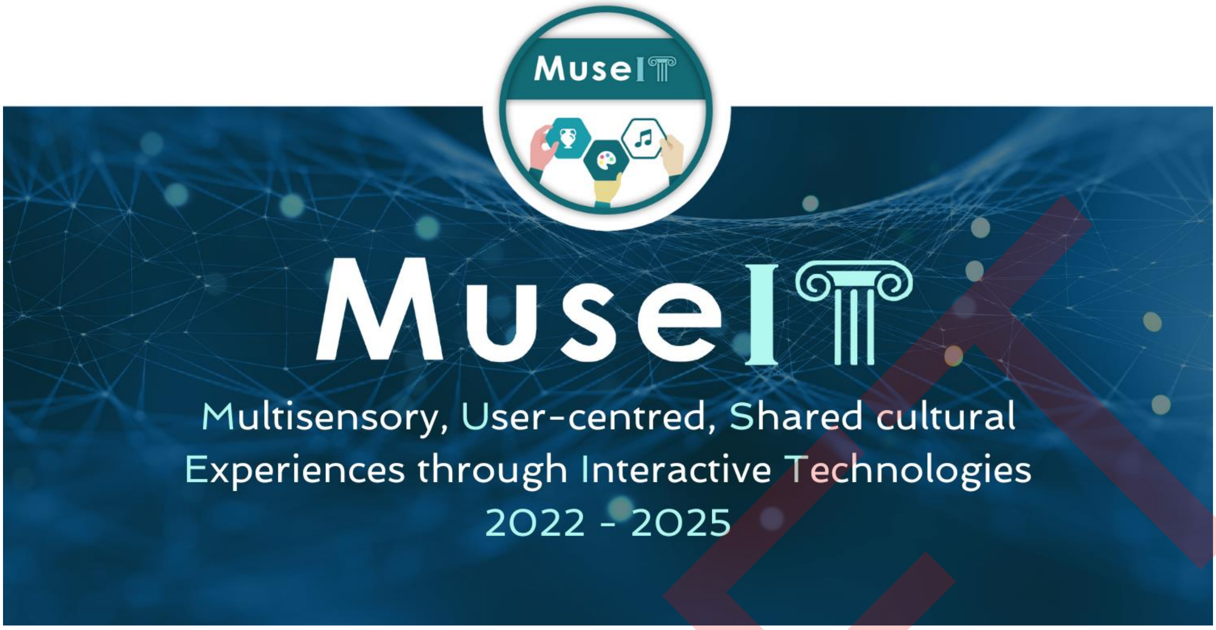
Ref. Ares(2024)2261549 - 26/03/2024



MuseIT Semantic Models & Semantic Infrastructure

Document Information

Project Name:	Multisensory, User-centred, Shared cultural Experiences through Interactive Technologies
Project Acronym:	MuseIT
Grant Agreement No:	101061441
Deliverable No:	D6.1
Deliverable Name:	MuseIT Semantic Models & Semantic Infrastructure
Work Package:	WP6
Task:	6.1, 6.2
Dissemination Level:	PU
Deliverable Type:	R
Lead Organization:	CTL
Lead Member:	Stelios Kontogiannis
Submission month:	M18
Date:	31 March 2024

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Document history

Revision	Date	Description / Reason of change	Submission by
v0.1	10-01-24	Structure proposal and initial draft	Stelios Kontogiannis
v0.2	07-02-24	First draft for internal review	Stelios Kontogiannis
v0.3	28-02-24	Second draft addressing review comments	Stelios Kontogiannis
v0.4	22-03-24	Final draft addressing the PMB review comments	Stelios Kontogiannis
v1.0	28-03-24	Final draft submitted to the EU	Renata Sadula

Authors

Partner	Name(s)
CTL	Stelios Kontogiannis, Nikolas Petrou, Georgia Christodoulou, Pavlos Kosmides
KCL	Nitisha Jain, Jongmo Kim, Chiara Di Bonaventura, Albert Merono Penuela, Barbara McGillivray
HB	Sándor Darányi
KNAW-DANS	Vyacheslav Tykhonov

Contributors

Partner	Contribution type	Name
CERTH	Review	Panagiotis Petrantonakis
SHMU	Review	Moa Johansson
EXUS	Review	Roberto Maffulli

Glossary

Acronym	Definition
ADHD	Attention Deficit Hyperactivity Disorder
AI	Artificial intelligence
AML	AgreementMakerLight
APD	Average Pairwise cosine Distance
API	Application Programming Interface
AR	Augmented Reality
B	Billion
BBC	British Broadcasting Corporation
Bi-LSTM	Bidirectional Long Short-Term Memory
CH	Cultural Heritage
CHA	Cultural Heritage Assets
CNN	Convolutional Neural Networks
CRM	Conceptual Reference Model
DCMC	Demonstration by CTL in 2nd Consortium Meeting in Cyprus
EDM	Europeana Data Model
FAIR	Findable, Accessible, Interoperable, Reusable
JSON	JavaScript Object Notation
JSON-LD	JavaScript Object Notation for Linked Data

Acronym	Definition
ADHD	Attention Deficit Hyperactivity Disorder
KG	Knowledge Graph
KRR	Knowledge Representation and Reasoning
LLM	Large Language Models
LM	Language Model
LSCD	Lexical Semantic Change Detection
LSTM	Long Short-Term Memory Networks
M	Million
NLP	Natural Language Processing
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OWL	Web Ontology Language
RAG	Region Adjacency Graphs
RDF	Resource Description Framework
REST API	Representational State Transfer Application Programming Interface
SDMX	Statistical Data and Metadata Exchange
SemKG	Semantic Knowledge Graph
SPARQL	SPARQL Protocol and RDF Query Language
SSD	Semantic Shift Detection
t-SNE	t-distributed Stochastic Neighbor Embedding
TMS	The Museum System
TTR	Type-Token Ratio
UIS	UNESCO Institute of Statistics
UNESCO	United Nations Educational, Scientific and Cultural Organization
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
VM	Virtual Machine
VR	Virtual Reality
W3C	World Wide Web Consortium
WADL	Web Application Description Language
WP	Work Package
WWW	World Wide Web

Table of contents

Executive Summary

The MuseIT project, funded by the European Union under Grant Agreement No. 101061441, has reached a critical milestone with the completion of Deliverable D6.1, "MuseIT Semantic Models & Semantic Infrastructure." This deliverable, positioned within Work Package 6 and addressing Tasks 6.1 and 6.2, is pivotal in realizing the project's ambition to foster multisensory, user-centered shared cultural experiences through cutting-edge interactive technologies.

D6.1, which is central to WP6, establishes ontological frameworks that guarantee semantic compatibility and promote the long-term digital preservation of cultural material. In addition to investigating the encoding of ontologies to reflect the temporal dynamics and cross-modal character of cultural narratives, the deliverable emphasizes the creation of particular ontologies for cultural heritage representation. This project is essential for incorporating multimodal data into Knowledge Graphs and expanding the cultural heritage space with cutting edge uses like generative AI models for music and audio content.

Looking ahead, the MuseIT will delve deeper into the potential of generative AI, aiming to revolutionize the accessibility and engagement of cultural heritage content. The roadmap outlined in D6.1 not only highlights the significant achievements of WP6 but also charts a course for future initiatives designed to transform the management and preservation of cultural heritage through digital solutions.

The main focus of future work includes expanding the CH assets ontology to encompass a wider array of cultural heritage aspects, enhancing semantic interoperability with broader cultural heritage and academic databases, and leveraging AI-driven personalization to offer more engaging and individualized cultural heritage explorations. Moreover, the project will continue to explore semantic drift within the cultural heritage domain to ensure the relevance and accuracy of cultural heritage representations in changing linguistic and cultural contexts. Through these initiatives, MuseIT aims to further enrich the cultural heritage domain with digital solutions that enhance accessibility, understanding, and preservation.

1. Introduction

In the evolving field of cultural heritage (CH), Work Package 6 (WP6) of the MuseIT project plays a crucial role in managing and integrating both tangible and intangible cultural assets. Its main goal is to establish ontological frameworks that ensure semantic interoperability and promote the long-term digital preservation of cultural assets, including both originally digital content and digitized versions. Essential to this effort are the development of digital rights management tools, as well as the design and implementation of archival processes and repository management systems.

Central to WP6, Task 6.1 focuses on creating specific ontologies for cultural heritage representation. This effort is key to building a structured, semantically rich environment that effectively captures the complexity of cultural assets.

Task 6.2 addresses the challenge of encoding ontologies in a way that reflects the temporal dynamics and cross-modal nature of the cultural narrative. It is divided into two main parts: the first emphasizes the use of multisensory data within Knowledge Graphs (KGs), incorporating sensor data, text, and visual information, and examines the integration with other technologies such as virtual reality (VR) and music, which are crucial to the MuseIT project. The second part investigates the evolution of meanings within cultural heritage texts, aiding in understanding how linguistic and cultural shifts affect the interpretation of large text corpora and artefact collections.

Within WP6, significant progress includes the development of a CH assets ontology that now integrates video metadata, ensuring alignment with W3C standards.

Looking forward, WP6 plans to explore the potential of generative AI models for creating audio and music content, which could significantly enhance the cultural heritage domain.

The roadmap for Deliverable D6.1 is as follows:

- Related Work: A comprehensive review of existing literature and projects relevant to the goals of D6.1.
- Knowledge Representation: Detailed exploration of ontological frameworks and their application in capturing the complexity of cultural heritage.
- Decision Making using Semantic Rules: An analysis of how semantic rules can support decision-making processes within the context of cultural heritage management.
- Future Work: An outline of planned advancements and explorations in the field, particularly the application of generative AI models for audio and music creation.

This structured outline for D6.1 not only highlights the achievements and contributions of WP6 but also sets a clear direction for future initiatives, aiming to transform the management and preservation of cultural heritage through innovative digital solutions.

2. Related work

The exploration of related work provides a foundation for understanding the current state of research and practice within the field of cultural heritage (CH) digital preservation and ontology development. This section delves into the theoretical background that informs the development and application of ontologies in cultural heritage, drawing upon a wide range of interdisciplinary studies and technological advancements.

2.1. Theoretical background

The theoretical underpinnings of using ontologies for cultural heritage representation stem from the fields of knowledge representation, semantic web technologies, and cultural studies. Ontologies, in the context of information science, are formal representations of a set of concepts within a domain and the relationships between those concepts. They are used to model domain knowledge in a structured and machine-interpretable format, facilitating semantic interoperability among diverse systems and datasets.

In exploring the application of ontologies within the realm of cultural heritage, it is crucial to delve into the theoretical background that underpins this approach. This involves examining several key areas that collectively form the foundation of ontological frameworks for cultural heritage representation. These areas include the principles of knowledge representation in cultural heritage, the role of semantic web technologies, the insights offered by cultural studies and interdisciplinary approaches, and the lessons learned from previous ontological frameworks in the CH domain. Each of these components contributes to our understanding and implementation of ontologies in cultural heritage as follows:

- Knowledge Representation in Cultural Heritage: At the core of ontological frameworks is the representation of complex and nuanced cultural information in a standardized, accessible manner. The theoretical basis for this approach draws from knowledge representation and reasoning (KRR) principles, which focus on modelling abstract concepts and their interrelations in a way that computers can process. This involves defining entities, such as cultural artefacts, events, or practices, and the semantic relationships that connect them, enabling rich, contextualized understanding of cultural data.
- Semantic Web Technologies: The development of ontologies for CH heavily relies on semantic web technologies, which aim to create a universal framework that allows data to be shared and reused across application, enterprise, and community boundaries. RDF (Resource Description Framework), OWL (Web Ontology Language), and SPARQL (SPARQL Protocol and RDF Query Language) are among the key technologies that underpin the semantic web, providing the tools to construct and query ontologies. These technologies enable the linking of cultural heritage information in a way that is both semantically rich and globally interoperable.
- Cultural Studies and Interdisciplinary Approaches: The theoretical background also encompasses insights from cultural studies, emphasizing the importance of understanding cultural assets not just as physical or digital objects, but as carriers of meaning, identity, and history. This perspective informs the ontological modelling process, ensuring that ontologies capture the multifaceted nature of cultural heritage, including its intangible aspects such as practices, languages, and traditions. Interdisciplinary approaches, integrating insights from anthropology, history, art history, and information technology, are crucial in developing ontologies that are both technically robust and culturally informed.
- Previous Ontological Frameworks in Cultural Heritage: A review of existing ontological frameworks in the CH domain reveals a variety of approaches to modelling cultural information. Projects such as CIDOC-CRM (Conceptual Reference Model) and Europeana Data Model (EDM) have set precedents for comprehensive ontological structures that facilitate the integration, sharing, and preservation of cultural heritage data across different platforms and institutions. Analysing these models provides valuable insights into best practices, challenges, and opportunities for innovation in the development of new ontological frameworks for cultural heritage.

In conclusion, the theoretical background of related work in cultural heritage ontologies is rich and multifaceted, drawing on principles from knowledge representation, semantic web technologies, cultural studies, and interdisciplinary research. This foundation not only informs the development of ontological frameworks but also guides the strategic integration of technological and cultural perspectives, aiming to enhance the accessibility, understanding, and preservation of cultural heritage in the digital age.

2.2. Digital Gateways to Cultural Heritage: Platforms and Databases

Access to comprehensive presentations and databases significantly enhances our understanding and appreciation of cultural heritage assets. This section introduces several key platforms that aggregate and provide access to a multitude of cultural heritage items from European museums, galleries, libraries, and archives. Europeana[1] stands out by offering millions of digitized items, including paintings, photographs, and manuscripts, facilitated through a REST API that supports various search and exploration functionalities.

Similarly, the Getty Research Portal and Getty Museum API connect users to a vast array of art history texts, rare books, and cultural heritage data, adhering to the Linked.Art standard for data integration and access. These tools leverage modern web standards such as REST endpoints, IIIF for images, and SPARQL for complex queries, ensuring that cultural heritage data is both accessible and interoperable.

UNESCO's World Heritage Centre further contributes to this landscape by providing detailed information on global cultural and natural heritage sites, supported by an API that requires a subscription key for access. This initiative aligns with efforts to make heritage data more available and usable for various stakeholders. The UNESCO Intangible Cultural Heritage list is a crucial initiative aimed at preserving and recognizing cultural practices, expressions, knowledge, and skills worldwide that form a part of humanity's heritage. This list includes diverse elements such as traditional music, dance, festivals, culinary traditions, and crafts, reflecting the rich cultural diversity across the globe. For instance, the list features practices from a wide range of countries, including traditional singing, culinary practices like the making of couscous in the Maghrebi cuisine, and unique cultural celebrations such as the summer solstice fire festivals in the Pyrenees[2]. The connection between the UNESCO Intangible Cultural Heritage list[3] and MuseIT is profound. MuseIT's focus on integrating both tangible and intangible cultural assets digitally aligns well with UNESCO's efforts to safeguard intangible cultural heritages. By leveraging technology, such as Knowledge Graphs, virtual reality (VR), and multisensory data integration, MuseIT contributes to the preservation and accessibility of cultural narratives and practices. This technological approach not only aids in documenting and preserving these heritages but also in making them more accessible and engaging to the public, thereby supporting UNESCO's objectives of promoting cultural diversity and intercultural dialogue.

In addition to these platforms, many museums, libraries, and archives maintain their online collections and presentations, enriching the digital cultural heritage space. These resources collectively facilitate a deeper engagement with cultural heritage, offering tools and data that support research, education, and public enjoyment of the world's cultural treasures.

[1] https://europeana.eu

[2] https://ich.unesco.org/en/lists
[3] https://en.wikipedia.org/wiki/UNESCO_Intangible_Cultural_Heritage_Lists

2.2.1 Europeana

Europeana is a platform that provides access to millions of cultural heritage items from European museums, galleries, libraries, and archives. It includes paintings, photographs, manuscripts, and more.

The Europeana REST API[4] provides access to a vast collection of over 50 million cultural heritage items from major museums and galleries across Europe. It has evolved to include various specialized APIs for different purposes. For simple searches, the Search API is suitable, while the SPARQL service allows for in-depth exploration of structured metadata. The Record API retrieves metadata for a single item, and the OAI-PMH Service facilitates harvesting the entire Europeana repository. The Entity API offers contextual information like Topics, Persons, and Places. Additionally, the Annotations API allows users to contribute information about items available on Europeana.

2.2.2. Getty Research Portal

Getty Research Portal[5] provides access to digitized art history texts, rare books, and related literature from various institutions.

The Getty Museum API[6] is based on the Linked.Art standard and offers access to cultural heritage data, including objects, places, documents, groups, persons, exhibitions, and activities. The API utilizes REST endpoints, IIIF standards for images, ActivityStreams for change tracking, and a SPARQL endpoint for graph queries. Entities in the collection are accessible via URLs, providing JSON documents with links to related entities. The API is designed for tasks such as getting records, tracking changes, and asking questions about the collection. While there's no current provision for a complete list or data download, these features are on the roadmap. Records, including objects like Van Gogh's Irises, persons like Vincent van Gogh, and places like West Pavilion, Gallery 204, can be accessed through specific URLs. The API's model is detailed on the linked.art website, with implementation specifics in the API Reference section. The data available is generally under CC0, and a GUI for SPARQL queries is provided.

2.2.3. UNESCO

UNESCO World Heritage Centre[7] provides information on cultural and natural heritage sites globally. It includes descriptions, images, and documentation for each site.

To use the UNESCO World Heritage Centre UIS API[8], a subscription key should be obtained through a sign-up process. The API is based on the Linked.Art standard and supports SDMX RESTful API Specifications. The API endpoint definition is available in WADL and Swagger formats. The Query Builder helps create valid SDMX-REST URLs.

---

4 https://pro.europeana.eu/page/intro
5 https://portal.getty.edu
6 https://data.getty.edu/
7 https://whc.unesco.org/
8 https://apiportal.uis.unesco.org/getting-started

Additionally, institutions such as museums, libraries, and archives often have their own online collections and presentations showcasing cultural assets. Several world-leading software solutions that museums and archives commonly use include:

- PastPerfect[9] is widely recognized for its comprehensive capabilities in managing collections and contacts. It's used by over 11,000 museums and offers both desktop and cloud-based solutions to accommodate various organizational needs, streamlining processes like acquisition, loans, cataloging, and donor management.

- TMS Collections by Gallery Systems[10] offers a web-based platform that delivers industry-leading functionality for efficient and user-friendly collections management. It supports a wide range of cultural institutions, including museums, archives, universities, libraries, and galleries, covering various types of collections from fine art to natural history.

- Zetcom's MuseumPlus[11] is chosen by more than 6,000 users in 1,000 institutions globally for collection management. It provides web-based software solutions tailored to the diverse needs of museums and archives, ensuring professional management and documentation for collection data and web content.

These software solutions play a pivotal role in the digital preservation and accessibility of cultural heritage, supporting UNESCO's mission by facilitating the management and presentation of both tangible and intangible cultural assets. By leveraging such technologies, institutions can enhance their contributions to cultural preservation, enabling broader public access and engagement with heritage globally.

3. Knowledge representation

In order to understand the concept of knowledge representation, it is essential to reference a key study in the field (Davis, Shrobe, & Szolovits, 1993), which delineates its five distinct roles: a surrogate, a set of ontological commitments, a fragmentary theory of intelligent reasoning, a medium for efficient computation, and a means for human expression. Knowledge representation is vital for several reasons, enhancing our ability to digitally handle, share, and interpret the complex data associated with our cultural heritage.

First, it serves as a digital stand-in for real-world objects and phenomena, allowing us to digitally represent and manipulate cultural assets and sensory data. Imagine converting a tangible artifact or an intangible practice into a digital model that we can examine, share, and analyze. It establishes a common framework or a set of rules that define the key elements within the cultural heritage domain and how they interact. This is similar to setting ground rules for a discussion to ensure everyone is on the same page about the topics and terminology being used. Knowledge representation helps infer new knowledge from existing information, acting as a basis for intelligent reasoning. This enables us to draw connections, identify patterns, and uncover new insights from the cultural data we have, enhancing our understanding and appreciation of cultural heritage. It organizes data in a way that computers can efficiently process. This means we can quickly analyze, compare, and share large volumes of cultural heritage information, leveraging computing power to discover new relationships

[9] https://museumsoftware.com/
[10] https://www.gallerysystems.com/solutions/collections-management/
[11] https://www.zetcom.com/en/

and insights. Finally, it facilitates human expression, providing a structured way to communicate complex ideas and stories about cultural heritage. This bridges the gap between raw data and meaningful narratives, helping to convey the significance and stories behind cultural artifacts and practices.

Central to these efforts are ontologies, structured frameworks that include categories (classes) for objects, concepts, and events; the ways these categories are related (relations); and specific examples within each category (instances). Ontologies help organize and link the diverse and complex data related to cultural heritage, making it easier to share, understand, and preserve this information for future generations. By leveraging knowledge representation, we can deepen our understanding of cultural heritage, making it more accessible and engaging while ensuring its preservation and appreciation through advanced technologies.

3.1. Knowledge for cultural assets

The MuseIT project is dedicated to enhancing the engagement and understanding of CH through digital innovation. At the core of our endeavor is a sophisticated data processing pipeline designed to transform diverse CH data into a rich, multi-layered digital representation. This pipeline integrates several key steps, each with a distinct purpose and contribution towards realizing our vision of a dynamic, accessible cultural heritage experience. Below is an overview of these steps, detailing why each is necessary and what it aims to achieve:

- **Cultural Heritage Assets Ontology Creation**: The foundation of our project is the development of a custom CH assets ontology. This ontology is crucial for organizing and categorizing CH data semantically, making it possible to store, retrieve, and interact with information in a meaningful way. It addresses the need for a standardized framework that captures the complexity of CH assets, including both tangible and intangible elements, and accommodates the requirements of diverse users, including those with disabilities.
- **Standards and Interoperability**: Adhering to Semantic Web standards and principles ensures our data is findable, accessible, interoperable, and reusable (FAIR). By aligning with established protocols like RDF and OWL, and integrating with ontologies such as CIDOC-CRM, Dublin Core, and Schema.org, we ensure our project's outputs are compatible with the broader digital ecosystem. This interoperability is key to facilitating the sharing and integration of CH data across different platforms and systems.
- **Ontology Alignment, Matching, and Integration**: To enrich our ontology and extend its utility, we engage in alignment processes with high-level ontologies. This step is about creating a comprehensive digital schema that not only serves our specific project needs but also aligns with global standards. The alignment enhances the expressiveness of our ontology, enabling it to represent a broader range of CH narratives and connections.
- **Use of Large Language Models** (LLMs): LLMs play a transformative role in structuring unorganized data into a format that aligns with our ontology. By processing text-based CH artifacts through models like the Mistral dolphin variant, we can generate structured, semantic representations of cultural narratives. This step is vital for incorporating vast amounts of textual data into our Knowledge Graph, making them accessible and interpretable in the context of VR experiences.
- **Data Enrichment through OpenRefine**: With OpenRefine, we refine and enhance our dataset, linking it to external repositories like Wikidata to add depth and context. This process enriches our CH assets with additional metadata, connecting them to a broader knowledge network. It

ensures that our VR experiences offer not just visual engagement but also educational value, providing users with a comprehensive understanding of each artifact.

- **Demonstration of Application**: The culmination of our data processing efforts is showcased in the integration of technology and CH assets to create immersive video and VR experiences. This step demonstrates the practical application of our processed data, highlighting how digital innovation can evoke emotional responses and foster a deeper connection with cultural heritage.

Through this pipeline, the MuseIT project aims to construct a digital environment where cultural heritage is not only preserved but brought to life through interactive technologies. Each step contributes to building a layered, nuanced representation of CH that is accessible, engaging, and informative for a wide audience.

3.1.1. Cultural Heritage Assets Ontology

The MuseIT project intends to provide cultural heritage protection, preservation, conservation, and safeguarding through digital intervention. Additionally, throughout the project, multi-layered, multisensory representations of the CHA will be designed, in order to accommodate the needs of individuals with disabilities by providing the opportunity of transforming unimodal cultural asset representation into several modalities.

To achieve these objectives, the CHA ontology emerges as a crucial element, tasked with capturing the vast domain of CHAs. The development of the CHA ontology is aimed at semantically representing a wide array of CHAs, enabling efficient data storage and retrieval, facilitating discovery and accessibility on digital platforms, and ensuring interoperability across different digital systems and archives.

To construct the ontology, we employed a hybrid approach that combines both top-down and bottom-up methodologies. This dual strategy is particularly effective in complex projects, as it integrates the benefits of each approach to ensure comprehensive coverage from general concepts to specific details.

The top-down approach begins with the definition of the broadest categories and works its way down to more specific concepts. In the context of our ontology, this meant first establishing the overarching categories of Cultural Heritage Assets, such as tangible (e.g., buildings, artifacts) and intangible (e.g., traditions, languages) CHAs. This approach is akin to creating an outline or a blueprint that sketches out the major sections of a building before detailing the individual rooms. For instance, using the top-down method, we identified 'Tangible CHAs' and 'Intangible CHAs' as primary classes within the ontology. This initial step set the general framework, providing a structured view of the different types of cultural heritage that the MuseIT project aims to encompass. Following this, the bottom-up approach was utilized to fill in the details within the framework established by the top-down method. This involves starting from specific instances or detailed data points and grouping them into broader categories. It's comparable to furnishing and decorating the rooms of the building once the overall structure is in place. A practical example of the bottom-up approach in our project was the identification and classification of specific types of CHAs, such as 'Funerary Sculptures' and 'Murals,' based on the concrete data and assets we planned to include in the ontology. These specific classes were then organized under the appropriate broader categories defined in the top-down phase.

Finally, after the classes were established using this hybrid approach, we defined corresponding properties to further describe and interconnect these categories, enriching the ontology's ability to represent the diversity and complexity of cultural heritage assets.

3.1.2. Standards and Interoperability

The Semantic Web is described as "a web of data that can be processed directly and indirectly by machines", according to Berners-Lee et al. (2001). Particularly, it is an extension of the World Wide Web (WWW), where web resources are enhanced with machine-readable semantic data to define their meaning. Since the information is now defined and explicitly linked, the Semantic Web adheres to the FAIR principles, ensuring that metadata and data are Findable to both humans and machines. Additionally, these resources are Accessible, with the metadata remaining retrievable even after the elimination of the associated data. The principle of Interoperability is also upheld, facilitating the seamless integration of diverse data sources. Finally, all the data and metadata become reusable across various settings as they are well-described.

The World Wide Web Consortium (W3C)[12], founded in 1994 and led by web inventor Tim Berners-Lee, played a pivotal role in shaping the Semantic Web. W3C developed standards such as the Resource Description Framework (RDF) and Web Ontology Language (OWL) to make web data machine-readable and understandable. For instance, RDF provides a description of relationships between data, whereas OWL has constructs for building vocabularies and interpretation rules. Hence, these standards are essential for relating different data types and content across the web and making the internet more than just a browsing platform. This approach makes the web more efficient for users, while it opens new possibilities for automated data analysis and artificial intelligence applications.

3.1.2.1. Ontologies in the Cultural Heritage Domain

In the digital landscape of CH, the significance of ontologies like CIDOC-CRM, Dublin Core, and Schema.org cannot be overstated. Each serves a critical role: CIDOC-CRM provides a comprehensive framework for documenting the complex relationships and histories of cultural artefacts, making it indispensable for museums and academic research. Dublin Core simplifies the categorization of digital assets, ensuring that videos, images, and documents are easily discoverable and accessible. Schema.org, on the other hand, extends the structuring of web content beyond cultural heritage, enabling a more semantic and interconnected web experience.

This requirement underpins the development of a unique MuseIT ontology, designed to provide a detailed and granular model that encapsulates the rich diversity of cultural heritage data, including aspects that are less emphasized in established frameworks. For example, while CIDOC-CRM is invaluable for documenting the complex relationships and histories of physical artifacts, it may not fully address the intricacies of multisensory data or the nuanced needs of individuals with disabilities. The MuseIT ontology is crafted to bridge these gaps, offering a more granular approach that encompasses the broad spectrum of cultural heritage, including its intangible and sensory dimensions.

To achieve this while ensuring broad applicability and interoperability, the MuseIT project embraces the concept of 'alignment.' This involves mapping and integrating concepts and terms from our custom ontology to those in CIDOC-CRM, Dublin Core, and Schema.org. Such strategic alignment allows us to build upon the solid foundations of these established ontologies, enhancing our framework's capability to document and link diverse types of cultural heritage data without duplicating existing efforts. This synergy is essential because it guarantees that our ontology satisfies both the requirements of the MuseIT project and international standards, allowing for a smooth integration into the larger ecosystem of digital cultural material.

Illustrated in Figure 1, this process begins with the initial design of the MuseIT ontology, where we identify opportunities for alignment and integration. Panel (a) presents an overview of the ontology's

[12] www.w3.org/standards/

class hierarchy before alignment, setting the stage for detailed documentation and semantic connections. Panel (b) zooms in on a specific example, illustrating the placement of 'Video' under 'IntangibleAsset,' showcasing how detailed categorization within our ontology enhances the representation of diverse cultural heritage assets.

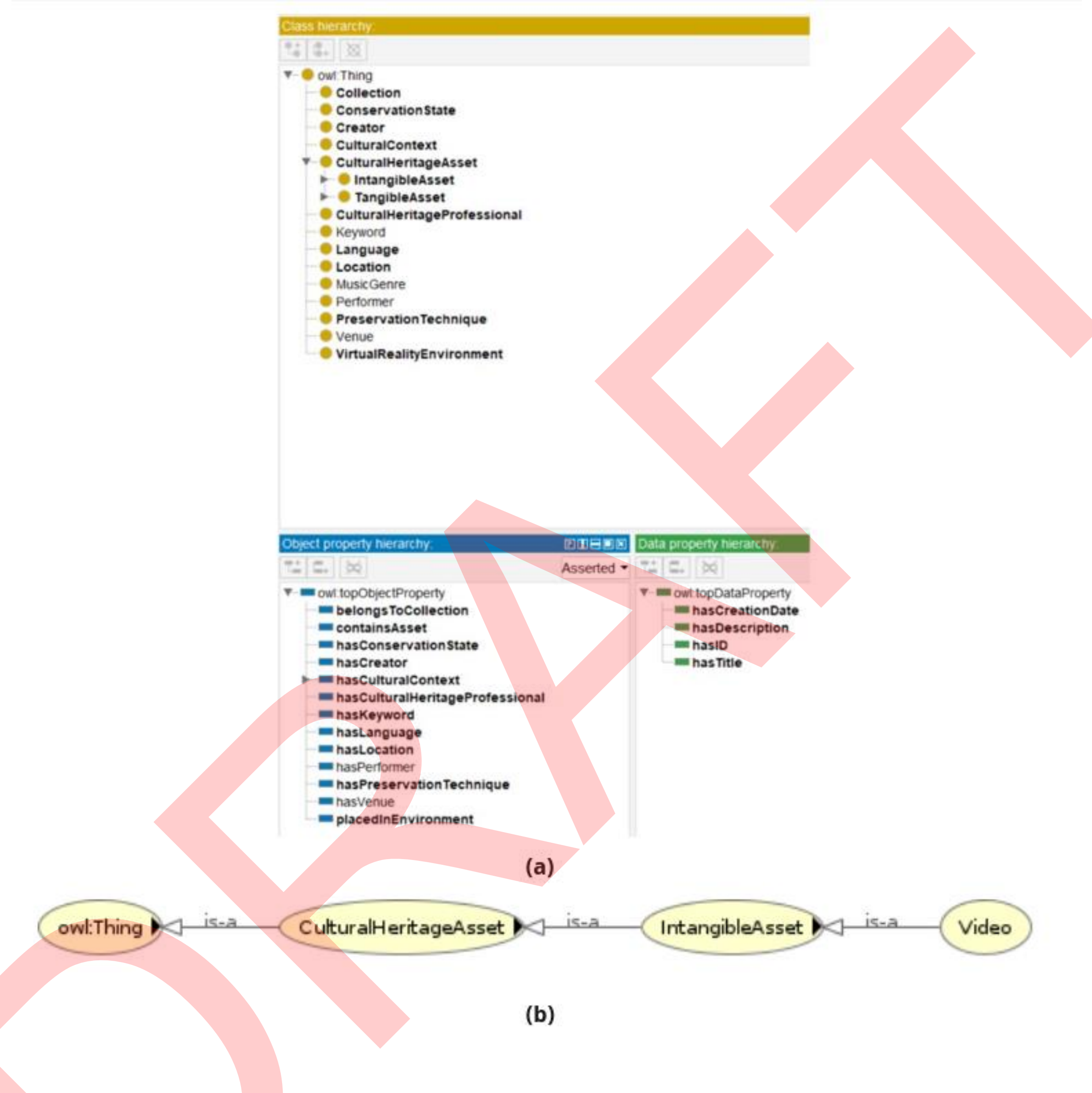


Figure 1: (a) The class hierarchy within the ontology (b) A segment of the ontology's class structure, illustrating the 'is-a' relationships defining 'CulturalHeritageAsset' as a superclass of 'IntangibleAsset' which is a superclass of 'Video'

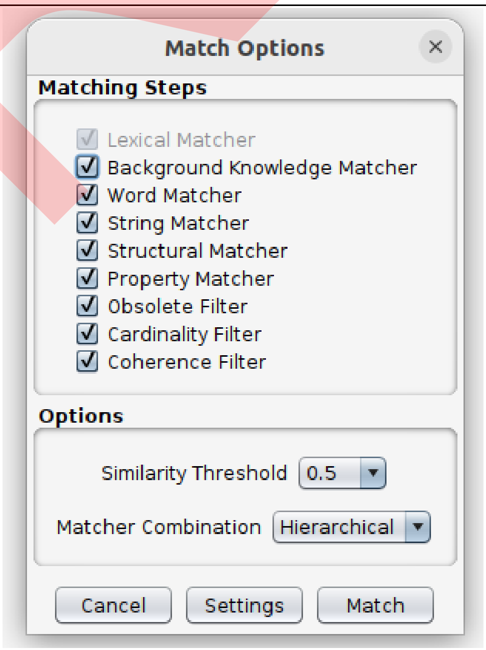
The development of a custom ontology does not occur in isolation. It represents an opportunity to merge the strengths of existing ontologies while filling in the gaps specific to the MuseIT project's objectives. This synthesis not only aids in the precise representation of complex CH data but also ensures the project's data ecosystem remains compatible with global standards, facilitating data sharing and interoperability across platforms. Integrating a custom ontology with elements from CIDOC-CRM, Dublin Core, and Schema.org also positions the MuseIT project at the forefront of digital heritage preservation. It enables a richer, more interactive user experience, bridging the gap between traditional cultural heritage understanding and the possibilities afforded by modern technology. Through this integration, the project aspires to create a multi-layered, multisensory experience that not only educates but also engages users on an unprecedented level.

In essence, the creation and refinement of a custom ontology within the MuseIT project are about more than just data organization—it's about creating a digital ecosystem that can adapt to the evolving needs of both the project and its end-users, ensuring that cultural heritage is not only preserved but also made vibrant and accessible in the digital era. In the next chapter, we will delve into the ontology alignment process, detailing how these integrations and alignments are conducted to achieve such a comprehensive digital ecosystem.

3.1.2.2. Ontology Alignment, Matching and Integration

The term ontology alignment or ontology matching refers to the process of establishing correspondences between ontological concepts. In general, given two ontologies O_1 and O_2, a mapping among entities e_1 and e_2 from O_1 and O_2 respectively is stated as $\langle id, e_1, e_2, r, n \rangle$, where 'id' is an identifier of the mapping, r represents the relation among e_1 and e_2, which can be equivalence (=), disjointness(⊥), subsumption/less general (⊑,≤), etc. and n it is a confidence measure that gives the correspondence among the two entities (n e [0,1]).

The AgreementMakerLight (AML)[13] tool was utilized for the ontology alignment process. AML, known for its robustness and efficiency in ontology alignment tasks, can be accessed at its GitHub repository. This tool facilitated the alignment of ontologies by providing a specialized platform for this purpose. The initial step of the procedure was to ensure that all the classes and properties of the CH asset ontology were well structured and described, in order to obtain the high-level ontologies that will be aligned with them. Once the AML tool is set up, the ontologies are loaded with the CHA ontology to be the source and the high-level target ones and the alignment is performed. The specific tool goes through multiple matching steps such as word, structural matcher, cardinality filter and others. Additionally, it provides multiple ways to customize the matching procedure, with the most important to select what filters to perform and how sensitive the similarity threshold will be (Figure 2).



[13] https://github.com/AgreementMakerLight/AML-Project

Figure 2: AML tool displaying matching steps and options for ontology alignment.

Following this automated alignment, a thorough manual review of the resulting mappings has to be conducted to ensure accuracy and relevance. The resulting mappings are suggested entities of the source and target ontologies that are equivalent and the percent of the confidence measure which is greater than the threshold that was defined (Figure 3). The matching procedure that was executed was set as a threshold measure the 0.5 and go through all the matchers and filters that are provided by the tool in order not to miss any potential alignment.

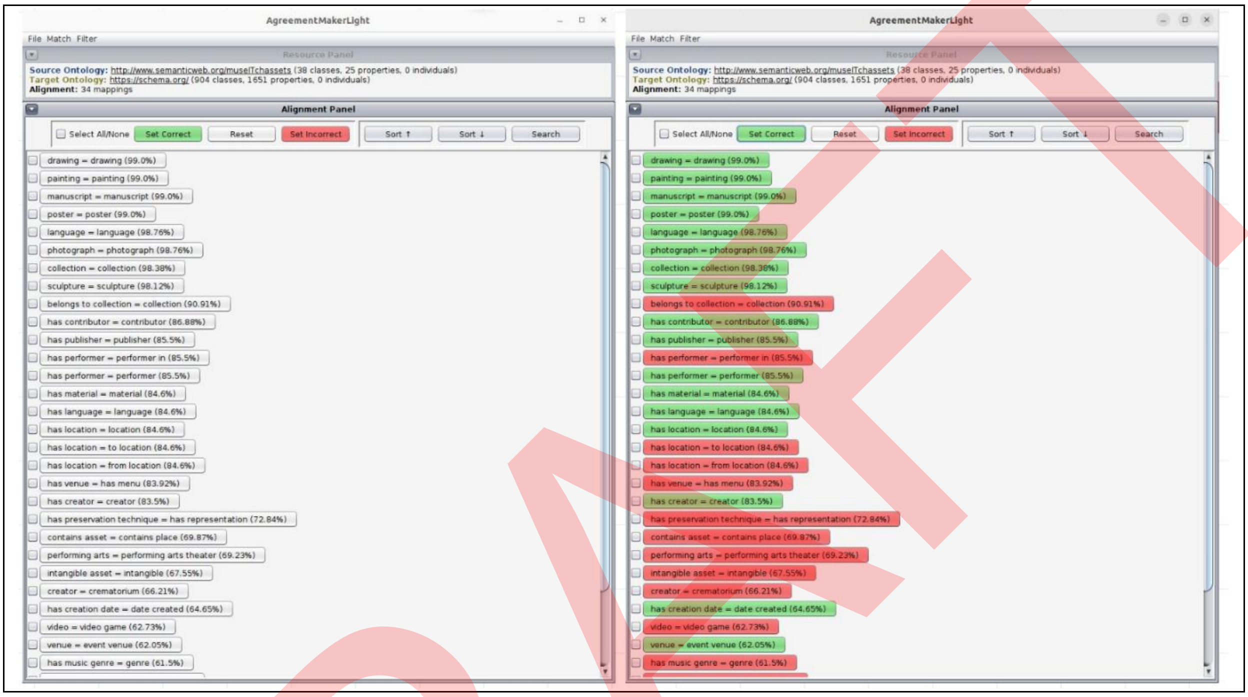


Figure 3: Alignment Mappings Before and After Manual Review.

As presented in Figure 3, the left panel displays the initial suggested mappings by the AML tool, along with confidence levels. The right panel shows the mappings post-manual review, where green indicates confirmed correct alignments and red denotes incorrect ones, based on a confidence threshold of 0.5.

After validating the mappings, the correct mappings can be exported in an RDF file format (Figure 4). The integration of the two ontologies was finalized through an automated process. A Python script was employed to add a new triplet for each output mapping. This included the subject, which is the entity of the CH asset ontology (either a class or property), and the object, which is the corresponding entity in the target ontology. The predicate was set as either owl:equivalentClass or owl:equivalentProperty, depending on the nature of the entity. Figure 4 shows an example of the updated ontology, with highlighted an instance where owl:equivalentClass is utilized. The described procedure was followed successively two times, one for each high-level ontology, schema.org and Dublin Core.

Figure 4: Exported RDF Mappings and Ontology Integration. On the left, the exported RDF file showcases the validated mappings post-manual review. On the right, we see the integrated ontology with highlighted an example of established equivalence, as automated by a Python script, demonstrating the successful alignment between the CH asset ontology and the high-level ontologies.

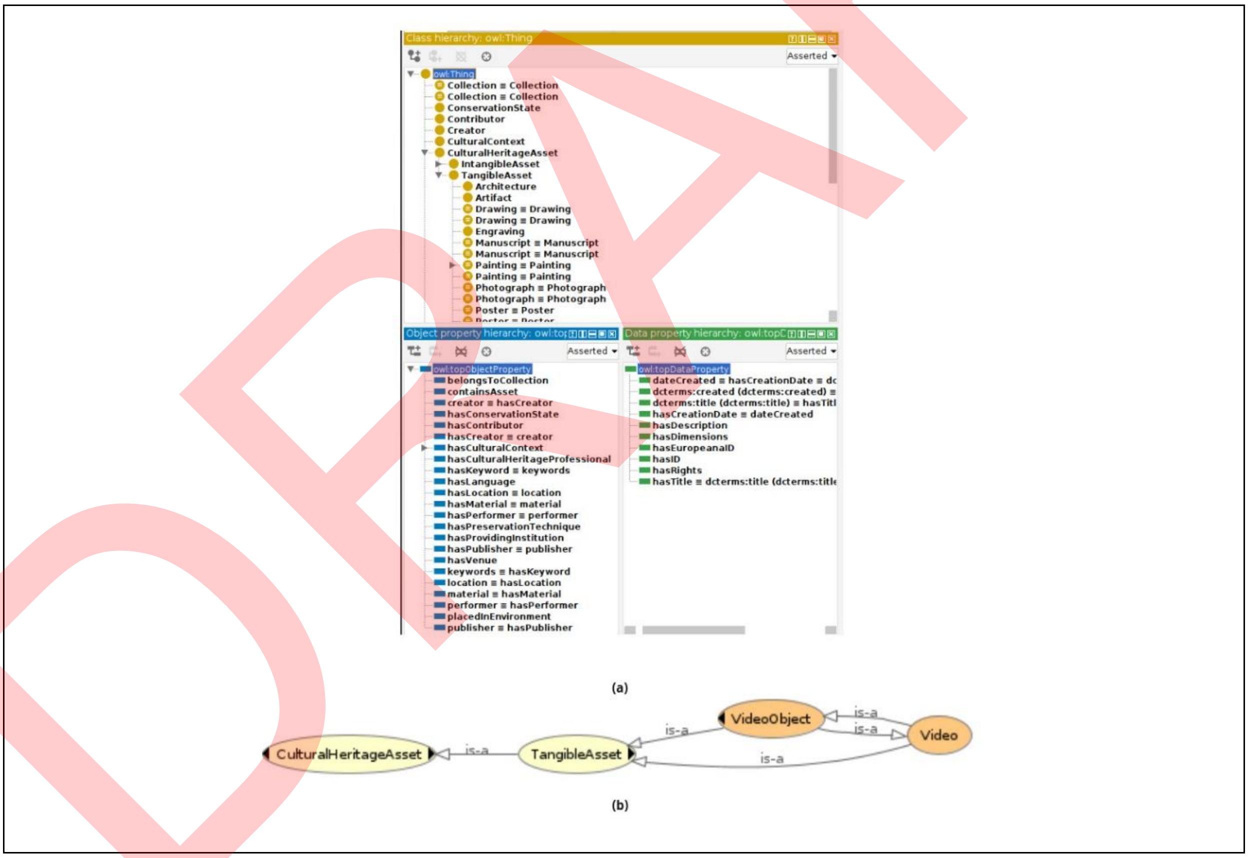


Figure 5: Enhanced Ontology Structure Post-Alignment. (a) The expanded class hierarchy and object properties within the ontology, show the alignment with standardized vocabularies. (b) A segment of the ontology's class structure, illustrating the refined subclass relationships, showing the "VideoObject" class from schema.org ontology aligned with the "Video" class from CH ontology

Figure 5 demonstrates the advanced state of our custom ontology after the alignment process. In panel (a), we see the enriched class hierarchy that now includes a more detailed breakdown of cultural heritage assets, as well as object and data properties that have been synchronized with external standards. Panel (b) provides a closer look at the specific enhancements made, such as the addition of 'VideoObject', which has been aligned with the "Video" class of the CH ontology.

3.1.3. Demonstration of Application

During the second Consortium meeting of the MuseIT project, which was hosted in Cyprus, attendees were shown a demonstration by CTL, illustrating the integration of technology and cultural assets to create music that reflects users' emotional experiences (DCMC).

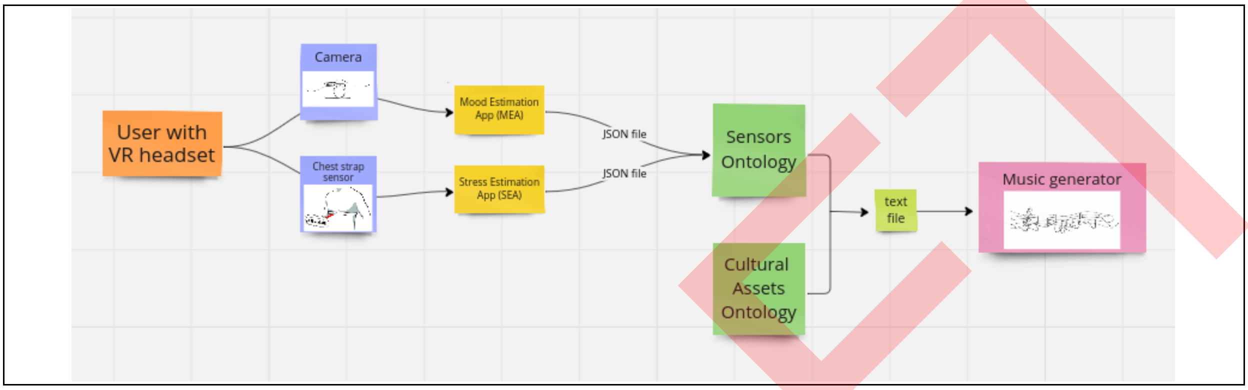


Figure 6: DCMC illustration

As depicted in Figure 6 users engages with content obtained from CH assets through an immersive VR experience while a camera captured their facial expressions for mood analysis and a chest strap monitored the stress level (more details in D5.2). Then, the biometric data is stored in a dedicated sensor ontology for subsequent processing. Subsequently, based on the sensor data along with the semantic information from the CH assets, a text is generated, as explained in section 4.1.2. The resulting text is fed into an AI music generation platform, MusicGen[14], to compose a unique musical piece tailored to the emotional responses elicited by the CH content.

The data used in the DCMC, consisted of YouTube videos which were carefully selected based on their topic to target specific emotions: sad, happy and neutral. An example to target sad emotions is a war movie scene chosen, accompanied by a classical adagio music piece. To induce happiness, lively dance videos and promotional videos for holiday destinations were selected. Neutral emotion was targeted through documentary scenes showcasing cultural sites and traditions from different countries. All the selected videos were edited to last approximately two minutes.

After defining and processing the dataset, video information such as title, music genre and keywords were manually extracted and structured into a JSON file format (Figure 7).

[14] https://arxiv.org/abs/2306.05284

```
"body": {
    "chassets": [
        {
            "id": 1,
            "title": "In memory of 11 September 2001 - Samuel Barber - Adagio for strings, op.11",
            "keywords": ["tragic", "deaths", "disaster", "war scene", "orchestra", "classical music"],
            "music_genre": ["classical work", "requiem", "lament"]
        },
        {
            "id": 2,
            "title": "Adagio for Strings, scene from Platoon movie",
            "keywords": ["tragic", "deaths", "disaster", "war scene", "orchestra", "classical music"],
            "music_genre": ["classical work", "requiem", "lament"]
        },
        {
            "id": 3,
            "title": "The Lion King - Simba finds Mufasa",
            "keywords": ["death", "tragic", "lies", "animals", "jungle"],
            "music_genre": ["jungle music (Drum and Bass)", "tribal music", "ambient music", "wildlife", "nature-inspired soundscapes"]
        },
        {
            "id": 4,
            "title": "Pharrell Williams - Happy WE ARE FROM PARIS",
            "keywords": ["France", "Paris", "dance", "sightseeing", "tour", "happy people"],
            "music_genre": ["Pop", "Uptempo soul", "Neo Soul"]
        },
        {
            "id": 5,
            "title": "Oh My Greece | Unlock the Feeling",
            "keywords": ["Greece", "summer", "islands", "sea", "beach"],
            "music_genre": ["greek music", "soft pop"]
```


Figure 7: Part of videos' Structured Data

Subsequently, this JSON-formatted file was processed by CASPAR[15], a CTL's framework designed to transform JSON data into RDF triplestore by automatically generating and executing SPARQL queries, to effectively populate the CH asset ontology - more details in Section 3.2.3. Figure 8 provides a visual representation of this process, illustrating the transformation using the graffoo[16] tool, where instances are mapped to their respective classes and properties.

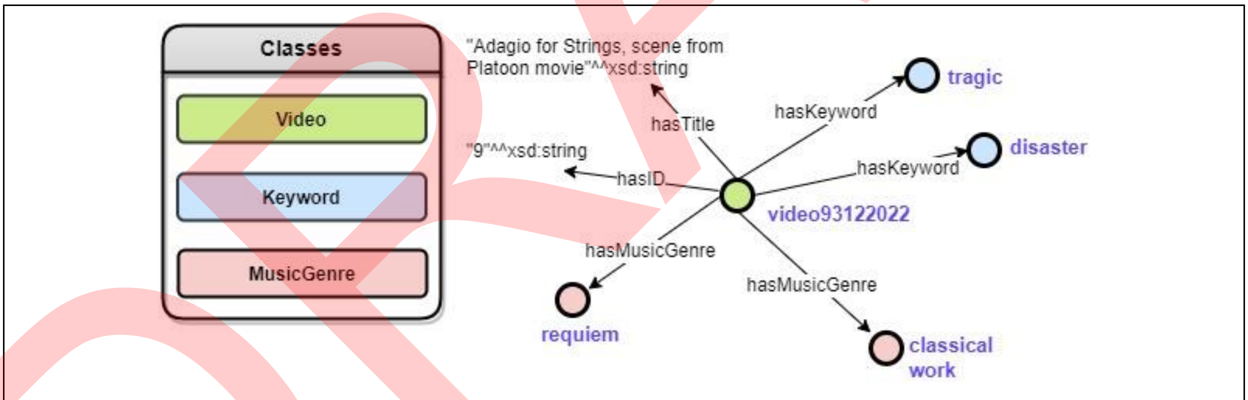


Figure 8: Illustration of RDF formed data

A comprehensive breakdown of the demonstration process, including the methodologies, tools, and outcomes, subsequent chapters will delve into detailed discussions, providing a deeper insight into each stage of the presented DCMC.

3.1.4. Integrating Multisensory Data for Enhanced VR Experiences

For the VR experience outlined in WP4, organizing data is crucial to enhancing understanding and engagement within virtual environments. For further details, refer to deliverable D4.1. Continuously modelling this data into ontologies and knowledge graphs is key, as it structures the data into a comprehensive, interconnected network. The importance of modelling data into ontologies and

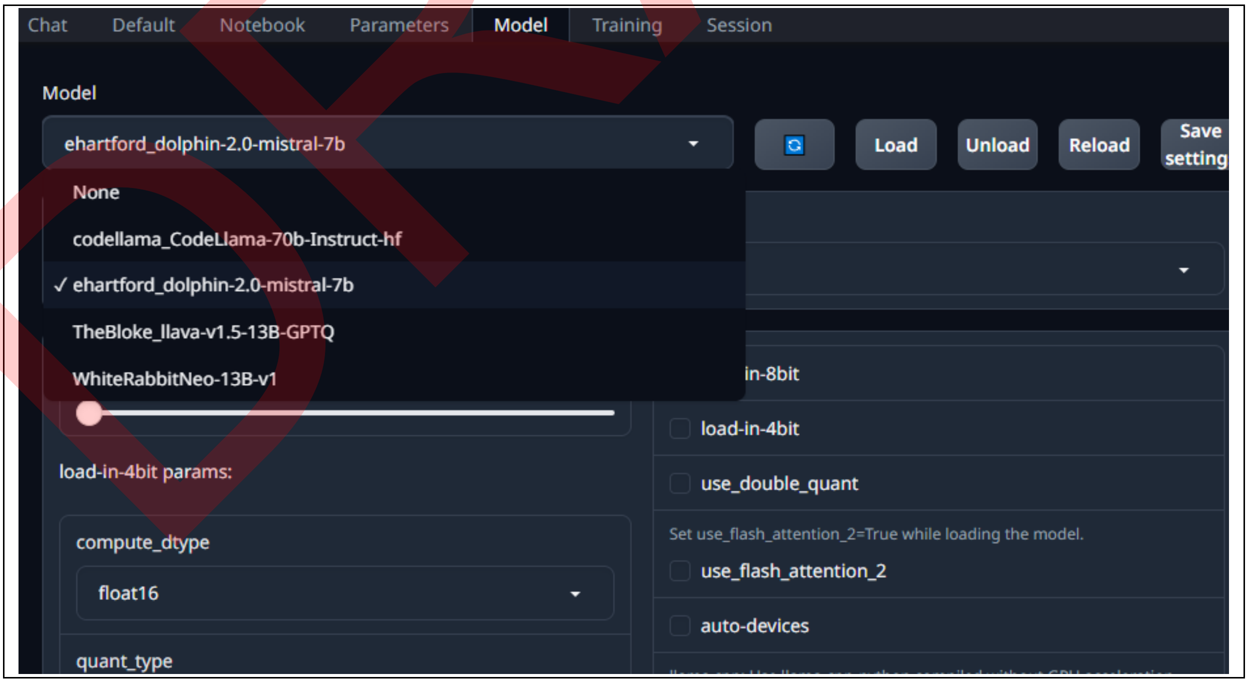
[15] https://caspar.catalink.eu/
[16] https://essepuntato.it/graffoo/

knowledge graphs lies in their ability to represent complex relationships and entities in a way that is both machine-readable and understandable to humans. This dual capability is crucial for developing VR experiences that are not only visually engaging but also deeply educational and contextually rich. By employing these models, the VR experience built in WP4 could provide users with a layered understanding of cultural heritage artifacts, showing not just the artifacts themselves but also their historical context, cultural significance, and connections to other artifacts and themes. This approach enriches the user experience, making it a more meaningful exploration of cultural heritage.

3.1.4.1. LLMs enhance data formation

Large Language Models (LLMs) have revolutionized the way we approach data structuring, offering sophisticated tools to analyze, interpret, and organize vast amounts of information. These models are capable of understanding and generating human-like text, which allows them to process unstructured data—such as free text from articles, social media posts, or documents—and convert it into structured formats that are easier to analyze and work with. For instance, LLMs can identify key elements within a text, such as names, dates, and specific facts, and organize these into a structured database or JSON format. This capability significantly streamlines the process of turning raw data into organized datasets that can be efficiently used in various applications, from research to product development.

In our project, in order to shape our data, we utilized LLMs via the straightforward interface provided by the "oobabooga" repository on GitHub[17]. The setup was executed on a Virtual Machine (VM) located within the CTL premises server. This configuration allowed us to access, load, run, and fine-tune models from Hugging Face[18]—a popular platform that hosts a wide range of pre-trained LLMs. Hugging Face serves as a central hub for the AI community, offering tools and models that facilitate machine learning research and development. It's renowned for its comprehensive collection of LLMs that can be adapted for diverse tasks, including but not limited to, text generation, sentiment analysis, and language translation. For our purposes, we tested several models, such as codellama_CodeLlama-70b-Instruct-hf, ehartford_dolphin-2.0-mistral-7b, and TheBloke_llava-v1.5-13B-GPTQ, ultimately selecting the Mistral dolphin variation for our specific needs (Figure 9).



[17] https://github.com/oobabooga/text-generation-webui
[18] https://huggingface.co/

Figure 9: Configuration and Model Selection Process on CTL's "oobabooga" implementation.

A pivotal element of our methodology was utilizing a Python script to interface with the Mistral model's API. We submitted our unstructured data to the LLM and received well-organized data in JSON format, as illustrated in Figure 10 and Figure 11.

1_1	The plan of the village of Rome. Engraving, c.1565, Unknown artist, Rodolfo Amedeo Lanciani (contributor), 280x430 mm, Library of Archeology and Art History of Rome	Plan of a part of the city of Rome oriented north down. On the left stretch is the Tiber River, with the inscription "Mensura et canne romane 50" (scale and measuring rod), with a small ladder next to it. Near the river is the Castel Sant'Angelo, above an indication of the Trastevere district. In the center, the S. Spirito hospital and the Vatican basilica are clearly visible.	It partially recalls the map of Rome published in: Giulio Ballino, "De' disegni delle più illustri città, et fortezze del mondo...", In Vinegia, appresso Bolognino Zaltieri, 1569. The presumed date of publication can be deduced from the fact that in the plan, the Castel Sant'Angelo appears with the final layout of Pius IV, who died in 1565.

Figure 10: Example of unstructured data before processing


```
    "chassets": [
        {
            "mainSubject": "The plan of the village of Rome",
            "type1": "Engraving",
            "creator/artist": "Unknown artist",
            "contributor1": "Rodolfo Amedeo Lanciani",
            "date": "c.1565",
            "dimensions": "280x430 mm",
            "currentLocation": "Library of Archeology and Art History of Rome",
            "description": "The plan of the village of Rome. Engraving, c.1565, Unknown artist,  Rodolfo Amedeo Lanciani  (
            "keywords": [
                "Engraving",
                "Renaissance - Early Modern",
                "Rome oriented north",
                "Tiber River",
                "Castel"
            ],
            "EuropeanaID": "446/BVE0709804",
            "CERTHID": "1_1",
            "Unit": "Unit 1.1: Mapmakers of the past.  View on 4 maps of Rome depicting different aspects of the city",
            "id": 1
        },
        {
            "mainSubject": "Rome",
            "type1": "Engraving",
            "contributor1": "Rodolfo Amedeo Lanciani",
            "contributor2": "Ferdinand Bertelli",
            "date": "1567",
            "dimensions": "200x270 mm",
```


Figure 11: Example of structured data after LLM processing

This phase, involving the strategic guidance of the LLM to structure our data—known as prompt engineering[19]—is crucial. It involves constructing the right queries so the model delivers data in the precise format we require. This approach transformed the raw data regarding the VR experiments (Figure 10) into a structured format that aligns with our project's requirements (Figure 11), particularly facilitating the OpenRefine[20] process detailed in the subsequent section.

3.1.4.2. Data enrichment through OpenRefine

Following the structuring of our data, we leveraged OpenRefine, a powerful tool adept at refining and cleaning complex datasets. OpenRefine excels in transforming and enhancing messy data into a structured, accessible format. One of its standout features is the reconciliation capability, which we utilized to align our cultural heritage data with corresponding Wikidata entries, thereby infusing our dataset with additional context and depth.

This integration with Wikidata not only augmented our dataset with essential information but also embedded each cultural heritage artifact within a vast network of knowledge. Through this process,

[19] https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-prompt-engineering
[20] https://openrefine.org/

our artifacts are no longer standalone pieces of information but are intricately connected within a larger informational web. This network includes details on historical significance, creation context, and connections to related artifacts. For instance, as demonstrated in Figure 12, the location "via Caffaro" associated with the cultural asset id "CERTH_VR1_11" was enriched with a Wikidata link. This linkage enabled us to automatically enrich our knowledge graph with detailed information from Wikidata, such as the city and country where "via Caffaro" is located, among other potential data points available in Wikidata.

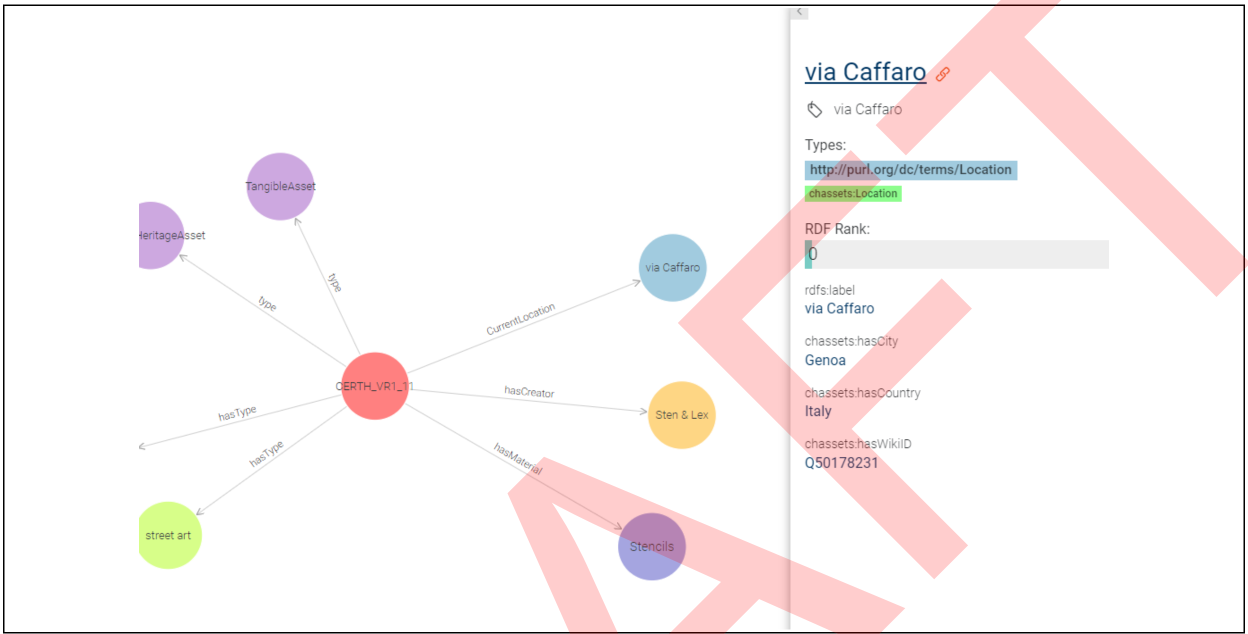


Figure 12: Linking 'via Caffaro' to Wikidata for Enhanced Cultural Asset Context

The outcome of this enrichment process was the transformation of structured data into .ttl (Turtle) files. These files were then seamlessly integrated into the cultural heritage assets ontology hosted on Ontotext's GraphDB[21]. This step not only enhanced the data quality but also enriched the virtual reality experience, offering users a more comprehensive and engaging exploration of cultural heritage assets. Through OpenRefine and the strategic use of Wikidata, we've significantly deepened the informational richness of our dataset, making each virtual asset a gateway to a broad spectrum of knowledge and historical context.

3.1.5. Future work

Future work includes expanding and enhancing the capabilities of the CH assets ontology. Since our goal is to provide a comprehensive vocabulary of classes and properties to accurately and thoroughly describe the various modalities of cultural assets, we will update the ontology as needed. Additionally, complex semantic relationships between various modalities will be established.

In order to improve data interoperability and further ease reuse across various platforms and applications, we also intend to further annotate our datasets. Besides this, aligning the current ontology with openly available models like CIDOC-CRM is necessary, to make the ontology more accessible and useful to third-party applications. We also plan to build links between the entities of our ontology and the corresponding Wikidata entities, to further improve the semantic context with

[21] https://graphdb.ontotext.com/

external, verified data. Lastly, the ontology will be also populated with more CH assets that will be used across the different tasks of the project.

3.2. Knowledge for multi-sensory data

3.2.1. Sensors Ontology Development

To create an engagement experience with CH assets, which either is an interaction with a VR environment, or by watching other relevant visual content, we recognized the necessity of a solid foundation that could interpret and classify user responses (Figure 13).

Figure 13: Engagement Experience with CH Assets through Visual Content at CTL

In developing our unique sensor ontology, we aimed to tailor a solution precisely fitted to the innovative multi-sensory VR experience we envisioned. Recognizing the limitations of existing ontologies to fully capture and categorize the nuanced emotional feedback from our custom sensors, we decided to build our ontology from the ground up. This decision allowed us to specifically address the data generated by our sensors, which not only detect emotions like happiness, sadness, and neutrality as well as stress levels for instance "stress", "no stress".

The decision to develop a standalone sensor ontology, rather than merely extending the existing CH ontology with additional properties, was multifaceted:

- Complexity and Specificity: The sensor data encompasses detailed emotional and physiological responses that require a level of granularity and specificity not present in general CH ontologies. A separate sensor ontology allowed us to tailor classifications and relationships precisely to the nature of the sensor data, ensuring accurate interpretation and processing.
- Modularity and Flexibility: Creating a distinct ontology for sensor data promotes modularity, allowing for easier updates, maintenance, and scalability. As sensor technology evolves or as we gather new types of emotional response data, the sensor ontology can be updated independently without impacting the structure or integrity of the CH ontology.
- Integration and Cross-Referencing: By developing separate ontologies and then integrating them through cross-referencing—particularly with CH artifacts like videos (Section 3.2.2)—we create a dynamic and flexible model. This approach allows us to map complex relationships between the sensor data and CH content, enriching the user's engagement experience by tailoring it based on emotional and physiological feedback.

In developing this ontology with Protege[22], we defined classes and subclasses to capture the complexity of human emotions in a format our system could understand. This carefully structured

[22] https://protege.stanford.edu/

ontology ensures reliable processing and categorization of sensor data, accommodating a wide range of user interactions and content types. Such an approach guarantees that the demo is not only immersive but also emotionally resonant, reflecting the depth and breadth of human experience.

The 'Video' class became the focal point of this ontology, serving as the bridge between the VR content and the user's emotional state (Figure 14).

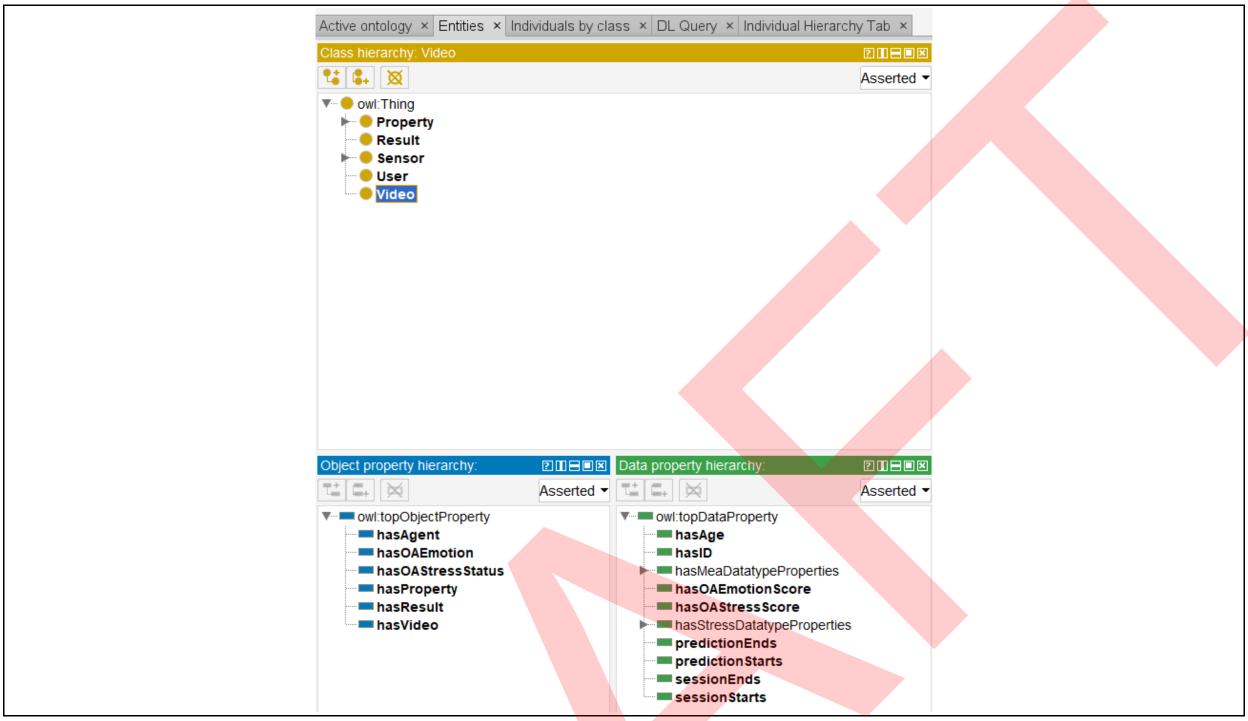


Figure 14: Sensor Ontology Example in Protege

For each video, sensor data was collected and categorized, allowing us to create a profile of emotional responses associated with specific content. This data was crucial, as it informed the adaptive aspects of the VR experience, such as music selection and environmental changes, which could enhance or alter the user's mood.

The ontology's structure also considered the temporal aspect of emotions, recognizing that user reactions could fluctuate throughout the VR experience. By tracking these changes over time, we could understand not just the immediate emotional response, but also the emotional journey of the user.

3.2.2. Integration with Cultural Heritage Data

Integrating the sensor ontology with our existing cultural heritage data significantly enriched the knowledge graph of the MuseIT project. The data integration process linked live emotional feedback from users, captured through our custom sensors, with static data about cultural heritage assets. This linkage was achieved using SPARQL rules, particularly employing the owl:sameAs property to associate videos within the cultural heritage ontology to the sensor data instances (Figure 15).

Connect instances of videos... | Retrieve sensorsIDs | Retrieve chassetsIDs | Quer

```sparql
PREFIX museITsensors: <http://www.semanticweb.org/museITsensors#>
PREFIX museITchassets: <http://www.semanticweb.org/museITchassets#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>

INSERT {
  ?se owl:sameAs ?ch .
}
WHERE {
    ?se rdf:type museITsensors:Video ;
        museITsensors:hasID ?id1 .
    ?ch rdf:type museITchassets:CulturalHeritageAsset ;
        museITchassets:hasID ?id2 .
    FILTER (?id1 = ?id2)
```

Figure 15: SPARQL query to associate Video instances from the two ontologies

This merging process was facilitated by SPARQL rules that matched sensor data to the corresponding cultural heritage content. The owl:sameAs property played a significant role in this process, as it allowed us to indicate that two distinct URIs actually referred to the same thing—such as a video in the cultural heritage ontology being the same as a video instance in the sensor ontology (Figure 16).

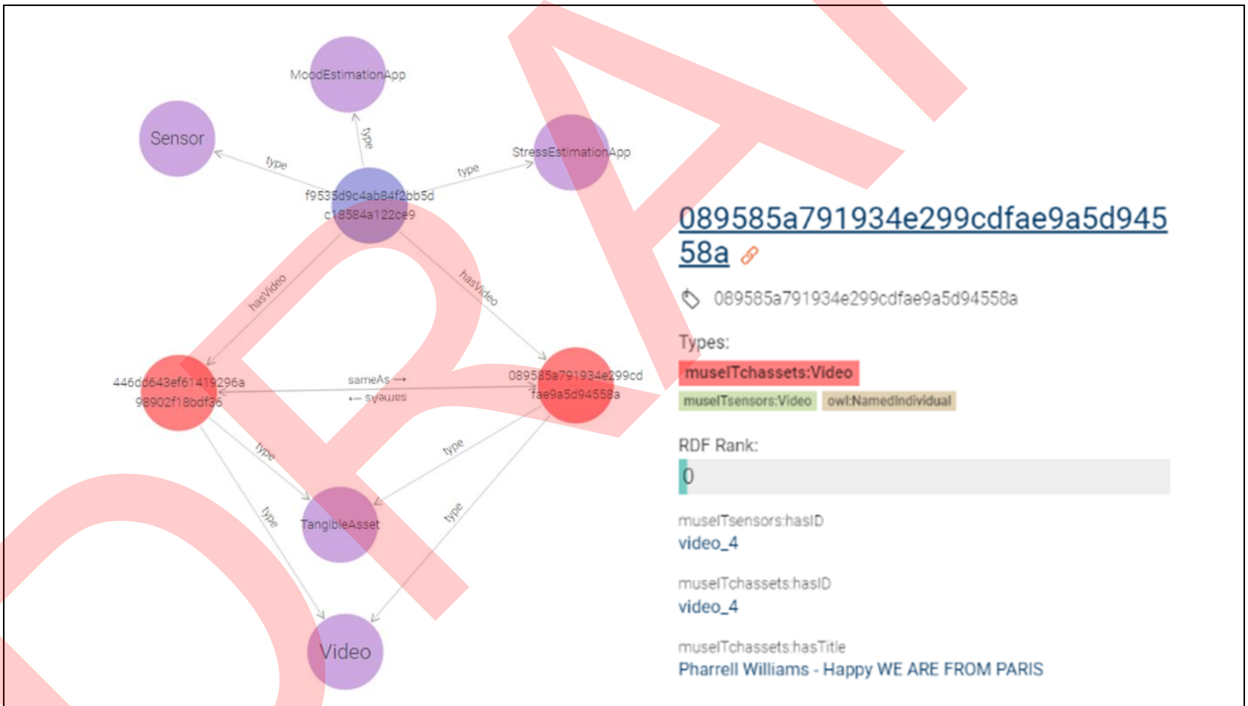


Figure 16: Integration in GraphDB - Mapping CH and Sensor Ontologies

The goal of this integration was straightforward: to deepen our understanding of user experiences and ultimately create a new form of music derived from this video asset, considering the metadata of the asset such as genre, keywords, title, age, and the end-users' senses, namely mood estimation and stress estimation while watching the video. For example, a user's joyful reaction to a "happy" video like Pharrell Williams' "Happy WE ARE FROM PARIS"[23] could not only enhance the user's engagement through personalized content recommendations but also be fed to AI tools that generate music based

[23] https://www.youtube.com/watch?v=hZ5rR0WlEkQ

on those data. This integration allows for the creation of unique musical pieces that mirror the emotional undertones identified in the user's responses, leveraging the metadata such as genre, keywords, and user mood. By utilizing AI in this manner, the platform can offer a truly bespoke experience, crafting soundscapes that reflect the user's current emotional state or even counterbalance their stress levels, thereby enriching the interaction with cultural heritage content in a deeply personal and innovative way.

These techniques, which are detailed in section 4, brought a new dimension to how cultural heritage content can be presented and experienced in the digital era. The subsequent section will delve into how we populated the knowledge graph with sensor data based on the framework provided by the sensors ontology.

3.2.3. Ontology population with Sensor Data

For the MuseIT project, enriching our Knowledge Graph with real-time sensor data is pivotal in creating an adaptive and immersive cultural heritage experience. To achieve this, we selected CASPAR, a transformative framework known for its efficiency in integrating sensor data into KGs. It efficiently converts JSON data into RDF triplestore, utilizing automatic generation and execution of SPARQL queries. This enables the immediate translation of sensor readings into a semantic format that GraphDB can interpret and store, thus significantly enhancing the ontology with real-time data.

The use of CASPAR is motivated by its modular design, which supports various connectors as depicted in the overall architecture in Figure 17.

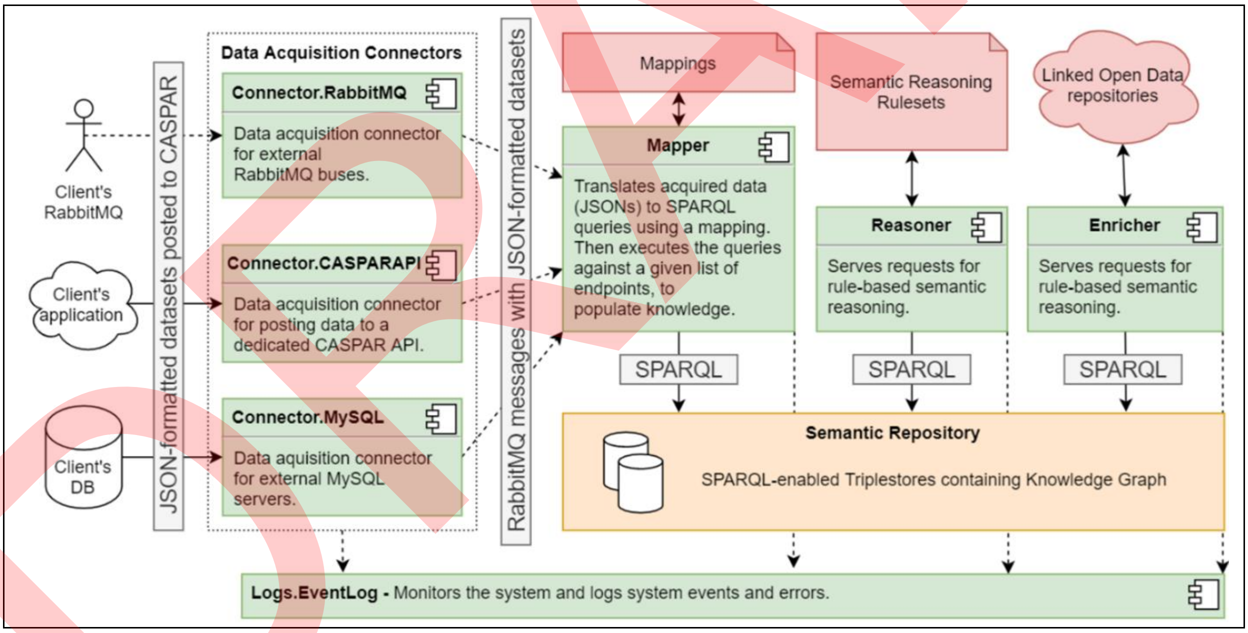


Figure 17: CASPAR's overall architecture

Specifically, the RabbitMQ Connector plays a crucial role in retrieving data from a network of sensors, which are then fed into CASPAR's Mapper component. This setup is ideal for handling our custom sensor data as it provides the flexibility to define unique JSON schemas for mood and stress estimation, an example of which is depicted in Figure 18.

```
"header": {
    "client_key": "museit_client_mea_key"
},
"body": {
    "mea": [
        {
            "id": "22556765356",
            "videoID": "9",
            "sessionStarts": "2023-06-07T09:30:00Z",
            "sessionEnds": "2023-06-07T09:30:20Z",
            "overallEmotion": "sad",
            "overallScore": 50.0,
            "user": [
                {
                    "id": "testuser",
                    "age": 28
                }
            ],
            "observations": [
                {
                    "normAvgConfHappy": 1.88,
                    "normAvgConfNeutral": 417.07,
                    "normAvgConfSad": 47.70,
                    "normAvgConfOther": 1.23,
                    "predictionStarts": "2023-06-07T09:30:00Z",
                    "predictionEnds": "2023-06-07T09:30:05Z",
                    "totFramDuration": 16,
                    "emotion": "sad"
                },
                {
                    "normAvgConfHappy": 56.88,
                    "normAvgConfNeutral": 12.07,
                    "normAvgConfSad": 7.70,
                    "normAvgConfOther": 8.23,
                    "predictionStarts": "2023-06-07T09:30:05Z",
                    "predictionEnds": "2023-06-07T09:30:12Z",
                    "totFramDuration": 7,
                    "emotion": "neutral"
                }
            ]
        }
    ]
}
```


Figure 18: Example JSON of Mood Estimation Algorithm

For each sensor data type, we have designed a specific mapping within CASPAR. These mappings guide the Mapper on how to translate the incoming JSON data into the corresponding ontological structure within GraphDB (Figure 19).

```
"templates": [
    {
        "prefixes": [
            {
                "prefix": "rdf",
                "namespace": "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
            },
            {
                "prefix": "museITsensors",
                "namespace": "http://www.semanticweb.org/museITsensors#"
            }
        ],
        "individuals": [
            {
                "path": "mea.[*]",
                "namespace": "http://www.semanticweb.org/museITsensors#",
                "classes": ["museITsensors:MoodEstimationApp"],
                "properties": [
                    {
                        "predicates": ["museITsensors:hasID"],
                        "object": {
                            "path": "mea.[*].id",
                            "datatype": "xsd:string"
                        }
                    },
                    {
                        "predicates": ["museITsensors:hasOAEmotionScore"],
                        "object": {
                            "path": "mea.[*].overallScore",
                            "datatype": "xsd:decimal"
                        }
                    },
                    {
                        "predicates": ["museITsensors:hasAgent"],
                        "object": {
                            "path": "mea.[*].user.[*]"
                        }
                    }
                ],
                "update_on": [{
                        "predicates": ["museITsensors:hasID"],
                        "object": {
                            "path": "mea.[*].id",
                            "datatype": "xsd:string"
                        }
                    }]
            },
            {
                "path": "mea.[*].videoID",
                "namespace": "http://www.semanticweb.org/museITsensors#",
                "classes": ["museITsensors:Video"],
                "properties": [
                    {
                        "predicates": ["museITsensors:hasID"],
```

Figure 19: Mapping Corresponding to Figure 18 schema

The process benefits from RabbitMQ's messaging capabilities, where each new JSON data packet is instantaneously transformed into RDF and populates the GraphDB hosted at CTL's premises (Figure 20).

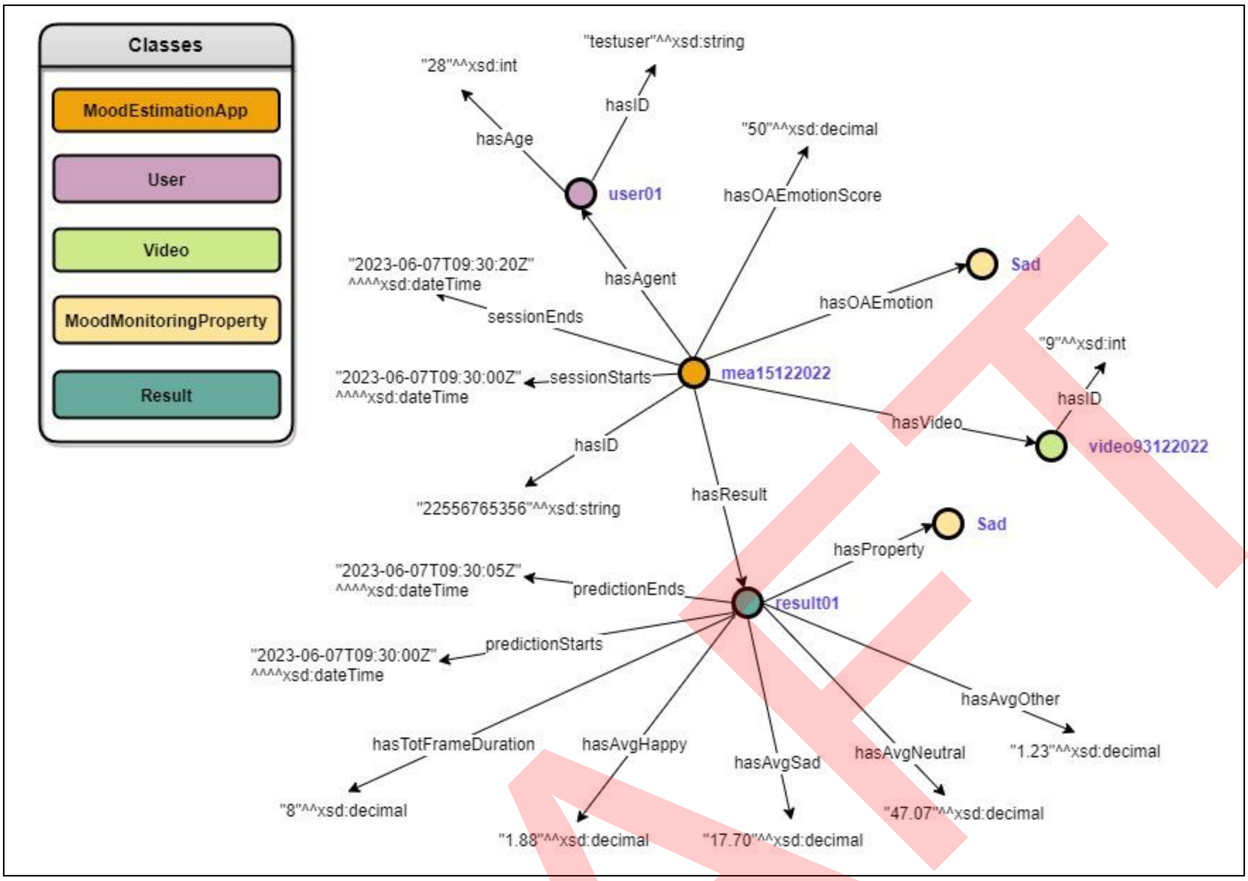


Figure 20: Illustration of Data Transformation via CASPAR for data in Figure 18

The automation brought by CASPAR through RabbitMQ ensures that as soon as new sensor data arrives, it is immediately processed. This smooth process not only minimizes the time lag between data capture and data utilization but also negates the need for manual data entry, reducing potential human error and ensuring a consistent flow of information into our system. CASPAR's ability to handle large volumes of data with its sophisticated mapping and automation processes makes it an indispensable tool in achieving the MuseIT project's goal of creating a dynamic and responsive visual experience, where every change in sensor data is reflected in real-time within the ontology, providing an ever-evolving narrative of user interaction.

3.3. Analysis of semantic drift in cultural heritage texts

3.3.1. Introduction

The ongoing evolution of technology, social norms, and language poses a constant threat to the preservation of digital artefacts over time. Whether it's the rapid advancements in technology or the changing landscape of societal norms, digital artefacts encapsulating cultural expressions are at risk of losing relevance or undergoing shifts in meaning. All language, including the language that we typically use to refer and describe cultural heritage, is subject to change at various levels (lexical, syntactical, semantic, pragmatic, etc.) as time passes by and human societies use and transform the language for different uses. This challenge is particularly pronounced in the context of multisensory experiences, which, whether transformed through technological progress or rooted in primal forms, are inherently susceptible to alterations in their semantics. Such alterations add a layer of complexity to their representation in a multimodal context, where various sensory elements converge.

Effectively addressing these challenges requires a comprehensive approach that goes beyond mere documentation of changes. Understanding, detecting, monitoring, measuring, and interpreting the evolution of semantic content within digital artefacts are critical aspects of ensuring their continued significance. This nuanced perspective becomes essential to maintaining the authenticity and relevance of cultural expressions in the face of dynamic technological landscapes and evolving societal norms. In navigating this complex interplay between technology, cultural elements, and language, we can better preserve and represent our cultural heritage in the digital age.

3.3.1.1. Measuring changes in language

While language is often mistakenly perceived as a stable and unchanging structure, it is, in fact, constantly evolving and adapting to the needs of its users. The semantics of words in a language shift due to influences from social practices, events, and political circumstances (Keidar et al., 2022), (Azarbonyad et al., 2017), (Castano et al., 2022). Understanding these shifts is crucial for grasping the dynamic nature of language and its intricate relationship with social and cultural elements. The task of Semantic Shift Detection (SSD) in Natural Language Processing (NLP) focuses on detecting, interpreting, and assessing potential changes in the meaning of a word over time.

In this work, we identify the changes in the semantics of words or phrases in the context of detection of offensive language. As explained in previous works (Hoeken et al., 2023), (McGillivray et al., 2022), changes in language semantics over time can influence what is considered offensive.

Words or phrases that may not have been considered offensive in the past can undergo changes in meaning, leading to their inclusion in offensive language categories. Similarly, terms that were once offensive might experience shifts in meaning, potentially becoming less offensive or even acquiring positive connotations. Analyzing and understanding these shifts in meaning are essential for developing effective NLP models that can accurately identify and respond to offensive language in different contexts.

3.3.1.2. Offensive language in disability domain

The detection of offensive language within the context of disabilities is a critical task with significant societal implications and of particular relevance to MuseIT: enabling a broader access to cultural heritage for all entails a careful choice of language that is mindful of disability and that acknowledges that its adequacy might change over time. Individuals with disabilities often face unique challenges and sensitivities, and the inappropriate use of language can contribute to the perpetuation of stereotypes, discrimination, and stigmatization (Andrews, Powell, & Ayers, 2022). Detecting offensive language in this domain involves the development of algorithms and models capable of recognizing and flagging content that may contain derogatory terms, slurs, or insensitive language targeting individuals with disabilities.

An example of offensive words experiencing semantic drift in the context of disabilities is the term “lame". Historically, this word was primarily associated with physical disabilities or injuries affecting a person's ability to walk or move normally[24]. However, over time, the word has undergone semantic drift, and in certain contexts, it has evolved to be used informally or colloquially to describe something as uninteresting, unimpressive, or subpar.

In this case, the semantic drift has led to a shift in the word's meaning from a physical disability context to a more casual and potentially derogatory usage. While some people may use it without intending to offend, it can perpetuate negative stereotypes and contribute to an insensitive language environment. Recognizing and addressing such instances of semantic drift is crucial for promoting awareness and fostering a more inclusive and respectful language use, particularly in discussions related to disabilities.

Another example of offensive words experiencing semantic drift in the domain of disabilities is the term “retarded." Originally used as a clinical term to describe individuals with intellectual disabilities[25], the word has undergone a significant shift in meaning over time. Due to misuse and derogatory appropriation, it has evolved into a derogatory slang term to insult someone's intelligence or mock their abilities.

3.3.1.3. Task Description

In this study, we concentrate on identifying offensive or derogatory language within the realm of disability discourse. While our current analysis is synchronous in nature, meaning it examines language use within a specific timeframe, the framework we have developed holds the potential for seamless extension into a diachronic evaluation. By doing so, we can gain valuable insights into how language evolves and changes over time within the context of disability discourse.

3.3.2. Related Works

3.3.2.1. Semantic Shift Detection

The Semantic Shift Detection (SSD) or Lexical Semantic Change Detection (LSCD) in Natural Language Processing (NLP) is defined as the task of identifying target words that change meaning over time (Schlechtweg et al., 2020). In MuseIT, we defined semantic drift or semantic change as the evolution that occurs in the meaning of some words or concepts over a course of time. Traditionally, the basic approach, a model-agnostic approach, to detect semantic changes involves finding the difference between two sets of embeddings from different embedding models trained with different corpora, at different time periods (Goel & Kumaraguru, 2021). In similarity comparison, there are two representative methods based on cosine similarity: Inverted cosine similarity and Average Pairwise cosine Distance (APD) between embeddings (Kutuzov, Velldal, & Øvrelid, 2022). This approach aims to detect lexical semantic change of a target word based on cosine similarity-based measures; thus it is limited in considering dynamic word distribution according to the relationships among them or the patterns of the corpus. On the contrary, the model-specific approach relies on unsupervised methods

---

24 https://ht.ac.uk/category/?type=search&qsearch=lame&word=lame&label=&category=&oef=&oel=&startf=&startl=&endf=&endl=&currentf=&currentl=&year=&twoEdNew=&twoEdUpdated=&page=1#id=9510

25 https://ht.ac.uk/category/?type=search&qsearch=retarded&word=retarded&label=&category=&oef=&oel=&startf=&startl=&endf=&endl=&currentf=&currentl=&year=&twoEdNew=&twoEdUpdated=&page=1#id=80273

such as clustering (Montariol, Martinc, & Pivovarova, 2021), topic modeling (Kutuzov, Velldal, & Øvrelid, 2022), or the vector space model (Loureiro et al., 2022), (Artetxe, Labaka, & Agirre, 2018). This approach is able to reflect the word distribution of a corpus sensitively, but its performance in SSD is unstable due to its sensitivity to data patterns rather than focusing solely on lexical semantic changes.

In terms of embedding methods, they can be categorized into token/type embedding and contextual embedding methods. Token/type embedding methods were proposed in the early stages of the SSD field to utilize simple NLP embedding techniques such as Word2Vec or GloVe. Contextual embedding methods such as BERT or ELMo have gained significant attention due to their generalization and predictive performance in NLP tasks (Martinc, Novak, & Pollak, 2019), (Laicher et al., 2021). However, many studies have reported that contextual embedding methods have not been explored sufficiently, and their performance is still limited in detecting semantic drift. (Kutuzov, Velldal, & Øvrelid, 2022) analyzed a problem of contextual embedding methods that could often be fuzzy in distinguishing changes between word distributional change and diachronic semantic shifts (Kutuzov, Velldal, & Øvrelid, 2022). (Schlechtweg et al., 2020), (Goel & Kumaraguru, 2021) described another reason that is the training process with a lack of historical corpora and proper utilization of it. Recently, many studies have been paying greater attention to using contextual embedding and pre-trained language models to detect semantic drift over time (Laicher et al., 2021), (Zhou, Tahmasebi, & Dubossarsky, 2023), (Laurino, De Deyne, Cabana, & Kaczer, 2023), (Card, 2023), (Periti et al., 2023). Most of them attempt to fine-tune the pre-trained models under the sense of semantic drift with historical datasets. Additionally, some studies have adopted the concept of word association and semantic graphs to focus on diachronic word meaning and avoid the influence of neighboring words that are irrelevant for detecting semantic drift in the contextual embedding task.

3.3.2.2. Hate Speech Detection

With the rapid growth of social media and online communication platforms, hate crime on these platforms is drastically increasing (MacAvaney et al., 2019). Detecting hate speech has become more complex due to the transition from traditional media to social media platforms. Social media's characteristics, such as fragmented and brief text, dynamic content, rapid dissemination, and large-scale reach, pose challenges for identifying hate speech effectively. The fast-paced nature of social media means that harmful content can spread swiftly, making timely detection crucial to mitigate its impact. (Gröndahl et al., 2018), (Mossie & Wang, 2020). Traditional approach based on keywords and lexicon has a fundamental limitation for interpreting short and varying posts of social media (Gitari, Zuping, Damien, & Long, 2015). To gain robust and generalized predictive performance, machine learning-based hate speech detection has been mainly proposed, recently. The studies in early stage have proposed to train machine learning models in the NLP such as Support Vector Machine, Wor2Vec, and LSTM (MacAvaney et al., 2019), (Alrehili, 2019). These studies have proved its performance for detecting hate speech for large-scale online social media, but they suffer from the lack of label information and low generalized performance (Mozafari, Farahbakhsh, & Crespi, 2020).

To address of the lack of label information, transfer learning with pre-trained model has gained a great attention (Mozafari, Farahbakhsh, & Crespi, 2020), (Yuan et al., 2019), (Toraman, Şahinuç, & Yilmaz, 2022), (Zia, Castro, Zubiaga, & Tyson, 2022). (Mozafari, Farahbakhsh, & Crespi, 2020) has proposed BERT-based four fine-tuning strategies and architectures with Nonlinear layer, Long Short-Term Memory Networks (LSTM), and (Convolutional Neural Networks) CNN. (Yuan et al., 2019) designed three steps that are pre-trained, shared, and task-specific classification (hate speech detection) using Bi-LSTM model. (Toraman, Şahinuç, & Yilmaz, 2022) constructed large-scale tweet datasets for hate speech detection in English and Turkish with human labeled tweets. (Zia, Castro, Zubiaga, & Tyson, 2022) has proposed a novel pipeline consisting of zero-shot, cross-lingual transfer learning based on pseudo-label and transformer language model.

3.3.3. Data Preparation

3.3.3.1. Dataset

In this study, our focus lies on examining a dataset pertinent to the discourse surrounding disability. It is notable that acquiring data in this field presents challenges, due to the limited participation from individuals with disabilities in research or survey initiatives. Addressing this issue, Palonis et al. (Palonis et al., 2023) have curated a substantial anonymized dataset sourced from social media platforms, predominantly Reddit, called Addrec. This dataset encompasses discussions on three distinct topics: ADHD, blindness, and disability at large. The selection of social media as a data source offers a unique advantage, as it captures spontaneous and candid conversations from diverse individuals, thus providing a nuanced understanding of the disability discourse. The dataset's value is highlighted by the quality of comments and its inclusivity, making it a good candidate for our semantic analysis of the disability discourse.

The dataset comprises publicly available comments from the three topics spanning from January 1st, 2015, to December 31st, 2019. Before any further filtering steps are applied, it contains a total of 1,526,980 comments.

Though the dataset contains a large number of samples, among them are irrelevant and non-informative posts for semantic drift detection in the domain of disability-related hate speech. In order to obtain a set of useful samples, we employed several data preparation steps that are described in the rest of this section. As a first step, we performed a keyword-based filtering.

3.3.3.2. Disability terminology-based filtering

Given the social media origin of the dataset, not all texts were pertinent to analyzing offensive language in the discourse. Some posts were brief responses or unrelated to meaningful conversations. To isolate relevant texts from Reddit comments, we employed filtering techniques based on keywords related to disabilities.

These keywords were compiled from various reputable sources to ensure a comprehensive coverage of relevant terms and phrases. We gathered terminology from research papers in esteemed journals (Walsh, Peterson, & Judkins, 2014), university-affiliated centers[26] (Burgstahler & Comden, 1994),

[26] https://cdsc.umn.edu/cds/terms

independent social organizations[27,28,29] and legal guidelines[30]. This approach aimed to encompass a diverse range of language associated with the topic of disabilities.

A total of 110 relevant keywords were gathered for filtering the dataset. These keywords were then utilized to match against the lemmatized texts of the Addrec dataset. Comments that included any of these keywords or phrases were retained, while those that did not were discarded. Following this filtering process, the dataset contained 59,639 comments or posts. The keywords, along with the count of their appearance in the three topics, is shown in the below table.

Table 1. Keywords with their frequencies

ADHD		Blind		Disability	
keyword	**appearance**	**keyword**	**appearance**	**keyword**	**appearance**
slow	15455	slow	434	slow	467
crazy	14879	crazy	159	attack	460
attack	7466	dumb	113	crazy	421
dumb	6240	attack	108	cripple	295
insane	3936	idiot	63	dumb	203
idiot	3450	insane	58	idiot	153
cripple	1131	lame	18	insane	146
nuts	744	cripple	13	retard	71
retard	660	moron	10	lame	44
moron	583	nuts	9	crip	42
lame	574	retard	8	moron	28
retarded	322	psycho	5	nuts	26
psycho	230	retarded	3	retarded	25
maniac	227	maniac	2	crippled	17
lunatic	133	midget	1	psycho	16
imbecile	27	deformed	1	lunatic	5
demented	24	wacko	1	maniac	3
wacko	22	stutterer	1	deformed	3
crippled	19	crip	1	midget	3
deformed	18			deranged	2
disfigure	13			disfigure	1
midget	12			unsound	1
deranged	11			imbecile	1
unsound	10			demented	1
stutterer	4				
brain-damaged	4				
feeble-minded	3				
Total	56197		1008		2434

[27] https://www.adl.org/sites/default/files/documents/2022-07/disability-glossary_3.pdf

[28] https://archive.anti-bullyingalliance.org.uk/sites/default/files/field/attachment/Ato-Z-of-Offensive-language-FINAL.pdf

[29] https://www.nsta.org/glossary-disability-terminology

[30] https://www.courts.ca.gov/partners/documents/7-terminology.pdf

3.3.3.3. Sentiment Analysis

After the filtering by keywords, we obtained a subset of the dataset that could be used for the annotation task of hate speech detection over time.

However, the annotation task is notably labor-intensive and exhausting. Furthermore, identifying hate intention in social media posts can be very implicit and latent due to their brevity and lack of context, making it challenging to detect meaning changes over time due to the rapid evolution of word meanings. To alleviate the difficulties of the annotation task, we further employed two types of indicators to support the annotation process.

The first indicator is the predicted sentiment label and score based on TimeLMs[31], which is a sentiment analysis model within a set of diachronic language models trained on Twitter data. TimeLMs was selected as a supportive model to predict sentiment labels because it is a comprehensive model that includes various natural language tasks, including offensive detection tasks on diachronic Twitter sets, aligning well with the scope and objectives of our dataset. Sentiment analysis is a conventional NLP task that involves analyzing people’s impressions regarding various topics or subjects, typically aiming to detect the positive and negative sentiments expressed by the writer through the analysis of written text. It's important to note that a negative sentiment label doesn't necessarily indicate hate or offensiveness in a post; however, we believe it provides valuable context for annotating hate speech by offering a sense of sentiment.

3.3.3.4. Relevance analysis

The second indicator is the classification result indicating whether a post is relevant to the domain of disability, predicted using the Llama-2-13b model[32]. Due to the absence of specialized models in a wide range of domain classifications, such as the disability domain, we utilized a Large Language Model (LLM) that demonstrates best performance independent of specific datasets. Llama2's open nature not only reduces licensing and API costs but also offers flexibility for various manipulations like free fine-tuning and parameter freezing. This adaptability makes Llama2 a suitable choice for future research and expandability in this domain.

Post domain classification, particularly considering relevance to disability, presents a challenging task due to the brevity of most posts, lack of context, and highly varied word distribution, making it difficult to train a classification model effectively. Additionally, there is a lack of concrete pre-trained models, methods, or label sets for detecting disability relevance. To address this issue, we utilized the Llama-2-13b model to assess the relevance of a post to disability. While this approach may have inaccuracies, it helps to avoid the need for manual investigation of the entire dataset.

[31] https://doi.org/10.48550/arXiv.2202.03829

[32] Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., ... & Scialom, T. (2023). Llama 2: Open foundation and fine-tuned chat models. arXiv preprint arXiv:2307.09288.

3.3.3.5. Dataset Statistics

The statistics of the final datasets, including keyword filtering, sentiment analysis, and relevance prediction, are summarized in Table 1. As shown in Table 1, the ADHD subset of the dataset is considerably larger in scale compared to the blind and disability datasets. In terms of the Type-Token Ratio (TTR), the posts in the ADHD dataset appear to be simpler, with relatively fewer token types compared to the blind and disability datasets. This observation is consistent when examining the average tokens and types columns. While the ADHD dataset comprises numerous posts, including very short ones, the blind and disability datasets contain denser posts that are relatively longer than those in the ADHD dataset.

Regarding sentiment ratios, the ADHD dataset demonstrates a mixed sentiment without a strong bias towards negativity. The blind dataset appears to be biased towards non-negative sentiments, whereas the disability dataset is evidently biased towards negative sentiments. In terms of relevance assessment, we calculated the relevance ratio of posts evaluated as relevant to the disability domain to those deemed irrelevant (Table 1). Since the relevance ratio in Table 2 is higher than 10 for all three datasets, there are overwhelmingly more relevant posts than irrelevant posts. The majority of posts across the three datasets exhibit strong relevance to the domain of disability.

Table 2. Statistics of final datasets: The Type-Token Ratio (TTR) is calculated as the ratio of tokens to types, multiplied by 1000. Average tokens and types represent the mean number of tokens and types respectively. The sentiment ratio is determined by the ratio of negative to positive sentiments, while the relevance ratio is computed as the ratio of relevance to irrelevance.

Corpus	Posts	Tokens	Types	TTR	Avg. tokens	Avg. types	Sentiment ratio	Relevance ratio
ADHD	56197	10812473	5613877	519.2	192.4	99.9	1.06	13.64
Blind	1008	214930	108452	504.59	213.22	107.59	0.74	10.33
Disability	2434	527366	264690	501.91	216.67	108.75	1.51	14.6

3.3.4. Method

For the offensive detection task, we initially had multiple types of labels such as hate speech, offensive language, hate speech targeting disability, neutral, etc. However, for the purpose of this deliverable, we have simplified the task by treating it as a binary classification problem. We annotated the datasets by collapsing all offensive-related labels into the category 'offensive', while retaining the other category as 'neutral'. This simplification allows us to better focus on considering temporal aspects in future work.

To test our datasets, we adopt two types of offensive detection methods. The first type is pre-trained model-based methods for detecting offensive content in social media. HateXplain is a pre-trained model for hate speech detection with manually annotated datasets collected from Twitter (Mathew et al., 2021). TimeLMs is a set of pre-trained models specialized in diachronic Twitter data such as hate and offensive detection, sentiment analysis, etc. (Loureiro et al., 2022). The other type is Large Language Model (LLM)-based methods with zero-shot and few-shot learning approaches. When there is sufficient labeled information available, we could choose different approaches like fine-tuning.

However, due to the limited scale of labeled datasets that we have available for use as the test set, we opted for zero-shot and few-shot learning approaches that do not require a training set. Due to the huge success of LLM models, the adoption of using LLMs is expanding in various NLP tasks. The advantages of LLMs in NLP tasks include relieving the necessity of a large-scale labeled dataset for training due to their high generalization performance and ability to handle multiple NLP tasks.

We adopt Llama2 7b and 13b models as offensive detection methods. These models can serve as suitable base models due to their predictive performance across various natural language tasks, making them versatile models. Additionally, existing methods often struggle to perform well on long texts, such as our dataset, as they are typically designed and trained for short texts like Twitter posts. To obtain desired classification results from LLMs, it is essential to design appropriate prompts. We conducted several tests to constrain the answer patterns of Llama2 models to only one word, 'Neutral' or 'Offensive,' as a classification result. The designed prompt for zero-shot and few-shot learning is depicted in Figure 1. Zero-shot is a prediction task without any additional examples for classification. Few-shot means using a small number of examples to provide helpful information for classifying datasets to models, thereby enhancing the understanding of the test dataset.

Llama2 models, particularly the 7b model, tend to generate overly verbalized text. To avoid overly verbalized text, we designed simple and short prompts with minimal constraints such as 'please respond with only one word.' Furthermore, we found that a higher temperature parameter could cause unexpected answers not desired as class labels, since the temperature parameter adjusts the diversity of generated text in the Llama2 model. For example, when we set the temperature parameter to 0.7, the Llama2 model sometimes produces undesired text such as '[no response],' 'Neutral/offensive,' or irrelevant texts. We empirically found that '0.2' is a reasonable value for the temperature parameter to obtain class labels from the Llama2 model. Additionally, we set the value of the repetition penalty to '1.19' to obtain clean labels and avoid noisy labels such as '\n offensive' or 'NeutralNeturalNN'.

Zero-shot prompt

```
###Below is a text that requires classification. Your task is to determine whether the text is offensive or
not. Please respond with only one word: "Neutral" or "Offensive".
###Text: {input_text}
###Classification result:
```

Few-shot prompt

```
###Below is a text that requires classification. Your task is to determine whether the text is offensive or
not. Please respond with only one word: "Neutral" or "Offensive".
An example of text to be classified as "Neutral" : {Text annotated as Neutral}
An example of text to be classified as "Offensive" : {Text annotated as Offensive}
###Text: {input_text}
###Classification result:
```

Figure 21. Prompt design for zero-shot and few-shot learning of offensive detection.

Lastly, we adopt the instruct-based fine-tuned Llama2 model with Knowledge Graphs (KGs) for abusive language detection (KG-Llama2-7b). This model enhances the knowledge of LLMs by providing descriptions of entities extracted from KGs in a prompt. We selected this model to assess the knowledge-enhancement capabilities of LLMs on different datasets. Furthermore, it utilizes the HateXplain dataset as the training dataset, making it a good model for comparison with the HateXplain model on our datasets.

3.3.5. Evaluation

In this section, we explain the process by which we obtained ground truth labels and measured the performance of the models for the detection of offensive language.

3.3.5.1. Annotated Label Information

The Addrec dataset, which is focused on disability discourse, aligns well with our project's objectives. However, it lacks explicit labels for offensive language. To accurately assess the effectiveness of the discussed models, we conducted manual evaluations. To accomplish this, we formulated an annotation task aimed at gathering manual labels from collaborators within the MuseIT project. These labels indicated whether a given text sample (which corresponds to a post or comment in the dataset) could be categorized as offensive or derogatory towards people with disabilities disability community, or if it was neutral. An additional option allowed annotators to mark a sample as irrelevant if it did not pertain to the topic. Only the relevant samples were utilized in the evaluation process and subsequent score calculations.

A total of 100 samples were randomly selected, each drawn in equal proportions from the filtered datasets related to Disability, ADHD, and Blindness. These samples were then distributed among 10 annotators for labeling purposes. Each sample underwent scrutiny from three distinct annotators.

To gauge the agreement among annotators, we computed the inter-annotator agreement using the Fleiss' Kappa metric [1]. Fleiss kappa is a statistical measure used to assess the reliability of agreement between multiple raters when categorizing items into multiple categories. This metric provided a quantitative measure of the level of consensus among the annotators regarding the classification of the text samples.

The overall and the label-wise scores for the Fleiss' Kappa metric are shown in the below Table 3

Table 3. Inter-annotator agreement scores

Label	Score
Overall	0.267
Offensive	0.142
Neutral	0.249
Irrelevant	0.33

The overall Fleiss kappa score of 0.267 indicates a relatively low level of agreement among the annotators. Upon closer examination of the individual category scores, it becomes evident that the agreement varies across different label categories. The score for offensive labels, at 0.142, suggests particularly poor agreement among annotators when identifying offensive content. This indicates a greater subjectivity or ambiguity in determining which samples were offensive in the dataset.

On the other hand, the agreement scores for the neutral and irrelevant categories are 0.249 and 0.330 respectively. While these scores are higher than that for offensive labels, they still fall within the range indicating only fair to moderate agreement. It suggests that even for less contentious categories like neutral or irrelevant, there remains a notable degree of discrepancy among annotators in their assessments.

To resolve the conflicts in the context of the low agreement scores, the samples were assigned the offensive label whenever any one annotator deemed it as such. For the other categories, the labels were chosen with a majority selection. After discarding the samples regarded as irrelevant by most of the annotators, 80 samples with their ground truth labels were used for the experiments and evaluations, the details of which are described in the next sections.

3.3.5.2. Experimental Results

We obtained classification results with the annotated datasets (as detailed above) and state-of-the-art offensive detection methods: Llama2-7b (zero-shot), Llama2-13b (zero-shot), HateXplain, OD-TimeLMs (subtask model for offensive detection), Llama2-7b-FL (few-shot), Llama2-13b-FL (few-shot), KG-Llama2-7b, and KG-Llama2-7b-FL. We tested these models for the offensive detection task without fine-tuning process for our dataset. For the evaluation of classification results, we adopted traditional measures: accuracy (Acc), precision (Prec), recall (Rec), and F1-score (F1). The results for different models are shown in Table 4.

Table 4. Offensive detection result for Addrec dataset with manual annotations as true labels

	Acc	Precision	Recall	F1-score
Llama2-7b	0.43	0.127272727	0.4375	0.197183099
Llama2-13b	0.18	0.15625	0.9375	0.267857143
HateXplain	0.16	0.16	1	0.275862069
OD-TimeLMs	0.16	0.16	1	0.275862069
Llama2-7b-FL	0.49	0.127659574	0.375	0.19047619
Llama2-13b-FL	0.23	0.141176471	0.75	0.237623762
KG-Llama2-7b	0.3	0.153846154	0.75	0.255319149
KG-Llama2-7b-FL	0.31	0.115942029	0.5	0.188235294

In Table 4, the Llama2-7b-FL model exhibits the lowest recall rate among all models, indicating a bias toward the 'Neutral' class and a struggle to perform effectively when faced with imbalanced data. Moreover, it suggests inadequacy in accurately identifying minority classes, such as offensive content.

Both HateXplain and TimeLMs failed to address the issue of skewed class distribution within our annotated labels. These models, trained on short texts from Twitter, lacked the capacity for generalization to longer texts, as present in our dataset. Language models performed comparatively better, while the Llama2-7b and 13b models performed reasonably well overall, as anticipated, the 7b model demonstrated instability, notably during few-shot learning scenarios.

Considering overall metrics, KG-Llama2-7b and KG-Llama2-7b-FL demonstrated the most stable and satisfactory performance. However, similar to the original Llama2-7b, KG-Llama2-7b-FL easily exhibited bias, resulting in higher recall at the expense of accuracy improvement. In other words, KG-Llama2-7b-FL, like the original Llama2-7b-FL, readily presented biased results and failed to find an appropriate balance point between accuracy and recall. The most stable and balanced model is KG-Llama2-7b. From the results, it is evident that it yields sufficiently stable results for imbalanced datasets and there is a good balance point between accuracy and recall.

3.3.6. Qualitative Analysis

In this section, we delve into instances where the models produce incorrect predictions and explore potential reasons behind them.

While the models generally demonstrate proficiency in identifying offensive text, a notable number of false positives are observed in the results.

For instance, several examples deemed neutral by annotators were categorized as offensive by all models except llama2_7b_pred and llama2_7b_FL_pred. This discrepancy may stem from the presence of specific keywords (marked in red) that are commonly linked with offensive language, leading the models to misinterpret the broader context and misclassify the texts as offensive.

> I've thought about a Braille display, but they are incredibly expensive and I doubt I could have it subsidized, I'm terribly slow at reading Braille, and the fact that even an 80-cell Braille wouldn't be enough to display a full terminal line would only slow me down further.

> You have two options. You can use direct typing the way I did (see below), or you can try out FlickType.
>
> The way I learned direct touch typing was with a lot of preparation. Before iOS 8 when it was added, I used touch typing for 4 years and never let myself make a mistake or drag my finger to find a letter. If I mistyped a letter in a word, I deleted the whole word and started over. It was hellishly frustrating, but I could almost immediately type with direct touch and not make any errors that autocorrect wouldn't fix. After a while, touch typing just felt slow and restrictive. I did have a Speed Dots protector on my phone for about six months, so that might have also helped me learn the positions of the keys a bit better. As far as I know, you can still get one.
>
> FlickType is the lazy option, but it is very good and much more forgiving. Even with the practice I've had, I prefer to use it, though I'm thankful I forced myself to memorize the keyboard and didn't grow up with any other option for typing on the screen..

In some cases, all the models incorrectly labeled text as offensive even though human annotators marked it as neutral. Such an example is shown below.

This suggests that the models may have been exposed to derogatory texts about blindness during training, leading them to overgeneralize and classify innocuous statements as offensive. Since large language models operate as black boxes, understanding their behavior is challenging, particularly with a limited sample of annotated texts.

He is blind from birth. I think he just got out of the habit of opening them. He opens them in one video, it is crazy how difficult it is for him, cause he never uses those muscles.

Also some blind people's eyes hurt. I mean, blind people are blind for actual medical reasons, many of which are not comfortable.

Light sensitivity is also pretty common. Most blind people can at least detect light and it is pretty common for it to hurt their eyes.

One thing I would like to know is if you have absolutely zero vision, could your eyes still hurt in bright light? That would be weird. u/fastfinge might know.

3.3.7. Conclusion

The intersection of cultural heritage and accessibility presents a pressing need for effective offensive and semantic drift detection, particularly within the context of diachronic data analysis. Despite its critical importance, the advancement of research in this field has been impeded by the scarcity of annotated datasets tailored to specific cultural heritage contexts and accessibility needs. Existing datasets are designed for general-purpose offensive language detection, overlooking the nuanced requirements of cultural heritage preservation and accessibility considerations. Moreover, the time intervals analyzed in these datasets are often too brief to capture the significant semantic shifts that occur over longer periods within cultural heritage contexts

To address these limitations, we utilized the Addrec dataset, which focuses on disability discourse from Reddit and performed manual annotations for a subset of the data. We conducted several processes to filter out irrelevant samples and supported the annotation task with sentiment analysis and the Llama2-13 model-based disability relevance prediction.

The experimental results pertaining to the classification task have shown that our dataset could be effectively used for offensive detection in the disability domain. Furthermore, it contains timestamp data for all posts, allowing us to expand the offensive detection task to include offensive and semantic drift detection for disability-related content.

However, the dataset has several limitations in its current form. Firstly, although we attempted to filter out irrelevant posts related to disabilities using selected keywords, we still found irrelevant or insufficient posts remaining for analysis in offensive detection. Secondly, the number of annotation samples is relatively small due to the effort of manual labelling, and the ratio of samples marked as offensive is quite small while also exhibiting a low inter-annotator agreement. Thirdly, some baseline models need to be fine-tuned for our dataset to properly evaluate its performance.

In the future work, we aim to address the identified limitations systematically. In collaboration with our partners at Stanford and HB, we are investigating alternative sources of disability-specific datasets such as newspaper articles archives or scientific publications from Web of Science (https://www.webofscience.com/wos) and PubMed (https://pubmed.ncbi.nlm.nih.gov/). To obtain a larger and more qualified set of annotated samples, we are collaborating with various experts in the field of disabilities. Once we have completed the additional annotation tasks, we plan to fine-tune the baseline models accordingly. Furthermore, we are considering the use of larger LLM models, such as Llama2-70b, to better estimate the relevance of posts to disabilities and the informativeness of posts to be classified. Finally, we intend to expand our experiments and evaluations to include offensive and semantic drift detection in the context of disabilities.

Finally, it is important to note that the task of offensive language detection that we have presented in this work is inherently subjective, influenced by cultural, social, and contextual factors. To address this subjectivity, incorporating a broad spectrum of linguistic expressions and iteratively refining models with input from individuals within the disability community is essential and it can enhance accuracy and inclusivity. This is planned in association with the different partners in the form of co-design activities in the future.

3.4. Knowledge from social media and the web

DANS is building a realtime collection of news on disabilities published in the MuseIT Dataverse and available for further analysis being done by KCL and CTL. It includes multilingual materials harvested from the Web on keyword ‘disability’ and processed by Natural Language pipelines in order to extract corresponding information like title, summary, keywords, resource name and authors. To avoid copyright issues there is no full text archived so this information isn’t available for the direct analysis and not included in metadata records. Dataverse also stores screenshots of all news taken automatically, with some limitations related to bot protection, and they’re available as links in the metadata record of every news item.

The archiving of materials on disabilities from social media is limited to the registration of public links with some description taken from public materials published on Twitter and Facebook, which is related to the license limitations of the corresponding social media platforms.

Example of the record in JSON-LD format is available below:

@context	"http://schema.org"
@type	"Dataset"
@id	"https://doi.org/10.5072/FK2/NJGV8A"
identifier	"https://doi.org/10.5072/FK2/NJGV8A"
name	"Monaco children learn about disability at school - Monaco Tribune"
creator	
0	
@type	"Person"
affiliation	

@type	"Organization"
name	"MuseIT"
name	"MuseIT"
author	
0	
@type	"Person"
affiliation	
@type	"Organization"
name	"MuseIT"
name	"MuseIT"
datePublished	"2024-01-19"
dateModified	"2024-01-19"
version	"1"
description	"Around 1,500 pupils of all ages are taking part in the week-long disability awareness campaign ©Unsplash Community and solidarity are the watchwords of the initiative, which seeks to promote more inclusion at school. The second edition of 'La Semaine de l'Ecole inclusive' (Inclusive Schools Week) has been taking place from 15 to 19 January. Over the 5 days, 78 classes from across the Principality and children from the Prince Albert II Leisure Centre have been taking part in workshops on the theme of disability. These are run in small groups and cover the different types of disability: motor, sensory and cognitive."
keywords	
0	"Arts and Humanities"
license	"https://now.museum/api/datasets/:persistentId/versions/1.0/customlicense?persistentId=doi:10.5072/FK2/NJGV8A"
includedInDataCatalog	
@type	"DataCatalog"
name	"Root"
url	"https://now.museum"
publisher	
@type	"Organization"
name	"Root"

provider	
@type	"Organization"
name	"Root"
distribution	
0	
@type	"DataDownload"
name	"screenshot_0.2439866814411784.jpg"
encodingFormat	"image/jpeg"
contentSize	671946
description	""
contentUrl	"https://now.museum/api/access/datafile/31037"

This export is ready to be ingested by available triple stores such as GraphDB, Jena Fuseki or Virtuoso, and further analysis could be done by using SPARQL queries.

4. Decision making using semantic rules

In the MuseIT project, the combination of semantic technologies with cultural heritage and sensor data analysis leads to a unique outcome: the generation of music. This approach brings together smart data management and informed decision-making to produce music that reflects the diverse experiences users have with cultural heritage content and their reactions to it. At the foundation of our work is the integration of two kinds of data: the detailed historical and cultural context of heritage assets and the real-time emotional and physiological feedback from users through sensors. This integration enables us to understand both the content that users are interacting with and their reactions to it as it happens. By applying semantic rules to this combined data, our system can intelligently generate music that mirrors the user's emotional state and the cultural context of the content they are engaging with. For example, if sensor data shows a user feeling a sense of awe while exploring an ancient architectural site in a virtual reality environment, the system could generate a piece of music that enhances that sense of awe, using musical elements that are associated with the site's cultural background. This method of generating music dynamically makes the experience of exploring cultural heritage richer and more personal, blending visual and informational experiences with auditory feedback.

4.1. Rule-based semantics

Our primary goal is to leverage rule-based semantics to generate music that reflects dynamic and static data during engagements with cultural heritage (CH) assets. These engagements can occur within a virtual reality (VR) environment or through the exploration of related visual content. At the heart of this endeavor is our Knowledge Graph (KG), a semantic data integration platform that serves two main functions:

- Dual Ontologies: We maintain separate but interconnected ontologies for sensors and cultural assets, enabling us to cover the spectrum of data types and interactions within our system.
- Data Integration: By merging heterogeneous data items under a unified semantic model, we ensure that diverse information sources are cohesively structured and accessible.

Rule-based semantics serve as the crucial mechanism through which our platform interprets and acts upon the rich data provided by the cultural heritage and sensor data ontologies. Through the implementation of rules, our system can make intelligent decisions based on the user's interaction with content and their exhibited emotions. This proactive approach allows for a deeply personalized user experience, whereby the platform can present content that resonates on an informative and emotional level, in real time.

4.1.1. Semantic Rules in DCMC

In the DCMC, as illustrated in Figure 6, we showcased two specially designed ontologies; the first captures the intricacies of cultural heritage assets, while the second focuses on sensor data (as discussed in sections 3.1 and 3.2). As mentioned in deliverable D5.3 and elaborated in section 3.2, these ontologies were populated with data from two types of sensors—one for mood assessment and the other for stress level monitoring (referenced in section 3.2).

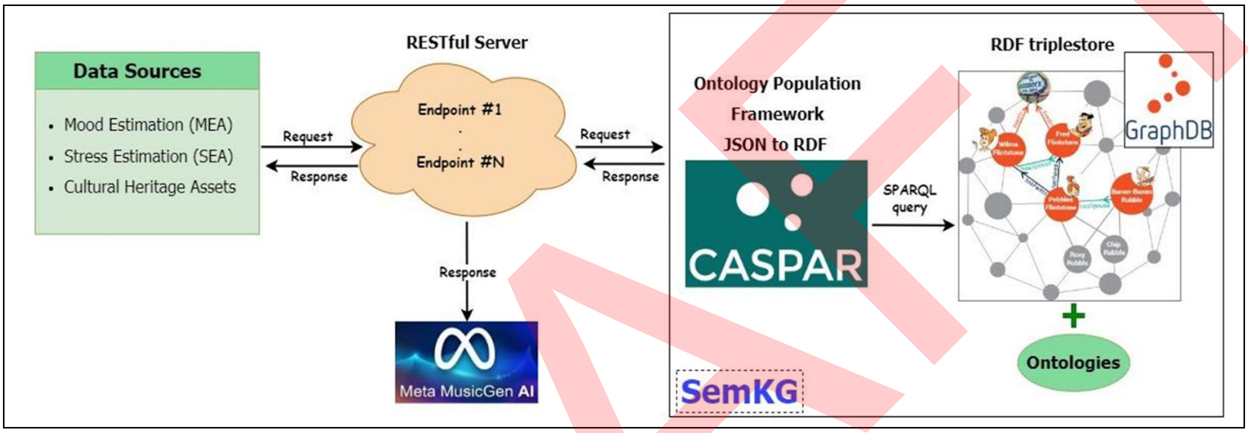


Figure 22: DCMC components illustration

Figure 22 presents a detailed diagram of how different components interact within the system. On the left, we introduce the data sources: cultural heritage content and readings from two sensors. This information is transferred to the Semantic Knowledge Graph (SemKG) through a RESTful server, using REST API connections. The SemKG incorporates CASPAR, GraphDB, and two ontologies, facilitating the structured storage and management of data. As shown in Figure 23, SPARQL queries are then executed within this environment to merge data from these varied sources.

```
PREFIX museITsensors: <http://www.semanticweb.org/museITsensors#>
PREFIX museITchassets: <http://www.semanticweb.org/museITchassets#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT ?id ?overallEmotion ?overallStressStatus (GROUP_CONCAT(DISTINCT ?musicgenre; separator=",") as ?concatenatedGenres)
       (GROUP_CONCAT(DISTINCT ?keyword; separator=",") as ?concatenatedKeywords)
WHERE {
    ?sensor rdf:type museITsensors:Sensor;
         museITsensors:hasID ?id;
         museITsensors:hasVideo ?video;
         museITsensors:hasOAEmotion ?moodProperty;
         museITsensors:hasOAStressStatus ?stressProperty.
    ?video museITchassets:hasKeyword ?chkeyword;
           museITchassets:hasMusicGenre ?genre.
    ?genre rdfs:label ?musicgenre.
    ?moodProperty rdfs:label ?overallEmotion.
    ?stressProperty rdfs:label ?overallStressStatus.
    ?chkeyword rdfs:label ?keyword.
}
GROUP BY ?sensor ?overallEmotion ?overallStressStatus ?id
```

Figure 23: SPARQL query generating input for MusicGen

The combined data string, as exemplified in Figure 24, is then fed into Meta's Music Gen. This system is specifically engineered to create melodies that reflect the integrated data, transforming abstract information into a harmonious musical expression.

	id	overallEmotion	overallStressStatus	concatenatedGenres	concatenatedKeywords
1	"session_01"	"neutral"	"calm"	"classical work,requiem,lament"	"tragic,deaths,disaster,war scene,orchestra,classical music"
2	"session_03"	"sad"	"calm"	"traditional Japanese music,electronic music sounds"	"Japan,tradition,Japanese technological advances/trends"
3	"session_04"	"sad"	"calm"	"jungle music (Drum and Bass),tribal music,ambient music,wildlife,nature-inspired soundscapes"	"tragic,death,lies,animals,jungle"
4	"session_06"	"happy"	"calm"	"Pop,Uptempo soul,Neo Soul"	"France,Paris,dance,sightseeing,tour,happy people"

Figure 24: Example of Combined Data String for Melody Generation in Meta's Music Gen

Communication between these components and the process of generating melodies is managed through the RESTful server. For more information on Meta's Music Gen and its melody generation capabilities, with further details provided in the next chapter.

One of the objectives was to study viewer interactions with cultural content, as well as their concurrent emotional and stress responses. This information was leveraged to make strategic decisions on the platform. For example, by understanding viewer reactions, we could automatically select music that not only matched the video content but also aligned with the viewer's emotional state, thus enriching the user experience.

With SPARQL, we extracted valuable insights, such as:

- Calculating the average emotional and stress responses for each video.
- Identifying sessions with heightened stress levels to inform content strategy and support mechanisms.
- Pinpointing users who frequently displayed certain emotions, which could be used to customize content offerings.
- Listing the top 10 cultural heritage assets that frequently induced happiness, contributing to content promotion strategies.
- Mapping the range of stress responses among users, which provided a comprehensive view of content interaction and user engagement.

This exploration into rule-based semantics and the integration of cultural heritage and sensor data showcases the potential of semantic technologies in enhancing digital experiences. Through the thoughtful application of these technologies, we've laid the groundwork for a system that not only interprets complex data but also transforms it into an engaging, emotional journey for users. As we continue to refine our approach and delve deeper into the capabilities of Meta's Music Gen in the following section, we anticipate further advancements in how we interact with and understand cultural heritage in the digital age.

4.1.2. Music Generation enhanced by DCMC

As it was previously mentioned, the above-mentioned process aimed to generate music based on Dynamic and Static data. For the music creation aspect, we've chosen to employ Meta's MusicGen, which was launched in mid-2023. MusicGen represents an advanced, controllable text-to-music model designed for the generation of music based on specific conditions. It functions through a single Language Model (LM) that processes multiple streams of compressed, discrete representations of music, namely tokens. This model differentiates itself from previous models by employing a singular-stage transformer LM and employing efficient token interleaving patterns. This method removes the necessity to use multiple models in sequence, such as hierarchical or upsampling models. MusicGen's training utilized 20,000 hours of music, incorporating both an internal collection of 10,000 high-quality music tracks from Meta and music data from ShutterStock[33] and Pond5[34]. The research paper titled "Simple and Controllable Music Generation[35]" details a thorough empirical analysis by the creators, including both automated and human assessments.

The rationale behind our decision to adopt MusicGen encompasses several compelling factors. Primarily, the unique nature of the data contained within a Knowledge Graph, combined with the potent capabilities of text-based data, necessitated a text-to-music generation approach. MusicGen excels in this domain. By leveraging MusicGen's specific model strategies, it is capable of producing "high-quality" music samples that are conditioned on either text descriptions or melodic inputs, thus offering improved control over the music it generates based on specific textual inputs such as emotions and keywords. In addition, the analysis presented by the authors of MusicGen, demonstrates that their method outperforms existing baseline models on a recognized text-to-music benchmark. Some generated music samples, and a comparison with previous works, are presented in the official sample page[36].

In order to acquire a meaningful representative text that will serve as an input to the MusicGen model for a specific session, the metadata contained in the Knowledge graph is obtained with the use of the developed Restful API (Figure 22). Specifically, this text generation phase, consists of the following steps:

1. **Establish connection:** Connect to the RESTful server's endpoint.
2. **Fetch session's record:** Request and retrieve session's metadata based on a specific session ID.
3. **Extract Keywords:** Randomly sample m keywords from a total of n keywords in the metadata record, where $m<n$, $n \in \mathbb{Z}$ and $m \in \mathbb{Z}$.
4. **Select a Music Genre:** Randomly select a single music-genre from the music genres associated with the assets showcased in the specific session
5. **Create Base String:** Combine the selected keywords and genre into a base string.
6. **Modify String Based on Emotion:** If the predicted emotion is significant (not 'neutral'), append it to the string.
7. **Adjust Tempo Based on Stress:** Append "normal tempo" for calm stress levels or "fast tempo" for high stress levels to the string.
8. **Obtain MusicGen's Final Input String:** Complete the process and return the constructed string.

An illustration of an example of the text generation, is depicted in Figure 25

[33] https://www.shutterstock.com
[34] https://www.pond5.com/
[35] https://arxiv.org/abs/2306.05284
[36] https://ai.honu.io/papers/musicgen/

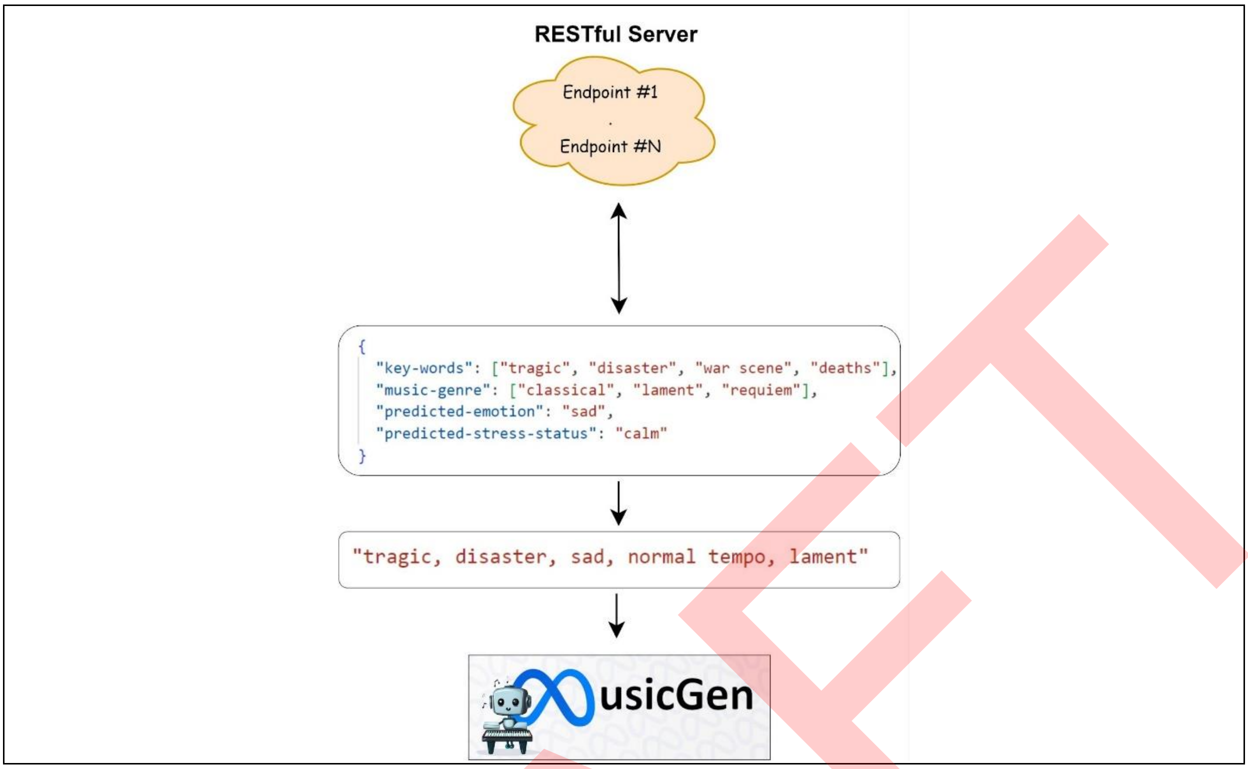


Figure 25: DCMC MusicGen

To employ MusicGen, Audiocraft, PyTorch library was utilized. Audiocraft is a more general library for audio processing and generation with deep learning. It features inference and training code for state-of-the-art AI generative models producing high-quality audio, including MusicGen. MusicGen's API offers a selection of ten distinct pre-trained models, each differing in parameter size, illustrating a common trade-off between model size and the quality of output. Furthermore, a crucial part that the API provides is to set the generation parameters. There are many arguments that their role could serve as hyperparameters as well (e.g. *top_k* or *top_p*) but the most crucial arguments that a user needs to be aware of are:

- **duration** : Denotes the duration of the generated waveform
- **extend_stride**: When doing extended generation (i.e. more than 30 seconds), stride denotes how much we should extend the audio each time. Larger values will mean less context is preserved, and shorter values will require extra computations).

MusicGen requires a GPU with at least 16 GB of memory for running inference with the medium-sized models (~1.5B parameters). Based on the official documentation, it is recommended to utilize a GPU with 16 GB of memory, but smaller GPUs are able to generate short sequences, or longer sequences with a smaller model.

In the development of our DCMC, the selection of the pre-trained model and hyperparameters was meticulously tailored to match our specific needs and limitations. Specifically, the pre-trained models available during the time of our demonstration were limited to four distinct options:

- small: 300M parameters, text to music only[37]

[37] https://huggingface.co/facebook/musicgen-small

- medium: 1.5B parameters, text to music only[38]
- melody: 1.5B parameters, text to music and text+melody to music[39]
- large: 3.3B parameters, text to music only[40]

Our hardware configuration includes an NVIDIA GeForce RTX 3060 GPU, equipped with 12 GB of memory. Given this constraint on memory capacity, we opted for the medium-sized text to music model. This choice was based on the model creators' recommendation, highlighting it as the optimal balance between output quality and computational demand [TODO: Add reference]. Furthermore, our experimentation revealed significant increases in GPU memory consumption when the **extend_stride** hyperparameter was used, occasionally resulting in system instability and crashes. Consequently, to mitigate these issues, we have decided to cap the music generation **duration** at the maximum allowable length of 30 seconds and to forego the use of stride extension.

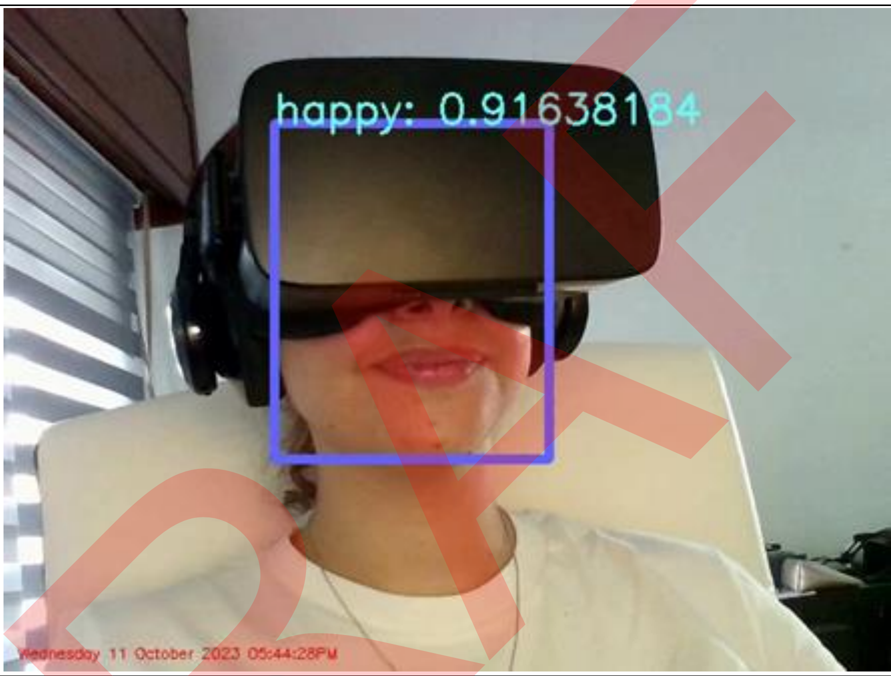


Figure 26: DCMC - Mood Estimation Sensor

Ultimately, the above-mentioned process leads to the composition of a thirty second music piece, based on a single session, which is an interaction of a user with a cultural heritage experience. In addition, the result of this composed music will be strongly based on the dynamic interaction of the users with the VR environment or video as shown in Figure 26. Furthermore, the objective of this process is to study the cross-modal encoding of ontology content, by encapsulating this emotional CH experience through the modality of audio, which is part of the work of T6.2. Finally, by obtaining these melodies, the enrichment of CH content is achieved.

4.2. From evolving semantics to multimodal representations

In 3.3 above, semantic drifts and their measurement gave us an example how changes in word meaning may lead to problems of understanding and interpretation of important concepts expressed by keywords. Next, we show how, due to interaction between collection development and acquisition

[38] https://huggingface.co/facebook/musicgen-medium
[39] https://huggingface.co/facebook/musicgen-melody
[40] https://huggingface.co/facebook/musicgen-large

policies, evolving semantics in longitudinal datasets about CH artefacts subtly influences understanding. This is the reason why, to address both descriptive language and inclusion at the same time, we decided to use robust LLMs to generate haptic translations of artefact captions, converting natural language into a haptic language to accompany their experience. This decision was in line with applying LLMs in Knowledge Engineering both worldwide and in this project (Minae et al., 2024). Our experimental solutions contribute to the Muse IT ontology.

4.2.1. Monitoring evolving topics in cultural heritage artefact collections

That second kind of "semantic dynamics" goes back to the changing topical composition of collections, leading to a period-specific, relativistic rather than absolute, notion of concept similarity expressed by distances in vector space. This symptom becomes apparent for instance from the acquisitions statistics of the Tate dataset, used below for exemplification. With data-driven museums in the intersection of data science and collection development being no more a rarity (Daish, 2017), through longitudinal studies it is possible to identify indicators pointing to the evolution of, e.g., discourse surrounding cultural heritage items, and provide an estimate of trends relating to represented items and creators (Tonkin et al., 2018).

Because term and document similarities are expressed by distances in vector space, but such distance structures have their origins in the time-dependent proportions between the number of artefacts vs. subsets indexed by certain keywords, our working hypothesis was that, given period-specific fluctuations in acquisition data, index term displacements will be observed in the analysed period. This followed from the concept and the importance of the *tfidf* measure to compute groups of items with related content (Spärck Jones, 1972). To that end we processed the dataset and designed a workflow to extract results to test our hypothesis.

4.2.2 Dataset

Tate, jointly with the National Galleries of Scotland, holds the national collection of British art from 1500 to the present day, plus international modern and contemporary art. The collection embraces all media, from painting, drawing, sculpture and prints to photography, video and film, installations and performance. The metadata for 69.202 artworks we used for analysis was published in 2014 in JSON format as open data for research and development purposes[41]. In accord with the nature of their 19th century holdings, the first chronological half of the dataset is dominated by the Turner Bequest (1856) that added approximately 30.000 items (in other counts, 41.000 including attributions of the "after" type[42]), works of art on paper including watercolours, drawings, plus 300 oil paintings. In the dataset, 53.698 records were timestamped. The artefacts were indexed by Tate's bespoke hierarchical subject index which has three levels, from general to specific index terms. These will be referred to as L1, L2 and L3, with an example entry under the heading Explore in the footnote.[43] The extracted dataset-specific subject index had 16.189 index terms.

---

[41] https://github.com/tategallery/collection

[42] https://github.com/parkan/collection-sans-turner

[43] https://www.tate.org.uk/art/artworks/turner-self-portrait-n00458

The subject index was originally created and developed alongside the digitisation of Tate's collection (a process which began in the late 1990s) as a means of providing extra keys into the collection by enabling visitors to search artworks via subject as well as artist name or artwork title. The design of the hierarchical structure and initial tagging of the bulk of collection artworks was carried out by a team recruited for this specific purpose, with Tate's curatorial team acting as advisers where necessary. The initial structure and key terms were based partly on a previously developed card index which Tate's Information team had compiled in response to popular enquiries by in-gallery visitors.

4.2.3 Experiment design

With two acquisition peaks in the dataset spanning altogether a hundred years, the number of incoming artefacts was 33.625 between 1795-1845, and 12.756 between 1960-2009. Subject indexing happened on three levels, the upper subject level having 21 persistent index terms present over all ten observation epochs vs. 22 in the second period. On a more granular, second level, the respective number of concepts used for indexing was 142 and 177, with 225 and 288 index terms on the lowest, conceptually most detailed subject level. For a proof-of-concept example, we created a period-specific series of 10 binary matrices over 5 years each to record only upper level (L1) index term occurrence rates over the respective document sets. As the numbers demonstrate, fluctuation in the proportions of subject matter was not the exception but the rule (Table 1, Figure 27).

Term level_epoch	L1_0	L1_1	L1_2	L1_3	L1_4	L1_5	L1_6	L1_7	L1_8	L1_9	Cumulated
Index terms	1795-1799	1800-1804	1805-1809	1810-1814	1815-1819	1820-1824	1825-1829	1830-1834	1835-1839	1840-1845	term use
architecture	1165	1227	826	1014	4168	2470	3219	4305	2308	2312	23014
belief	36	141	151	27	61	40	105	51	39	58	709
concepts	19	67	236	70	82	60	153	40	73	48	848
emotions	19	67	236	70	82	60	153	40	73	48	848
fiction	8	68	139	20	29	49	133	154	157	29	786
history	12	30	11	5	22	73	16	55	32	28	284
ideas	19	67	236	70	82	60	153	40	73	48	848
interiors	73	22	68	13	87	42	123	84	51	39	602
leisure	29	44	83	43	69	97	119	102	52	35	673
literature	8	68	139	20	29	49	133	154	157	29	786
nature	1412	2131	1245	1463	4462	2500	3412	4239	2281	2894	26039
objects	151	224	206	203	437	347	305	312	175	115	2475
occupations	95	100	193	91	187	209	208	230	134	56	1503
pastimes	29	44	83	43	69	97	119	102	52	35	673
people	442	619	797	317	1143	956	1031	1068	636	401	7410
personifications	23	43	219	255	405	319	205	170	49	155	1843
places	1118	1519	824	981	4408	1848	2105	2830	1913	1015	18561
religion	36	141	151	27	61	40	105	51	39	58	709
society	333	614	472	471	926	1497	955	1173	560	429	7430
symbols	23	43	219	255	405	319	205	170	49	155	1843
work	95	100	193	91	187	209	208	230	134	56	1503
Terms/epoch	*5145*	*7379*	*6727*	*5549*	*17401*	*11341*	*13165*	*15600*	*9037*	*8043*	
Artefacts/epoch & total	1776	2614	1944	1894	5801	3298	4546	5592	2825	3335	**33625**

Table 1: Level 1 index term use describing the collection between 1795-1845.

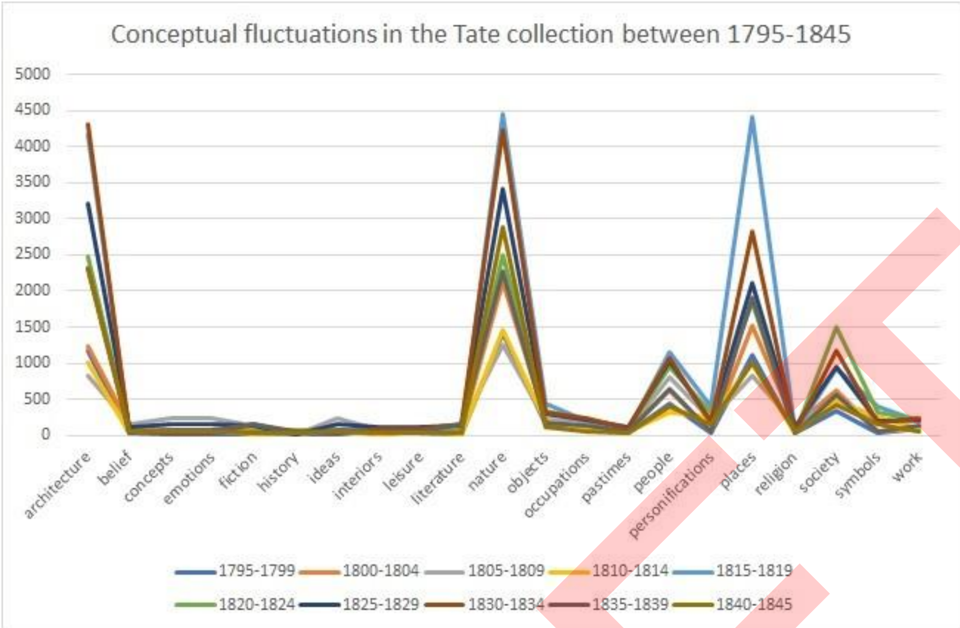


Figure 27: Conceptual foci of the dataset in the specified period.

Using the Orange data mining software (Godec et al., 2019), we built a workflow to analyse and visualise the 1795-1845 subcorpus by Hierarchical Cluster Analysis (HCA, Ward method), heatmaps, distance maps and contour maps, one of each for every 5 year timestep. For the contour maps, we extracted the *xy* coordinates by t-SNE (Van der Maaten & Hinton, 2008), and for the *z* coordinate added cluster-specific silhouette values computed by *k*-means (MacQueen, 1967). Finally, the *xyz* coordinate values indicating period-specific locations of index terms, including their distance structure, were visualised by 3dField[44], using the inverse distance interpolation method for contour drawing.

4.2.4. Results and evaluation

In the period in focus, the L1 level index terms most used were `nature`, `architecture`, `places` and `society`, expressing the topical profile of artefacts in the subset (Table 2). However, referring back to Table 1, within that period this profile was evolving over ten epochs, with relative stress on different concepts within any and all of them.

[44] https://3dfmaps.com/

(a) Not ranked		*(b) Ranked*	
architecture	23014	nature	26039
belief	709	architecture	23014
concepts	848	places	18561
emotions	848	society	7430
fiction	786	people	7410
history	284	objects	2475
ideas	848	personifications	1843
interiors	602	symbols	1843
leisure	673	occupations	1503
literature	786	work	1503
nature	26039	concepts	848
objects	2475	emotions	848
occupations	1503	ideas	848
pastimes	673	fiction	786
people	7410	literature	786
personifications	1843	belief	709
places	18561	religion	709
religion	709	leisure	673
society	7430	pastimes	673
symbols	1843	interiors	602
work	1503	history	284

Table 2: Level 1 index terms used in the analysed period in alphabetical vs. ranked order.

Visual inspection of the ten respective heatmaps (Figure 28) vs. contour maps for the first 15 years of collection development (Figure 29) confirmed that artefact vs. index term number and specificity changes resulted in different content distributions and topographies, with the implication to perceive related content such as word synonymy or artefact similarity in flexible ways. This confirmed the working hypothesis.

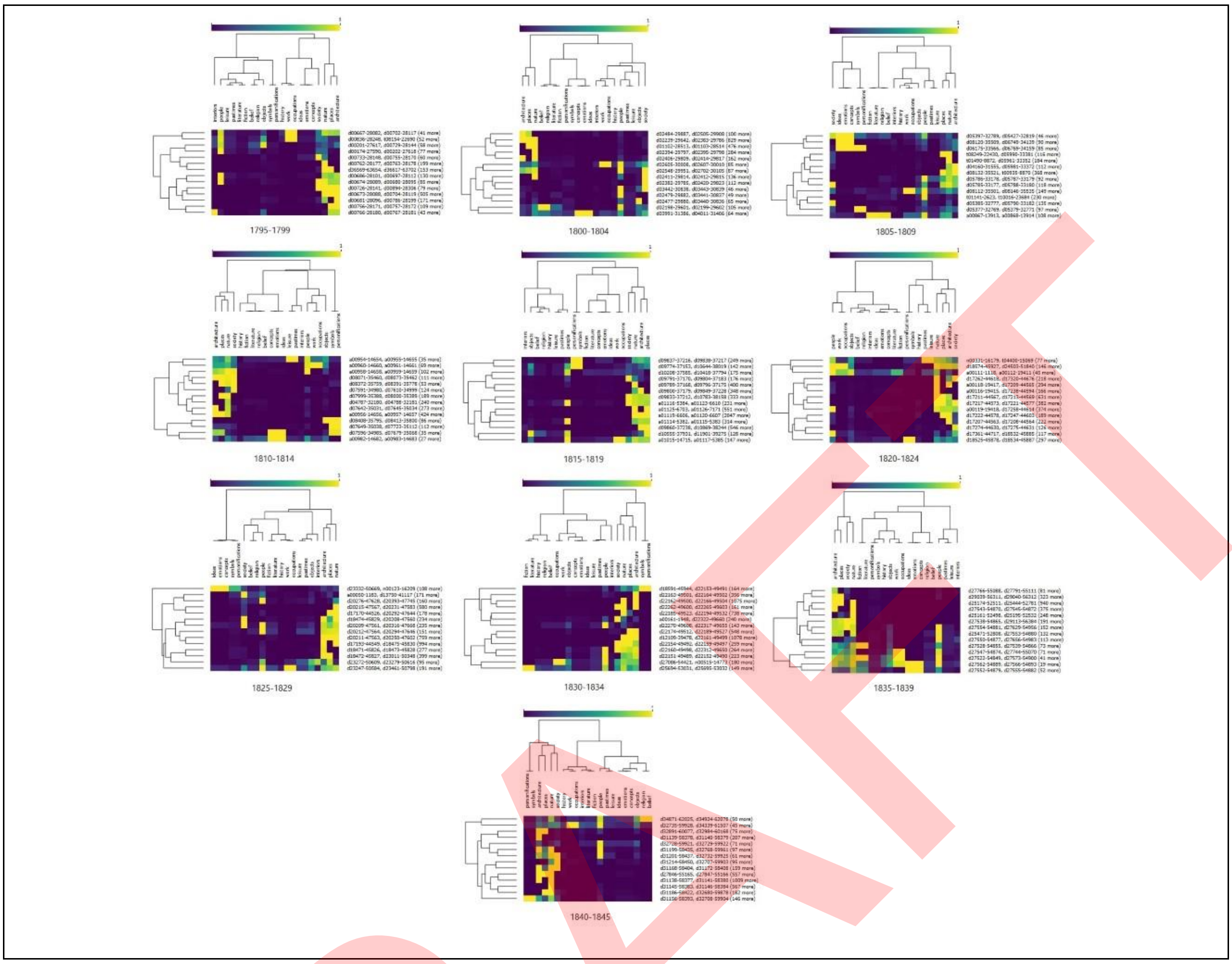


Figure 28: Distance-based heatmaps of collection development manifest differences in conceptual composition over time.

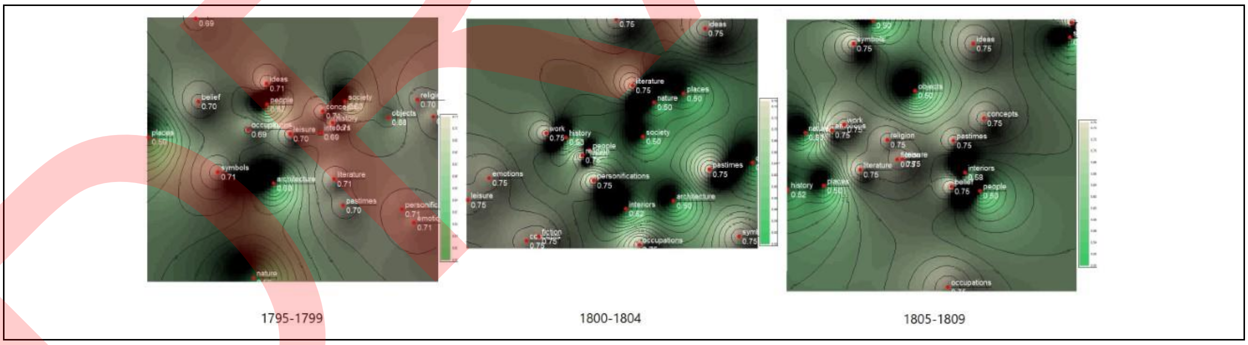


Figure 29: Three contour maps for 1795-1809 illustrate the changing semantic relatedness of high-level index terms.

4.3.1. Ongoing and future work: experimental considerations for new types of metadata

The above experiment exposed the need for robust statistical solutions when it comes to the experimental conversion of natural language-based content descriptions to modal equivalents. The solution proved to be to apply LLMs. With a proof-of-concept approach in mind both to compute new kinds of metadata for future purposes, and at the same time enable new kinds of sonic and tactile content experience for users with sensory deprivation, we continued to develop the ideas introduced in the first Technical Report, Part B as a particular way to integrate semantic content with emotional and neurophysiological signals.

We have been employing a two-pronged approach:

Track A: Enrich paintings with modal augments to 'translate' the visual experience

- Direction 1: J.M.W. Turner's and J. Sorolla y Bastida's seascapes turned into soundscapes to add the feeling of presence for users with visual impairment. Sound effects were retrieved from the BBC archive[45].
- Direction 2: paintings processed by a combined workflow of object detection (Redmon & Farhadi, 2018; Marinescu et al., 2020) to label entities present in them, contour detection to demarcate regions, finally colour detection within those regions by Region Adjacency Graphs (RAG) (Van der Walt et al., 2014). The extracted data can be fed to the KG. Both tracks are work in progress for UC1.

Track B: Modulated artefacts for the platform (AR & VR)

We realized that pretrained GloVe word embeddings (Pennington et al., 2014) can be converted to modal transcripts in basically two ways:

- Map a limited number of extreme coordinate values to a respective number of vibration actuators in a haptic grid for output. The same procedure can be employed to express semantic content as atonal chords, or emission/absorption lines in the visible spectrum. This scheme has importance first and foremost for the HaptiDesigner Toolkit (Olson & Järvoll, 2022) in T3.3 and Actronika's design solutions.
- Map the complete set of coordinate values for any word embedding vector to oscillation space by amplitude modulated frequencies. This track plans to couple neurophysiological signals, e.g., from AI-driven music generation, with concept signals in stimulus-response scenarios. Paving the way for such experiments, recently Lewis (2023) has shown that specifically engineered textile artefacts react to changes in the electromagnetic field.

5. Future work

The MuseIT project stands as a testament to the significant strides made in the realm of cultural heritage, showcasing the fruitful integration of semantic technologies, ontology development, and the pioneering application of generative AI for the creation of audio and music. This initiative has not only enriched the cultural heritage domain but has also set the stage for further exploration and innovation.

In the pursuit of enhancing ontology development and management, the project will focus on the continuous refinement of the CH assets ontology to integrate emerging cultural heritage concepts and digital artifacts. Emphasizing the expansion of the ontology to include intangible cultural heritage aspects like folklore and digital storytelling will be paramount. Furthermore, the development of intuitive semantic annotation tools is planned to ease the contribution of metadata by cultural heritage professionals and the public, thereby enriching the ontology's comprehensiveness and accuracy.

Advancing semantic interoperability remains a core objective. The project aims to foster data exchange and enrichment across various platforms and disciplines by enhancing semantic interoperability with other cultural heritage and academic databases. Expanding the use of linked open data standards is

[45] https://sound-effects.bbcrewind.co.uk/

also on the agenda, to connect with a broader ecosystem of cultural heritage information on platforms such as Dataverse, Wikidata, and DBpedia.

Interactive and immersive experiences are also a focal point, with efforts geared towards implementing AI-driven personalization algorithms. These algorithms will curate individualized cultural heritage journeys based on users' interests, cultural backgrounds, and learning goals, aiming to provide a more personalized and engaging exploration of cultural heritage.

In the realm of research and development, the project will continue its exploration into semantic drift and the evolution of language within the cultural heritage domain. Developing dynamic models capable of adapting to changing linguistic and cultural contexts will ensure the relevance and accuracy of cultural heritage representations. Additionally, the advancement of AI and machine learning for cultural heritage research will proceed, utilizing these technologies for pattern recognition, predictive analysis, and uncovering hidden connections within large cultural heritage datasets.

By embarking on these future directions, the MuseIT project aims to further elevate the accessibility, understanding, and preservation of cultural heritage, leveraging innovative digital solutions. These efforts promise not only to build on the project's existing accomplishments but also to unlock new avenues for discovery and engagement within the cultural heritage domain.

References

Davis, Randall, Howard Shrobe, and Peter Szolovits. "What is a knowledge representation?." AI magazine 14.1 (1993): 17-17.

Berners-Lee, Tim, James Hendler, and Ora Lassila. "The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities." Linking the World's Information: Essays on Tim Berners-Lee's Invention of the World Wide Web. 2023. 91-103.

Keidar, D., Opedal, A., Jin, Z., Sachan, M.: Slangvolution: A causal analysis of semantic change and frequency dynamics in slang. arXiv preprint arXiv:2203.04651 (2022)

Azarbonyad, H., Dehghani, M., Beelen, K., Arkut, A., Marx, M., Kamps, J.: Words are malleable: Computing semantic shifts in political and media discourse. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. pp. 1509–1518 (2017) Castano, S., Ferrara, A., Montanelli, S., Periti, F., et al.: Semantic shift detection in vatican publications: a case study from leo xiii to francis. In: CEUR WORKSHOP PROCEEDINGS. vol. 3194, pp. 231–243. CEUR-WS (2022)

Hoeken, S., Spliethoff, S., Schwandt, S., Zarrieß, S., Ala¸cam,¨O.: Towards detecting lexical change of hate speech in historical data. In: Proceedings of the 4th Workshop on Computational Approaches to Historical Language Change. pp. 100–111 (2023)

McGillivray, B., Alahapperuma, M., Cook, J., Di Bonaventura, C., Merono-Penuela, A., Tyson, G., Wilson, S.: Leveraging time-dependent lexical features for offensive language detection. In: Proceedings of the The First Workshop on Ever Evolving NLP (EvoNLP). pp. 39–54 (2022)

Andrews, E.E., Powell, R.M., Ayers, K.: The evolution of disability language: Choosing terms to describe disability. Disability and Health Journal 15(3), 101328 (2022)

Schlechtweg, D., McGillivray, B., Hengchen, S., Dubossarsky, H., & Tahmasebi, N. (2020). SemEval-2020 task 1: Unsupervised lexical semantic change detection. arXiv preprint arXiv:2007.11464.

Goel, A., & Kumaraguru, P. (2021, May). Detecting Lexical Semantic Change across Corpora with Smooth Manifolds (Student Abstract). In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 35, No. 18, pp. 15783-15784).

Kutuzov, A., Velldal, E., & Øvrelid, L. (2022). Contextualized language models for semantic change detection: lessons learned. arXiv preprint arXiv:2209.00154.

Montariol, S., Martinc, M., & Pivovarova, L. (2021, June). Scalable and interpretable semantic change detection. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 4642-4652).

Loureiro, D., D'Souza, A., Muhajab, A. N., White, I. A., Wong, G., Anke, L. E., ... & Camacho-Collados, J. (2022). TempoWiC: An evaluation benchmark for detecting meaning shift in social media. arXiv preprint arXiv:2209.07216.

Artetxe, M., Labaka, G., & Agirre, E. (2018). A robust self-learning method for fully unsupervised cross-lingual mappings of word embeddings. arXiv preprint arXiv:1805.06297.

Martinc, M., Novak, P. K., & Pollak, S. (2019). Leveraging contextual embeddings for detecting diachronic semantic shift. arXiv preprint arXiv:1912.01072.

Laicher, S., Kurtyigit, S., Schlechtweg, D., Kuhn, J., & Walde, S. S. I. (2021). Explaining and improving BERT performance on lexical semantic change detection. arXiv preprint arXiv:2103.07259.

Zhou, W., Tahmasebi, N., & Dubossarsky, H. (2023, May). The Finer They Get: Combining Fine-Tuned Models For Better Semantic Change Detection. In Proceedings of the 24th Nordic Conference on Computational Linguistics (NoDaLiDa) (pp. 518-528).

Laurino, J., De Deyne, S., Cabana, Á., & Kaczer, L. (2023). The pandemic in words: Tracking fast semantic changes via a large-scale word association task. Open Mind, 1-19.

Card, D. (2023). Substitution-based semantic change detection using contextual embeddings. arXiv preprint arXiv:2309.02403.

Periti, F., Picascia, S., Montanelli, S., Ferrara, A., & Tahmasebi, N. (2023). Studying Word Meaning Evolution through Incremental Semantic Shift Detection: A Case Study of Italian Parliamentary Speeches. Authorea Preprints.

MacAvaney, S., Yao, H. R., Yang, E., Russell, K., Goharian, N., & Frieder, O. (2019). Hate speech detection: Challenges and solutions. PloS one, 14(8), e0221152.

Gröndahl, T., Pajola, L., Juuti, M., Conti, M., & Asokan, N. (2018, January). All you need is" love" evading hate speech detection. In Proceedings of the 11th ACM workshop on artificial intelligence and security (pp. 2-12).

Mossie, Z., & Wang, J. H. (2020). Vulnerable community identification using hate speech detection on social media. Information Processing & Management, 57(3), 102087.

Gitari, N. D., Zuping, Z., Damien, H., & Long, J. (2015). A lexicon-based approach for hate speech detection. International Journal of Multimedia and Ubiquitous Engineering, 10(4), 215-230.

Alrehili, A. (2019, November). Automatic hate speech detection on social media: A brief survey. In 2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA) (pp. 1-6). IEEE.

Mozafari, M., Farahbakhsh, R., & Crespi, N. (2020). A BERT-based transfer learning approach for hate speech detection in online social media. In Complex Networks and Their Applications VIII: Volume 1 Proceedings of the Eighth International Conference on Complex Networks and Their Applications COMPLEX NETWORKS 2019 8 (pp. 928-940). Springer International Publishing.

Yuan, L., Wang, T., Ferraro, G., Suominen, H., & Rizoiu, M. A. (2019). Transfer learning for hate speech detection in social media. arXiv preprint arXiv:1906.03829.

Toraman, C., Şahinuç, F., & Yilmaz, E. H. (2022). Large-scale hate speech detection with cross-domain transfer. arXiv preprint arXiv:2203.01111.

Zia, H. B., Castro, I., Zubiaga, A., & Tyson, G. (2022, May). Improving zero-shot cross-lingual hate speech detection with pseudo-label fine-tuning of transformer language models. In Proceedings of the International AAAI conference on web and social media (Vol. 16, pp. 1435-1439).

Palonis, B., Dobesh, S.J., Bellscheidt, S., Mkaouer, M.W., Liu, Y., Elglaly, Y.N.: Large-scale anonymized text-based disability discourse dataset. In: Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility. ASSETS '23, Association for Computing Machinery, New York, NY, USA (2023). https://doi.org/10.1145/3597638.3614476, https://doi.org/10.1145/3597638.3614476

Walsh ES, Peterson JJ, Judkins DZ; Expert Panel on Health Care Disparities Among Individuals With Disabilities. Searching for disability in electronic databases of published literature. Disabil Health J. 2014 Jan;7(1):114-8. doi: 10.1016/j.dhjo.2013.10.005. Epub 2013 Oct 17. PMID: 24411515.

Burgstahler, S., Comden, D.: Disabilities, opportunities, internetworking and technology (do-it) on the electronic highway. In: Proceedings of the First Annual ACM Conference on Assistive Technologies. p. 153–156. Assets'94, Association for Computing Machinery, New York, NY, USA (1994). https://doi.org/10.1145/191028.191075, https://doi.org/10.1145/191028.191075

Mathew, B., Saha, P., Yimam, S. M., Biemann, C., Goyal, P., & Mukherjee, A. (2021, May). Hatexplain: A benchmark dataset for explainable hate speech detection. In Proceedings of the AAAI conference on artificial intelligence (Vol. 35, No. 17, pp. 14867-14875).

Loureiro, D., Barbieri, F., Neves, L., Anke, L. E., & Camacho-Collados, J. (2022). Timelms: Diachronic language models from twitter. arXiv preprint arXiv:2202.03829.

Daish, A. (2017). Data-Driven Museums. 12th International Digital Curation Conference. The British Museum.
https://www.dcc.ac.uk/sites/default/files/documents/IDCC17~/presentations/DataDrivenMuseumsAliceDaishIDCC17.pdf

Godec, P., Pančur, M., Ilenič, N. *et al.* Democratized image analytics by visual programming through integration of deep models and small-scale machine learning. *Nat Commun* **10**, 4551 (2019). https://doi.org/10.1038/s41467-019-12397-x

Lewis, E. (2023). Radiant Textiles : Designing electromagnetic textile systems (PhD dissertation, Högskolan i Borås). Retrieved from https://urn.kb.se/resolve?urn=urn:nbn:se:hb:diva-29751
MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability.* Vol. 1. University of California Press. pp. 281–297.

Marinescu, M., Reshetnikov, A., & Lopez, J. (2020). Improving object detection in paintings based on time contexts. *International Conference on Data Mining Workshops (ICDMW)*, Sorrento, Italy, 2020 pp. 926-932. doi: 10.1109/ICDMW51313.2020.00133

Minaee, S., Mikolov, T., Nikzad, N., Chenaghlu, M., Socher, R., Amatriain, X., & Gao, J. (2024). Large Language Models: A Survey. *arXiv preprint arXiv:2402.06196*.

Olson, N. & Järvoll, J. (2022). Haptic Pattern Designer Toolkit – HaptiDesigner: Software and Hardware for Creation of Actuation Patterns. In 16th International Conference on Universal Access in Human-Computer Interaction, UAHCI 2022 Held as Part of the 24th HCI International Conference, HCII 2022 / [ed] Antona M., Stephanidis C., Springer Science+Business Media B.V., 2022, p. 489-509.

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global Vectors for Word Representation. In EMNLP-2014, pp. 1532–1543.

Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv:1804.02767.

Spärck Jones, K. (1972). A Statistical Interpretation of Term Specificity and Its Application in Retrieval. *Journal of Documentation*. 28 (1): 11-21. CiteSeerX 10.1.1.115.8343. doi:10.1108/eb026526. S2CID 2996187.

Tonkin, E.L., Tourte, G.J.L., Gill, A. (2018). Crowd Mining Applied to Preservation of Digital Cultural Heritage. In: Vermeeren, A., Calvi, L., Sabiescu, A. (eds) Museum Experience Design. Springer Series on Cultural Computing. Springer, Cham. https://doi.org/10.1007/978-3-319-58550-5_6 .

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, *9*(11).

Van der Walt, S., Schönberger, J.L., Nunez-Iglesias, J., et al. (2014). scikit-image: Image processing in Python. PeerJ 2:e453(2014). https://doi.org/10.7717/peerj.453