b01",
27                  "pname":"itMakesMeFeel.emotions",
28                  "pvalue":{
29                      "Sadness":0.93
30                  },
31                  "context":"application",
32                  "category":"interest"
33              }
34          ]
35      },
36      "itRemindsMeOf.emotions":{
37          "N72IRgFy":[
38              {
39                  "id":"63384f412899916df970879f",
40                  "userid":"N72IRgFy",
41                  "origin":"35362",
42                  "source_id":"fake90e6d701748f08514b01",
43                  "pname":"itRemindsMeOf.emotions",
44                  "pvalue":{
45                      "Joy":1.0
46                  },
47                  "context":"application",
48                  "category":"interest"
49              }
50          ]
51      }
52  }
53
```

Figure 2: This excerpt represents the interaction between three different museum visitors and the selected artworks during their visit. The user-generated emotional responses fall into three categories: 'itMakesMeThink-About.emotions', 'itMakesMeFeel.emotions', and 'itRemindsMeOf.emotions'. Each category contains entries grouped by user IDs, where each entry records the emotions experienced by the user. These entries include details like a unique ID, the context, the origin of the data, and the specific emotions with their associated intensity values.

.

Table 2

Description of the JSON Dataset 'ugcContributions.json'.

Field Name	Description
userid	A unique identifier for each user. It is used to track the contributions made by individual users.
id	The ID uniquely identifies a story and is automatically generated by the GAMGame app each time a new story is created. In Figure 2, user with userid BTKF72et created the story with is 632dab3fcfb08e51f124bcef, which can contain a minimum of 2 and a maximum of 3 artworks selected by the user (each artwork is uniquely identified by the "origin" attribute contened in the dataset "artworks.json" shown in Figure **??**).
origin	Represents the source or context of the emotional response, likely linking back to a specific artwork selected by the user. In particular, in Figure 2, the user with userid BTKF72et selected the painting with id 39347 (Dans mon pays by Marc Chagall[4]), generating the emotion Sadness with an intensity of 0.93.
source_id	An identifier linking the response to the source of the stimulus, possibly an internal reference within the application.
source	Indicates the origin of the data, often marked as "fake" in this dataset, suggesting placeholder or anonymized data.
pname	The parameter name, typically set to `itMakesMeThinkAbout.emotions`, (Used to describe the emotions evoked by a memory when admiring the painting) or `itRemindsMeOf.emotions` (indicating to describe the emotions evoked by the type of memories that the painting brings to mind) and finally `itMakesMeFeel.emotions` (used to convey the emotions that are expressed and the feelings experienced when admiring that particular painting).
pvalue	Contains the actual emotional response data, represented as a dictionary with keys being emotion names (e.g., `sadness`, `joy`) and values being the intensity of these emotions.
context	Describes the context in which the data was collected, usually set to "application", indicating it was gathered through the GAMGame app.
datapoints	Numerical value representing the aggregation level of the data, often set to 0 in this dataset.
category	Indicates the type of data, commonly set to "interest", reflecting the user's engagement or emotional response to the content.

Table 3

Description of the JSON Dataset "ugcUsers.json".

Field Name	Type	Description
userid	String	A unique identifier for each user. It is used to associate demographic data and preferences with individual users.
id	Integer	A unique identifier for each demographic entry associated with a user.
origin	String	Indicates the origin of the data, often set as "User", indicating that the data was input by the user.
source_id	String	An identifier linking the demographic entry to its source within the application.
source	String	Indicates the origin of the demographic data, often labeled as "fake", suggesting placeholder or anonymized data.
pname	String	The parameter name describing the demographic aspect being recorded, such as Gender, Age, RelationshipWithArt, etc.
pvalue	String	Contains the actual demographic data, with values corresponding to the parameter name, such as "Male" for Gender, "20-30 age" for Age, etc.
context	String	Describes the context in which the demographic data was collected, typically set to "application".
datapoints	Integer	Numerical value representing the aggregation level of the data, often set to 0 in this dataset.
category	String	Indicates the category of data, which in this dataset is consistently "demographics", reflecting that the entries are related to user demographic information.

Acknowledgments

References

[1] E. Daga, L. Asprino, R. Damiano, M. Daquino, B. D. Agudo, A. Gangemi, T. Kuflik, A. Lieto, M. Maguire, A. M. Marras, et al., Integrating citizen experiences in cultural heritage archives: requirements, state of the art, and challenges, ACM Journal on Computing and Cultural Heritage (JOCCH) 15 (2022) 1–35.

[2] L. Bruni, E. Daga, R. Damiano, L. Diaz, T. Kuflik, A. Lieto, A. Gangemi, P. Mulholland, S. Peroni, S. Pescarin, A. Wecker, Towards advanced interfaces for citizen curation, in: AVI2CH Workshop on Advanced Visual Interfaces and Interactions in Cultural Heritage colocated with 2020 International Conference on Advanced Visual Interfaces (AVI 2020), 2020.

[3] R. Damiano, T. Kuflik, A. J. Wecker, M. Striani, A. Lieto, L. Bruni, N. Kadastik, T. A. Pedersen, Exploring values in museum artifacts in the spice project: A preliminary study, in: Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization, UMAP '22 Adjunct, Association for Computing Machinery, New York, NY, USA, 2022, p. 391–396. URL: https://doi.org/10.1145/3511047.3537662. doi:10.1145/3511047.3537662.

[4] C. Kreps, Curatorship as social practice, Curator: The Museum Journal 46 (2003) 311–323. doi:https://doi.org/10.1111/j.2151-6952.2003.tb00097.x.

[5] D. P. McAdams, Narrative Identity, Springer, New York, NY, 2011, pp. 99–115. doi:10.1007/978-1-4419-7988-9_5.

[6] L. E. Bruni, Cultural narrative identities and the entanglement of value systems, De Gruyter Mouton, Berlin, Boston, 2021, pp. 121–146. doi:doi:10.1515/9783110662900-006.

[7] A. Lieto, M. Striani, C. Gena, E. Dolza, A. M. Marras, G. L. Pozzato, R. Damiano, A sensemaking system for grouping and suggesting stories from multiple affective viewpoints in museums, Hum. Comput. Interact. 39 (2024) 109–143. URL: https://doi.org/10.1080/07370024.2023.2242355. doi:10.1080/07370024.2023.2242355.

[8] A. Lieto, G. L. Pozzato, M. Striani, S. Zoia, R. Damiano, Degari 2.0: A diversity-seeking, explainable, and affective art recommender for social inclusion, Cognitive Systems Research 77 (2023) 1–17. URL: https://www.sciencedirect.com/science/article/pii/S1389041722000456. doi:https://doi.org/10.1016/j.cogsys.2022.10.001.

[9] G. Jiménez-Díaz, B. Diaz-Agudo, L. E. Bruni, N. Kadastik, A. Follo, R. Damiano, M. Striani, A. Sanchez-Martin, A. Lieto, Interpretable clusters for representing citizens' sense of belonging through interaction with cultural heritage, J. Comput. Cult. Herit. (2024). URL: https://doi.org/10.1145/3665142. doi:10.1145/3665142, just Accepted.

[10] R. Plutchik, Emotions: A general psychoevolutionary theory, Approaches to emotion 1984 (1984) 2–4.

[11] A. Lieto, G. L. Pozzato, S. Zoia, V. Patti, R. Damiano, A commonsense reasoning framework for explanatory emotion attribution, generation and re-classification, Knowl. Based Syst. 227 (2021) 107166. doi:10.1016/j.knosys.2021.107166.

A. Online Resources

From this GitHub repository, you can download our dataset in JSON-LDH format. Although the dataset currently lacks machine readable semantic metadata, the vocabulary illustrated in this paper for data

description is aligned with the SPICE Ontology Network (SON)[5] in compliance with the FAIR paradigm.

- GitHub

[5]https://github.com/spice-h2020/SON

- GitHub