Metadata as Semantic Palimpsests: The Case of PHAIDRA@unipd

Anna Bellotto and Cristiana Bettella(✉)

The University Library Centre "ULC", University of Padova,
Via Anghinoni, 3, 35121 Padua, PD, Italy
{anna.bellotto,cristiana.bettella}@unipd.it

Abstract. This paper illustrates the experience of the Library System of the University of Padova in reviewing the data model of Phaidra (Permanent Hosting, Archiving and Indexing of Digital Resources and Assets), the digital repository for the long-term management and preservation of digital objects in place since 2010, whose system was created and developed by the University of Vienna. In order to provide better informational representation and visualisation of data, both in terms of metadata quality and display, this re-examination consisted in a critical analysis of the foundational metadata profile of Phaidra, its mapping and conversion into the Dublin Core metadata schema (Dublin Core Metadata Element Set 1.1) and, at prototype level, into the Metadata Object Description Schema (MODS). This paper discusses the evidence of the identified solutions being guided by two core principles: on the one hand, the distinctive valorisation of the dual analogue-digital nature of the Phaidra cultural heritage object, on the other, the metadata reuse in the visual function for the graphic updating of the new web interface, which is being done in order to encourage the discovery, even serendipitously, of its content by the digital researcher. Finally, the presentation considers the development activities being carried out by the Phaidra working groups of the Universities of Padova and Vienna, focused on the semantic evolution of the concept of metadata to open data, by presenting here an unpublished example of the Simple Knowledge Organization System (SKOS) prototype and last, but not least, suggesting the definition of a new Phaidra data model.

Keywords: Cultural heritage object metadata · Crosswalk · Data model · Web of data

1 Introduction

Phaidra (Permanent Hosting, Archiving and Indexing of Digital Resources and Assets) is the platform of the University Library System of the University of Padova for the long-term archiving of digital objects and collections, currently hosting a vast range of 390,000 digital objects including antiquarian books, manuscripts, photographs, wall charts, maps, learning objects, films, archival material and museum objects [1].

Designed and developed by the University of Vienna[1] [2] beginning in 2008 based on the digital architecture of the Fedora open source system, Phaidra was adopted by the

[1] Phaidra has arisen from the cooperation between the Computer Centre of the University of Vienna, the University Library of Vienna and the Centre for Teaching and Learning. Project management is located at the University Library.

P. Manghi et al. (Eds.): IRCDL 2019, CCIS 988, pp. 167–184, 2019.
https://doi.org/10.1007/978-3-030-11226-4_14

University of Padova in 2010, when the two institutions signed a bilateral collaboration agreement which led to the creation of the organisational structure and the formation of the Phaidra.org network infrastructure [3], which gradually continued to grow, both locally and internationally, by hosting cultural institutions such as Galleries, Libraries, Archives, and Museums, also known by the acronym GLAM [4, 5] (See Fig. 1).

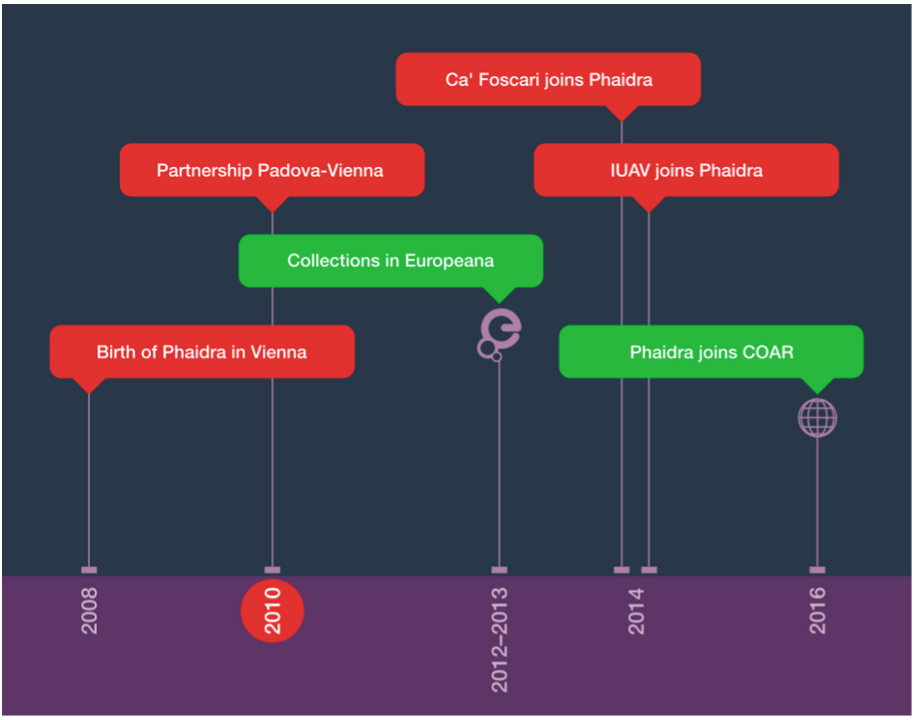

Fig. 1. Phaidra@unipd.it timeline.

The distinctive characterisation of the Paduan instance of Phaidra since its inception has on the one hand been the illustrative and heterotopic[2] valorisation of the heterogeneous richness of the University's digital collections of cultural heritage, such as

[2] Heterotopic, from heterotopia, a placeless place that refers to all other spaces, to every conceivable space, according to the meaning that the philosopher Michel Foucault gave to the medical term in a conference held in March 1967 and published later under the title "Des espaces autres" in the magazine Architecture-Mouvement-Continuité, n. 5, October 1984 (translated as: "Of other spaces" in Diacritics, Vol. XVI, n. 1, 1986, available online at https://foucault.info/documents/heterotopia/foucault.heteroTopia.en/). The extension of the spatial concept of heterotopia to libraries and digital collections – the library is among the Foucauldian examples of heterotopia – would deserve a reflection of its own with respect to the objective that we propose in this paper. See: Bruno, G. Atlas of Emotions. Journeys in Art, Architecture, and Film, Verso (2002) and Wikipedia entry https://en.wikipedia.org/wiki/Heterotopia_(space).

those from departments and research centres, archives and museums, as well as from libraries' digitisation projects [6]. On the other hand, it acts as an attraction for other academic institutions in the region[3], triggering a virtuous cycle of cultural and technological infrastructural osmosis which has conferred upon Phaidra, in addition to its primary function as a Digital Asset Management system, as well as the inter-institutional aggregator, an organisation that collects and aggregates, creates and administers metadata from multiple content providers.

Given the definition of aggregator, Phaidra serves at the same time as a service provider, through its portal and Web API [4, 7][4], and as a data provider to external service providers, exposing its metadata through the OAI-PMH protocol [8, 9]. The expansion of this aggregative and meta-aggregative function of heterogeneous metadata from similarly heterogeneous origins has raised the urgent need for a critical analysis of the foundational data model of Phaidra Universität Wien metadata (hereinafter UWmetadata), both from the point of view of its mapping and conversion into the Dublin Core metadata scheme aimed at its publication in the OAI-PMH Phaidra data provider, as well as the visualisation and presentation of data in the Phaidra web interface [10].

This presentation provides evidence of the solutions which were identified and their outcomes, aimed on the one hand at the distinctive valorisation of the dual analogue-digital identity of the Phaidra cultural heritage object, highlighting its profile from the authorial point of view (people → Who), from the physical-digital materiality of the work being described (works → What), from the space-time dimension (Where and When) and from the traceability of the provenance; and, on the other hand, it is aimed at metadata reuse as a visual function and for accessibility to content, used in the conception of the new graphic design of the web interface, which is being done in order to encourage discovery, even serendipitously, of the content found in Phaidra by digital researchers and browsers.

Coherent to a process intended to be evolutionary and seamless, the implementation project of the Metadata Object Description Schema (MODS) standard in Phaidra will also be illustrated. Through the prototype processing of mapping from UWmetadata, this has allowed for a retrospective and recursive review of some coding choices previously defined in the mapping between the UWmetadata schema and the Dublin Core elements set, resulting in the updating to a new version of the profile and regeneration of Dublin Core metadata in the Phaidra web interface [11].

Finally, we intend to highlight how the MODS implementation is also functional in the prospects of ongoing and future development of the Phaidra platform, focused on Resource Description Framework (RDF) migration and adoption of standards and their technologies of the Semantic Web, in line with the semantic evolution of the concept of metadata to linked open data, meant as a new informational entity which necessarily

[3] The Universities of Ca′ Foscari and IUAV of Venice are a part of Phaidra since 2014.

[4] "A set of public APIs, REST-compliant, used to provide search services, content management of digital objects and sampling and handling of metadata, is also available on Phaidra. Anyone who wants to develop an application that presents digital objects in a customised way, can freely use these APIs. An example of such an application is the Collection Viewer, developed for sharing and browsing in digital collections on Phaidra in an external site through embedding" [4] (See: http://bibliotecavallisneri.cab.unipd.it/collezioni-digitali/zoologische-wandtafeln-von-prof-dr-paul-pfurtscheller).

eludes the identity and exclusiveness of a single catalogue, a single digital repository, or a single digital library system, by presenting here an unpublished example of the Simple Knowledge Organization System (SKOS) prototype and last, but not least, suggesting the definition of a new Phaidra data model.

2 The PHAIDRA_DC Profile (and PHAIDRA_MODS Profile)

The foundational data model of Phaidra, called Universität Wien metadata and abridged as UWmetadata, informs the design of Phaidra metadata, both in terms of representation of the values as well as the description of the contents [12]. It is the result of the expansion of the IEEE Learning Object Metadata (LOM) standard (IEEE 1484.12.1 – 2002), and the combination of elements of different metadata namespaces. This places UWmetadata among the examples of application profile (AP), which for example include the data profile of the portal for CulturaItalia PICO, the data profile of the portal for Europeana E(uropeana)D(ata)M(odel) and the former E(uropeana)S (emantics)E(lements) as well as the application of EDM in the context of the Digital Public Library of America and the German Digital Library.

In general, the LOM schema is a data model, usually encoded in XML, used to describe learning objects or digital resources for educational purposes such as learning supports. LOM, as its UWmetadata application profile, structures the metadata in accordance with a hierarchy of elements defined in nine top-level categories, and containing groups of attributes in a tree structure. In the following images (Figs. 2, 3 and 4), see respectively the LOM conceptual map [13], the explanation of its top-level categories and an XML snippet of UWmetadata schema:

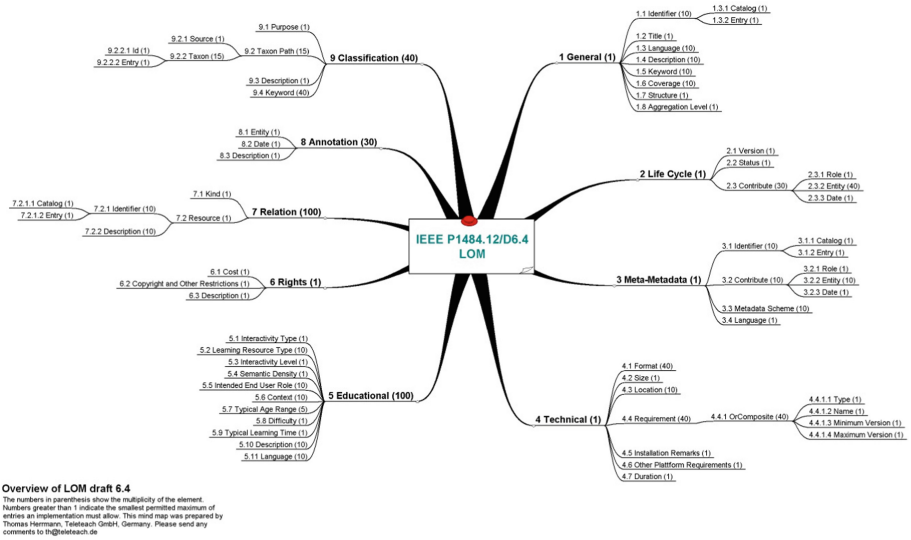

Fig. 2. Overview of LOM. Mind map prepared by Thomas Herrmann (Source: [13])

LOM top level category	Explanation
General	This category groups the general information that describes the learning object as a whole.
Lifecycle	This category describes the history and current state of the learning object and those entities that have affected the learning object during its evolution.
Meta-metadata	This category describes how the metadata instance can be identified; who created this metadata instance; and how, when, and with what references.
Technical	This category describes the technical requirements and characteristics of the learning object.
Educational	This category describes the key educational or pedagogic characteristics of the learning object.
Rights	This category describes the intellectual property rights and conditions of use for the learning object.
Relation	This category defines the relationship between a learning object and other learning objects, if any.
Annotation	This category provides comments on the educational use of the learning object, and information on when and by whom the comments were created.
Classification	This category describes where the learning object falls within a particular classification system. To define multiple classifications there may be multiple instances of this category.

UWmetadata extends the LOM schema with further additional categories aiming to represent primary data stored in Phaidra (Contextual allegation and Provenience respectively)[5], digital books (Digital book)[6], and electronic theses (eThesis)[7]:

[5] The Institute History of Arts at the University of Vienna asked for these new sections [12].

[6] Connected to The Vienna University Library eBook on demand (EOD) service: https://bibliothek.univie.ac.at/en/eod.html. Moreover, the Digitalbook sub-elements, if completed, have a bearing on the way the book information frame appears in the Book Viewer, for instance: Place of publication, Publisher, Date of publication, Catalogue URL, Number of pages, or volume no. By way of example, see: https://fc.cab.unipd.it/fedora/objects/o:387557/methods/bdef:Book/view?language=en#page/1/mode/2up.

[7] The eThesis category has not been implemented at Phaidra@unipd since theses are currently hosted at the EPrints repositories Padua@thesis (http://tesi.cab.unipd.it/) and Padua@research (http://paduaresearch.cab.unipd.it/), the latter specifically devoted to doctoral dissertations.

UWmetadata	Explanation
Contextual allegation	This category defines the physical description of primary data
Provenience	This category defines the provenance of primary data
Digital Book	This category describes the bibliographic information of digital books

Fig. 3. LOM (Source: [14]) and UWmetadata top-level categories.

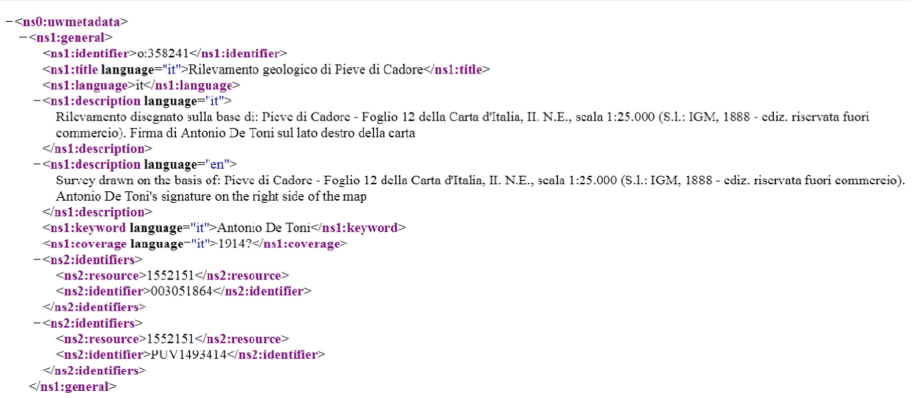

Fig. 4. UWmetadata_XML: General.

The analysis of mapping between the UWmetadata schema source and the Dublin Core target schema has established, as each of metadata crosswalk activities, the mapping of the correspondences of elements, the syntax and semantics of the two schemas involved, adopting a relative translation mode, namely trying to map each source element into at least one of the target elements, in order to avoid as much as possible any loss of information recorded in the source schema.

Each correspondence has also determined, depending on the encoding purposes, an inter-schema relationship such as one-to-one, one-to-many, many-to-one, and many-to-many, *a fortiori* if one considers the conversion of a descriptive schema organised in blocks, or nuclei, semantic and nested like UWmetadata in a flat schema such as the Dublin Core metadata schema.[8]

[8] There is not a full overlapping among the two metadata schemas. For instance, the unrefined element `<dc:relation>` encoded the system of relationships currently handled in Phaidra, which are not represented in UWmetadata (See: https://github.com/phaidra/phaidra-api/wiki/Relations).

By way of example, the authorial entity has been defined both from the point of view of the intellectual level of contribution in the creation of the described resource (Creator, Contributor) and from that of the mode and form established for the value recorded in the source elements[9]:

UWmetadata_ Source	**1. Lifecycle** **Creator, Contributor_Person**

```
<ns1:lifecycle>
      <ns1:contribute seq="">
            <ns1:role>PHAIDRA role code for Creator, Contributor, person</ns1:role>
            <ns1:entity seq="">
                  <ns3:firstname>Firstname</ns3:firstname>
                  <ns3:lastname>Lastname</ns3:lastname>
                  <ns3:type>person</ns3:type>
            </ns1:entity>
            <ns1:date>Date expressed as YYYY</ns1:date>
      </ns1:contribute>
</ns1:lifecycle>
```

Creator, Contributor_Institution

```
<ns1:lifecycle>
      <ns1:contribute seq="">
            <ns1:role>PHAIDRA role code for Creator, Contributor, institution</ns1:role>
            <ns1:entity seq="">
                  <ns3:institution>Institution</ns3:institution>
                  <ns3:type>institution</ns3:type>
            </ns1:entity>
            <ns1:date>Date expressed as YYYY</ns1:date>
      </ns1:contribute>
</ns1:lifecycle>
```

[9] The value encoded in <ns1:role> is taken from Phaidra Roles Vocabulary, which relates to the role of the entities contributing to the creation of the (analogue or digital) resource (See further: Towards a semantic data modelling).

DC_Target	<dc:creator>ns3:lastname, ns3:firstname (ns1:role eng)</dc:creator> <dc:contributor>ns11:lastname, ns11:firstname (ns10:role eng)</dc:contributor>
	<dc:creator>ns3:institution (ns1:role eng)</dc:creator> <dc:contributor>ns11:institution (ns11:role eng)</dc:contributor>

Example	<dc:creator>Pellegrini, Giovan Battista</dc:creator> <dc:creator>Germania : Wehrmacht : Propaganda Staffel</dc:creator> <dc:contributor>Mari, Mario (Illustrator) </dc:contributor> <dc:contributor>Università di Padova - Centro di Ateneo per le Biblioteche (Digitiser)</dc:contributor> <dc:contributor>Facoltà di Scienze matematiche, fisiche e naturali (Curator)</dc:contributor> <dc:contributor>Bodrero, Emilio (Former owner)</dc:contributor> People Pellegrini, Giovan Battista (Author) Germania : Wehrmacht : Propaganda Staffel (Author) Mari, Mario (Illustrator) Università di Padova - Centro di Ateneo per le Biblioteche (Digitiser) Facoltà di Scienze matematiche, fisiche e naturali (Curator) Bodrero, Emilio (Former owner)

Each inter-schema encoding relationship has been described and reviewed following a system of standard analytical matrix accompanied by the following informational units: standard textual descriptions and definitions of the elements adapted to the Phaidra informational context; their unique identification (Uniform Resource Identifier, URI) and possible classification into the corresponding elements refined by DCMI Metadata Terms; the compulsoriness and replicability of conversion encoding; the version of the mapping, namely replacement of an earlier version; notes where the reasons for the encoding choices adopted are critically discussed, also in terms of historical stratification of the mapping, together with the formalisation of any logical conditions to be complied with for a correct writing of the conversion code; the exemplary extrapolation of XML snippets from source and target elements; the label

adopted in the display element and its real representation from the point of view of the form prescribed for the value of the reference being converted [10]:

- Name
- Label (ita)
- Label (eng)
- Defined by
- Term_URI
- Definition (ita)
- Definition (eng)
- Refined by
- Obligation & Occurence
- Mapping version
- Replaces
- Comment (ita)
- UWmetadata_Source
- DC_Target
- Visualizzato come (ita)
- Visualised as (eng)
- Esempio
- Example.

As a reflection of the inter-schema complexity of mapping of the two schemas involved, we can use the example of encoding the Dublin Core Date element compared to its specular version in the temporal elements of the MODS standard, from which, however, the Dublin Core mapping has recursively derived benefits both in terms of information quality of the metadata and in their presentation.

2.1 Date

In the context of PHAIDRA_DC profile, the Date element <dc:date> translates a temporal event related to the lifecycle of the resource, certain or inferred, in terms of the date of publication (dateIssued) or creation (Created), which can be expressed either in exact form or as a time interval, open or closed, according to the formats that conform to the ISO 8601 Date and time format standard.

UWmetadata sources for coding the Date element are:

1. sub-element <ns1:date> of Contribute <ns1:contribute> in Lifecycle <ns1:lifecycle>
2. sub-elements <ns10: date_from><ns10:date_to> of Contribute <ns10:contribute> in Provenience <ns10:provenience>
3. the element <ns12:releaseyear> in Digitalbook <ns12:digitalbook>.

The conversion to <dc:date> of sub-elements <ns10:date_from> <ns10:date_to> from Contribute <ns10:contribute> in Provenience <ns10: provenience> was introduced in version 1.1 2018, on the basis of the encoding example of MODS temporal sub-elements, in order to permit the allocation of uncertain or inferred dating to the described resource.

In fact, MODS provides granular and distinctive representation about the temporalisation of the resource by hosting in the upper element <originInfo> the sub-elements specific to each type of dating: <dateIssued>, <dateCreated>, <dateValid>, <dateModified>, <copyrightDate>, <dateOther>, partially also represented by qualified terms of the Dublin Core vocabulary.

By contrast, encoding of the Date <dc:date> element not only translates the values of heterogeneous source elements, but also if, and only if, there are the predetermined conditions, which are in order:

1. <ns12:releaseyear> of <ns12:digitalbook>. If <ns12:releaseyear> is not filled, then:
2. <ns1:date> of <ns1:contribute> in <ns1:lifecycle> with <ns1:role> code 47 equal to Publisher type institution (<ns3:type>institution </ ns3:type>) OR with <ns1:role> code 1557141 of Printer type person or institution. If these elements are not populated, then:
3. the first occurrence <ns1:date> of <ns1:contribute> in <ns1:lifecycle> with <ns1:role> code 1552095 same as Author, whether the entity is a person or an institution. If these elements are not populated, then:
4. the first occurrence of <ns1:date> of <ns1:contribute> of <ns1:lifecycle> if the value of <ns1:role> has different code than 47 Publisher institution type, and from 1557141 Printer to code 1552154 Author of the digitisation, whether the entity is a person or an institution.

If the conditions described in paragraphs 1–4 are not met, then they take the values of the sub-elements Date from <ns10:date_from>; Date to, Date up to <ns10:date_to> of Contribute <ns10:contribute> in Provenience, which can represent an uncertain or inferred date of the resource, according to the following modes and forms:

5(a) <ns10: date_from> and <ns10: date_to> in <ns10:provenience>, where the two dates are equivalent if precise dates are being attributed (eg.: 1950)

5(b) <ns10: date_from> and <ns10: date_to> in <ns10:provenience>, where the two dates differ if it intends to assign a set interval of dates (eg.: 1950–1960)

5(c) <ns10: date_from> in <ns10:provenience> if an open date interval (eg.: 1950-) are being attributed.

If the conditions set out in paragraphs 1–5 are not met, the <dc:date> is omitted, and the date of publication of the digital object in PHAIDRA recorded in <ns1:upload_date> of <ns1:lifecycle> is not published.

See the example of display of the target elements, which significantly highlights the translational operation described, in particular by observing the outcome displayed:

DC_Target	<dc:date>ns12:releaseyear</dc:date> If not then: <dc:date>ns1:date</dc:date> → *as expressed by conditions 2) and 3)* If not then: <dc:date>ns10:date_from=ns10:date_to</dc:date> → *as expressed by conditions 5a)* If not then: <dc:date>ns10:date_from-ns10:date_to</dc:date> → *as expressed by conditions 5b)* If not then: <dc:date>ns10:date_from-</dc:date> → *as expressed by conditions 5c)*

Example	Date 1822 1820-1830 1822-

2.2 Some Remarks in the Margin

The mapping and conversion of UWmetadata to Dublin Core is *de facto* a chance conversion, sometimes acrobatic, aimed at informing a qualified and comprehensive informative aggregate in the conceptual simplification of the Dublin Core element, which has made it possible at once to identify the functional requirements needed to accommodate schemas of refined metadata, such as for example the MODS standard, whose prototype mapping analysis has made further refinement possible in the Dublin Core encoding choices.

Additionally, it has enhanced precision in the representation of data and information content of Phaidra by disclosing its data profile and conceptualising the source model of data by qualifying and formalising its core elements. It also encouraged reflection in evolutionary terms of its ontological structure, outlining where possible the classes of Persons, Works, Space, and Time.

It has clarified and characterised the different levels and the interdependence of the dimensional combination of physicality and digitality of the information content (and knowledge) conveyed by the Phaidra cultural heritage object.

It has, in particular, made it possible to experience and evaluate the vital importance of standards, the adoption of which creates interoperability, leading to the decontextualisation of data from the scope of their original creation, accelerating their exposure and potential for reuse in different contexts and by different services, with the valorisation and development of the informative and cognitive value of which they are a memorial device.

It has helped to strengthen the methodological attitude for a correct reading of the data, or of reading of the data as "semantic palimpsests" given the stratified and permanent coexistence of heterogeneous data models, and their evolutionary and generative function in terms of schema, structure, profile, model, in aggregate form and as *corpora* of data, stressing the knowledge that each mapping activity is an inter-data conceptual negotiation which implies, more than anything else, a mutual understanding and compromise (See Fig. 5).

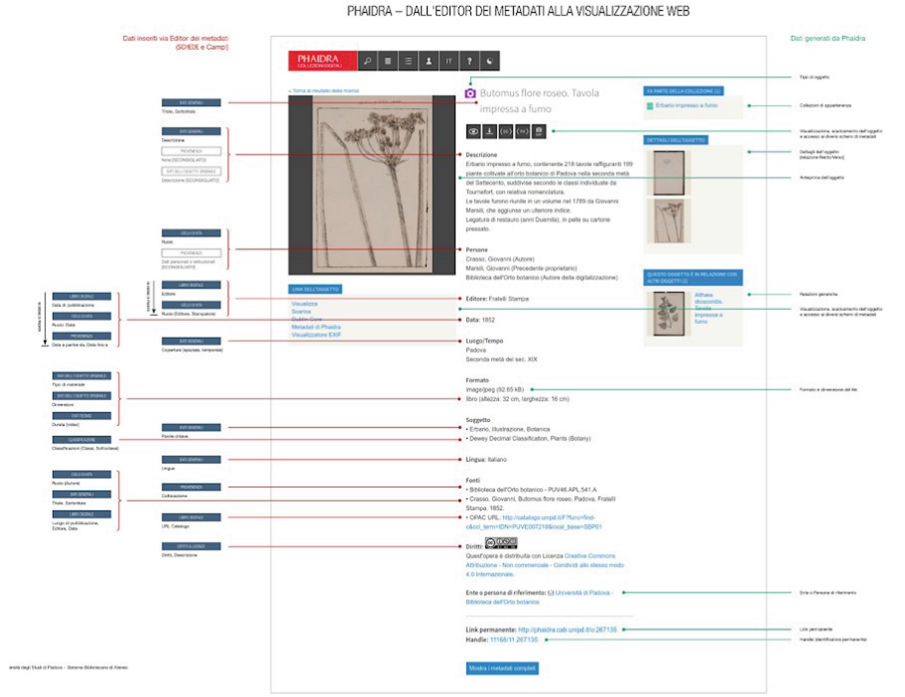

Fig. 5. From the Metadata editor to the web visualisation (http://phaidra.cab.unipd.it/static/campi-di-phaidra.pdf).

3 Towards a Semantic Data Modelling

"In libraries, metadata is a first-class object. Metadata doesn't just describe a library resource, it also connects that resource to entities such as people, places, events, and to other resources" [15]. The persisting validity of this statement in the context of digital repositories can be easily let it be understood. Rather, it is here considered of greater value to focus on how the technical set of established standards and technologies, collectively referred to as "Linked Data", are able to implement and increase the faceted connective potential of metadata.

Since 2016, within a broader workflow pursued by the team of the University of Vienna targeting an enhanced interoperability and accessibility of Phaidra, the development and implementation of a Classification Server has been assessed as one task needed to be carried out. Consisting of an external component to the digital asset management system, the Classification Server would handle and provide in one place all relevant thesauri, classification schemes, taxonomies and controlled vocabularies potentially needed for the assignment of well-defined metadata to digital objects [16] (see Fig. 6). The ideal goal of the Classification Server would be to ensure that during the upload of new items to the repository, when the user adds metadata information to these resources, as well as during the search for objects, when the user types terms in order to search for specific digital objects, Phaidra's users could have access to the main Knowledge Organization Systems (KOSs) as much as easy, comfortable and user-friendly as possible [16]. The opportunity represented by this tool to remarkably enrich both the platform usability and metadata interoperability of the repository was recognised by the working group of the University of Padova as well; this team thus decided in the last few months to start its own narrower implementation analysis of the new component in its local management system.

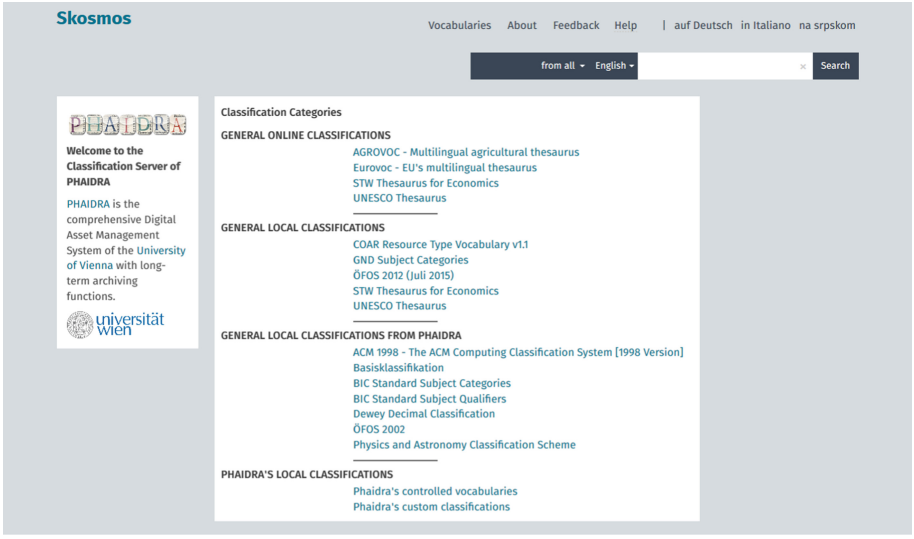

Fig. 6. Opening page of the Classification Server (Source: [19]).

From a technical point of view, the required import format for the various controlled vocabularies and classification schemes to be imported and locally managed by the database and the web application that form up the Classification Server architecture (see Fig. 7) is Simple Knowledge Organization System (SKOS) data model. This is defined as a data-sharing standard recommended by the W3C community for representing in Semantic Web and Linked Data technologies context, that is in Resource Description Framework (RDF) language, the many existing systems used to organise information. Based on the assumption that the Knowledge Organization Systems share a similar structure and content, SKOS allows them to be transformed from isolated and stand-alone entities of organized information into a global machine-readable network of highly integrated conceptual schemes, which are thus publishable in the Web, shareable, readable, wider re-usable and therefore automatically (and much more meaningfully) discoverable by software applications [17, 18].

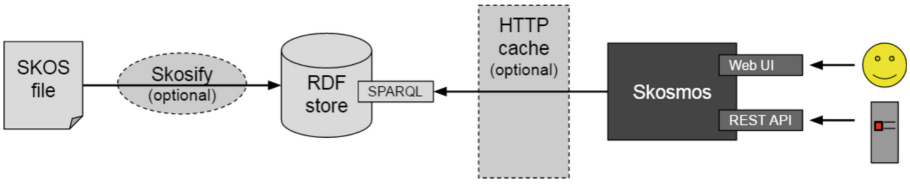

Fig. 7. Tehnical architecture of the Classification Server (Source: [20]).

To date, within Phaidra archival system, the most concrete attempt towards the use of standardised values as metadata has resulted into two locally-implemented lists of descriptors intended to index the catalogued objects and to facilitate their easy retrieval. They are the Phaidra Type of Material Vocabulary[10] and the Phaidra Roles Vocabulary[11], relating respectively to the material of the resource and the role of the entities contributing to the creation of the (analogue or digital) object. Given this premise, the first phase of implementation of the Classification Server was thus focused on transferring these two (not machine-readable) local vocabularies to the Semantic Web technology context. Their representation in SKOS format was tested towards their future migration from data silos to largely linkable and shareable Linked Open Vocabularies.

[10] Phaidra Type of Material Vocabulary includes the following terms available to selection when cataloguing a resource: Arrangement, Article of periodical, Atlas, Book, Book part, Drawing, Image, Letter, Manuscript, Map, Negative, Object, Other, Painting, Periodical, Picture, Postcard, Poster, Print, Remote sensing image, Score, Slide, Sound recording, Video, Wallchart. The vocabulary also includes other currently hidden terms, made accessable on request [21].

[11] Phaidra Roles Vocabulary currently includes 29 terms: Architect, Arranger, Artist, Author, Calligrapher, Cartographer, Composer, Curator, Data contributor, Dedicatee, Digitiser, Dubious author, Editor, Editor of compilation, Engraver, Former owner, Graphic technician, Illuminator, Illustrator, Musician, Other, Photographer, Printer, Publisher, Scientific advisor, Sculptor, Thesis advisor, Transcriber, Translator, Videographer. In addition, similarly to the Phaidra Type of Material Vocabulary additional currently hidden entries are made available at the request of the individual cataloguer [21].

Overall, during the prototypal encoding of these two lists of terms, the essential emphasis of SKOS data model on semantics rather than on terminology enabled the problematic and poorly interoperable approach of the "term" to be replaced with the less ambiguous and much more effectively understandable notion of the "concept". The concept is what constitutes the fundamental element of SKOS vocabulary: it is an abstract unit of thought, i.e. an idea, a meaning, a class of objects or events, uniquely identified by an URI and independent from the multiform expressions used to label it in natural language [17, 18]. The example in Fig. 8 visually displays the aforementioned theoretical core of SKOS standard: the concept of "calligrapher" (`<skos:Concept>`), contained in Phaidra Roles Vocabulary, is noticeably distinct from its corresponding bilingual terminological expressions, both the preferred lexical label (`<skos:prefLabel>`) and the alternative one (`<skos:altLabel>`), as well as from the explanatory note about the intended meaning of the concept (`<skos:definition>`).

```
<skos:Concept rdf:about="http://phaidra.org/vocabularies/roles/calligrapher">
  <skos:inScheme rdf:resource="http://phaidra.org/vocabularies/roles"/>
  <skos:topConceptOf rdf:resource="http://phaidra.org/vocabularies/roles"/>
  <skos:prefLabel xml:lang="en">calligrapher</skos:prefLabel>
  <skos:prefLabel xml:lang="it">copista</skos:prefLabel>
  <skos:altLabel xml:lang="en">scribe</skos:altLabel>
  <skos:altLabel xml:lang="it">scriba</skos:altLabel>
  <skos:definition xml:lang="en">A person who inscribe or copy texts, especially
    those who transcribed, copied, and edited manuscripts before mechanical printing
    technology was developed. </skos:definition>
  <skos:definition xml:lang="it">Una persona che copia testi; si riferisce
    in particolare a chi ha trascritto, copiato e curato un manoscritto prima
    dell'invenzione della stampa. </skos:definition>
</skos:Concept>
```

Fig. 8. Modelling of the concept "calligrapher" in SKOS standard.

Nevertheless, within a scenario of future feasibility, the true most valuable enhancement in terms of data association and integration lies on the mapping of semantic relations between concepts. The logic of the "Web of Data", consisting in expressing the relationships among data in order to define and describe the same data, mirrors the crucial features of SKOS data model. In this context, asserting which relationships exist between concepts of a single concept scheme and, more importantly, which links could be established between concepts of two or more different (but still semantically-related) schemes, constitutes the real key advantage for the different communities of Knowledge Organization Systems. Indeed, SKOS mapping properties would enable information retrieval tools to make use of a widely disseminated and heterogeneous web of Knowledge Organization Systems, causing concepts even supposedly modelled according to dissimilar principles and coming from different contexts to be automatically connected, compared and matched [17].

According to these terms, the step of specifying the semantic relations for each concept of the two local vocabularies of Phaidra, whether they be associative, hierarchical or equivalent links, gained a remarkable weight. By way of example, at the time of a search by type of material of the digital objects stored in Phaidra, the SKOS-formatted concept "photograph" shown below (see Fig. 9) not only would let the

Phaidra search engine to suggest the user all the available materially-related resources, i.e. all the digital objects labelled as "microfilm" or "negative" (as encoded in the tag `<skos:narrower>`). Additionally, the accessibility and visibility of the resources existing in the preservation system of the University of Padova would be increased by the potential interlinking of their metadata with external datasets in the Web, described according to concepts relatively similar or equivalent to the aforementioned local term "photograph". In the snippet represented in Fig. 9 see respectively `<skos:closeMatch>`, referencing to the concept "Fotografie" in *Nuovo soggettario* edited by the National Central Library of Florence, and `<skos:exactMatch>`, linking for example to the concept "Photographs" in *Art & Architecture Thesaurus® (AAT)*.

```
<skos:Concept rdf:about="http://phaidra.org/vocabularies/typeofmaterial/photograph">
  <skos:inScheme rdf:resource="http://phaidra.org/vocabularies/typeofmaterial"/>
  <skos:topConceptOf rdf:resource="http://phaidra.org/vocabularies/typeofmaterial"/>
  <skos:prefLabel xml:lang="en">photograph</skos:prefLabel>
  <skos:prefLabel xml:lang="it">fotografia</skos:prefLabel>
  <skos:definition xml:lang="en">Still image produced from radiation-sensitive materials
    (sensitive to light, electron beams, or nuclear radiation), generally by means of the
    chemical action of light on a sensitive film, paper, glass, or metal. A photograph may
    be positive or negative, opaque or transparent. The concept includes photographs made by
    digital means. </skos:definition>
  <skos:definition xml:lang="it">Immagine prodotta da materiale sensibile alla radiazione
    (sensibile alla luce, fasci di elettroni, o radiazione nucleare), generalmente attraverso
    l'azione chimica della luce su una pellicola sensibile, carta, vetro, o metallo.
    Una fotografia potrebbe essere positiva o negativa, opaca o trasparente.
    Questo concetto include anche fotografie realizzate in modo digitale. </skos:definition>
  <skos:narrower rdf:resource="http://phaidra.org/vocabularies/typeofmaterial/microfilm"/>
  <skos:narrower rdf:resource="http://phaidra.org/vocabularies/typeofmaterial/negative"/>
  <skos:exactMatch rdf:resource="http://vocab.getty.edu/aat/300046300"/>
  <skos:exactMatch rdf:resource="http://rdaregistry.info/termList/RDACarrierEU/1025"/>
  <skos:closeMatch rdf:resource="http://purl.org/bncf/tid/1578"/>
</skos:Concept>
```

Fig. 9. Modelling of the concept "photograph" in SKOS standard.

3.1 Future Outcome

Despite the limited extent of this case study, it is well-known that standards as RDF, SKOS and SPARQL, key means as ontologies, controlled vocabularies and authority files, and back-end architectures such as triplestores and reasoners, are comprehensively pivotal. They all would be equally demanded for a long-term digital archive, as is Phaidra, to be migrated to the "Web of Data" in an actual and full manner. An unparalleled data interoperability and a serendipitous networked knowledge discovery would be the granted benefits (see Fig. 10).

But at least, this is the promising outcome, and efforts are now set on this future perspective. The whole work which has been done so far on the theoretical mapping of UWmetadata schema to MODS was undertaken by both the Universities of Padova and Vienna in the light of a broader outlook, namely the announcement of a development and implementation of a new version of Fedora web-architecture, i.e. Fedora 4, which would have stored metadata in RDF standard. Mostly on the basis of recommendations and discussions guided by an international community of institutions engaged in transitioning their MODS-based digital repository systems to RDF [22], both the

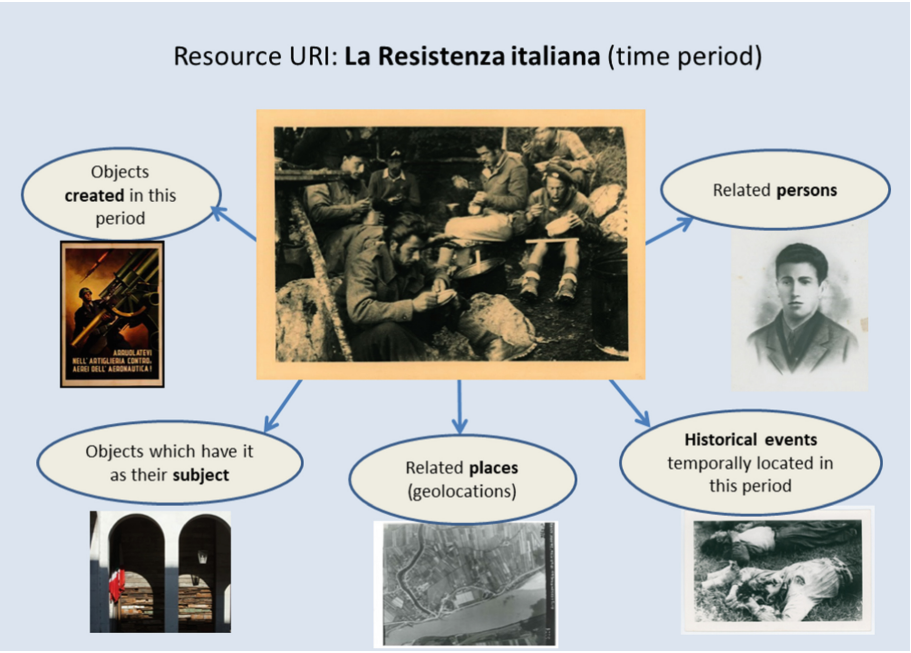

Fig. 10. Potential example of Semantic Web data in the digital repository of the University of Padova.

working groups of Phaidra are currently involved in outlining a new foundational data model represented in RDF triples. By the potential use of an arbitrary number of external vocabularies and ontologies, along with the possibility of RDF graphs to be extended with new nodes and new relationship types effortlessly, this new semantics-embedding model would be much more flexible, extendable to every future needs as well as highly interoperable.

In these terms, the forthcoming achievement of such a framework might thus exactly represent the most effective means to disclose the associative potential of metadata that was cited at the beginning of Sect. 3.

Acknowledgements. A pivotal thank you goes to the team of Phaidra, University of Padova – Lorisa Andreoli, Linda Cappellato, Yuri Carrer, Gianluca Drago, Giulio Turetta, and Antonella Zane – without whom the work illustrated in this paper would have never been existed.

References

1. https://phaidra.cab.unipd.it/
2. https://phaidra.univie.ac.at/
3. https://www.phaidra.org/community/phaidra-partners/

4. Andreoli, L., Carrer, Y., Drago, G., Turetta, G., Zane, A.: The user in focus: an inclusive approach to the presentation of digital collections of GLAM institutions. Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare **71**(1), 125–135 (2018)
5. Andreoli, L., Fornasiero, M., Menegazzi, A., Talas, S.: Defining University museums' objects for the web. In: "Turning Inside Out European University Heritage: Collections, Audiences, Stakeholders". Proceedings of the 16th Universeum, Athens, 11–13 June 2015 (2015)
6. https://phaidra.cab.unipd.it/info/impressum
7. https://github.com/phaidra/phaidra-api/wiki/Documentation
8. https://fc.cab.unipd.it/oaiprovider/
9. Turetta, G.: Esperienze e modelli di dati: il caso Phaidra. In: Bettella, C., De Robbio, A., Duzzin, M., Zane, A. (eds.) Corso di formazione La biblioteca dei dati (2017, unpublished)
10. Bettella, C., Andreoli, L., Cappellato, L., Carrer, Y., Drago, G., Turetta, G.: PHAIDRA_DC Metadata Element Set Version 1.1 (2018). http://phaidra.cab.unipd.it/static/phaidra_dc-metadata-element-set.pdf
11. Bellotto, A., Bettella, C.: PHAIDRA_MODS Metadata Element Set Version 1.0 (2018, unpublished draft)
12. Budroni, P., Höckner, M.: PHAIDRA, a repository project of the University of Vienna. In: iPRES 2010, 7th International Conference on Preservation of Digital Objects, Vienna (2010)
13. Budin, G., et al.: LOM Univie Schema, April 2006
14. Friesen, N., Fisher, S., Roberts, A.: CanCore Guidelines for the Implementation of Learning Object Metadata (IEEE 1484.12.1-2002), Version 2.0 (2004). http://cancore.athabascau.ca/en/guidelines.html
15. Powell, J., Hopkins, M.: A Librarian's Guide to Graphs, Data and the Semantic Web, p. 130. Chandos Publishing, Waltham (2015)
16. Kopácsi, S., Hudak, R., Ganguly, R.: Implementation of a classification server to support metadata organization for long term preservation systems. In: Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare Bd. 70, Nr. 2 (2017)
17. World Wide Web Consortium (W3C): SKOS Simple Knowledge Organization System Primer. Working Group Note (2009)
18. World Wide Web Consortium (W3C): SKOS Simple Knowledge Organization System Reference. W3C Recommendation (2009)
19. Skosmos Homepage. http://vocab.phaidra.org/skosmos/en/?clang=en
20. Suominen, O.: Publishing SKOS concept schemes with Skosmos. AIMS Webinar 6th April 2016, Slide 25 (2016)
21. Cappellato, L.: Linee guida per la compilazione dei metadati (2018, revised edition, forthcoming)
22. Samvera MODS to RDF Working Group: MODS to RDF Mapping Recommendations. v.0.1 - Draft for Review & Comment (2018). https://docs.google.com/document/d/1FZI8KJiW4qSKYUUKe0mAwqlx0ppVRFyPtsfLDqQE5T8/edit