# eXtreme Design for Ontological Engineering in the Digital Humanities with Viewsari, a Knowledge Graph of Giorgio Vasari's The Lives

Sarah Rebecca Ondraszek[1,*], Grischka Petri[1], Ulrike Blumenthal[3], Lisa Dieckmann[4], Etienne Posthumus[1] and Harald Sack[1,2]

[1]*FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany*

[2]*Institute of Applied Informatics and Formal Description Methods (AIFB) of KIT, Kaiserstraße 89, 76133 Karlsruhe, Germany*

[3]*Deutsches Historisches Institut Paris, 8 Rue du Parc Royal, 75003 Paris, France*

[4]*University of Cologne (prometheus - The Digital Image Archive), Albertus-Magnus-Platz, 50923 Cologne, Germany*

**Abstract**

Knowledge graphs (KGs) and ontologies have become valuable tools in the digital humanities (DH) for integrating interdisciplinary, diverse datasets and applying linked data and FAIR principles. In the form of a use case, this paper emphasizes the practical value of employing established best practices in ontological engineering, such as eXtreme Design (XD) and Ontology Design Patterns (ODPs), in the DH. As part of a broader project, it realizes the design and evaluation of a KG built upon Giorgio Vasari's seminal Renaissance text, *Lives of the Most Excellent Painters, Sculptors, and Architects*, a collection of artist biographies – the Viewsari KG. This way, applying an ontology engineering methodology in a real-world DH project functions as a metaphor for guiding future projects in the domain, emphasizing the importance of thoughtful ontology design for representing DH content.

**Keywords**

Ontology Design, eXtreme Design, Digital Humanities, Digital Art History, Knowledge Graphs

## 1. Introduction

The field of Digital Humanities (DH) is diverse, bringing together a wide range of methodologies and resources. This inherent heterogeneity necessitates effective management, retrieval, and storage of data. To address this growing need, many projects have turned to semantic technologies [1, 2]. However, this very heterogeneity creates a challenge. Multiple, often incompatible, standards dominate the field, hindering the efficient reuse of individual systems for new projects. This issue underscores the importance of open and FAIR (Findable, Accessible, Interoperable,

*SemDH 2024: First International Workshop of Semantic Digital Humanities, May 26–27, 2024, ESWC 2024, Hersonissos, Greece*

*Corresponding author.

sarah-rebecca.ondraszek@fiz-karlsruhe.de (S. R. Ondraszek); grischka.petri@fiz.karlsruhe.de (G. Petri); UBlumenthal@dhi-paris.fr (U. Blumenthal); lisa.dieckmann@uni-koeln.de (L. Dieckmann); etienne.posthumus@fiz-karlsruhe.de (E. Posthumus); harald.sack@fiz.karlsruhe.de (H. Sack)

0009-0003-7945-6704 (S. R. Ondraszek); 0000-0002-2548-449X (G. Petri); 0000-0002-9586-6039 (U. Blumenthal); 0000-0002-1708-7371 (L. Dieckmann); 0000-0002-0006-7542 (E. Posthumus); 0000-0001-7069-9804 (H. Sack)

and Reusable) research principles in DH. These principles, particularly reusability and interoperability, are equally crucial for the development of ontologies and knowledge graphs (KGs) within the field [3].

In a field of application as diverse as the DH, it is crucial to have flexible design methods, to provide reliable documentation of the development process, and to determine the effectiveness of design decisions. Methodologies like eXtreme Design (XD) adhere to standardized approaches from related domains like software engineering and propose a structured way to build an ontology and document the engineering process [4, 5].

Notably, while XD has demonstrably benefited related domains such as cultural heritage, it has yet to establish itself as a standard practice in the DH ontological engineering toolbox. This project aims to showcase its application to a cornerstone of art historical research, which demonstrates how KGs capture structured and rich, interdisciplinary databases effectively in DH contexts: Giorgio Vasari's *Lives of the Most Eminent Painters Sculptors and Architects* (*The Lives*) [6]. Originally written in Italian, it was published in 1550 and expanded in 1568. The text provides biographical descriptions of prominent Renaissance artists, including detailed descriptions of their lives, works, and artistic styles. Remaining a valuable resource for scholars and anyone interested in the period, Vasari's texts have significantly shaped the modern understanding of Renaissance art and artists [7]. The project's overarching goal is to develop a KG based on a historical social network automatically generated by Santini et al. from *The Lives* using named entity recognition, coreference resolution, and statistical association from the cooccurrence of names in paragraphs [8].

The Viewsari KG serves as a new gateway to Vasari's Renaissance, aiming to transform his chef-d'œuvre into a flexible and scalable graph-based knowledge representation of the text and its contents. Viewsari is intended as an integrated system of ontology, KG, and web application and, in the future, shall provide the possibility to explore and query information related to artists, their shared relationships, their artworks, and other related entities [9]. Viewsari makes use of the W3C's semantic web stack.[1] Additionally, connecting extracted data to provenance information (internal and external, e.g. secondary literature like the German Wagenbach edition[2]) should support digital hermeneutics and source criticism. Simultaneously, the project explores the integration of state-of-the-art ontological engineering methodologies, such as XD, into the DH domain. Recognizing the multidimensional nature of Vasari's work and the intricate interplay between extracted content and contextual information, it resembles the inherent nature of many DH projects. Thus, the process – involving collaboration with domain experts, who provide expertise that ensures the accuracy, relevance, and completeness of the ontological representations on which Viewsari is based – is a metaphor for how ontological engineering could work in the context of DH. This paper introduces only the initial ontology version, which describes foundational ontological concepts derived from the data foundation.

The two subsequent sections present the significance of Vasari's work for DH research and the data foundation for the Viewsari KG. Section 4 introduces related work on ontological engineering – XD in particular. Following, Section 5 presents a brief overview of DH-related projects that already apply best practices for ontology design. Section 6 focuses on the methodology

[1] https://www.w3.org/2004/Talks/1117-sb-gartnerWS/slide18-0.html
[2] https://www.wagenbach.de/buecher/edition-giorgio-vasari.html

related to the design process and project iterations. Finally, Section 7 discusses the usefulness of methods like XD in the DH, after which the conclusion summarizes this paper.

## 2. Relevance of Vasari's Œuvre for Knowledge Graphs and Digital Art Historical Research

It is hard to overestimate Vasari's influence on art history and the reputation that "his" artists still enjoy. It is a fact (in the following quote phrased by Lisa Pon) that since "its first publication in 1550, those seeking to learn about Italian Renaissance artists have read Vasari's Lives of the Artists" [10, p. 261]. Matteo Burioni has singled out three paradigmatic functions of Vasari's Vite for art theory and, in extension, the foundations of art history as a discipline. He mentions Vasari as a threefold model: of the artist biography, of a collection of artist portraits, and of the Renaissance as a historic concept [7, p. 278]. Pon also notes how copies of the book have left marks and annotations by several different readers: travelers who had visited paintings and places described by Vasari and writers commenting on (and questioning) Vasari's views [10, p. 275]. Despite the blatant anachronism it would constitute, it is tempting to regard these layers of comment, additions, and corrections as the inception of a demand for a knowledge graph. The sheer number of editions and translations of Vasari's text literally speaks volumes about this necessity. To address this task, it is first necessary to return to Vasari's text to analyze its structure and, in turn, the facts he describes. This concerns artists and works, patrons and events, the relationships between which are not always as clear-cut as Vasari presents. Art historical research has established the text in its multitudinous variants and commented on its content, confirming parts of the biographies and falsifying others. Despite all scholarly efforts spent on *The Lives*, art history still needs a structural presentation of its content, which systematically prepares the biographies for questions such as network research, the disclosure of dependencies, and organizational affiliations (courts, workshops, families, collections), but also seemingly simple analyses such as chronologies. A visual approach to Vasari's historiographical project, a kind of methodological x-ray, will unlock a critical potential of research that hitherto is burdened with either the weight of the tradition of the *Lives* or the hermeneutic thicket of Vasari's narrative interests. Connections between artists, artworks, patrons, and historical events that may not be immediately apparent from traditional textual analysis will come to light, and we can expect to uncover relationships between various entities in the art world, facilitating deeper insights into the construction of artistic styles and the evolution of social as well as technical practices. Stephen J. Campbell has formulated a research program to understand Vasari's text as an all-encompassing system of disparate and contending positions, asking for a sense of "historical implicatedness" [11, p. 50]. Indeed, Viewsari is conceived as a tool to bring us closer to that understanding, to enhance access and discovery of the many perspectives of the text, and to unlock new dimensions of understanding.

## 3. Data Foundation

In this section, the data foundation for Viewsari is described.

In parallel to [8], the Viewsari project uses the translated English edition of *The Lives* for all knowledge extraction and engineering steps. Gaston C. Du Vere wrote said translation in 1912 based on the 1568 edition. A digital edition of Du Vere's translation is available in ten volumes, each containing a subset of all biographies [6].

To facilitate entity extraction, experiments on both entity recognition and entity linking were conducted, and a set of annotation guidelines was additionally developed [12]. In a corresponding GitHub repository, a sample set of the following entity types is available: persons, organizations, places, and miscellaneous (all following the CoNLL 2003 dataset, partially linked to Wikidata[3]), as well as artwork references, motifs, terms, and dates. Motifs are "recurring subject[s], theme[s], or idea[s] in art. "[4] For persons, the Index of Names was used to capture all occurrences in the continuous text. The Index of Names is an appendix for *The Lives*. It lists all occurring persons in the work and is included in the original Italian and translated versions. In addition, relationships between artists were modeled through co-reference resolution and statistical association from the cooccurrence of names in paragraphs to create the social network.

The results are available in tabular form. For each volume of Du Vere's translation, a CSV file was created, containing entries for the cooccurrences. The data includes 673 cooccurrences and 1.073 persons, from which 312 appear in the relevant cooccurrences. Given the experimental approach of [8, 12], the quantity of other extracted entities is lower: for example, so far, the data basis covers only 133 recognized artworks and 311 motifs.[5][6][7]

## 4. Related Works on Ontological Engineering

In this section, the paper briefly outlines best practices in ontological engineering, focusing on eXtreme Design and Ontology Design Patterns.

Ontological engineering means formalizing and building knowledge representations for specified domains [13, p. 1–4]. Design strategies influence the development, implementation, and usability of systems like KGs. They also guide the creation of formal representations within ontologies, such as classes, properties, instances, and relationships [13, pp. 1-7], [14, pp. 13–16], [15, p. 5]. Over the years, various methodologies, techniques, and tools have emerged to facilitate this process (for a detailed recapitulation see works like [16], [17], or [15]).

### 4.1. eXtreme Design and Ontology Design Patterns

The XD methodology was proposed as a dynamic approach to ontological engineering, inspired by agile software engineering methodologies. It is a user-centered design method. XD incorporates incremental, iterative workflow phases like requirements analysis, development, testing, and release. The methodology emphasizes the high involvement of stakeholders who will later use the ontology directly or indirectly [16, pp. 88–93]. XD captures requirements in user stories, which portray real-world scenarios of potential ontology content. Personas can complement

[3] https://www.wikidata.org/wiki/Wikidata:Main_Page
[4] https://blog.stephens.edu/arh101glossary/?glossary=motif
[5] https://paperswithcode.com/dataset/conll-2003
[6] https://github.com/ISE-FIZKarlsruhe/vasari_nlp/blob/main/data/works.csv
[7] https://github.com/ISE-FIZKarlsruhe/vasari_nlp/blob/main/data/motifs.csv

user stories, portraying extreme stereotypes of potential users [18, pp. 278–279]. User stories are the basis for competency questions (CQs). CQs are questions that an ontology must be able to answer and a tool for evaluation during later engineering phases [19, p. 659]. In XD, they are also the baseline for finding and creating ODPs [16, p. 91].

ODPs – content patterns (CPs) in particular – help to keep solutions sectioned, providing small and modular frameworks for particular use cases [4, p. 5]. ODPs are modeling solutions for recurring design problems in ontologies. They function like building blocks, providing a common foundation for ontology modules [20, p. 5]. ODPs concern different abstraction levels and can be categorized into six groups. An extensive description is given in [20]. CPs encode conceptual design patterns specific to domain classes and properties within ontologies [20, pp. 4–8]. Selected ODPs function like templates and can be adapted and specialized for the domain of the projects [4, p. 15]. Browsing ODPs is possible in repositories like OntologyDesignPatterns.org.[8]

## 5. eXtreme Design and Digital Humanities

Ontological engineering methodologies like XD already found application in cultural heritage and a few DH-related projects. The ArCo KG includes interconnected resources from Italian cultural heritage. Alongside the release of the KG, the team also published documentation artifacts of the design process. To match ArCo's dynamic nature, they chose XD as methodology [5]. Another example that makes use of ODPs and agile design is the DigiKar project, which captures geographical representations and relations to space from various sources in an ontology.[9] A similar ontology to Viewsari is the MEETUPs ontology. In the project, the team utilizes a KG to describe a historical social network of musicians, embedded into contexts of events. Their ontology describes historical encounters between persons, also with provenance description. For engineering, they use XD and NeOn [21].[10]

Nevertheless, design is not limited to formal ontologies representing a domain. Likewise, higher-level systems like virtual endpoints or web applications shape the ontology's content and provide use cases that influence the design. Thus, since ontologies are the basis for KGs and are often used by human users or integrated into applications, further technical requirements arise: approaches from human-computer interaction (HCI), user interface design (UI), and user experience (UX) come into play [23]. In the DH domain, portals for consortia such as NFDI4Culture, which is targeted towards the implementation of a KG-based research data infrastructure for the cultural domain, provide their data as KGs, and users can search via SPARQL queries that are closely related to the underlying ontologies.[11]

Polifonia, a project for the interconnection of resources in the music heritage domain, uses a faceted way of requirements engineering: they combine methods from UI and UX design with the ontological engineering approach to create an endpoint "specialised enough to answer complex questions but [...] usable by stakeholders with different backgrounds" [23, p. 1].

[8] http://ontologydesignpatterns.org/wiki/Main_Page

[9] https://digikar.eu/datenmodellierung/

[10] NeOn is not relevant to this paper but proposes another strategy for agile ontological engineering. The methodology framework is also introduced in [22].

[11] https://nfdi4culture.de/de/resources/dataservices-datasets/datenportale.html

# 6. Methodology (Design Process and Iterations)

This section outlines the methodology for ontology design. For creating the ontology behind the Viewsari KG, XD is the central methodology, including the reuse and proposal of ODPs. Visualizations for the requirements engineering process and the ontology formalization can be found in the GitHub repository.[12]

## 6.1. Requirements Engineering

Following the XD methodology, the first step of our workflow is the domain analysis. In terms of concepts for the Viewsari ontology, the currently available data resources play a key role. Globally, *The Lives* represents the subdomain – super-domains are 'art history' and the Italian Renaissance. Since all information originates from Vasari's written work alone, the actual scope is more concrete: Vasari's Renaissance.

The text, consisting of biographies, produces information about artists and their lives, combined with related artworks or motifs, enriched with relational information as given by [8]. This existing pool of information influences ontological design decisions and thus represents a set of something we call data requirements (DRs). The DRs dictate the relevant entity types that the ontology should capture, and thus influence the design choices.

Our starting point is the social network from [8, 12]. Initially, the focus is solely on people and their relationships with each other (their cooccurrences). Such a simple ontology would include classes for persons and properties representing shared relationships.

Yet, all extracted resources and their relations stem from one or more paragraphs in *The Lives*. Source criticism and context sensitivity are critical for interpreting results from distant-reading methods like entity extraction. Thus, a graph should give the provenance of a piece of information [2, pp. 1, 8]. The statements about relationships need to be connected to the source paragraph, the so-called positional information. This paragraph naturally also needs to be connected to the source it appears in, that source needs metadata information, and so on. In an RDF-based model, edge annotations are possible via reification, named graphs or RDF-star, or the insertion of a new class for said property.[13][14]

Additionally, not only persons are relevant: the initial database includes some extracted artworks, locations, motifs, and dates – these can contribute as context to relationships between artists. The DRs are summarized in Table 1.

Based on the domain analysis and the DRs, the next step includes verifying and specifying the previously drafted requirements in user stories, scenarios, personas, and CQs.

Since XD is a flexible approach, teams of stakeholders can vary in size for each project. Therefore, there is no predefined number that would define a representative group [16, see p. 87]. A team of five stakeholders was invited to participate in this phase; three established art historians and two students in art history. Introducing them to the project, we gave a brief overview of the goals and challenges in Viewsari during a user interview. To verify our research so far and to elicitate further feedback, we asked questions regarding classes and properties

[12] https://github.com/ISE-FIZKarlsruhe/viewsari-kg/tree/main
[13] https://www.w3.org/wiki/RdfReification
[14] https://www.w3.org/2022/08/rdf-star-wg-charter/

| (Data) Requirements | Concepts for Ontology |
|---|---|
| Representation of extracted entities in the social network: persons (artists) | Class for :Person, subclass :Artist Persons can be connected |
| Provenance information for the relations between persons | Reification of inter-person connections |
| Representation of all extracted entities: persons, organizations, places, artworks, motifs, terms, and dates | Class :Resource Subclasses like:Artwork or :Location |
| Relations between all entities are possible | Properties OR reification that describe these relationships |
| Provenance information for paragraphs and more top-level information about the work, connected to the extracted entities | A layer for metadata about the work, representing: :Book, :Volume, :Translation, :Author, etc.; relationships with extracted classes (e.g. persons) |

**Table 1**
Data Requirements. Requirements captured from the data foundation.

within the ontology. Moreover, we asked about the relevance of source material, corresponding metadata, and the importance of direct access to textual evidence. This helps to capture the perspective of domain experts, both on the ontology and web application levels.

From the user interviews, we collected the requirements in the form of general research scenarios, user stories, and adjacent personas. We used the personas to group related stories in a non-structured text of a maximum of 250 characters on a Canva board.[15] To each persona, we assign one research scenario and multiple user stories. Moreover, we introduce four top-level categories that later help to categorize the CQs: ontology-content requirements (OCoRs), system requirements (SRs) from and functional requirements (FRs) for the web application, and those that potentially match both. Another group of requirements concerns characteristics of the ontology (OChars). These are, for example, interoperability and reusability.

In total, we constructed four personas representing a traditional researcher in art history, a student, a professor in digital art history, and a computer scientist. For the art historian persona, we collected 15 user stories, for the student we collected six, four for the digital art historian, and seven for the computer scientist.

An example of an art historian user story is the following:

> As a researcher, I want to explore cooccurrences of artists mentioned in The Lives so that I can identify potential collaborations.

This story summarizes different requirements for content and functionality: the end-user needs representations of cooccurrences related to *The Lives*. This points to concepts for cooccurrences that connect artists, annotated with positional arguments. However, during the user interviews, the stakeholders also vetoed a more general class to represent persons besides artists (commissioners, clergy, etc.). This partly verifies our first approach in Table 1. Additionally, the story captures the need for exploration, so a tool that would support not only straightforward searches but also ways to navigate the resources without a fully-fledged research question in

[15] https://www.canva.com/design/DAF6nI8fbmY/N2fAZuySC_EjF5Cw-mZsJQ/edit

mind. The term 'potential collaborations' also hints at the need to support researchers in interpreting represented data. This could happen through provided context, additional information regarding the connection strength, or implied (visualized) clusters. Table 2 represents the main requirements, as collected in the art historian persona.

| **Main Requirements** |
|---|
| Relevant concepts: cooccurrences, resources like persons, artists, artworks, locations, historical events (based on dates or mentions in the text), document and text-specific metadata, like books, their volumes, editions and translations, authorship information, inter-text structure, like biographies, pages, paragraphs |
| Relevant attributes of resources: birth and death dates for persons, creation dates for artworks, provenance information |
| Incorporation of external sources connected to the data: Wikidata or other authority data |
| Typed relationships between persons, persons and artworks, artworks and locations, and other resources (e.g. mentorship) – different types of cooccurrences |
| Representation of biographies, occurrences in biographies |
| Support in searches, interpretation, and discovery through a visualization of resources and their connectedness |
| Semantic and visual searches |
| (Index) searches for extracted resources: artists, artworks, etc. |
| Exploration of resources, cooccurrences |
| Visibility of contextual implications |
| Provenance information for paragraphs and more top-level information about the work, connected to the extracted entities - direct links to the original text passages from The Lives (also to the Italian source text) |

**Table 2**
Requirements for Viewsari. Results from the requirements elicitation.

Most requirements are interdependent. Naturally, the future web application can only represent knowledge defined by the ontology, so, e.g. an index depends on the available classes. Likewise, the necessity to visualize a network influences modeling choices in terms of connections between resources so that the computing time is not too high.

Requirements from the professor of digital art history or the computer scientist perspectives are more conceptual and mirror this dependency.

> As a professor in digital art history, I want to use interactive visualizations within the Viewsari platform to explore relationships between artists, artworks, historical events, or locations so that I can let my students identify patterns and trends, like an accumulation of collaborations in a certain city or place.

This example shows the focus on interactive visualizations and web application interfaces, as well as the possible knowledge discovery for laypersons (students). The student's perspective shares this trait. From the computer science perspective, there is a focus on ontology characteristics, like interoperability and reusability of the ontology.

From these stories, we formalized CQs for the ontology content. However, we also included CQs that concern the other levels, so the DRs, the SRs, FRs, and ontology characteristics.

To exemplify our procedure, we refer to the first user story introduced beforehand. This story directs to the following CQs: *Who cooccurred with whom? Did artist XY cooccur with artist YX? Where in the text did artists XY and artists YX cooccur?* From the CQs, we constructed a foundational ontology for Viewsari. In future steps that are out of the scope of this paper, the CQs are also relevant for the ontology evaluation of the Viewsari KG.

## 6.2. An Ontology for Viewsari – Formalization

This subsection describes the ontology construction from CQs. It includes building classes and properties and ODPs to represent generalizable structures from the domain. The ontology built here is the first version and is still under development. A detailed visualization of the classes and their shared properties can be found in the GitHub repository.[16] They were generated using the Graphical Framework for OWL Ontologies (Graffoo).[17]

### 6.2.1. Classes

From a first cluster of related CQs, we deduced the necessity to consider different descriptive layers in the ontology. Here, we differentiate three layers, the work level, the instantiation level, and the content level. CQs voice the need to include metadata for the work (*The Lives*), its different editions, and translations. This promotes work-level and instantiation-level modules for Viewsari. For the work level, we introduced the `:Book` class. This class summarizes basic metadata about *The Lives*, like the creation year or the creator (similar to Functional Requirements for Bibliographic Record (FRBR) works, expressions, and manifestations).[18] This is necessary to provide references to Vasari's original text and its different editions in the future.

To represent the instantiation level, we constructed the classes `:Edition`, `:Translation`, and `:Volume` to give clarity about the provenance of information. This way, Du Vere's translation can be included in the KG and complemented with references to the original Italian version or other related editions.

For the content level, we refer to the previously introduced CQs, from which we could derive classes directly (also see Table 1). Accordingly, we crafted a super-class `:Resource` to summarize these entities, and specific subclasses `:Person` and `:Artist`, `:Artwork`, `:Location`, and `:HistoricalEvent`.

For representing associations between resources, the so-called cooccurrences, we introduced a class `:Cooccurrence`, that can be further specified in later stages with, for example, relationship types.

Modeling the connections between resources as a class instead of a property makes it possible to annotate the cooccurrences with additional information. This reflects the need to contextualize statements and bind cooccurrences to a positional argument. This would also be possible with other reification techniques, such as RDF statements, or the use of RDF-star. However, we decided against this, since the cooccurrences are a central concept of the Viewsari KG. They are not only a connection of at least two entities but carry specific semantic meaning, such as

[16] https://github.com/ISE-FIZKarlsruhe/viewsari-kg/tree/main?tab=readme-ov-file#formalization
[17] https://essepuntato.it/graffoo/
[18] http://www.sparontologies.net/ontologies/frbr

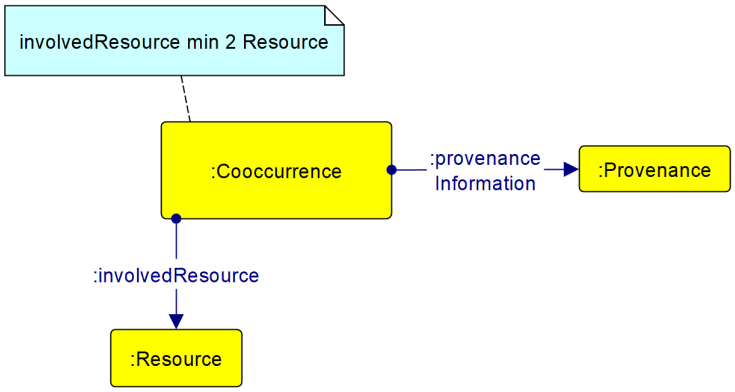


**Figure 1:** Coocurrence Pattern. A class that connects a minimum of two resources and is annotated with provenance information.

implicit indications of the role of this connection (art-artist, artist-artist), and can be used to construct a social network layer on top of the KG later on. A possible contextual argument is positional information, given here as `:Paragraph`, since the cooccurrences are based on statistical associations in paragraphs. From paragraphs, and upper structures, further document-related classes can be addressed, e.g. `:Page` and `:Biography`. The modeling of provenance information in Viewsari is inspired by existing ontologies and vocabularies, such as Historical Context Ontology [24], PROV(-DM) and the Web Annotation Vocabulary.[19][20] In the initial version presented in this paper, the mapping to said frameworks is still pending but planned for future steps. The goal is to map the properties present in Viewsari to equivalent ones in those.

### 6.2.2. Properties

For properties, a focus is on intra- and inter-level object properties. To describe the work level, persons can be assigned with the `:author` property to a `:Book` to describe authorship. Then, going from the work level to the instantiation level, additional properties are `:edition`, `:translation` and `:volume`. These can also have an `:author` or a `:translator`.

Content-wise, cooccurrences can be assigned to a minimum of two resources (e.g. artists) via `:involvedResource` and addressed to a paragraph with `:in`. Paragraphs are `:on` pages, those are again `:in` a volume. A biography is `:in` a volume and `:aboutArtist`, so related to the artist class. The properties create connections across and inside the modules. According to developments in the project, these can be adjusted, remodeled, or expanded. We also added inverse properties on the instantiation level, e.g. `:volumeOf` or `:translationOf`.

### 6.2.3. Existing ODPs and New Propositions

Mapping CQs to existing ODPs or proposing new patterns is crucial in XD [5]. These patterns can describe complex, recurring situations or single concepts.

[19] https://www.w3.org/TR/annotation-vocab/
[20] https://www.w3.org/TR/prov-dm/

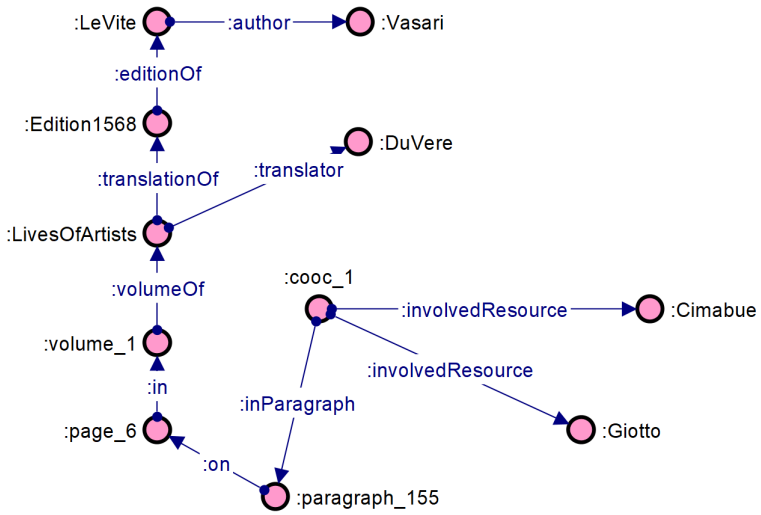


**Figure 2:** Example of a Coocurrence Pattern. An example of a cooccurrence between Giotto and Cimabue.

We reused the Persons and Location CP from submissions in the ODP repository. Persons describe natural or social persons, the element `:Person` is part of this CP and can be implemented as a class (like we did in Viewsari). The Place CP allows statements about places of things and includes properties like `:hasLocation` and `:isLocationOf`, as well as `:Place` as a specific class. We reused this pattern but renamed `:Place` to `:Location`. Moreover, we intend to reuse the Topic CP to represent topics and their relations to other resources. In our case, topics cover motifs or genres of artworks.[21][22][23]

Most importantly, we used the Provenance CP as inspiration for the representation of source evidence for extracted entities and cooccurrences. The solution in the repository does not fully match our use case, so we adapted it to the cooccurrence-in-paragraph situation.[24]

We propose the pattern shown in Figure 1 as a new CP to describe relations between resources in a given context. In particular, it allows for the traceability of automatically generated associations between entities in the CP. This is important for domains like network analysis.

In subsequent iterations and once the ontology has reached a more extensive level, we will add more ODPs and mappings from and to related existing ontologies like FRBR or CIDOC-CRM[25]. An example of the population of the model is given in Figure 2.

[21] http://ontologydesignpatterns.org/wiki/Submissions:Persons
[22] http://ontologydesignpatterns.org/wiki/Submissions:Place
[23] http://ontologydesignpatterns.org/wiki/Submissions:Topic
[24] http://ontologydesignpatterns.org/wiki/Submissions:Provenance
[25] https://www.cidoc-crm.org/

## 7. Challenges in the Digital Humanities: Best Practices as Solutions?

In the following discussion, the importance of adhering to best practices such as XD in DH is examined using the preceding ontological engineering process for the Viewsari KG as a case study.

Recognizing the unique hurdles inherent in DH projects hints at iterative, stakeholder-driven approaches to effectively address them. Vasari's *The Lives* helps to illustrate this process. Its rich, heterogeneous content mirrors the complexity and diversity of DH data. Modeling Vasari's narratives within Viewsari means addressing recurring challenges such as data extraction and integration, digital hermeneutics, and effective knowledge representation. Just like in Viewsari, DH projects often encompass a wide range of digital resources, including texts or images, archival collections, and objects sourced from libraries, museums, or digital repositories. When considered from a zoomed-in perspective, these datasets cover another layer of heterogeneous information about persons, locations, events, or the like. Managing, analyzing, and integrating these diverse resources necessitates specialized skills and methodologies tailored to the interdisciplinary nature of DH research. Standard practices and methods can help to ensure standardization, efficiency, and quality assurance in ontological engineering.
So, is XD a candidate for such a standard practice in DH ontological engineering?

This paper shows how this question could be answered with a "yes": Like most DH projects, the Viewsari KG brings together many different resources. The recognition of the unique hurdles inherent in the DH domain is a step that should be integrated into any project that aims to integrate semantic technologies. Specialized knowledge and needs must be validated via experts in the domain. These experts contribute requirements for concepts and relationships from said resources, also providing feedback on terminology and semantics specific to their domain. Designers and engineers complement this with expertise from the technological side. In addition, for applications transcending the formal ontology, a holistic view of different perspectives is particularly important: Which high-level systems will have access to the ontology? For DH ontologies and KGs, researchers should thus also consider UIs to facilitate user engagement with the ontology, which then in turn influences formal ontological design decisions [5].

Additionally, in the DH, knowledge about similar things can be distributed across different projects and infrastructures, or even come from data from different sources. Projects can work on similar topics and come up with completely different models of the same situation. However, understanding connections, causes, consequences, and correlations is crucial in humanities research. Comparative research – e.g., the depiction of the Italian Renaissance in Vasari's work compared to other available sources – requires flexibility in data models. To this end, ODPs are solutions that can be adapted to meet immediate needs and still be used as a common language to communicate similarities between different approaches [2], [19]. XD offers solutions to systematize and document the requirements analysis in an iterative way that harmonizes all perspectives and makes design decisions traceable [5]. Transparent communication of any decisions made while designing the system makes it more credible and could increase acceptance in the more conservative art historical community. While the art historical community is no more skeptical of digital approaches than other DH communities per se, Vasari's work is a

central pillar of the discipline. It is *the* source text of art historical research and has been involved in countless research endeavors. It is therefore important to engage with the community, to foster a constructive dialogue so that domain experts can add their expertise and contribute their nuanced interpretations, terminologies, and methodologies to the knowledge representations. The modular approach used in Viewsari allows for adaptability and scalability, addressing the heterogeneity of data in the DH and the dynamic nature of stakeholder requirements. This rhetoric could be generalized to a step valid to many DH projects: ontological engineering is a community effort and with principles like open science and FAIR as leitmotifs, it can foster interdisciplinary dialogues to enhance usefulness.

Beyond ODPs, using XD is also extremely flexible (agile, one might say) and allows for change at any stage. Modules and different layers in Viewsari represent various perspectives on the same data basis. They also capture domain-specific concepts and sub-concepts. Maintaining smaller modules over time is simpler than adjusting a large system. The modular design also reduces risks of redundancy and inconsistency, overcoming common challenges in KGs and ontologies within the DH domain: complex and diverse concepts can be modeled with existing ODPs or new patterns and shared with the community. Moreover, XD fosters continuous refinement and adaptation of ontological concepts, which adheres to the common problem of incompleteness and endlessness of research data ("Unabschließbarkeit") in the DH [5], [25], [26], [27].

Thus, iteration is key. For Viewsari, considering a broader audience as stakeholders and later integrating everything into a system with a web application that provides tools for visual exploration will change the relevance of existing elements and add requirements for new items. In a more rigid approach, this would mean reworking the entire ontology [5]. Since DH projects are subject to frequent changes due to the end of funding periods or fluctuating team composition, researchers should consider reliable structures that do not overturn all previous work when something needs to be adjusted. Taking Viewsari as an example, the user interviews made it clear that further data from *The Lives* is needed to complement the KG in a way that makes the work's complexity tangible. This includes relationship extraction, advanced named entity extraction for resources other than people and linking everything to external secondary resources for fact-checking.

Following XD – plus the dedicated documentation and stakeholder involvement – demonstrates a commitment to systematic engineering practices and quality assurance. In addition, best practices in ontological engineering advocate transparency, traceability, interoperability, and reusability at every step. Reusing existing ODPs and mappings to other ontologies promotes collaboration and thus reduces potential redundancies in similar domains.

## 8. Conclusion and Future Work

Integrating semantic technologies, in particular ontologies and KGs, into DH has manifested a possible solution to the challenges of effective data management, retrieval, and storage: structured data models for standardized knowledge representations or the possibility of implementing cross-disciplinary FAIR data are just two examples.

However, to go one step further and improve the communication of ontology design processes,

this project demonstrates how to bridge the gap between formalized approaches and the interdisciplinary nature of DH research by documenting each step of the engineering process behind the Viewsari KG. Viewsari, in the context of this project, functions as a case study for the utility of agile, use-case-driven engineering methodologies – in particular, XD – for the diverse, heterogeneous, and dynamic DH.

The project showcases how flexible and interactive fundamental semantic modeling can be when collaborating with stakeholders. This way, ontological design decisions can be made according to requirements from potential users, tailoring results for the community. The iterative design process and documentation of each development phase – from requirements engineering to the formalization of the first fundamental ontological aspects – shows how important transparency in ontological engineering is. This also concerns the traceability and interpretability of conceptualizations: ODPs and modularizations provide fast development solutions to problems in the DH domain, all the while enabling the proposal of new patterns with potentially interested other parties.

Viewsari utilizes best practices to promote standardized, high-quality ontological engineering in the DH. In future project steps, additional information extraction will expand the scope of Viewsari to include artworks, artist connections, motifs, and typed relationships. Accordingly, user and data requirements will evolve. In line with the methods used in this paper, the project will continue in iterative phases, soliciting stakeholder feedback and incorporating requirements from different levels (ontology content or web application interface).

Demonstrating the development process of a KG from Vasari's seminal work *The Lives* with XD offers a role model for future DH endeavors with similar goals.

## References


[1] S. Cristofaro, et al., Towards the representation of claims in ontologies for the digital humanities, in: Proceedings of the International Joint Workshop on Semantic Web and Ontology Design for Cultural Heritage co-located with the Bolzano Summer of Knowledge 2021, Vol. 2949 of CEUR Workshop Proceedings, CEUR-WS.org, 2021.
URL https://ceur-ws.org/Vol-2949/paper6.pdf

[2] A. Meroño-Peñuela, et al., Semantic technologies for historical research: A survey, Semantic Web 6 (6) (2014). doi:10.3233/SW-140158.

[3] M. D. Wilkinson, et al., The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data 3 (1) (2016). doi:10.1038/sdata.2016.18.

[4] E. Blomqvist, K. Hammar, V. Presutti, Engineering Ontologies with Patterns - The eXtreme Design Methodology., in: Ontology Engineering with Ontology Design Patterns - Foundations and Applications, 2016. doi:10.3233/978-1-61499-676-7-23.

[5] V. Carriero, et al., Pattern-based design applied to cultural heritage knowledge graphs, Semantic Web 12 (2) (2021). doi:10.3233/SW-200422.

[6] G. Vasari, Lives of the Most Eminent Painters, Sculptors and Architects, Project Gutenberg, 2008.
URL https://onlinebooks.library.upenn.edu/webbin/gutbook/lookup?num=25326

[7] M. Burioni, Giorgio Vasaris Vite, in: Handbuch Rhetorik der Bildenden Künste, De Gruyter, 2017. doi:10.1515/9783110331493-015.
[8] H. Sack, et al., Knowledge extraction for art history: the case of Vasari's The Lives of The Artists (1568), in: Conference on Digital Curation Technologies (Qurator), Conference on Digital Curation Technologies (Qurator), CEUR-WS.org, 2022. doi:10.34657/10668.
[9] J. Waitelonis, H. Sack, Towards exploratory video search using linked data, Multimedia Tools and Applications 59 (2) (2012). doi:10.1007/s11042-011-0733-1.
[10] L. Pon, Rewriting Vasari, in: The Ashgate Research Companion to Giorgio Vasari, Ashgate, 2014.
[11] S. J. Campbell, Vasari's Renaissance and Its Renaissance Alternatives, in: Renaissance Theory, Routledge, 2008.
[12] C. Santini, et al., Guidelines for the Annotation of ExtrART: Evaluation Dataset for Entity Extraction from The Lives Of The Artists (1550) (2023). doi:10.5281/zenodo.7991340.
[13] E. F. Kendall, D. L. McGuinness, Foundations, in: Ontology Engineering, Springer International Publishing, 2019. doi:10.1007/978-3-031-79486-5_1.
[14] E. F. Kendall, D. L. McGuinness, Before You Begin, in: Ontology Engineering, Springer International Publishing, 2019. doi:10.1007/978-3-031-79486-5_2.
[15] A. Gómez-Pérez, M. Fernández-López, O. Corcho, Methodologies and Methods for Building Ontologies, in: Ontological Engineering: With Examples from the Areas of Knowledge Management, e-Commerce and the Semantic Web, Springer London, 2004. doi:10.1007/1-85233-840-7_3.
[16] V. Presutti, et al., eXtreme design with content ontology design patterns, in: Proceedings of the 2009 International Conference on Ontology Patterns, Vol. 516 of WOP'09, CEUR-WS.org, 2009.
[17] C. Roussey, et al., An Introduction to Ontologies and Ontology Engineering, in: Ontologies in Urban Development Projects, Springer London, 2011. doi:10.1007/978-0-85729-724-2_2.
[18] P. T. A. Junior, L. V. L. Filgueiras, User modeling with personas, in: Proceedings of the 2005 Latin American Conference on Human-Computer Interaction, CLIHC '05, Association for Computing Machinery, 2005. doi:10.1145/1111360.1111388.
[19] A. De Nicola, M. Missikoff, R. Navigli, A Proposal for a Unified Process for Ontology Building: UPON, in: Database and Expert Systems Applications, Springer Berlin Heidelberg, 2005. doi:10.1007/11546924_64.
[20] A. Gangemi, V. Presutti, Ontology Design Patterns, in: Handbook on Ontologies, 2009. doi:10.1007/978-3-540-92673-3_10.
[21] A. M. Tirado, et al., Musical Meetups: a Knowledge Graph approach for Historical Social Network Analysis, in: ESWC 2023 Workshops and Tutorials. Semantic Methods for Events and Stories (SEMMES), Vol. 3443, CEUR-WS.org, 2023.
URL https://oro.open.ac.uk/88720/
[22] M. C. Suárez-Figueroa, A. Gómez-Pérez, M. Fernández-López, The NeOn Methodology for Ontology Engineering, in: Ontology Engineering in a Networked World, Springer Berlin Heidelberg, 2012. doi:10.1007/978-3-642-24794-1_2.
[23] G. Renda, M. Grasso, M. Daquino, From ontology design to user-centred interfaces for music heritage (2023). doi:10.48550/arXiv.2306.12973.

[24] M. Daquino, F. Tomasi, Historical Context Ontology (HiCO): A Conceptual Model for Describing Context Information of Cultural Heritage Objects, in: Metadata and Semantics Research, Springer, 2015. doi:https://doi.org/10.1007/978-3-319-24129-6_37.

[25] J. Wettlaufer, Der nächste Schritt? Semantic Web und digitale Editionen, Digitale Metamorphose: Digital Humanities und Editionswissenschaft Sonderband (2) (2018). doi:10.17175/SB002_007.

[26] V. Presutti, et al., Pattern-Based Ontology Design, in: Ontology Engineering in a Networked World, Springer Berlin Heidelberg, 2012. doi:10.1007/978-3-642-24794-1_3.

[27] E. Blomqvist, et al., Experimenting with eXtreme Design, in: Knowledge Engineering and Management by the Masses, Springer Berlin Heidelberg, 2010. doi:10.3233/978-1-61499-676-7-23.