Exploring the Evolution of Modern Art with ArtMovr and Linked data

Darshan Navadiya
Masters in Software Engineering
dnavadi1@asu.edu

Varun Menon
Masters in Software Engineering
vmenon12@asu.edu

Gokul Subramanian
Masters in Software Engineering
gsubra12@asu.edu

Vidit Sanghvi
Masters in Software Engineering
vmsanghv@asu.edu

Vamsi Krishna Yadav Loya
Masters in Software Engineering
vloya1@asu.edu

Abstract: Bridging the gap between the art and technology is a part of the new era of digital humanities where the beauty of human creativity and the power of advanced computing come together. With this, we introduce ArtMovr. ArtMovr is a cutting-edge web application that harnesses linked data and knowledge graphs to enable users to explore the vast Museum of Modern Art (MoMA) collection. With powerful search and exploration tools, artist insights, and interactive visualizations, ArtMovr allows users to delve into the rich tapestry of almost 200,000 artworks spanning 150 years. The application reveals intricate relationships between artists, artworks, and various concepts within the MoMA collection, providing valuable insights into the evolution and diversity of modern art. ArtMovr showcases the potential of linked data and knowledge networks to make art more engaging, informative, and accessible for art enthusiasts, researchers, and the public.

Index Terms—ArtMovr, Linked data, Museum of Modern Art (MoMA), The Metropolitan Museum of Art (The Met), Artist insights, Knowledge graphs, Collection, and Interactive visualizations.

I INTRODUCTION

As we stand on the brink of a new era where technology and art converge, a unique opportunity to explore and appreciate art in unprecedented ways emerges. The "ArtMovr" project is an initiative that combines art, technology, and knowledge representation. Its goal is to unlock the intricate world of art found in the Museum of Modern Art (MoMA) collection. MoMA, which was established in 1929, houses 200,000 artworks spanning 150 years. These include a range of expressions such as paintings, sculptures, films, designs, and media art. In this transformation era, ArtMovr aims to bridge the gap between appreciating art and exploring it through data-driven methods by utilizing linked data and knowledge graphs [6]. By developing a web application and a comprehensive ontology that seamlessly integrates MoMA's data ArtMovr provides users with a platform to navigate the museum's collection in a meaningful way that enhances their understanding and enjoyment. This project does not demonstrate the potential of linked data and knowledge networks. Also opens new avenues for exploring the world of art by making it more accessible and engaging for art enthusiasts, researchers and the wider public.

II PROBLEM DEFINITION

The challenge of navigating and comprehending the extensive Museum of Modern Art (MoMA) collection has spurred the creation of ArtMovr. With a collection boasting 200,000 artworks spanning 150 years, MoMA stands as a reservoir of cultural heritage [MoMA]. However, for users, exploring this vast collection efficiently presents a daunting task. ArtMovr addresses this challenge by aiming to bridge the gap between appreciating and exploring art through the strategic utilization of linked data and knowledge graphs.

Understanding the intricacies of art development, identifying distinct styles and movements, and unraveling the influences that have shaped artists over time is inherently complex [Art History]. Furthermore, conducting research on artist biographies, including their influences, and exploring works by the same artist within the MoMA collection can be a time-consuming and fragmented endeavor [Art Research]. ArtMovr seeks to streamline these processes by leveraging linked data and knowledge graphs to illuminate the interconnected relationships between artists, artworks, and associated concepts within the MoMA collection.

The primary objective is to enhance the accessibility, engagement, and informativeness of art exploration for diverse user groups, including art enthusiasts, researchers, and the wider public. By delving into the relationships embedded in the MoMA collection, ArtMovr aspires to offer a comprehensive understanding of the evolution and diversity of art [Linked Data]. The project strives to establish an informative art ecosystem that facilitates art discovery processes, encourages a deeper understanding of cultural heritage, and supports research initiatives in the realm of art [Cultural Heritage].

III LITERATURE REVIEW

III.1 Linked Open Data System of National Museums in Thailand

This paper discusses about a system called TOMS that is meant to help Thai national museums share information about objects that are important to their nation's history and heritage [1]. They start by discussing how hard it is to manage and share this data because it is stored in different systems and follows different rules. As an answer, they say to use Linked Open Data (LOD), a web based format that lets you share and connect data from different sources. The CIDOC Conceptual Reference Model (CIDOC CRM) [1] is a standard way to talk about museum data that TOMS uses. It has three layers: one for storing data, one for changing and processing data, and one for user interactions. The writers use a case study of three museums to show how well the system works and to show the benefits of using LOD for sharing cultural heritage data. To sum up, TOMS encourages museums to work together and share data. It can also help the creation of educational and cultural tools.

III.2 The Rijksmuseum Collection as Linked Data

This paper tells the museum's process of making its collections data available as LOD. LOD provides advantages like enhanced prominence, accessibility, and opportunities for novel research. The museum employs the Resource Description Framework (RDF) to disseminate data, establishing connections to other LOD datasets through the utilization of vocabularies such as SKOS and Dublin Core. The challenges encompass the need for specialized skills, absence of standardized practices, and the need to uphold data quality [2]. The museum is dedicated to the ongoing release of LOD and the exploration of novel approaches to enhance the museum's offerings and improve the visiting experience.

III.3 A machine learning framework for building linked open data

This publication introduces us to a Machine Learning (ML) [3] framework that converts museum collection data into a LOD, thereby enhancing its organization and accessibility. This framework comprises three components: firstly, it establishes connections between correlated records from diverse museum collections. Additionally, it retrieves data from the textual descriptions of museum artifacts. Finally, it constructs an organized depiction of the data. The researchers conducted a comprehensive evaluation of the framework using two museum collections and observed that it performed exceptionally well, exhibiting high accuracy. This approach aids museums in addressing the challenges they have while attempting to enhance the openness and connectivity of their data, mainly when the data is disorganized and lacking in completeness. The framework facilitates the comprehension of museum data by both machines and individuals, rendering it a valuable instrument for museums to disseminate their collections to researchers and the public.

III.4 Evaluation of Semantic Web Ontologies for Modelling Art Collections

This paper delves into the significance of ontologies in the Cultural Heritage (CH) domain, specifically for modeling art collections metadata [7]. The authors evaluate three prominent CH ontologies: CIDOC-CRM, European Data Model (EDM) [4][8], and VRA Core. Their assessment is based on criteria like cataloguing, hierarchy, computational efficiency, and user experience. The CIDOC-CRM stands out for its event-centric approach, allowing interrelation between entities like people, places, and time. EDM, developed for the European project, aligns with CIDOC-CRM, while VRA Core is tailored for visual culture works. The paper concludes by highlighting the strengths and limitations of each ontology in catering to art collections.

III.5 Museum Ontology-Based Metadata for Digital Collections

This paper dives into the intricacies of metadata creation and management for digital collections in museums [9]. The authors emphasize the importance of structured metadata for ensuring the accessibility, discoverability, and interoperability of digital assets. They introduce an ontology-based approach [5], leveraging the power of semantic web technologies to create context-aware metadata. The proposed ontology captures various facets of museum objects, including their provenance, historical significance, and relationships with other objects. By doing so, it offers a comprehensive view of the collection, allowing for more intricate searches and analyses. The paper also touches upon the challenges faced in ontology development, such as the need for domain expertise and the complexities of mapping legacy metadata to the new schema. However, the benefits, including enhanced user engagement and the potential for cross-museum collaborations, make the endeavour worthwhile. The authors conclude by highlighting some real-world implementations of their approach, showing its practicality and impact.

IV OVERALL APPROACH & HIGH-LEVEL DESIGN

Standard data models and vocabularies must be followed in the overall process of publishing data from modern and classical art collections as Linked Data in the “ArtMovr”. The accepted framework for these kinds of collections, which is extensively used by organisations throughout the world, is the CIDOC Conceptual Reference Model (CRM) [2]. A crucial first step in ensuring the ongoing relevance and accuracy of the Linked Data dataset is ensuring the conversion of collecting data into Linked Data at the source [1][3].

Regarding high-level design, the first major task is to recognize the collection’s entities, which include the items themselves, as well as their creators, history, and associated concepts. The relevant standard vocabularies, which frequently include the CRM, are then mapped to these entities and any additional vocabularies required for specific notions [6].

Since RDF is the industry standard format for publishing Linked Data, transforming what has been found into this format is necessary. Structured and connected data representation is made feasible by it.

In the end, open access should be utilized to make the data publicly available to every person in order to increase its value and accessibility. This implies that the Linked Data dataset ought to be publicly available to anyone, so promoting a more comprehensive and diverse application of the gathered data. For the dataset to be reliable and useful, high-quality data must also be incorporated, along with thorough cleaning, validation, and documentation. After completing these actions, the linked data can be made available by hosting it on a dedicated server or publishing it to a public repository.

V ONTOLOGY DESIGN AND VISUALIZATION

In the world of creating ontologies, we often work with three fundamental concepts: classes, object properties, and data properties. Classes represent entities that are commonly identified by their Internationalized Resource Identifiers (IRIs) and can be defined using RDF properties. The rdf:type attribute is used to indicate the association between a resource and a specific class. Object properties, on the other hand, establish connections between entities and represent relationships between them. Data properties, in contrast, relate to the connections between an instance of an entity and literal data values, such as numbers, text information, or dates. These properties help quantify and describe aspects of the entity.

A. Artist Ontology

The Artist class encompasses essential attributes that describe artists, including their gender, nationality, and role. The hasArtist object property links an Artwork instance to an Artist, indicating the creator of the artwork. The hasRole data property specifies the role of the artist, such as painter, sculptor, or musician. The Artist class also includes data properties to capture additional details:

- **hasAge**: Records the age of the artist
- **hasBornYear**: Indicates the year of the artist's birth
- **hasConstituentID**: Assigns a unique identifier to the artist
- **hasDeathYear**: Records the year of the artist's death

B. Artwork Ontology

The Artwork class serves as a central entity in the ontology, encapsulating the diverse characteristics of artworks. It is connected to other classes and incorporates various data properties to describe and categorize art pieces effectively. Subclasses of Artwork include ModernArt and ClassicalArt, which are also subclasses of the broader Classification class. This structure reflects the historical and stylistic categorization of the artworks.

Object properties for the Artwork class establish relationships with other entities within the ontology:

- **hasArtist**: Links an Artwork instance to an Artist, indicating the creator of the artwork.
- **hasClassificationType**: Connects Artwork to a Classification instance, defining the category or type of the artwork, such as a painting, sculpture, or installation.

Artwork is enriched with an array of data properties that provide descriptive details and measurable aspects of each artwork instance:

- **hasTitle**: Assigns a specific title to an Artwork, essential for identification and reference.
- **hasDate**: Records the creation date of the Artwork, which can be specific or span a range of years.
- **hasMedium**: Describes the materials and method used in the creation of the Artwork, from oil on canvas to mixed media installations.

- **hasCulture**: Captures the cultural context or origin of the Artwork, which can inform the viewer of its background and historical significance.
- **hasDimensions**: Comprises the measurable aspects of the Artwork, such as height, width, and depth, necessary for understanding the size and space it occupies.

C. Department Ontology

The Department class represents the departments within a museum or gallery that manage and maintain artwork collections. It serves as a connection point between artworks and their associated departments. The hasDepartment object property links an Artwork instance to a Department, indicating the department responsible for the artwork.

D. Connecting with Upper-Level Ontology

The integration of our ontology with the CIDOC CRM ontology serves as a crucial step in enhancing its scope and compatibility. By establishing connections with CIDOC CRM, specifically linking the 'Artist' class to 'Person' (E21 class) and 'Artifacts' to 'Human-Made Object' (E22 class), we ensure a logical framework for representation and knowledge sharing across diverse domains. This linkage not only provides a standardized structure but also facilitates interoperability, enabling seamless integration of our artwork ontology with broader knowledge bases. Such connections pave the way for a more comprehensive understanding and exchange of information while maintaining logical relationships within the semantic web framework.

VI VISUALIZATION OF ONTOLOGY

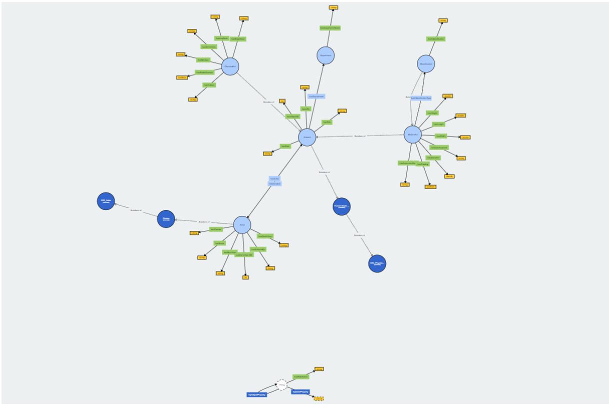

Figure 1: Final Ontology

VII DATA COLLECTION AND PREPROCESSING

Two main datasets for the "ArtMovr" web application: The Museum of Modern Art (MoMA)[1] collection and The Metropolitan Museum of Art (The Met)[2]. Over 470,000 artworks from throughout the world dating back 5,000 years are covered by the Met dataset, while the MoMA dataset has 140,848 records total that correspond to all the artworks that have been added to MoMA's collection and are listed in their database.

The data fetching process involves preprocessing and selecting specific columns from both datasets. From the MoMA dataset, the columns include Constituent ID, ArtistName, ArtistBio, Nationality, Gender, BornYear, DeathYear, Age, Title, Date, Accession, Classification, Department, DateAcquired, Cataloged, ObjectID, and URL. From The Met dataset, the columns include Is Highlight, Is Timeline Work, Is Public Domain, ObjectID, Department, Title, Culture, Period, Dynasty, Reign, Portfolio, Constituent ID, Artist Role, Artist Prefix, Display Name, ArtistBio, Artist Suffix, Artist Alpha Sort, Nationality, BornYear, DeathYear, Gender, URL, Artist Wiki-

[1] `https://github.com/MuseumofModernArt/collection/blob/master/README.md#the-museum-of-modern-art-moma-collection`

[2] `https://github.com/metmuseum/openaccess?search=1`

data URL, Object Date, Object Begin Date, Object End Date, Medium, Dimensions, Credit Line, and Geography Type.

An essential phase in the data collection process is data preparation. Data integration, cleaning and formatting the data, choosing appropriate columns for the ontology, and renaming columns to ensure consistency between datasets are all involved. This entails managing null values, eliminating superfluous columns, and transforming data formats to ensure cross-platform interoperability.

Constructing a successful ontology requires a thorough understanding of the dataset content. This entails determining the fields, structures, and kinds of data that are present in every dataset in addition to determining how to manage incomplete records from the MoMA dataset.

The preprocessed dataset forms the basis for developing an ontology that facilitates interaction, comprehension, and study of the extensive art collections from MoMA and The Met. Through the "ArtMovr" online application, this offers art fans, researchers, and the general public a seamless and relevant experience.

VIII IMPLEMENTATION PLAN

VIII.1 Ontology Development

Our ontology was diligently created using Protege, an industry-standard ontology creation tool. Protege provides a user-friendly interface for defining classes, properties, and connections within the ontology. It let us to build and structure the Artist and Artwork classes, as well as their associated characteristics such as 'hasAge,' 'hasBornYear,' 'hasDeathYear,' and 'hasTitle.' The links between entities and the attributes that define them were meticulously established to ensure a comprehensive depiction of the artistic domain.

VIII.2 RDF Triples and Knowledge Graph Development

After finishing the ontology in Protege, we used Stardog, a graph database technology, to generate RDF triples and build a knowledge graph. Stardog helped us convert our ontology into RDF triples, which allowed us to express entities and their relationships in a graph-based style. This allowed us to properly store and manage the interrelated data, laying the groundwork for semantic querying and inference.

VIII.3 Integration with Django Backend and SPARQL Queries

Our application is built on the Django backend, which is smoothly connected to run SPARQL queries using the SPARQLWrapper package. Because of this interface, our system was able to conduct SPARQL queries against the Stardog database, retrieving appropriate information based on user input. SPARQLWrapper improved communication between the frontend and the RDF knowledge graph by using Django's robust framework, ensuring efficient retrieval and processing of query data.

VIII.4 Frontend Implementation and User Interaction

ArtMovr's frontend architecture contains a user-centric design that offers a visually beautiful and intuitive user experience. The index.html page has a search box with radio buttons that allows users to search for either an artist's name or the name of an artwork. When a user searches for an artist, the artists.html page reveals detailed information on the artist, such as name, nationality, gender, constituent ID, birth year, death year, and a curated list of seven artworks by the artist. Searching for an artwork, on the other hand, gives its title, date, Object ID, URL, department, and the connected artist's name. Notably, the "View Artworks by Artist" button directs viewers to the artworks.html page, which displays a gallery of artworks by the selected artist, boosting the interactive discovery of artistic masterpieces.

Figure 2: Homepage

IX IMPLEMENTATION

We began with the development of an ontology in Protege, which defined classes, attributes, and relationships for artists and artworks. We then used Stardog to convert this into RDF triples, resulting in a well-connected knowledge graph. The Django-based backend used SPARQLWrapper to run SPARQL queries on the Stardog database, allowing for quick data retrieval.

The user interface has a search field for artists or artworks. Artist searches resulted in a detailed display that included information such as name, country, and artworks, whereas artwork searches revealed information such as title, date, and associated artist. Users might discover additional artworks by the same artist by navigating from artwork information.

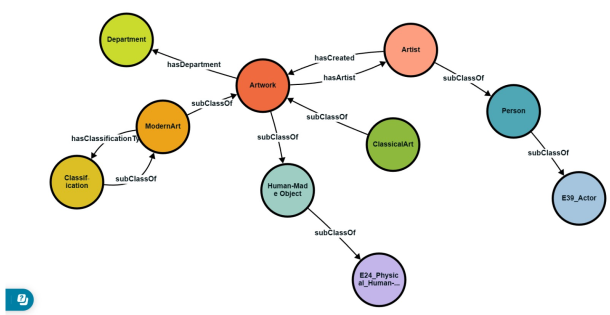

Figure 3: Knowledge Graph

This approach provided a smooth transition from Django user queries to SPARQL searches on Stardog, displaying information in an interesting, user-friendly manner for researching artists and artworks.

X SPARQL QUERYING

We used three queries:

1. The first query looks for information on artworks created by a specific artist.

2. The second query is similar to the first, except that it includes a placeholder for the artist's name.

3. The third question seeks information about a specific piece of art based on its title.

   SPARQL queries were used to obtain information and the outcomes were established.

XI CHALLENGES FACED

1. Linked Data Conversion: Ensuring the ongoing relevance and accuracy of Linked Data required converting collecting data into Linked Data at the source, posing a significant challenge.

2. Data Cleaning and Validation: Meticulous cleaning, validation, and documentation of high-quality data were essential to guarantee the reliability and usefulness of the linked data.

3. Designing and implementing the frontend for ensuring a visually appealing and intuitive user experience, especially in presenting detailed information about artists and artworks. Coordinating the frontend architecture with Django's backend and integrating SPARQL queries seamlessly for a responsive user interface posed challenges that were successfully addressed through collaborative efforts.

4. The integration of Stardog for RDF triples and knowledge graph development was a crucial step, but optimizing semantic queries for efficient data retrieval was a challenge.

XII ROLES AND RESPONSIBILITIES

1. Darshan Navadiya: In my role, I concentrated on the backend aspects of the project. I was

responsible for assisting and reviewing the updated OWL file. My primary focus was on backend development using Django to create an API connection with Stardog, facilitating the smooth integration of the knowledge graph with the frontend of the project.

2. Varun Menon: In my role, I was primarily dedicated to backend development, where I utilized Django to establish a seamless connection between our knowledge graph and the frontend. Also assisted in creating the knowledge graph, and its mapping. Additionally, I provided assistance in formulating and executing SPARQL queries.

3. Gokul Subramanian: In my role, I primarily focused on updating the ontology and establishing connections to an upper-level ontology. I played a key role in creating the knowledge graph and took responsibility for developing the frontend using HTML and CSS. Additionally, I contributed to the documentation by writing and updating sections of the project report.

4. Vidit Sanghvi: I was primary working on updating the OWL file in accordance with the professor's input, actively participating in the creation of a knowledge graph using Stardog, performing crucial data preprocessing operations and mapped data to the knowledge graph, and providing assistance in crafting SQL queries.

5. Vamsi Krishna Yadav Loya: In our project AtrMovr I played a major role in Data Preprocessing and Cleaning, updated the document in responce to previous feedback, and actively participating in the development of the front-end application. I also played a key role in articulating the project's limitations and future scope, providing valuable insights into potential challenges and opportunities for expansion and in addressing data quality issues, and contributing to the design and implementation of user interfaces.

XIII FUTURE SCOPE

1. Enhanced Interactivity: ArtMovr's future developments can prioritize enhancing user engagement. By introducing features like user-generated content, comments, and collaborative exploration, the platform aims to create a more dynamic and participatory experience. Users can contribute their insights, share perspectives, and engage in discussions, fostering a vibrant community of art enthusiasts.

2. Mobile Application: The development of a mobile application version of ArtMovr responds to the growing trend of mobile-centric interactions. This enhancement aims to make art accessible to users on the go, allowing them to explore and appreciate artworks anytime and anywhere. The mobile application would provide a seamless and convenient user experience.

3. Multilingual Support: To broaden its global reach, ArtMovr can introduce multilingual support, enabling users to access the platform in their preferred language. This inclusivity facilitates engagement from a diverse international audience, breaking language barriers and ensuring a more culturally diverse exploration of artworks.

4. Collaboration with Educational Institutions: By establishing collaborations with educational institutions, ArtMovr can become a valuable tool for academic research and art-related curricula. Such partnerships would elevate ArtMovr's role in supporting educational endeavors, providing students and researchers with a curated platform for in-depth art analysis and study.

XIV LIMITATIONS

1. Dynamic Collection Updates: The dynamic nature of museum collections, particularly at renowned institutions like the Museum of Modern Art (MoMA), presents a significant challenge for projects like ArtMovr. Ensuring that the platform stays current with the continuous changes to the MoMA collection involves several complexities. Museums often acquire new artworks, and existing ones may be removed or transferred for various reasons. The frequency of up-

dates to MoMA's collection can vary. Some museums update their collection databases regularly, while others may do so sporadically. MoMA's collection data may be in a specific format or adhere to particular standards. Changes in data formats or standards can complicate the extraction and integration process. Ensuring that ArtMovr's dataset reflects the current state of MoMA's collection requires an effective data synchronization process. Since, Users expect accurate and up-to-date information when exploring artworks on ArtMovr. Outdated or missing information can lead to a suboptimal user experience. Potential Considerations for the complexities mentioned above might include keeping ArtMovr updated requires a reliable mechanism for receiving timely information about these changes from MoMA. This might involve establishing a direct data feed or regular data synchronization processes. ArtMovr needs to establish a schedule for checking and updating its dataset. Frequent updates may strain resources, so finding a balance between real-time updates and a manageable update frequency is crucial. Implementing robust data synchronization protocols is essential. This might involve comparing the existing dataset with MoMA's database, identifying changes, and updating ArtMovr accordingly.

XV REFERENCES

[1] W. Chanhom and C. Anutariya, "Toms: A linked open data system for collaboration and distribution of Cultural Heritage Artifact Collections of National Museums in Thailand - new generation computing," SpringerLink, https://link.springer.com/article/10.1007/s00354-01900063-1 (accessed Oct. 22, 2023).

[2] C. Dijkshoorn et al., "The Rijksmuseum collection as Linked Data," Semantic Web, vol. 9, no. 2, pp. 221–230, Jan. 2018, doi: https://doi.org/10.3233/sw-170257.

[3] K. Dutia and J. Stack, "Heritage connector: A machine learning framework for building linked open data from museum collections," Applied AI Letters, vol. 2, no. 2, May 2021, doi: https://doi.org/10.1002/ail2.23.

[4] "Museum Ontology-Based Metadata," ieeexplore.ieee.org. https://ieeexplore.ieee.org/abstract/document/7439312 (accessed Oct. 23, 2023).

[5] Somesh Siddabasappa, Suyog Somesh Halikar and Ravikanth Dodda et al. Smart Museum Tour using Linked Open Data. 2022. DOI: 10.14236/ewic/ODAK22.6

[6] D. Liu, Antonis Bikakis, and A. Vlachidis, "Evaluation of Semantic Web Ontologies for Modelling Art Collections," Communications in computer and information science, pp. 343–352, Jan. 2017, doi: https://doi.org/10.1007/9783-319-67162-$8_3$4.

[7] V. de Boer et al., "Amsterdam Museum Linked Open Data," Semantic Web, vol. 4, no. 3, pp. 237–243, 2013, doi: https://doi.org/10.3233/sw-2012-0074 .

[8] Konstantinos Kotis et al. 'Children's art museum collections as Linked Open Data'. August 10, 2021, 60-70. https://doi.org/10.1504/IJMSO.2021.117107

[9] S. Angelis and K. Kotis, "Generating and Exploiting Semantically Enriched, Integrated, Linked and Open Museum Data," Communications in computer and information science, pp. 367–379, Jan. 2021, doi: https://doi.org/10.1007/978-3-030-71903-$6_3$4.

[10] A. Lara-Clares, Garcıa -SerranoA., and C. Rodrigo, "Enrichment of Accessible LD and Visualization for Humanities: MPOC Model and Prototype," Communications in computer and information science, Jan. 2017, doi: https://doi.org/10.1007/978-3-319-708638$_3$1.