Metainformation scenarios in Digital Humanities: Characterization and conceptual modelling strategies

Patricia Martin-Rodilla and Cesar Gonzalez-Perez

Version: accepted article

How to cite:

Copyright information:

Metainformation Scenarios in Digital Humanities: Characterisation and Conceptual Modelling Strategies

Patricia Martin-Rodilla[a,*], Cesar Gonzalez-Perez [b]

[a] CiTIUS center, Rúa de Jenaro de la Fuente Domínguez, 15782.
University of Santiago de Compostela, Spain

[b] Institute of Heritage Sciences (Incipit), Spanish National Research Council (CSIC)
Avda. Vigo s/n Campus Vida, 15705
Santiago de Compostela, Spain

patricia.martin.rodilla@usc.es, cesar.gonzalez-perez@incipit.csic.es

Abstract

Requirements for the analysis, interpretation and reuse of information are becoming more and more ambitious as we generate larger and more complex datasets. This is leading to the conceptual development and widespread use of information about information, often called metainformation (or metadata) in most disciplines. The Digital Humanities are not an exception. We often assume that metainformation helps us in documenting information for future reference by recording who has created it, when and how, among other aspects. We also assume that recording metainformation will facilitate the tasks of interpreting information at later stages. However, some works have identified some issues with existing metadata approaches, related to 1) the proliferation of too many "standards" and difficulties to choose between them; 2) the generalized assumption that metadata and data (or metainformation and information) are essentially different, and the subsequent development of separate sets of languages and tools for each (introducing redundant models); and 3) the combination of conceptual and implementation concerns within most approaches, violating basic engineering principles of modularity and separation of concerns. Some of these problems are especially relevant in Digital Humanities.

In addition, we argue here that the lack of characterization of the scenarios in which metainformation plays a relevant role in humanistic projects often results in metainformation being recorded and managed without a specific purpose in mind. In turn, this hinders the process of decision making on issues such as what metainformation must be recorded in a specific project, and how it must be conceptualized, stored and managed.

This paper presents a review of the most used metadata approaches in Digital Humanities and, taking a conceptual modelling perspective, analyses their major issues as outlined above. It also describes what the most

* Corresponding author. Tel.: +34 981 590962 Ext. 310; e-mail: patricia.martin.rodilla@usc.es

common scenarios for the use of metainformation in Digital Humanities are, presenting a characterization that can assist in the setting of goals for metainformation recording and management in each case. Based on these two aspects, a new approach is proposed for the recording and management of metainformation in the Digital Humanities, using the ConML conceptual modelling language, and adopting the overall view that metainformation is not essentially different to information. The proposal is validated in the Digital Humanities scenarios characterized with real external data.

1. Introduction

There is an increasing amount of humanistic data available everywhere (online repositories, research infrastructures, virtual museums and organizations, etc.) for analysis and interpretation. However, use (and reuse) of data in Digital Humanities is still very low [1-4]. This situation, among others, is usually explained by the need of rigorous approaches to express the various properties of the data that, hopefully, will bring new life to it. Some examples of this include recording who has created the data (i.e. its author), when (i.e. its time of creation), for what purpose (i.e. intent), with which quality criteria (i.e. reliability), etc. in order to interpret and re-interpret the data as needed. Data that describes properties of other data has been called "metadata", and has been vaguely defined as "data about data" or "data that describes other data" [5, 6]. In some cases, this "data" and "metadata" duality is replaced by the term "information" (referring to "data"), and "metainformation" (which is correspondingly defined as "information about information"). Although data and information usually correspond to different levels of complexity [7] in most information theories, we consider throughout this paper that the "meta-" relationship that connects metadata to data can also (and even better, as we will explain) be established between metainformation and information.

This need for a better representation of the properties of data has led us to use the term "metadata" (or "metainformation") to mean very different things. At least, two separate meanings of the term are clearly visible in the literature [8]:

- "Metadata" as synonym for the structure behind other data. For example, the specification of tables in a relational database (number of tables, what columns exist in each of them, what data type is used for each column, etc.) is often referred to as "metadata" of the data contained in the database.
- "Metadata" as a description that documents other data. For example, information or tags added to data records (such as rows in a database table) to document its author, creation time and reliability factor is also usually referred to as "metadata".

In the first case, metadata describes the structure or "shape" of the data, i.e. it determines the data scope and internal representation. In fact, by describing what columns a database table has, or what tables exist in the database, we are not only specifying how data is structured, but also what data we are considering to work with. In this regard, this first kind of "metadata" determines not only the data structure but also the domain of discourse that we are choosing to work with. Thus, metadata of the first kind is inextricably linked to the data it describes; we cannot strip a dataset of its metadata. A good example of this usage of the "metadata" term and conceptualization is the archaeological dataset metadata repository of ARIADNE [9] described in [10]

In the second case, contrarily, metadata constitutes an additional descriptive layer that can be added on top of existing data. By labelling a record with its author or its creation time, for example, we are enriching the dataset and informing future users about the recording conditions; but even if we do not annotate this information, the dataset and all of its data are still there. An example of this second case is the way in which the Dublin Core Metadata Element Set [11] works to describe the author or creation date of other data.

Given these major differences between the two senses of "metadata", we advise against using this term to refer to the first kind of situations as described above, and recommend to use it only for the second kind. To refer to the first sense, we can use terms such as "data schema", "information structure", "type model", or something similar.

Consequently, we can say that metadata refers to properties of the data that are documented to aid with its analysis and interpretation.

In addition to the terminological problem outlined above, three additional issues exist with current metadata approaches. Firstly, a very large number of them have appeared over the last few decades. Some target a very specific domain, whereas others are general in their scope. There are metadata approaches focusing on domains related to Digital Humanities, such as multimedia web elements [12]; online educational environments [13], geography [14] or cultural heritage [15], but we can also find approaches in more analytic domains, such as genetics, biology or medical information [16] or astronomy and meteorology [17]. All these approaches are extremely heterogeneous as to what they encompass, and make adoption decisions a very difficult task for tentative users. In this regard, we have reviewed in a previous work [8] 28 well-known metadata approaches for different domains (including multi-domain). Now, and due to our disciplinary background and domain interest, we re-analyzed them from the Digital Humanities perspective (yielding 23 that are directly applicable) and expanded the review to cover most of the Digital Humanities metadata ecosystem with a complete review of 30 Digital Humanities metadata approaches. We present here the details of the most common ones, including some approaches that have not been analyzed before. We also characterize the most common digital humanities metainformation scenarios of application.

As a second issue, most metadata approaches assume the previously indicated "data/metadata" duality, i.e. that metadata is an essentially different thing as compared to "plain" data. Therefore, an entire different modelling, encoding and management approach is often considered to be necessary for each metadata approach. As we will describe in detail in further sections, this assumption makes that many well-known metadata approaches contain are highly redundant in relation to already existing "non-meta" data modelling approaches. We call this aspect "model redundancy" in our analysis. We do not share this assumption, and instead work from a different premise: the "meta-" prefix in "metadata" is a qualifier of "data" and, therefore, metadata must be a particular kind of data. This constitutes a subsumption relation (in terms of Liskov's substitution principle [18]), so we must conclude that everything that we say about data, and every technology that we can employ on it, we also can say and employ on metadata. We explain in this paper how to put this principle into practice and what challenges it presents for expressing digital humanities metadata.

The third and final issue is that most metadata approaches contain in their specifications what we call a "conceptual/technical mixture". This means that they are mixing conceptual-level definitions with implementation-level concerns. As an example, some approaches include conceptual issues (such as whether to document the provenance of certain kinds of information) mixed with technical ones (such as whether to store it as Unicode or ASCII, for example). The conceptual/technical mixture of most metadata approaches represent a repeated violation of software engineering basic principles, such as modularity or separation of concerns, that serve to produce more robust and maintainable systems. We argue in this paper that a sound metadata approach should observe these engineering principles for better quality and reliability.

In summary, the proliferation of metadata approaches in Digital Humanities, together with the issues described above (data/metadata duality, model redundancy, and conceptual/technical mixture), is limiting the quality of the humanistic data and its potential for reuse, and neglecting well-known engineering good practices. We present here a complete survey that justifies this, as well as a proposal to improve this situation.

2. Metadata approaches in Digital Humanities

Many metadata approaches exist in Digital Humanities and related fields. In this section, we elaborate on our previous work [8] to provide a glimpse of the breadth and complexity of existing approaches through a literature and analytic review of metadata approaches as used in Digital Humanities. From the original 28 metadata approaches analyzed in our previous work, only 23 are directly applicable in Digital Humanities, either because they are specifically designed for it, or because they are multi-domain. For this paper, we have added 7 additional metadata approaches not previously revised.

As a starting point, we analyzed previously published aggregate studies on metadata approaches, and determined which approaches have been used or can be used in Digital Humanities fields. Four aggregate studies were considered:

- A complete survey is presented in [19] ("SICILIA" henceforth), including a compilation of 34 standards for semantic web including the MUSCLE [12] compilation for multimedia elements or *de facto* standards such as Schema.org [20].
- The most representative report from the point of view of standardization bodies is the NISO report [21] ("NISO" henceforth), which brings together 7 general standards for metadata representation, and which has worked as the seed for the ISO/IEC 11179 [22] international standard.
- Most Digital Humanities and information documentation (including archives and libraries, translation and language studies, communication areas and humanities and social sciences researchers) fields are using Riley's glossary as a metadata approaches list reference [23] ("RILEY" henceforth), which indexes, visualizes and briefly describes 104 metadata standards used by these professionals.
- Research Data Alliance (RDA) [24] is an international organization sponsored by the European Commission as well as some American and Australian government agencies pursuing research on data-driven initiatives, and currently having over 5100 members from 122 countries. RDA maintains an active community project [25] ("RDA" henceforth) for the indexation, implementation and maintenance of metadata approaches in multiple knowledge areas, with 39 registered metadata approaches in their catalogue [25].

In order to provide a better relevance, we have selected the metadata approaches that appear in at least two of the mentioned aggregate works, and which are specifically oriented to Digital Humanities fields or have a multi-domain purpose. This first filter yielded 23 metadata approaches. Although the four aggregate studies present some overlap regarding which approaches they describe, only one metadata approach appears in all four studies, and most pairs of studies have four to six approaches in common of a grand total of 156. Such a small overlap is also a reflection of the considerable fragmentation and proliferation of metadata approaches that we have mentioned before.

As a second step in our selection of metadata approaches, we have selected for analysis those metadata approaches that, without being present in at least two aggregate studies, are specifically mentioned in any of the aggregate studies as a metadata approach for humanities. This yielded an extra 7 approaches. Adding these guarantees that we are not leaving out approaches that are relevant to the Digital Humanities. In summary, we present here a review of 30 metadata approaches that are directly applicable to Digital Humanities fields.

The approaches were then classified regarding different criteria. Although there are many works that index or catalogue metadata approaches, there is not a consensus regarding classification principles or review guidelines. In addition, one of the most complete works in terms of metadata classification criteria is [26], which proposes a classification based on granularity and application domain. Using this study as a basis, we firstly organize approaches according to their application domain (e.g. semantic web, humanities, or even domain-independent), which we summarize as domain independent (I) or domain dependent (D). Then, we indicate the main field of application (e.g. cultural heritage, libraries, geospatial studies, etc.). In addition, granularity is quantified as the number of hierarchical levels in the metadata approach, and expressed as Global (G; general-purpose metadata approach), Local (L; domain-specific approach), Container (C; metadata aggregator) or Conceptual (CC; approach containing a rigorous schema definition).

Besides these criteria, we have also reviewed each metadata approach regarding the three issues that we identified in the Introduction: 1) data/metadata duality, 2) model redundancy, and 3) conceptual/technical mixture. These are expressed as H (high), M (medium) or L (low). In most cases we include an example or comment about the assigned value. Table 1 shows the results of the review of metadata approaches for Digital Humanities fields.

Table 1. The reviewed metadata approaches. Sources are indicated as R (RILEY), N (NISO), S (SICILIA) or D (RDA).

Name	Application Domain	Gty.	Specific Field	Source	Data/Metadata Duality	Model Redundancy	Conceptual/ Technical Mixture
BIBFRAME	D: Humanities	L	Cultural Heritage Libraries and Archives	N	H: Works vs. Other Aspects	H: RDF reuse	M: RDF conventions.
CDWA	D: Humanities	L	Cultural Heritage Museums	S	H: Data elements for display vs. for indexing	H: Overlapping with MARC, VRA Core, DC. Several Crosswalks.	L
CIDOC-CRM	D: Humanities	CC	Cultural Heritage Museums	R	L	M: Not Reuse Several Crosswalks.	L
DCAT	I: Semantic web	C	Catalogue interoperability	SD	L: (as RDF, SKOS)	H: RDF, DC, SKOS reuse.	H: (as RDF, DC, SKOS)
DDI	D: Humanities	L	Surveys	D	H: (as ISO/IEC 11179)	H: Based on ISO/IEC 11179. Several Crosswalks.	H: Documentation as Technical Specifications
DIG35	D: Images	L	Images	RS	H: Images vs. Metadata	L	M: XML-based implementation
EAD	D: Humanities	C	Record description	RS	M	M: Not Reuse. Several Crosswalks	M: XML-based implementation
FGDC/ CSDGM	D: Geography	L	Geospatial features	RD	H: Metadata Resource Type	H: ISO 19100 reuse. Some Crosswalks (specially with ISO 19115)	M: XML-based implementation
FOAF	I: Semantic web	G	People	RND	L: (as RDF, SKOS)	H: RDF, DC, SKOS reuse	H: e.g. URL-based specification elements
FRBR	I: Humanities	CC	Bibliography	RS	M	M: partial reference to MARC.	L
ISO 15836 (Dublin Core)	I	CC	Anything	RNSD	H: Resources vs. Descriptions	M: Not Reuse. Some overlapping with non-metadata languages. Several Crosswalks.	H: e.g. URL-based specification elements (URI…)
ISO 19115	D: Geography	L	Geospatial features	RSD	H: non-metadata packages vs. metadata information packages	M: CSDGM partial reuse.	H: e.g. URL-based specification elements (URI…)
ISO/IEC 11179	I	CC	Organizational information	NS	H: Data_Description package vs. Registration package	H: Not Reuse. High level of overlapping with non-metadata languages. Several Crosswalks.	M: e.g. references to storage formats.
ISO/IEC 13250 (Topic Maps)	I	C	Topical knowledge	RS	L	M: Not Reuse. Some overlapping with non-metadata languages.	L
Linked Data	I: Semantic web	G	Anything	RS	L	H: RDF Reuse	H: HTTP URI-based specification
LOM	D: Education	L	Educational resources	RS	H: Education Resource vs. Other Metadata Categories	M: based on DC.	L: (only XML-based technological recommendation)
MARC	D: Humanities	C	Bibliography	RS	H: MARC Content vs. MARC record	M	H: Defined MARC Formats (e.g. UNICODE codification)
METS	I: Humanities	C	Data collection	RS	M	H: Explicit redundancy identified e.g. with PREMIS	M: XML-based implementation, file types specification.
MIDAS	D: Humanities	L	Cultural Heritage Monuments, Sites, Events.	D	H: Themes vs. Information Groups	M: CIDOC-CRM partial reuse	M
MODS	D: Humanities	L	Cultural Heritage Libraries and Archives	S	M	H: a subset of MARC is included (reuse)	M: XML-based implementation

Name	Application Domain	Gty.	Specific Field	Source	Data/Metadata Duality	Model Redundancy	Conceptual/ Technical Mixture
OAI-ORE	I: Semantic web	G	Resource exchange formats	RSD	L: (as RDF)	H: RDF and Linked Data reuse	H: e.g. URL-based specification elements
OAIS	D: Humanities	C	Cultural Heritage Preservation protocols	RS	H: Content Information Type vs. Preservation Description Information	M: Not Reuse. Several crosswalks	L
ONIX	D: Book industry	C	Published material	RND	ONIX product vs. ONIX (product) record	M: Not Reuse. Some crosswalks (e.g. MARC, RDA).	M: XML-based implementation
RDF/OWL	I: Semantic web	C	Machine-readable semantics	NSD	L	H: OWL reuses RDF. RDF presents some overlapping with non-metadata languages. Several Crosswalks.	H: e.g. URI-based specification elements.
PMH	I: Semantic web	C	Digital Data collections	RS	H: Item vs. Metadata Record	H: Metadata descriptions is expressed in DC.	H: XML-based implementation, OAI-PMH requests are expressed as HTTP requests.
PREMIS	I: Humanities	G	Cultural Heritage Preservation of objects	RSD	M: Objects metadata dictionary.	H: For metadata descriptions is necessary other metadata standards (e.g. DC)	H: e.g. URL-based specification elements
RDA	D: Humanities	CC	Bibliography	RS	H: Elements vs. Recoding attributes	H: MARC, FRBR and AACR2 reuse. Several Crosswalks.	L
SKOS	I: Semantic Web	G	Taxonomies, Knowledge Organization Systems	RND	L	H: RDF and OWL Reuse. Several Crosswalks.	H: e.g. URI-based specification elements. Also reference to text encodings etc.
TEI	I: Humanities	L	Representation of Texts and marking-up	RS	M: Elements vs. Annotations	L	M: XML-based implementation
VRA Core	D: Images	C	Visual collections	RS	H: Elements vs. Attributes	H: Based on CDWA. Some crosswalks.	M: XML-based implementation

Two important aspects must be highlighted. First of all, if we focus on highly ad hoc metadata approaches for digital humanities, most of them are reusing domain-independent structures as a basis for metadata creation. For instance, RDA builds on CDWA, MARC, FRBR and AACR2 (older version of RDA metadata). This is also observed in important digital humanities projects, such as Europeana [27], whose metadata model (EDM) reuses up to 6 other metadata approaches referred to in Table 1 (RDF, OAI-ORE, SKOS, Dublin Core, CIDOC CRM and DCAT) [28, 29]. EDM is not included in any of the selected aggregated studies as it is considered an extra model layer overlapping these metadata approaches.

Secondly, if we focus on highly conceptual and more general-purpose metadata approaches, one approach (ISO 15836) is present in all the aggregate works considered, and only two of those classified as being of conceptual granularity (ISO 15836 and ISO/IEC 11179) are included in most of the aggregate works. These two constitute the approaches with the larger disciplinary scope and are amongst the most widely used (and replicated with model redundancy or reuse) [30], especially within the Digital Humanities [26].

Finally, we must also highlight that most of the metadata approaches reviewed provide from the Semantic Web application domain or reuse some approach in this domain.

For all these reasons, we have selected ISO 15836 and ISO/IEC 11179 for a deeper analysis, as well as the family of metadata approaches proposed by the Semantic Web domain. The following subsections present the results of this analysis.

2.1. ISO 15836 (Dublin Core)

ISO 15836, also known as Dublin Core, comprises a set of works carried out by the Dublin Core Metadata Initiative (DCMI), launched in 1995 to develop specialized metadata vocabularies that enable the construction of more advanced information search systems. Initially, Dublin Core specified 15 information descriptors that should be considered when documenting metadata, such as Creator, Date or Language.
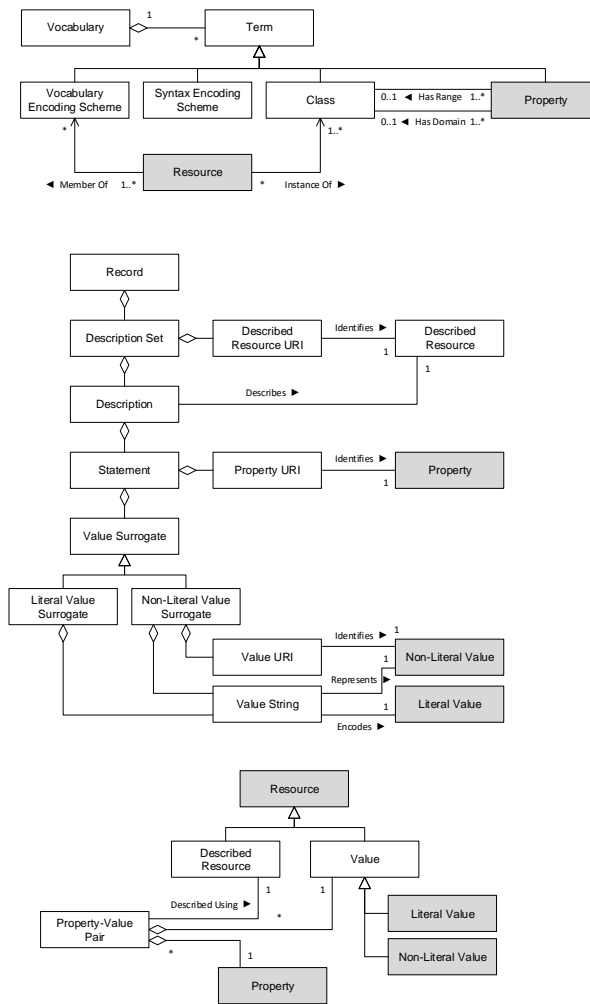
Fig. 1. Overall structure of the DCMI metamodel, including the three class families defined in the standard. Classes highlighted in grey work as connectors between the families. (After [10], with permission).

The initial success of Dublin Core led the DCMI to promote the creation of an abstract model as well as a language specification to express Dublin Core-compliant metadata. This language is specified as a metamodel expressed in UML [31], as the DCMI was targeting developers, designers and software scientists as relevant stakeholders. Since 2003, the term "Dublin Core" refers to the ISO 15836 international standard [32] for the description of information resources across domains; all types of resources are describable regardless of their format, area of expertise or cultural origin. Such a wide purpose and scope explains its presence in all the revised aggregate works and its wide adoption across disciplines [33]. Some domains where Dublin Core has been used include health systems [34], Earth sciences [35], social sciences [36] or humanities [30].

The DCMI metamodel is defined in three parts: the DCMI Resource Model, the DCMI Description Set Model, and the DCMI Vocabulary Model. Fig. 1 shows the major structure of the DCMI metamodel.

In the DCMI metamodel, Resources correspond to information entities about things in the world that are to be described. For example, a bibliographic catalogue entry in a library is a resource under DCMI. However, resources are defined in DCMI as "anything that might be identified, i.e. an image, a service, and a collection of other resources, human beings, corporations, etc.". This definition is alluding to the entities in the world, rather than to information records about them. However, metadata is supposed to be data about data, rather than data about entities in the world.

This is a significant contradiction and a potential source for confusion with the DCMI metamodel. Metadata itself corresponds to Descriptions, which are defined as "one or more statements about one, and only one, resource". This captures the essence of metadata satisfactorily, albeit being based on a poor definition of Resource. Finally, Vocabularies correspond to the words and terms that we use when constructing descriptions.

In addition to the confusing definition of *Resource*, there are some other conceptual weaknesses in the DCMI metamodel. Firstly, the organization of the metamodel in three parts indicates that information about entities in the world (i.e. data) is being treated separately from the information gathered about it (i.e. metadata). In other words, the *Resources* vs. *Descriptions* dichotomy shows the presence of the "data/metadata" duality issue previously explained. Consequently, under DCMI something is either data or metadata. This poses some serious problems that are discussed in the next section.

Secondly, the DCMI metamodel *Resources* model presents some overlapping with "non-meta" data modelling approaches, such as UML (e.g. Class).

Finally, the DCMI metamodel incorporates implementation-level constructs, such as URI-related classes, together with highly abstract ones, such as *Resource* or *Description*. This is exemplified by the definition of the *Literal Value* class: "A value which uses the Unicode string as a lexical form, together with an optional language tag or datatype, to denote a resource (i.e. "literal" as defined by RDF)" [27]. Making the definition of a conceptual-level class depend on specific implementation choices such as Unicode or RDF clearly contravenes the layering and separation of concerns principles in software engineering.

2.2. The ISO/IEC 11179 (Metadata Registry)

ISO/IEC 11179 Metadata Registry (MDR hereafter) is an international standard developed between 2005 and 2015 [22]. It aims at the representation of metadata structures in organizations as well as their exchange and integration in heterogeneous environments. As with Dublin Core, it is a general-purpose standard with applications in a range of disciplines. The metamodel of the standard is defined in UML as in the previous case. Good examples of ISO/IEC 11179 applications exist in e-government (including national health systems [34], legal or quality of service) or smart cities (including urban transport systems) [38].

The MDR metamodel is defined in six parts: ISO/IEC 11179-1:2015 Framework, ISO/IEC 11179-2:2005 Classification, ISO/IEC 11179-3:2013 Registry metamodel and basic attributes, ISO/IEC 11179-4:2004 Formulation of data definitions, ISO/IEC 11179-5:2015 Naming and identification principles and ISO/IEC 11179-6:2015 Registration. The first part maintains a registry of standard components and modifications. Use recommendations, naming and information management issues are explained in parts 2, 4 and 5, while part 3 corresponds to the formal specification of the standard metamodel. This constitutes the focus of our conceptual analysis here. Due to the large size of the metamodel, we focus only on the Data_Description package, which contains the classes that enable the description of specific metadata objects.

From a conceptual perspective, and as in the case of Dublin Core, some weaknesses appear for MDR. Firstly, metadata is conceived as a separate thing to the data that it describes. The "data/metadata" duality issue is evident by looking at the Data_Description package (Fig. 2) and some areas in the Registration package. To start with, the standard is providing a complex and large metamodel to represent information, despite the fact that a number of information modelling languages and approaches such as UML [31] or ConML [39] already exist. We argue that most of the constructs in Fig. 2 are redundant with UML, (illustrating the model redundancy issue mentioning a popular modelling language), and that UML will always make a better job at expressing information modelling concerns that MDR, which treats them as a small and non-central aspect. Furthermore, the fact that MDR provides its own mini-

metamodel to represent data on which metadata will be gathered indicates that data and metadata are treated, like in the case of Dublin Core, as separate and different things.
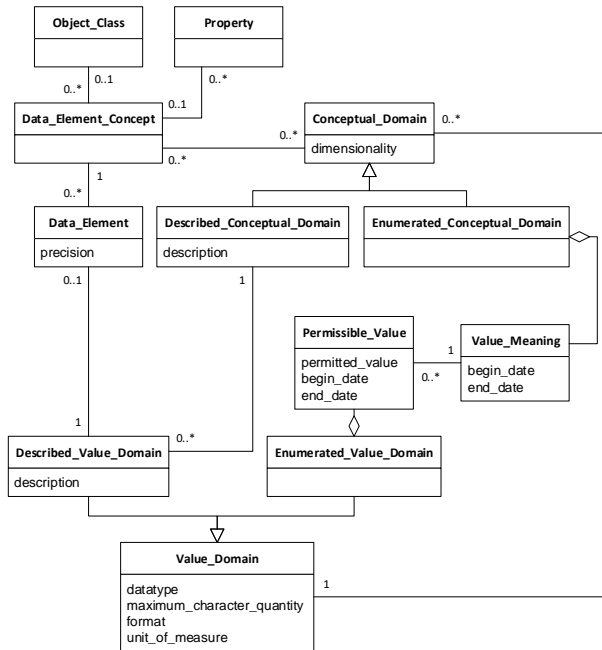

Fig. 2. Excerpt of the metamodel of the *Data_Description* package in MDR. (After [10], with permission).

2.3. Metadata Approaches Reviewed from Semantic Web

As we can see on Table 1, 7 metadata approaches were characterized in relation to Semantic Web technologies. In addition, most of the other approaches reuse some Semantic Web approaches in order to operate in the Semantic Web environment. Despite their diversity in scope, applications and goals, these 7 approaches (DCAT, FOAF, Linked Data, OAI-ORE, RDF/OWL, PMH and SKOS), as well as other related approaches (which are less commonly used or more specific, such as SHACL [28]), share similar behavior in terms of the issues being analyzed.

Firstly, and as expected, all of them are conceived form the point of view of the Semantic Web, which "provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries" [29]. The term "Semantic Web" refers to W3C's vision where "The ultimate goal of the Web of data is to enable computers to do more useful work and to develop systems that can support trusted interactions over the network" [29]. Thus, all these metadata approaches are more focused on data descriptions for machine processing and access than for human processing. This statement affects to the three issues analyzed in our review, as follows.

Firstly, the data/metadata duality does not constitute an important issue for these approaches, which present better behavior in this regard than previously described approaches. This particularity is explained by the absence of an explicit difference between data and metadata under the Semantic Web paradigm.

Secondly, all of them are particularly affected by the second issue of model redundancy. On the one hand, some of the approaches reuse others as they are built as a layered structure (such as OWL on top of RDF). In addition, there was (and still is) a huge emphasis by the Semantic Web community to achieve a high degree of interoperation between most of these approaches, and to build multiple crosswalks to other metadata approaches that would cover other areas. This is a necessary goal, but unfortunately it introduces multiple redundancies and dependencies that hinder the adoption and applicability of these approaches in general, and especially in the Digital Humanities metadata ecosystem.

Thirdly, Semantic Web metadata approaches present a high conceptual/technical mixture. This issue may not be seen as a weakness, but an intrinsic characteristic of the Semantic Web philosophy: if we need metadata approaches

highly readable by machines in order to transform the web into a huge and collective database, we need specifications that include technical and implementation-oriented aspects as well as conceptual ones. Thus, this third issue is only patent when humans try to construct, document or understand models (or ontologies) under a paradigm that is heavily machine-oriented. This fact comprises one of the strongest reasons for proposing a metainformation approach that is easier to understand by humans.

In summary, approaches within the Semantic Web ecosystem present some of the issues identified, especially due to the machine-oriented paradigm. Consequently, we argue that a human-oriented metainformation modelling approach would be useful as an initial step to improve the application of proven engineering principles (such as modularity, layering, or separation of concerns) to humanistic metainformation. Besides, non-Semantic Web-oriented approaches such as Dublin Core and MDR also suffer from these problems. In the following sections, we elaborate on these problems and describe the design principles for a better approach to metadata (or metainformation) management.

3. Problems with Digital Humanities metadata conceptualization

The two metadata standards described in depth, together with the metadata approaches from the Semantic Web ecosystem, exemplify the major issues presented in the Introduction, and which are shared by the vast majority of the 30 reviewed approaches. In this section we confront the evidence obtained from this analysis with the theoretical principles sketched in the Introduction.

First of all, there is a significant terminological and scope problem. For example, and as we discussed in the previous section, Dublin Core considers metadata to be the description of resources, where resources are physical entities in the world; but if this is metadata, what is plain data. In addition, the 30 reviewed approaches involve very different kinds of artefacts, ranging from genuine metadata schemata to data modelling languages to process recommendations. To clarify this confusion, we suggest the adoption of the following terminology and associated definitions (Fig. 3):

- A **data item** is anything stored in a computer system, and which describes some entity in the world.
- A **metadata item** is a data item that describes a data item.
- A **metadata schema** is a specification of the structure that metadata items about a particular data item must follow. In this manner, every metadata item conforms to a metadata schema.
- A **metadata language** is a formal language that allows us to construct metadata schemata. In this manner, every metadata schema is expressed in a particular metadata language.
- **Metadata management** is the process of using metadata languages to construct metadata schemata and/or to define metadata items conforming to the latter.
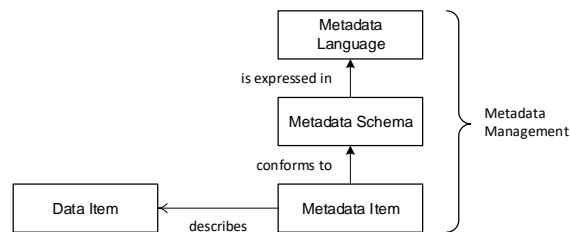

Fig. 3. Major concepts and relationships related to metadata.

Note that, in the definition of "metadata schema", we employ "schema" in the sense of the "shape" or structure of the information, which determines what scope is being considered and how it has been conceptualized. This often includes terminological (i.e. vocabulary) concerns as well as conceptual ones.

According to this conceptualization, the DCMI described in the previous section (Fig. 1) is a metadata language, although the original Dublin Core 15 information descriptors constitute a metadata schema. Similarly, the

Data_Description package in MDR (Fig. 2) constitutes a metadata language, whereas most of its *Registration* package constitutes a specification of a metadata management approach. Of the 30 reviewed approaches in the previous section, most of them constitute metadata schemata and metadata languages.

It is worth noting that the conceptualization proposed above is very permissible as to what may constitute a data item. In this regard, data items may correspond to M0 data (using OMG's parlance) such as objects or records in a database, but also to M1 data such as data structures, classes or other type-level specifications. Actually, data items may correspond to data at any M level. This makes sense as any kind of data maybe susceptible of being documented and described.

In addition, and as we hinted at in the Introduction, the "meta" relationship that exists between metadata and data should be better described in terms of information rather than data. This makes sense if we consider information to be "conveyed data", that is, data that constitutes part of a message sent from someone to someone else with the aim to transmit knowledge. When data is seen in this context (sender, receivers, intention), then the description inherent to the "meta" relationship can be better characterized and formalized, as "naked" data, without any context, is hardly describable. For this reason, we advise that "metainformation" should be a better term than "metadata" for our purposes.

Also, the relationship between the definitions of information (previously, data) item and metainformation (previously, metadata) item has the consequence that every metainformation item is also an information item. In other words, *MetaInformationItem* may be seen as a subtype of *InformationItem* (Fig. 4) as an immediate and first attempt to conceptualize a solution.
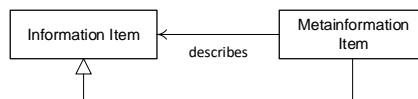

Fig. 4. Metainformation items as a particular kind of information items.

This fits the logic described in previous sections that metainformation is not essentially different to (plain) information, but a particular kind of it. What's more, we argue that metainformation is not even a subtype of information, but a role that information may take in certain situations (Fig. 5).
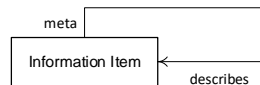

Fig. 5. Metainformation as a particular role that information may play in certain situations.

Consider the following scenario. A team of museum curators, art restorers and researchers catalogue a number of paintings using a computer system. The data that they enter constitutes plain information, since it describes entities in the physical world, the paintings. The professionals also record some metainformation, such as who entered each painting record, their role or working position in the team, when this work was done, and whether information about the paintings was obtained by looking at the physical paintings or from a pre-existing catalogue. All these information items describe the painting information, and hence we call them "metainformation" (Fig. 6 A). However, let's imagine that, at some point in the future, someone decides to carry out a research project on the cataloguing practices being employed, and takes the metainformation entered by the first team as his/her object of study. In other words, metainformation items about the paintings constitute the research project's primary data. The researcher may subsequently add their own metainformation about it; for example, they may record what quality or reliability each metainformation item has, what kind of process was followed by the person who entered it, whether the curator employed mechanisms to protect confidential information, etc. These are, in fact, meta- metainformation items in relation to the original painting data (Fig. 6 B), but only metainformation items in relation to the new study. In other words, the "meta-" prefix is relative, and we can easily conceive chains of meta- relationships between information items of arbitrary length. In Fig. 6 B, object *pmi* is both plain information and metainformation, depending on how we look at it. For the first team carrying out the cataloguing, it is metainformation about the painting information; for

the researcher carrying out the second study, it is information being described by further metainformation. Hence our argument that metainformation is not a base type (in terms of [30]), but a role. As a consequence, we argue that a metainformation (or metadata) schema is not essentially different from a plain information (or data) schema, but just one for which the conforming information (or data) plays the role of metainformation (or metadata) in some scenarios. Similarly, a metainformation (or metadata) language is just an information (or data) language for which the resulting schemata happen to be used for metainformation (or metadata). There are some works in the literature that support this relative character of the "meta-" prefix (specially in Digital Humanities, such as [31]), although they do not deal with metainformation conceptualization as we do here.
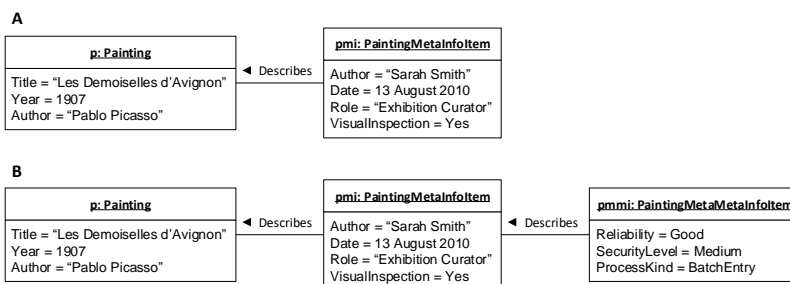

Fig. 6. Metainformation and meta-metainformation items. In A, *pmi* constitutes metainformation about *p1*. In B *pmmi* constitutes metainformation about *pmi*.

We must conclude, then, that having a "meta" stack like that in Fig. 3 does not make much sense and that, as hinted at from the beginning of this paper, *metainformation is just a particular role that information may play in specific scenarios*. According to this, we state the following design principles for a better metadata/metainformation management approach:

1. Since metainformation is not essentially different to plain information, specifically-designed complete metainformation languages are likely to be unnecessary. Regular information modelling languages should generally suffice, complemented as needed by some specific mechanisms to capture the "describes" relationship in Fig. 5.
2. Since information items in Fig. 5 may exist at any M level, metainformation should be applicable to any of them. In particular, metainformation should be applicable to any information regardless of whether it is a type (M1), an instance (M0), or anything else.
3. Since metainformation is a role that information plays in some scenarios (Fig. 5), the context of these situations must be taken into account. For this reason, we prefer to raise the level of abstraction and discuss the "describes" relationship in terms of information rather than data.
4. Since the "describes" relationship captures a conceptual link between a piece of information and another piece of information describing the latter, we claim that implementation issues on how to encode, store, transmit or process this information should be treated as a separate concern and excluded from a metainformation management approach. In addition, and since metainformation is information, the usual encoding, storage, transmission and processing mechanisms that are used for information should in general be applicable to metainformation as well.

4. Characterization of scenarios

The conceptual problems described in the previous section, and illustrated through our analysis in section 2 (DCMI, MDR and Semantic Web metadata approaches), are especially relevant when we work in the Digital Humanities. The main reasons include (but are not limited to) information domain aspects, methodological reasons, and disciplinary traditions [32]. We have identified the following as major (but not only) reasons:

- The Digital Humanities, as a relatively new and interdisciplinary field [33], presents significant institutional fragmentation, both in funding programs and in information custody and usage policies. This situation entails the

frequent need to integrate heterogeneous sources and metainformation schemata from different projects, initiatives and organizations.

- It is a common situation to work in the same informational contexts at several user levels: e.g. cultural heritage information can be retrieved and used by interested non-expert users, treated for pedagogical or informational purposes, used in managerial contexts for decision making, and/or studied for research purposes, among other possibilities. Each of these contexts can shift the focus of what constitutes metainformation as opposed to (plain) information.

- This variable focus, together with the high importance of subjectivity in most humanistic disciplines, are particularly relevant in the context of methodological and "best practices" work [34, 35]. Digital Humanities also involve a wide range of self-reflexive works on how the discipline is working to produce new knowledge from data. As in the example on painting cataloguing from the previous section, these works usually imply changes in what is considered metainformation in each case, not only in the curation subdomain, but also in any other disciplines such as archaeology, anthropology, history, or geography, where emphasis in methodology is often found.

- Current trends are promoting the reuse of information in Digital Humanities fields [33], which adds a temporal component to the previously identified aspects: what may have been considered information or metainformation in the past may change when reused in a different project or with a different purpose.

In this context, there are many proposals for the representation and effective use of most of the metadata reviewed approaches (Table 1), including those that have been created *ad hoc* for Digital Humanities [26, 36], or even more generalist ones that have been adopted in Digital Humanities [14, 27]. As a result of this review, we define in Table 1 four types of scenarios for the conceptualization, modelling and use of metainformation in Digital Humanities, according to the purpose of the metainformation being defined.

Table 1. Digital Humanities Metainformation Scenarios

Scenario	Main Purpose
1. Indexing and searching	Metainformation is mainly used as a reference to index and search the associated information. It works as a reference guide to implement information retrieval mechanisms in software systems.
2. Representation	Metainformation is mainly used to obtain an ontological or conceptual description of the associated information.
3. Integration	Metainformation is mainly used to connect or reconcile different information sources or structures.
4. Knowledge generation	Metainformation is mainly used as a basis for deep analysis and inference, and the creation of new knowledge.

This characterization is fully described and exemplified below, where scenario types are listed from lowest to highest requirements in terms of conceptualization and exploitation of the necessary metainformation.

4.1. Indexing and searching

Indexing and searching scenarios employ metainformation with the ultimate purpose to build information retrieval mechanisms in the form of software systems. Thus, these systems are exclusively based on the chosen metadata

schema, and the data items collected are expressed following this metadata schema. Good examples include databases, web services or web applications for searching and retrieving information about objects in heritage collections [49, 50] and online museums [50-52], learning objects in educational contexts [13], or multimedia objects in institutional repositories [53, 54]. We also include here all the scenarios related to metadata tagging and indexing defined for Semantic Web purposes [55-57], previously commented, such as automatically finding web resources. This type of scenarios concentrates most of the metadata purposes in the works revised, comprising an area so extensive and fragmented that a full account is impossible in this paper.

Indexing and searching scenarios generally focus on an extensive description of the information resources at an entity level (what we would call an object level in object-oriented terms), often involving small metainformation schemata and large volumes of stored data items. Despite the complexity and potential expressiveness, their final purpose only focuses on common searching and filtering tasks based on the chosen metadata schema. Subsequently, the software systems that are built usually implement retrieval tasks only (in contrast to more demanding scenarios that we describe below, with demands related to software exploitation or automatic or semiautomatic software reasoning from the metainformation contained and their underlying schema i.e. [58, 59]).

It is important to highlight that some of these works, such as [49], although referring to themselves as "metadata solutions", only work at the level of what we have previously called "data item" or "information item". Interesting examples can be found in [49], which allow us to continue with the example presented in Fig. 6. In [49], two complete searching scenarios are shown in relation to an art repository. The scenarios correspond to search operations by painting author (e.g. "Picasso" as in our previous example) and painting style. As explained above, the definition of any information item as either metainformation or plain information is relative, and depends on the particular usage situation. Accordingly, both the author and the style of a painting are susceptible of being considered information or metainformation items, depending on what the task at hand involves. However, in these examples (and similar ones reviewed), the information that describes the physical thing being documented (a painting) is called metainformation, giving raise to the paradox that we have metainformation without information. The systematic use of confusing terminology like this (as opposed to what we propose in Section 3, or something similar), together with the conceptual problems described for DCMI (the metadata schema chosen in this particular work), mean that many of the works in the literature present this confusing situation. We return below to these indexing and searching scenarios to describe how this confusion can be avoided by the application of the metainformation approach proposed here.

4.2. Representation

Secondly, there are scenarios whose main purpose does not focus on the search for information, but rather on obtaining a complete representation of the context of the information items being managed. We call these representation scenarios. Here, and in addition to search/retrieval tasks, comprehensive information about items must be obtained, recorded and managed as an ontological or conceptual image of the reality being dealt with. This may involve complex information structures at the metainformation level, including multiple interconnected items. For example, and following our previous example, not only we want to find "Picasso" as a textual term in the metainformation, but we also want the metainformation to resolve the involved ambiguities (as also described in [49]). For example, are we looking for paintings painted by Picasso? Or for paintings that were owned by Picasso? Or for paintings that are currently in the Picasso Museum? In the first case (paintings authored by Picasso), we may employ a simple search/retrieval approach as detailed earlier. However, in the second and third cases, we need to introduce conceptual distinctions between the author of a painting, its owner, and its location, as well as a temporal aspect so that questions can be posed on who owned what paintings when, or where is a painting now and where it was in the past.

In some representation scenarios, a structure of discrete metainformation items (entities and relationships) may be needed to express the necessary information, deviating from the simple usage of simple metainformation items as monolithic blobs to that store name-value pairs. Some examples of how these scenarios have been addressed in Digital Humanities can be found in relation to CIDOC CRM and its extensions [60-62]. However, some of the problems that we detected previously have not been solved in these scenarios, especially 1) the relative character of the "meta-" prefix depending on the situation, and 2) the changes in this conception throughout the life cycle of the information;

in fact, the literature only offers partial solutions for sub-domains or specific vocabularies, as is the case of extensions for the curation sub-domain [63].

4.3. Integration

Thirdly, a quite different type of scenario emerged in our revision, concentrating a wide range of current and significative works in Digital Humanities. We call them integration scenarios, and their main purpose is the conceptual reconciliation of different metainformation schemata (e.g. from different organizations or projects [64, 65]). Some well-known examples can be found in initiatives such as DARIAH [66], Europeana [27], CARARE [65], Bamboo [67] or ARIADNE [9]. In integration scenarios, the focus is the construction of metainformation schemata or languages, or sometimes the development and validation of metainformation management practices, always with a mapping or interoperability purpose. The improvement of the expressive capabilities of metainformation schematas is usually not a goal. Following our previous example, the objective would no longer be to find "Picasso" in the dataset, or to resolve ambiguities about its semantics; rather, the objective would be to map existing metainformation schemas in order to allow, for example, that concepts such as "Painter", "Author", "Owner", or "Museum"", or relationships such as "painting X was painted by author Y", can be understood and transferred across multiple metainformation schemata. These mappings can act as an infrastructure for the federation of previous types of scenarios (indexing/search and representation), so that users can tag, search, index, retrieve and resolve semantic ambiguities across heterogeneous datasets.

Although recent years have seen remarkable developments in relation to this type of scenario, many of the existing proposals constitute ad hoc solutions for the integration and interoperability of very specific schemata [33], and cannot be extrapolated to work as generally applicable solutions. The proposal that we advance in this paper, however, can, as we describe in further sections.

4.4. Knowledge generation

Finally, we find knowledge generation scenarios, having the fundamental purpose to generate new knowledge from the available metainformation and associated information [7, 58, 59]. In this context, "knowledge generation" refers to the process by which a human being, from the moment he/she comes in contact with metainformation, and by way of cognitive processes based on analytical reasoning and "upward" changes in the level of abstraction, obtains previously unknown results. This process has been described in literature [7] as a layered-based model, and can be found in any discipline and using different kinds of data and information as a basis. Note that, in this fourth scenario, metainformation usually plays a more central role than in previous scenarios. Following our "Picasso" example, we would not be interested in searching and retrieving information that we already have about Picasso's works, nor in solving ambiguities about whether a certain work is painted by Picasso or is located in the Picasso Museum, nor in integrating heterogeneous sources about Picasso. Rather, the main purpose would be to discover something new about Picasso by using the metainformation that we have about Picasso's paintings. This is usually accomplished by formulating an initial hypothesis and then composing a research question to that can be supported or falsified. For example, let's imagine that we have carefully described pictorial techniques and sub-techniques, textures, restoration materials used, dates of manufacturing, places where the paintings were painted and stored, their owners, etc. as metainformation in our Picasso repository. This would allow us to test hypotheses in relation to potential patterns in this information when analyzed as a whole. For example, it is documented in the literature that certain color choices during some periods of Picasso's work are possibly related to the painter's location, such as cities with more or fewer hours of sunlight. Thus, there seems to be a difference in color choices between works created at night mostly using artificial light, as opposed to works conceived during his Mediterranean stays. Other possible hypotheses could be related to the discovering of certain patterns to defend authorship of works: Does this specific painting respond to the characteristics of a work by Picasso, being able to affirm that it is painted by him or by a close disciple, for example.

This kind of hypothesis-based research requires that the metainformation schema being used, as well as the metainformation itself, are carefully conceptualized. Having a clear research goal in mind before metainformation is created and recorded can also help in being able to successfully address research questions like these. Also, situations like these often involve subjectivity and temporality aspects as intrinsic properties of the information being described.

That is why, without prejudice to the previous scenarios characterized, knowledge generation scenarios represent the ideal use case to apply the approach presented here, as we illustrate in Section 6.

This characterization of scenarios covers, as far as we know, the major kinds of purpose and usage in relation to conceptualization of metainformation in Digital Humanities, according to the bibliographical review as described in Section 2. It should be noted that, like in most characterization studies, our proposal here is driven by one specific criterion; in our case, the purpose of metainformation usage. As far as we are aware, there are no similar studies in the Digital Humanities literature (although similar criteria-based analyses exist in other fields, such as [26]). Likewise, the described scenarios should not be understood as isolated categories, but as four ideals between which real-world situations often occur. In fact, most projects in Digital Humanities, such as those referenced above [9, 27], address many, if not all, of the four scenario types described, although most place the emphasis in a specific one.

The main contribution of the characterization study presented here is the ease with which the purpose of creating and recording metainformation can be defined when starting a new project involving metainformation in Digital Humanities. This, in turn, helps clarify the requirements and decisions that must be made to answer crucial questions such as: 1) what things we are going to consider information, and what metainformation, 2) what metainformation schemata we are going to use (an existing one, an extension, our own), 3) if we are creating a metainformation schema, what metainformation language and what metainformation management techniques should we use to maintain and exploit the metainformation

The next section describes how we have applied the design principles outlined in Section 3 to construct a metainformation approach based on the ConML conceptual modelling language, and how this approach can help us answer these questions and be used to record metainformation in the different scenarios identified.

5. Proposed solution

ConML [37, 38] is a conceptual modelling language especially designed to be affordable to non-specialists in information technologies, free from implementation details, and supportive of "soft" aspects such as temporality, vagueness and subjectivity, which are very relevant to the humanities [59]. Superficially, ConML resembles other well-known object-oriented modelling languages such as UML [39], having as major modelling primitives those of *Class*, *Attribute*, *Association* and *Object* (Fig. 7). ConML is capable of capturing information in any domain at a conceptual level, very much as UML does for software and systems specifications.
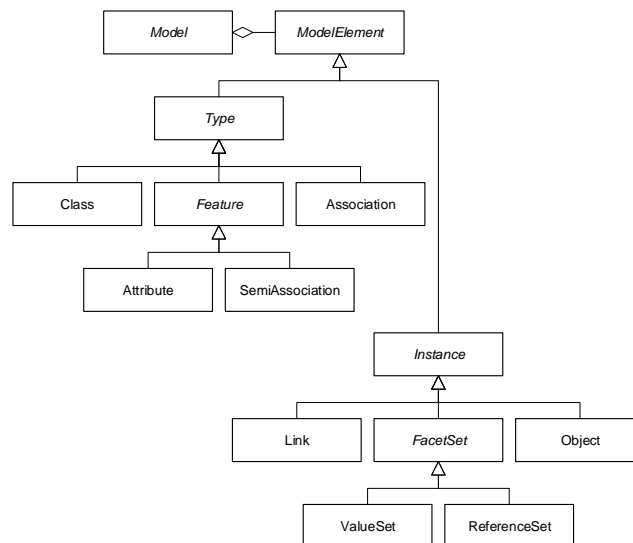

Fig. 7. Overview of the ConML metamodel.

A comprehensive description of ConML is out of the scope of this paper. However, its major benefits for conceptual modelling in digital humanities can be readily summarized:

- It is oriented towards conceptual modelling, rather than the specification of software systems as are other popular languages such as UML [39].
- It is simple enough to be quickly learnable by non-experts in software engineering, which compose the majority of practitioners and researchers in the digital humanities [40].
- It supports features that are absent from UML and other languages, but which are essential for the humanities, such as the modelling of temporality, subjectivity, multilingualism, and ontological as well as epistemic vagueness [40].
- It offers rich and formally-supported model extension features, so that models can be tweaked and adjusted to specific needs while guaranteeing their Liskov compatibility [18, 41]

The proposal described here builds on top of ConML by adding the minimum set of language features so that the modelling of metainformation is fully supported according to the design principles described in Section 3. Only one extra association is required in the metamodel to implement metainformation (Fig. 8).
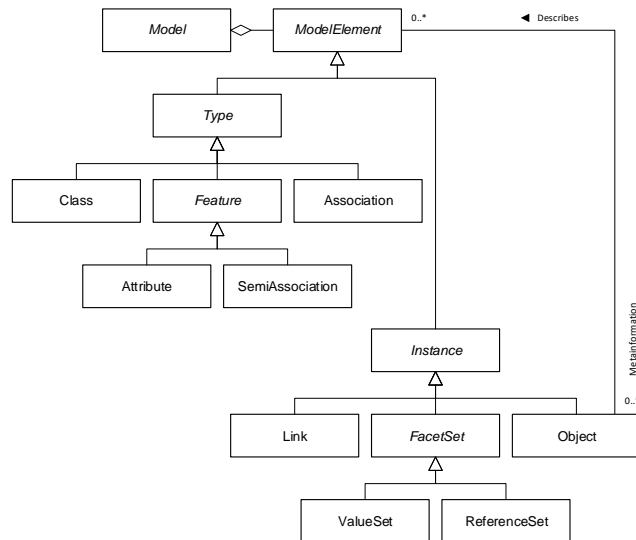

Fig. 8. Overview of the ConML metamodel, after having added the *Describes* association to implement metainformation.

As shown in Fig. 8, a *Describes* association has been added from *Object* to *ModelElement*. This association captures the fact that, in ConML, an object may describe a number of model elements, whether they are types (M1) or instances (M0). Adding this association has two interesting consequences. Firstly, it "reuses" a class in the metamodel (namely, *Object*) to implement metainformation items. In other words, metainformation items (as per Fig. 3) are objects in ConML, very much like (plain) information items. This satisfies design principle number 1 in Section 3, i.e. that metainformation is not essentially different from information. The second consequence of adding the association is that chained metainformation relationships are allowed (as shown in Fig. 6 B), because the model element described by an object can be another object which, in turn, describes another one, and so on and so forth.

Design principle 2 is also satisfied, since the target of the *Describes* association, *ModelElement*, is a very abstract class from which both types and instances specialize. This means that objects acting as metainformation may be applied to anything in a model regardless of its M level, including other objects, particular values, classes, attributes, associations, etc.

In addition, design principles 3 and 4 in Section 3 are also satisfied. Principle 3 is fulfilled since the *Describes* association happens between high-level, conceptual model elements (such as objects and other subtypes of *Model¬Element*) rather than individual information atoms. This allows for a better contextualization of the information being described as compared to other approaches that are less conceptual and more implementation-oriented. Also in this regard, principle 4 is satisfied since no extra constructs are introduced in the modelling language to describe how metainformation is encoded, stored, transmitted or processed. Since metainformation in our proposal

is composed of instances of *Object*, it is open to being treated as any other collection of objects, and stored in databases, encoded as XML, or processed in any meaningful manner. Our metainformation proposal does not pose any limitations on this.

It is also important to note that, in this proposal, metainformation items may be embedded in the same model where information items are being described, or exist as part of a different model. The extension mechanisms of ConML [38] (section 4.7) allow for a model to refer to model elements in another model. Although metainformation objects can exist in the same metamodel being described, using a separate model results in better modularity.

So far, we have been using regular link notation to depict instances of the *Describes* association (e.g. in Fig. 6). However, we must realize that *Describes* in the ConML metamodel (Fig. 8) is not the same as an association, for which there is a separate class, *Association*. In fact, a link between objects captures the fact that the entities represented by the connected objects are linked, whereas *Describes* captures the fact that the connected model elements themselves, rather than the represented entities, are connected.
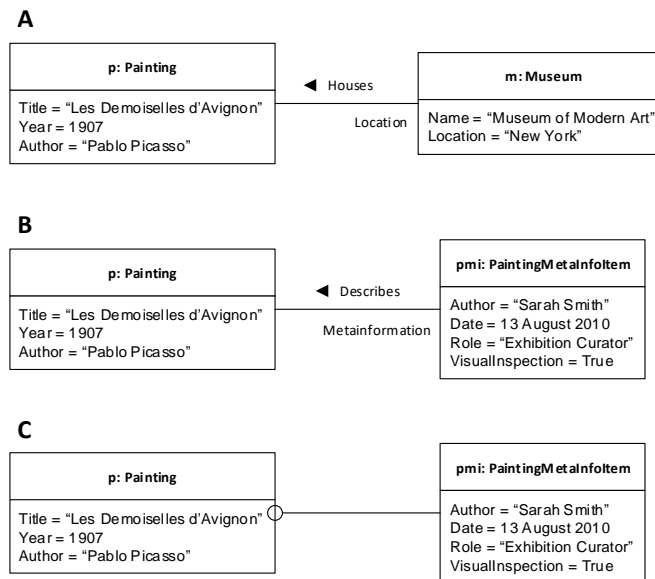

Fig. 9. Semantics of links and *Describes* relationships, expressed in ConML. In A, a link is shown. In B, a *Describes* relationship is shown using link notation (which is incorrect). In C, the same *Describes* relationship is shown, but using specific notation.

In Fig. 9 A, the *Houses* link means that the museum represented by object *m* houses the painting represented by object *p*. In Fig. 9 B, however, *Describes* means that the object *pmi* itself describes the object *p* itself; nothing is stated about the represented physical things. In other words, links happen between entities in the world and are represented in a model through the associated formal constructs, whereas metainformation *Describes* relationships happen model elements themselves, and have no counterpart that is external to the model. To highlight this difference, we propose that a special graphical device is used to depict *Describes* relationships. In Fig. 9 C, the same *Describes* relationship is shown, but using a line connecting the involved model elements, plus a small circle overlapping the target one. This circle can be read as "applies to". Hence, Fig. 9 C can be read as "the *pmi* metainformation item applies to object *p*". We said earlier in this section that *Describes* relationships connect metainformation objects to the model elements that are being described. These model elements may be objects, but also of many other types (see Fig. 8). Thus, the notation just described can be easily employed in a variety of situations involving not only objects as targets, but also classes, values or elements of other kinds. Fig. 10 shows some examples. Here, metainformation object *pmi1* applies to the *p* object, while metainformation object *pmi2* applies to the *Restored = Yes* value within the object. Note how the latter is displayed with the small circle next to the attribute value, rather than intersecting the object box border. Additionally, metainformation object *cmi* applies to the *Painting* class. Note that *cmi* is of a different type (*ClassMetaInfoItem*, rather than *PaintingMetaInfoItem*) as compared to the previous two metainformation

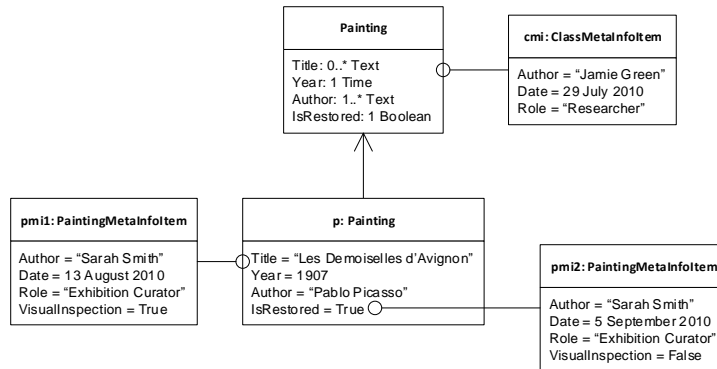objects, since different metainformation is required to describe a class and a painting object.

Fig. 10. Metainformation objects being applied to different model elements: a class (*cmi*), a whole object (*pmi1*), and a single value within an object (pmi2).

As a final note, we must say that the metamodel of ConML is much larger than what is depicted by Fig. 8. A most relevant aspect to this paper that is not shown in the figure is the fact that class features (attributes and semi-associations) can be marked as temporal and/or subjective, so that the corresponding object facets (values or references to other objects) can be tracked over time and across individual perspectives. A comprehensive description of ConML's capabilities for temporality and subjectivity modelling can be found in [59]. As far as the proposal presented in this paper is concerned, these features of ConML mean that metainformation objects are not "frozen" in time or across points of view but, on the contrary, they can be versioned, like any other objects, to describe model elements at specific moments or according to someone's specific views. This adds an extra layer of expressiveness to our metainformation proposal.

6. Applying the ConML metainformation approach to Digital Humanities scenarios

As described in the previous section, our proposal takes the relative character of the "meta-" prefix as a basis, and employs the subjectivity and temporality modelling capabilities of ConML to express these aspects as part of metainformation. This makes ConML [39, 68] an ideal language to support complex scenarios in Digital Humanities [69, 70, 71]. We must emphasize, however, that ConML is a general-purpose language and not specifically tied to the humanities, so the proposed metainformation approach can be used across disciplines. In this section we apply the proposal to the four characterized scenarios in Digital Humanities. In addition, we validate the proposal in the most complex scenarios (scenario 3 and 4) implementing the ConML class diagrams of the examples in a relational database on SQLite and SQLLite Studio [72, 73] with real data, and performing the subsequent queries for illustrating how metainformation works on each context. The data sources used are available external information about Picasso's artworks, for ensuring the independence between our proposal (and our implementation) and the data used in the validation.

It is important to highlight that the metainformation proposal explained here is independent of the implementation, due to is expressed in ConML at a conceptual level. Said that, ConML implementations have been tested in relational contexts [74], so we also have chosen a relational implementation for validating the ConML metainformation proposal.

6.1. Indexing and searching

As we explained in Section 4, it is common for this type of scenarios to present some confusion regarding the use and exploitation of information, as searching, indexing and retrieval tasks are often practically duplicated in many systems for information and metainformation items. We illustrate here how to avoid this confusion using the proposed

approach.

In indexing and searching scenarios, most of the necessary attributes are usually defined as part of the information rather than metainformation, as the main goal is to find information in a future software system. For example, a common search in a possible information software system about paintings could be "Find all the paintings having Pablo Picasso as an author". As we explained in Section 4, we decided to include the author as an information attribute, because it is a property of the actual painting rather than information about information (see Fig. 11). In this case, the software system is not using metainformation for searching, indexing and retrieving Picasso's paintings, but information describing the physical entities themselves.

However, let's imagine that, instead of all Picasso's paintings, we wanted to find what paintings have been catalogued by the exhibition curator Sarah Smith. In this case, we would have to work with metainformation, as the name of the person who created an information object has been modelled as Author in class Painting¬Meta¬Info¬Item. Interestingly, in both search cases we employ the same strategy: find objects having a specific content in their values. Since our approach unifies information and metainformation, no separate search mechanisms are necessary for each.
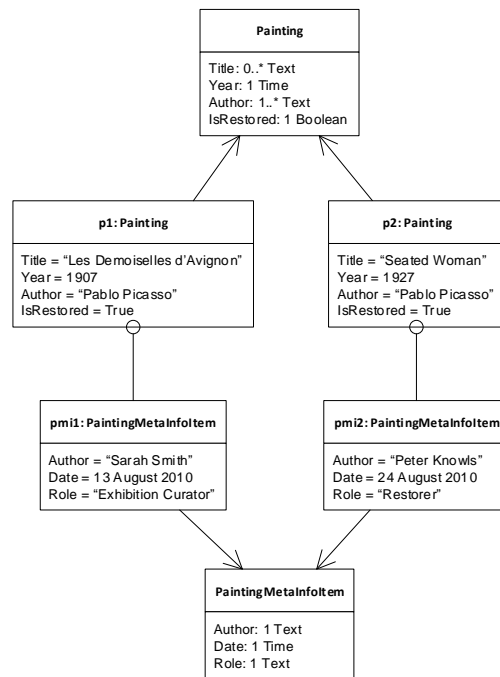

Fig. 11. Model for an indexing and searching scenario, using the proposed approach.

6.2. Representation

A good example of a representation scenario would be to "Find all the paintings catalogued by staff from the same organization as Sarah Smith". In this case, it is desirable that we have well organized metainformation describing the person who did the cataloguing of each painting as well as their organization. Fig. 12 shows a possible model for this scenario using ConML notation. Note that, by searching within the metainformation objects, we can first determine what is the organization of Sarah Smith, then find all the instances of Painting¬Meta-Info¬Item having the same value for Organization, and finally obtain the information items (i.e. painting objects) described by said metainformation objects.
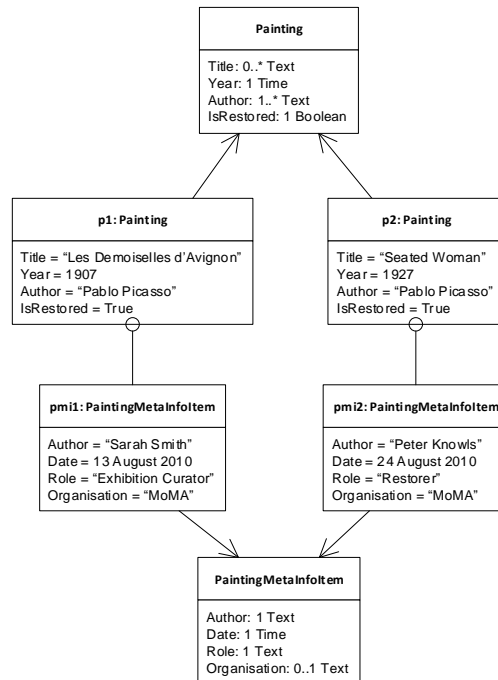
Fig. 12. Model for a representation scenario, using the proposed approach.

6.3. Integration

Thirdly, we address the integration scenarios, in which the main purpose is the federated work of several heterogeneous information sources regarding metainformation about entities that are common to all sources. Instead of illustrating this scenario with an example, we have performed a validation with two real data sources of information, each having a different conceptual schema: public catalogues of art works exposed by the Metropolitan Museum of Art (Met) [75] and the Museum of Modern Art (MoMA) [76], both situated in New York. In both catalogues, we can find information about paintings catalogued by a multidisciplinary team, but the structure of the information is expressed differently. We have imported both catalogues to a relational database following a different conceptual schema for each source, in order to illustrate the integration scenario from different conceptual schemas. For the Met catalogue, we found a lot of information about the artists (Bio, activity periods, etc.), so we have created a separate class for the *Artist* information, maintaining the artwork information in the *Met_ArtWork* class. In the case of the MoMA, the conceptual schema is imported in a unique class MoMA_*PaintingArtwork* (Fig. 13). We performed a study about the Picasso paintings registered in both catalogues (1309 registries for the MoMA, 176 registries for the Met), trying to study together some series of the painter that involves artworks of the two catalogues. Specifically, we decide to define a metainformation class *PaintingMetaInfoItem* and create the subsequent table in our database to group these series of artworks, collecting information about dates of begin and end of the series, the dimensions of the artworks involved following a standard ISO criteria (for instance with the purpose of printing the artworks in posters) and also cataloguing information about the date of our study and our role and name.

The final database (called *Scenario3Picasso*) presents a total of 5 tables: 4 tables correspond to the four classes involved (*Artist*, *Met_ArtWork*, *MoMA_PaintingArtWork* and *PaintingMetaInfoItem*), and an extra table to implement the relation *CreatedBy*, following previous specification of relational implementation of ConML class diagrams [74], and following the same criteria for information and metainformation involved on the scenario. Note that, due to any *PaintingMetaInfoItem* is describing metinformation related to at least one or various artworks objects, each metainformation object registries the identifier(s) of the referred objects for tracking.

Integration occurs by applying the same metainformation item to multiple information items, one for each source. This is depicted in Fig. 13 by a single *pmi* object connected to both *p1* and *p2* painting objects, belonging to the same

serie (*Toros en Vallauris*). The fact that both *p1* and *p2* share a common metainformation item signals that they both represent the same entity in the real world (the same series in our example), thus achieving integration.
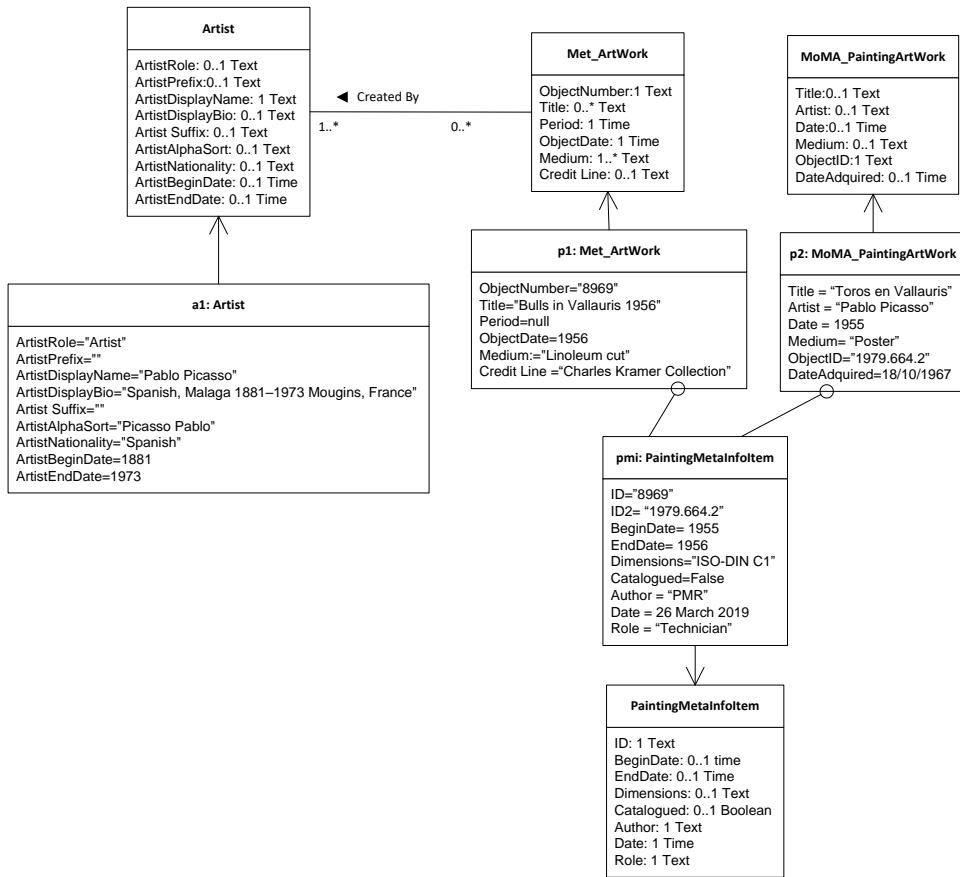

Fig. 13. Model for the integration scenario, using the proposed approach, and Met and MoMA real data sources.

This would allow us to carry out simple queries on the shared metainformation, as described above for indexing and searching scenarios. In this case, we have tested our relational database accessing by SQL queries to the shared metainformation, by searching all artworks involved in the "*Toros en Vallauris"* series. Specifically, the query searches for information dependent of the museum source (the credit line in the case of Met and the Date of acquisition in the case of MoMA) plus the recent dimensions value given in our study (as a part of the common metadata information of both artworks objects). Thus, a simple SQL query with a join structure between the relational tables is enough to consult information and metainformation at the same time from both catalogues:

SELECT Met_ArtObject.CreditLine, MoMA_PaintingArtWork.DateAdquired, PaintingMetaInfoItem.Dimensions FROM Met_ArtObject, PaintingMetaInfoItem, MoMA_PaintingArtWork WHERE Met_ArtObject.Title='Bulls in Vallauris 1956' and Met_ArtObject.ObjectNumber=PaintingMetaInfoItem.ID and MoMA_PaintingArtWork.ObjectID=PaintingMetaInfoItem.ID2

Fig. 14 shows the query executed in the SQLite Studio console and how the system recovers the shared metainformation from both catalogues (matching with the case exemplified in Fig. 13). In this figure, the system recovers information from both catalogues and metainformation associated to the series "*Toros en Vallauris*", all at once. This approach can be elaborated by adding extra properties to the metainformation in order to support, for example, representation scenarios as well or to expand the number of Picasso's series studied.

It is important to note that the proposed scenario shown in Fig. 13 involves significant simplifications as compared to the real-world situation. Firstly, we have simplified in the diagram the number of attributes of both catalogues for "Toros en Vallauris" for space reasons. It is possible to see the complete implementation in [77]. Secondly, a practical application of this model would likely involve the use of ConML's temporality modelling mechanisms [59] in order to keep track of what museum each painting is in over time. Said that, this validation shows how the ConML metainformation proposal integrates both external Digital Humanities data sources treating metainformation as "regular" information included in the querying logic of the system, avoiding the needs of specific querying infrastructures for accessing to metainformation and performing hybrid queries that involve information and metainformation at the same time.
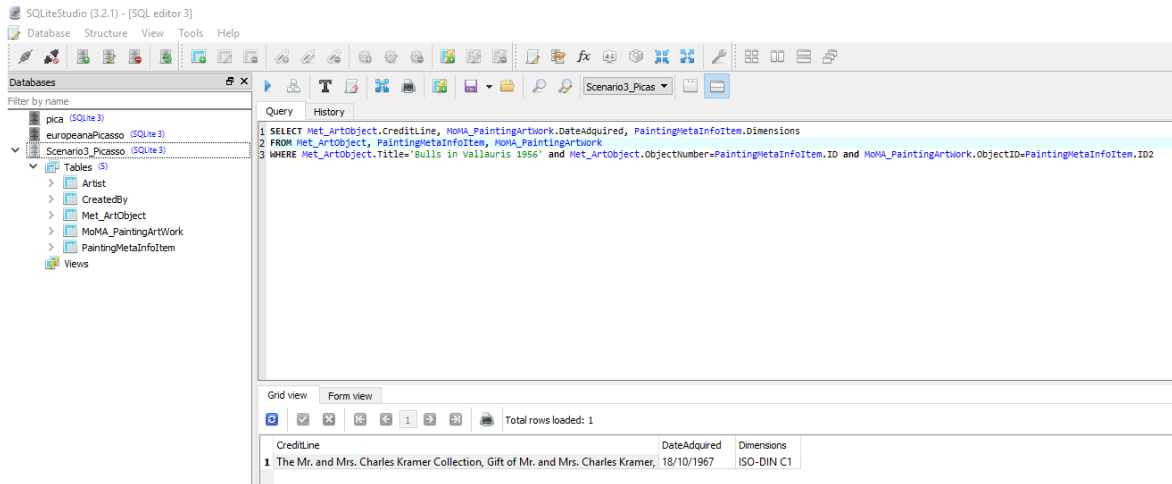

Fig. 14. Execution of integration SQL query, recovering information and metainformation from both Met and MoMA data sources for the Picasso's series "*Toros en Vallauris*".

6.4. Knowledge generation

Finally, the main purpose of knowledge-generation scenarios is to generate new knowledge from the available metainformation and associated information. Let's imagine that we want to test the following research hypothesis: "Certain Picasso's color choices are related with sunlight hours of his place of residence, and there is a correspondence between paintings that were made in darker places or painted at night with artificial light, as compared to works painted during his Mediterranean stays". Thus, we need information about Picasso's artworks, the locations involved, and the colors assigned to the different artworks. In order to validate this scenario with real external data, we have imported in a relational database 50 artworks assigned to Picasso in the Europeana repository [46], originally obtained by Europeana's API in JSON format. Contrary to other sources used in the scenario 3, Europeana provides information about the colors assigned to each artwork in their repository system.

Fig. 14 shows that the place of work has been captured for each painting, together with an interpretation of what the dominant colors are. Since this interpretation may have a subjective bias and is likely to be dependent on the technician's background and expertise, metainformation items are used to describe who determined the dominant colors, what role they play in the organization, and what degree of certainty (between 0 and 10) they assign to their analysis. In this case, we have found information about colors from 34 Europeana artworks to the 50 artworks recovered. We have completed the rest of artwork's information about colors assigning by ourselves, but indicating in different metainformation objects our authority and a lower level a certainty. Thus, the final database (called *EuropeanaPicasso_Scenario4*) presents 2 tables: *Painting* and *PaintingMetaInfoItem*.

Note that, as opposed to the examples for previous scenarios, the metainformation items in Fig. 14 apply to the *Colours* values rather than the complete *Painting* objects. In order to implement it, it is necessary to assign an identifier

for each value of the *Colours* attribute. This identifier is referred on each metainformation object described to maintain tracking.
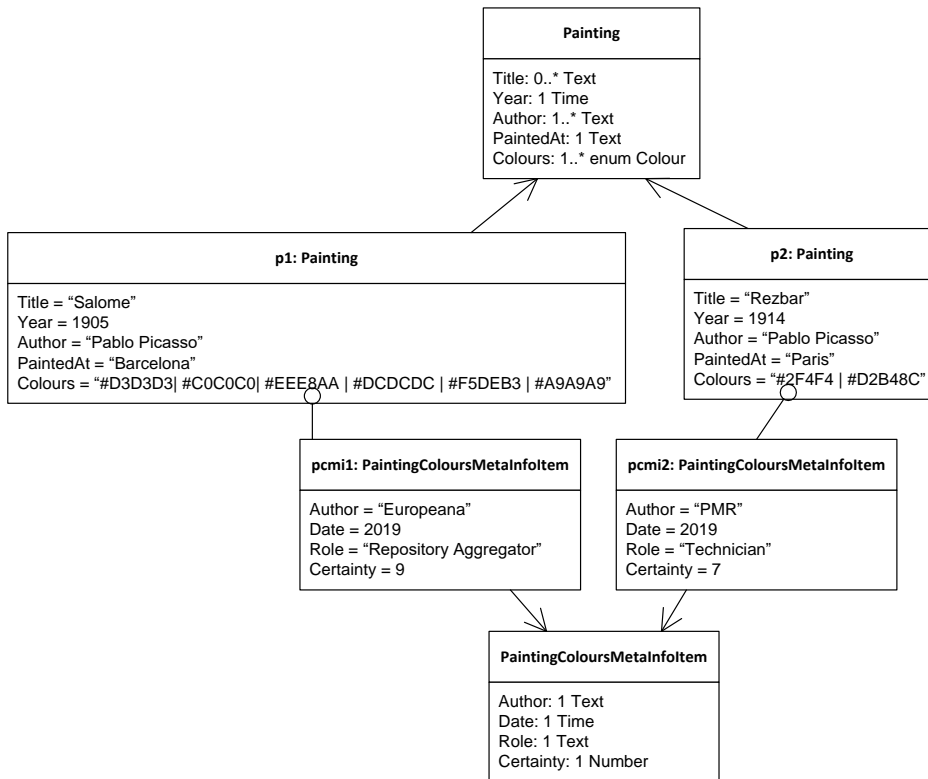

Fig. 15. Model for a knowledge generation scenario, using the proposed approach, and Europeana real data source extraction.

In this case, we have tested our relational database accessing by SQL queries to the metainformation. Specifically, the query searches for artworks painted in a specific location (i.e. Barcelona) and their colors specified by Europeana with a certain degree higher than 8. Thus, a simple SQL query with a join structure between the relational tables is enough to consult information and metainformation at the same time from both catalogues:

SELECT DISTINCT Painting.Title, Painting.Artist, PAinting.Year, Painting.Colours, Painting.PaintedAt, PaintingMetaInfoItem.Author FROM Painting, PaintingMetaInfoItem WHERE PaintedAt= 'Barcelona' and PaintingMetaInfoItem.Author= 'Europeana' and PaintingMetaInfoItem.Certainty > 8

Then, we also asked about artworks with other location and certainty level, and in this case including our own color characterization.

SELECT   DISTINCT   Painting.Title,   Painting.Artist,   Painting.Year,   Painting.Colours,   Painting.PaintedAt, PaintingMetaInfoItem.Certainty   FROM   Painting,   PaintingMetaInfoItem   WHERE   PaintedAt= 'Paris'   and PaintingMetaInfoItem.Author= 'PMR'

The execution of these queries allows us to test the hypothesis only with or without Europeana color assignment, trying different locations etc. Note that, in order to test the hypothesis, similar queries construction for other locations and predominant colors is needed. Fig 16 and Fig. 17 show the execution of both queries in the SQLite Studio console and how the system recovers the information asked (matching with the case exemplified in Fig. 15).
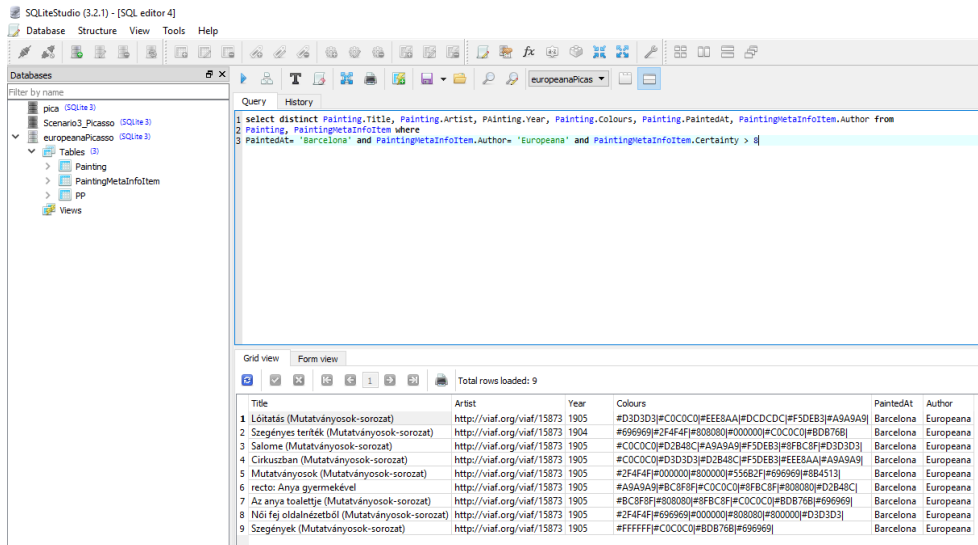

Fig. 16 Execution of knowledge generation first SQL query, recovering information and metainformation from Europeana data source for the colors assigned by Europeana to Picasso's artworks painted at Barcelona.

Using this approach, we would be able to document many paintings, and then assess what kind of correlation exists between place of painting and dominant colors, using the role of the technician and their stated certainty as weighing factors. This would yield an empirical measure that should work as a starting point to support or reject the hypothesis.
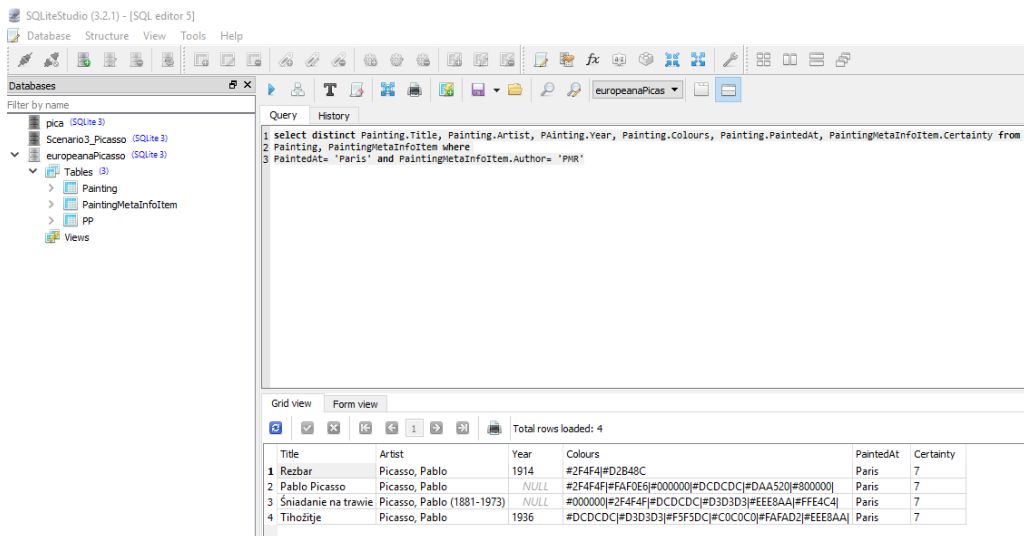

Fig. 17 Execution of knowledge generation second SQL query, recovering information and metainformation from Europeana data source for the colors assigned by us to Picasso's artworks painted at Paris.

7. Conclusions and future steps

In this paper we have dealt with the panorama of metadata in Digital Humanities. We have presented a review of 30 existing approaches, describing their most salient features, and selecting some of them for a deeper characterization and critique. Using this as a motivational starting point, and establishing some design criteria to develop a better approach, we explain how ConML incorporates a "Describes" relationship to implement metainformation. This implementation adopts the view that metainformation is information playing a specific role; that "meta" chains of arbitrary length must be supported to address relevant situations; and that conceptual concerns about metainformation must be kept separate from implementation issues such as the necessary storage or transmittal mechanisms. In addition, we have identified four major scenarios for metadata/metainformation use in Digital Humanities, and have described how our proposed approach can deal with them, testing the most complex scenarios 3 and 4 in an implementation with real data from external sources in digital humanities.

It is important to note that, in terms of metamodelling, our proposal has a minimal impact, as only one association in the ConML metamodel is necessary to introduce it. This means that other languages willing to adopt a similar solution would be able to do so with very few ripple effects on their metamodel. In this regard, it would be also interesting to evaluate the implementation of similar solutions on top of other modelling languages, especially those that lack a specific orientation towards metadata representation, such as conceptual graphs [78] (commonly used in RDF-based metadata applications).

A number of aspects remain to be developed. First of all, we are aware that the proposed solution works better when information and metainformation can be shaped as necessary from the beginning, in order to be adjusted to the task at hand, and that retrofitting metainformation as an afterthought to existing datasets is often a cumbersome and time-consuming task. However, this problem is inherent to working with data, and any other metainformation approach would suffer from these issues. In this regard, we argue that the proposal presented here, if anything, should be simpler to employ on legacy datasets given the facts that it introduces only a minimal set of new concepts to whatever modelling language is employed, and that the same tools and techniques that are employed for information can also be used for metainformation.

In addition, we are considering implementation options for metainformation propagation rules, which we have not dealt with in this paper. For example, does metainformation applied to an object (see e.g. pmi in Fig. 13) apply automatically to the contained values. Preliminary work seems to indicate that the propagation of metainformation (for example, from an object to its values or references) needs explicit regulation, so we are working on how to incorporate this to the current approach.

In summary, we have shown in this paper that a simple and expressive approach to metainformation can be used to address the major scenarios in Digital Humanities. We hope this will contribute to diminishing the proliferation of metadata "standards", improving the existing approaches and mitigating the issues found on them.

Acknowledgements

References

[1] M. Barbera, Linked (open) data at web scale: research, social and engineering challenges in the digital humanities, JLIS. it, 4 (2013) 91.
[2] C.L. Borgman, The digital future is now: A call to action for the humanities, Digital humanities quarterly, 3 (2009) 233.
[3] K.R. Eschenfelder, M. Caswell, Digital cultural collections in an age of reuse and remixes, Proceedings of the Association for Information Science and Technology, 47 (2010) 1-10.

[4] I. Faniel, E. Kansa, S.W. Kansa, J. Barrera-Gomez, E. Yakel, The challenges of digging data: a study of context in archaeological data reuse, in: Proceedings of the 13th ACM/IEEE-CS joint conference on Digital libraries, ACM, Indianapolis, Indiana, USA, 2013, pp. 295-304.

[5] J. Radatz, A. Geraci, F. Katki, IEEE standard glossary of software engineering terminology, IEEE Std, 610121990 (1990) 3.

[6] D.P. LLC., Glossary of Software Engineering terms, in, 2005.

[7] R.L. Ackoff, From data to wisdom, Journal of Applied Systems Analysis, 16 (1988) 3-9.

[8] C. Gonzalez-Perez, P. Martin-Rodilla, An Alternative Approach to Metainformation Conceptualisation and Use, in: H.C. Mayr, G. Guizzardi, H. Ma, O. Pastor (Eds.) Conceptual Modeling: 36th International Conference, ER 2017, Valencia, Spain, November 6–9, 2017, Proceedings, Springer International Publishing, Cham, 2017, pp. 92-105.

[9] web site at http://ariadne-infrastructure.eu/ (accessed on 26/04/2016).

[10] A. Felicetti, C. Meghini, C. Papatheodorou, J. Richards, ARIADNE D3.4: Final Report on Standards and Project Registry, in, FP7 Research Programme ARIADNE Project, 2016. .

[11] Dublin Core Metadata Element Set, version 1.1: Reference Description, web site at http://dublincore.org/documents/dces/ (accessed on

[12] O. Salvetti, G. Pieri, M. Di Bono, WP9: A review of data and metadata standards and techniques for representation of multimedia content, MUSCLE. Network of Excellence FP6-5077-52, (2004).

[13] W. Hodgins, E. Duval, Draft standard for learning object metadata, IEEE, 1484 (2002) 1-2002.

[14] ISO 19115-1:2014 Geographic information -- Metadata, web site at https://www.iso.org/standard/53798.html (accessed on

[15] ISO 21127:2006 Information and documentation -- A reference ontology for the interchange of cultural heritage information, web site at https://www.iso.org/standard/34424.html (accessed on

[16] C.F. Taylor, D. Field, S.-A. Sansone, J. Aerts, R. Apweiler, M. Ashburner, C.A. Ball, P.-A. Binz, M. Bogue, T. Booth, Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project, Nature biotechnology, 26 (2008) 889-896.

[17] L. Olsen, Directory Interchange Format (DIF): writer's guide, in.

[18] B.H. Liskov, J.M. Wing, A behavioral notion of subtyping, ACM Transactions on Programming Languages and Systems (TOPLAS), 16 (1994) 1811-1841.

[19] M.-A. Sicilia, Handbook of metadata, semantics and ontologies, World Scientific, 2013.

[20] A. Tort, A. Olivé, A computer-guided approach to website Schema. org design, in: International Conference on Conceptual Modeling, Springer, 2014, pp. 28-42.

[21] NISO, Understanding metadata, National Information Standards, 20 (2004).

[22] ISO/IEC 11179, Information Technology -- Metadata registries (MDR), web site at http://metadata-standards.org/11179/ (accessed on

[23] J. Riley, Glossary of metadata standards, J. Riley, 2010.

[24] Research Data Alliance, web site at https://www.rd-alliance.org/ (accessed on

[25] RDA Metadata Directory, web site at http://rd-alliance.github.io/metadata-directory/standards/ (accessed on

[26] E. Mendez, S. Van Hooland, Metadata typology and metadata uses, Handbook of metadata, semantics and ontologies, 1, (2013).

[27] Europeana project, web site at http://www.europeana.eu/ (accessed on 26/04/2016).

[28] Europeana Data Model Documentation, web site at https://pro.europeana.eu/resources/standardization-tools/edm-documentation (accessed on

[29] EDM Presentation, web site at https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_slides_130714.ppt (accessed on

[30] D.J. Macêdo, M. Shintaku, R.F. De Brito, Dublin core usage for describing documents in Brazilian government digital libraries, in: International Conference On Dublin Core And Metadata Applications, 2015, pp. 129-135.

[31] ISO/IEC, Information technology -- Object Management Group Unified Modeling Language (OMG UML) Part 1: Infrastructure. ISO/IEC 19505-1:2012., in, 2012.

[32] A.S. Maganto, J.N. Iso, D. Ballari, Normas sobre Metadatos (ISO 19115, ISO 19115-2, ISO 19139, ISO 15836), Mapping (123), (2008) 48-57.

[33] J.-r. Park, E. Childress, Dublin Core metadata semantics: An analysis of the perspectives of information professionals, Journal of Information Science, 35 (2009) 727-739.

[34] W.D. Robertson, E.M. Leadem, J. Dube, J. Greenberg, Design and Implementation of the National Institute of Environmental Health Sciences Dublin Core Metadata Schema, in: International Conference on Dublin Core and Metadata Applications, 2001, pp. 193-199.

[35] I. Mougenot, J.-C. Desconnets, H. Chahdi, A DCAP to promote easy-to-use data for multiresolution and multitemporal satellite imagery analysis, in: International Conference on Dublin Core and Metadata Applications, 2015, pp. 10-19.

[36] M.C. Malta, A.A. Baptista, C. Parente, A DCAP for the Social and Solidarity Economy, in: 2015 Proceedings of the Interna5onal Conference on Dublin Core and Metadata Applications, DCMI, 2015, pp. 20-29.

[37] A. Powell, M. Nilsson, A. Naeve, P. Johnston, DCMI abstract model, (2007).

[38] B.E. Bargmeyer, D.W. Gillman, Metadata standards and metadata registries: An overview, in: International Conference on Establishment Surveys II, Buffalo, New York, 2000.

[39] ConML Technical Specification. ConML 1.4.4, web site at http://www.conml.org/Resources_TechSpec.aspx (accessed on

[40] Shapes Constraint Language (SHACL), web site at https://www.w3.org/TR/shacl/ (accessed on

[41] W3C, W3C Semantic Web Online Web Page, (2017).

[42] G. Guizzardi, G. Wagner, A Unified Foundational Ontology and some Applications of it in Business Modeling, in: CAiSE Workshops (3), 2004, pp. 129-143.

[43] J.H. Lee, J.T. Tennis, R.I. Clarke, M. Carpenter, Developing a video game metadata schema for the Seattle Interactive Media Museum, International Journal on Digital Libraries, 13 (2013) 105-117.

[44] A. Liu, The state of the digital humanities: A report and a critique, Arts and Humanities in Higher Education, 11 (2012) 8-41.

[45] C. Warwick, M. Terras, J. Nyhan, Digital humanities in practice, Facet Publishing, 2012.

[46] A. Benardou, P. Constantopoulos, C. Dallas, D. Gavrilis, Understanding the information requirements of arts and humanities scholarship, International Journal of Digital Curation, 5 (2010) 18-33.

[47] T. Blanke, M. Hedges, Humanities e-Science: From systematic investigations to institutional infrastructures, in: e-Science (e-Science), 2010 IEEE Sixth International Conference on, IEEE, 2010, pp. 25-32.

[48] M. Doerr, The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata, AI Mag., 24 (2003) 75-92.

[49] G. Schreiber, A. Amin, L. Aroyo, M. van Assem, V. de Boer, L. Hardman, M. Hildebrand, B. Omelayenko, J. van Osenbruggen, A. Tordai, Semantic annotation and search of cultural-heritage collections: The MultimediaN E-Culture demonstrator, Web Semantics: Science, Services and Agents on the World Wide Web, 6 (2008) 243-249.

[50] Y. Wang, N. Stash, L. Aroyo, P. Gorgels, L. Rutledge, G. Schreiber, Recommendations based on semantically enriched museum collections, Web Semantics: Science, Services and Agents on the World Wide Web, 6 (2008) 283-290.

[51] E. Hyvönen, E. Mäkelä, M. Salminen, A. Valo, K. Viljanen, S. Saarela, M. Junnila, S. Kettula, MuseumFinland—Finnish museums on the semantic web, Web Semantics: Science, Services and Agents on the World Wide Web, 3 (2005) 224-241.

[52] M. Patel, M. White, N. Mourkoussis, K. Walczak, R. Wojciechowski, J. Chmielewski, Metadata requirements for digital museum environments, International Journal on Digital Libraries, 5 (2005) 179-192.

[53] C. Tsinaraki, E. Fatourou, S. Christodoulakis, An ontology-driven framework for the management of semantic metadata describing audiovisual information, in: CAiSE, Springer, 2003, pp. 340-356.

[54] A. Chianese, F. Marulli, F. Piccialli, I. Valente, A novel challenge into multimedia cultural heritage: An integrated approach to support cultural information enrichment, in: Signal-Image Technology & Internet-Based Systems (SITIS), 2013 International Conference on, IEEE, 2013, pp. 217-224.

[55] M.J. Addis, K. Martinez, P. Lewis, J. Stevenson, F. Giorgini, New ways to search, navigate and use multimedia museum collections over the web, (2005).

[56] M. Dragoni, E. Cabrio, S. Tonelli, S. Villata, Enriching a Small Artwork Collection Through Semantic Linking, in: International Semantic Web Conference, Springer, 2016, pp. 724-740.

[57] C. Binding, K. May, D. Tudhope, Semantic interoperability in archaeological datasets: Data mapping and extraction via the CIDOC CRM, in: Research and Advanced Technology for Digital Libraries, Springer, 2008, pp. 280-290.

[58] K. Ivanova, E. Velikova, P. Stanchev, I. Mitov, Automated Metadata Extraction from Art Images.

[59] P. Martín-Rodilla, Software-Assisted Knowledge Generation in the Archaeological Domain: A Conceptual

Framework, in: B.W.E. Marta Indulska (Ed.) 25th International Conference on Advanced Information Systems Engineering (CAiSE 2013): Doctoral Consortium, Valencia, Spain, 2013.

[60] C. Kakali, I. Lourdi, T. Stasinopoulou, L. Bountouri, C. Papatheodorou, M. Doerr, M. Gergatsoulis, Integrating Dublin Core metadata for cultural heritage collections using ontologies, in: International Conference on Dublin Core and Metadata Applications, 2007, pp. 128-139.

[61] J. Svensson, A. Edman, F. Bengtsson, J. Zelin, F. Gröndahl, Modeling museum context in CIDOC CRM using relational databases, (2006).

[62] I. Karvasonis, Study of the heterogeneity in cultural databases and transformation of examples from CIMI to the CIDOC CRM, in, Technical Report 291, ICS-FORTH, 2001.

[63] Digital Curation Centre, University of Edinburgh, web site at http://www.dcc.ac.uk/resources/metadata-standards/list (accessed on

[64] R. Denaux, C. Dolbear, G. Hart, V. Dimitrova, A.G. Cohn, Supporting domain experts to construct conceptual ontologies: A holistic approach, Web Semantics: Science, Services and Agents on the World Wide Web, 9 (2011) 113-127.

[65] V. Charles, A. Isaac, K. Fernie, C. Dallas, D. Gavrilis, S. Angelis, Achieving interoperability between the CARARE schema for monuments and sites and the Europeana Data Model, arXiv preprint arXiv:1306.2869, (2013).

[66] Digital Research Infrastructure for the Arts and Humanities (DARIAH) web site at https://dariah.eu/ (accessed on 26/04/2016).

[67] Bamboo project, web site at http://www.projectbamboo.org/ (accessed on 26/04/2016).

[68] ConML Technical Specification. ConML 1.4., web site at http://www.conml.org/Resources_TechSpec.aspx (accessed on 30/03/2016).

[69] C. Gonzalez-Perez, P. Martín-Rodilla, Teaching Conceptual Modelling in Humanities and Social Sciences, Revista de Humanidades Digitales, 1 (2017) 408-416.

[70] C. Gonzalez-Perez, P. Martín-Rodilla, Integration of Archaeological Datasets through the Gradual Refinement of Models, in: F. Giligny, F. Djindjian, L. Costa, P. Moscati, S. Robert (Eds.) 21st Century Archaeology: Concepts, Methods and Tools - Proceedings of the 42nd Annual Conference on Computer Applications and Quantitative Methods in Archaeology, Archaeopress, 2015, pp. 193-204.

[71] C. Gonzalez-Perez, A conceptual modelling language for the humanities and social sciences in: RCIS'12: Sixth International Conference on Research Challenges in Information Science, Xplore, Valencia, Spain, 2012, pp. 1-6.

[72] SQLite database system. https://www.sqlite.org/index.html

[73] SQLite Studio software https://sqlitestudio.pl/index.rvt

[74] C. Gonzalez-Perez, Information Modelling for Archaeology and Anthropology: Software Engineering. Principles for Cultural Heritage. Springer, 2018.

[75] MET Metropolitan Museum of Art's CC0 select datasets. https://github.com/metmuseum/openaccess , 2019.

[76] MoMA The Museum of Modern Art (MoMA) Collection, https://github.com/MuseumofModernArt/collection, 2019.

[77] P. Martin-Rodilla, ConML Metainformation. https://github.com/patrimrodilla/PicassoMetaInformation, 2019.

[78] J.F. Sowa, Conceptual structures: information processing in mind and machine, (1983).