Enhancing semantic expressivity in the cultural heritage domain: exposing the Zeri Photo Archive as Linked Open Data

Marilena Daquino, marilena.daquino2@unibo.it - Department of Classical Philology and Italian Studies, University of Bologna, Italy
Francesca Mambelli, francesca.mambelli6@unibo.it - Federico Zeri Foundation, University of Bologna, Bologna, Italy
Silvio Peroni, silvio.peroni@unibo.it - Digital And Semantic Publishing Laboratory, Department of Computer Science and Engineering, University of Bologna, Bologna, Italy
Francesca Tomasi, francesca.tomasi@unibo.it - Department of Classical Philology and Italian Studies, University of Bologna, Italy
Fabio Vitali, fabio.vitali@unibo.it - Digital And Semantic Publishing Laboratory, Department of Computer Science and Engineering, University of Bologna, Bologna, Italy

The description of cultural heritage objects in the Linked Open Data (LOD) perspective is not a trivial task. The process often requires not only to choose pertinent ontologies, but also to develop new models that preserve the maximum of information and express the semantic power of cultural heritage data. Indeed, data managed in archives, libraries and museums are complex objects themselves, that need a deep reflection on even non conventional conceptual models. Starting from these considerations, this paper describes a research project: to expose as Linked Open Data the vastness of one of the most important collections of European cultural heritage, the Zeri Photo Archive. We describe here the steps we undertook to this end: we developed two *ad hoc* ontologies for describing all the issues not completely covered by existent models (the *F Entry* and the *OA Entry Ontology*); we mapped into RDF the descriptive elements used in the current Zeri Photo Archive catalog into RDF, converting into CIDOC-CRM, SPAR Ontologies and the two new aforementioned models the source data based on the Italian content standards Scheda F (Photography Entry, in English) and Scheda OA (Work of art Entry, in English). Finally, we created an RDF dataset of the output of the mapping that could show a result capable to demonstrate the complexity of our scenario.

RASH: https://w3id.org/people/essepuntato/papers/zeri-and-lode-jocch.html

Categories and Subject Descriptors: Information systems [**Information systems applications**] Digital libraries and archives; Information systems [**World Wide Web**]: Web data description languages—*Web Ontology Language (OWL)*; Computing methodologies [**Artificial intelligence**]: Knowledge representation and reasoning—*Ontology engineering*; Applied computing [**Arts and humanities**] ; Applied computing [**Computers in other domains**] Digital libraries and archives

Additional Key Words and Phrases: Scheda F, Scheda OA, mapping CIDOC-CRM

ACM Reference Format:

Marilena Daquino, Francesca Mambelli, Silvio Peroni, Francesca Tomasi, Fabio Vitali. Enhancing semantic expressivity in the cultural heritage domain: exposing the Zeri Photo Archive as Linked Open Data *jn* 1, 1, Article 1 (January 2XXX), 22 pages.
DOI: http://dx.doi.org/10.1145/0000000.0000000

1. INTRODUCTION

Libraries, archives and museums are facing a substantial transformation in managing cultural heritage data. In particular, the attempt to express the latent semantic power of data – by creating relations and interconnections between different kind of cultural objects owned by similar cultural institutions – is considered a shared approach in the community. The idea to define an infrastructure for sharing metadata about such objects is needed so as to create comprehensive tools for researchers, to provide innovative environments for users' learning, and, finally, to increase the impact of culture in society. Cultural heritage institutions are dealing with two urgent issues: on the one hand, they need to provide a complete and exhaustive semantic

description of their data; on the other hand, they have to open up their data to interchange, interconnection and enrichment.

The urgent need for such scenarios has been addressed in the context of a particular project, *PHAROS: An International Consortium of Photo Archives* [Inge Reist et al. 2015]. PHAROS is one of the first concrete steps towards the creation of an actual digital infrastructure of the notable photographic archives of works of art in Europe and the United States of America. The Consortium enables the active collaboration between the institutions responsible for fourteen photo archives so as to create a common platform for research on images and metadata of Western and non-Western works of art in all media. One of the most important institutions included in the Consortium is the Zeri Foundation (http://www.fondazionezeri.unibo.it/en). Its Photo Archive is planned to be one of the first assets to be included in PHAROS.

In this article we introduce a project called*Zeri and LODE*, which addresses the enhancement of existing metadata collections of the Zeri Photo Archive by means of Semantic Web technologies, in order to make them easily discoverable, interchangeable, and interlinked with other existing and relevant sources.

Our sources were the collections owned by one of the most important art historian of the XX century: Federico Zeri (1921-1998). His extensive library of art books, auction catalogs and individual photos of monuments and artworks was his main instrument of work. He began to collect them in the 1940s and, over time, had it become "the world's largest private archive of Italian paintings", an essential reference work for the historical sequencing of out-of-context works. At the time of his death, the archive included more than 46.000 volumes, 37.000 auction catalogs, 60 periodicals and 290.000 individual photographs. In order to preserve his bequest and put it to best use, the Zeri Foundation was set up in his name at the University of Bologna, and it has come to be recognized as one of the most important research and training centres for art historians in the world.

Among the activities of its first years, the Zeri Foundation undertook the cataloguing of Zeri's repository, employing to that end two Italian metadata content standards, *Scheda F*, for *Scheda di fotografia* (*photograph*, available at http://www.iccd.beniculturali.it/index.php?it/473/standard-catalografici/Standard/10) and *Scheda OA*, for *Scheda Opera d'Arte* (*work of art*, available at http://www.iccd.beniculturali.it/index.php?it/473/standard-catalografici/Standard/29), which will be respectively called *F Entry* and *OA Entry* in the following, both issued by the *ICCD* (*Istituto Centrale per il Catalogo e la Documentazione*, Central Institute for the Cataloguing and Documentation) of the Italian Ministry of Cultural Heritage. As a side note, the ICCD coordinates the research in the definition of cataloging standards for all kinds of cultural objects belonging to the archaeological, architectural, artistic and ethno-anthropological areas of interest of the Ministry.

The *Scheda F* includes more than 300 fields recording information about photographers, history of photo production, publication, preservation, changes of location, exhibitions, materials and techniques, reference bibliography, attached documentation and technical data derived from the philological analysis of the item. In the context of the Zeri Photo Archive, the *Scheda OA* refers to the content of the photography, i.e. the work of art. In particular, it includes more than 200 fields that offer a detailed description of the object depicted and provides similar information to the ones described in the *Scheda F*.

The effort in cataloguing Zeri's collection compliantly with these two Italian standards has resulted in the Zeri Photo Archive catalog, which is stored in a relational database and accessible through a web interface (http://catalogo.fondazionezeri.unibo.it).

Because of the national and international relevance of this photo archive, the complexity recorded in its heterogeneous catalog, and the explicit request from the PHAROS Consortium, we have started to work on these data as a use case for defining a common model between institutions that could use different descriptive standards for photographs. In particular, our main goal was the development of appropriate Semantic Web models and technologies so as to provide a representation of the Zeri Photo Archive as Linked Open Data (LOD) .

The PHAROS Consortium suggested the CIDOC-CRM conceptual reference model [Patrick Le Boeuf et al. 2015] as required representational framework for data to be shared within the Consortium, given its explicit characterization for cultural heritage entities and its widely recognized international standing. CIDOC-CRM has been widely adopted in Italy in the past years and it was considered by PHAROS members a good starting point for addressing dialogue between data sources – even if no photo archive had been described by means of it yet. Thus, the original plan was to map the Scheda F and Scheda OA fields with CIDOC-CRM properties, so as to represent the metadata of the Zeri Photo Archive into machine-readable form according to the required standard.

Unfortunately the full mapping was not possible, since many properties of photos and works of art described by Scheda F and Scheda OA, that are necessary for the full and complete description of Zeri's catalog, do not have a natural representation in CIDOC-CRM. For instance, a clear representation of subjective attributions (e.g. authorship attributions, objects appellation, influence between works, and dating) should be handled appropriately by means of dedicated entities, and should also consider the provision of data integration among stakeholders – who may catalog the same object providing contradictory information. Moreover, the network of people and organizations involved in the lifecycle of such cultural objects with different roles required to be represented precisely, together with an explicit representation of the influence a cultural object may have on the conception of another one (e.g. a copy or a derivative work). While all these data are crucial for enabling further relations to otherwise not discoverable information, they cannot be represented in CIDOC-CRM without ambiguities.

So as to address all the aforementioned issues, we focussed on the creation of:

(1) two ontologies, i.e. the F Entry Ontology and the OA Entry Ontology, to represent all missing information from CIDOC-CRM;

(2) two documents exemplifying the mappings and alignments from Scheda F and Scheda OA to CIDOC-CRM and our ontologies, so as to provide a complete representation of the description requirements for photographs and works of art;

(3) an RDF dataset, published as Linked Open Data, describing a huge amount of the available F/OA entries in the Zeri's catalog according to the aforementioned ontologies.

Even if the new entities introduced in the ontologies we have developed mainly comes from terms of two specific national standards (i.e. Scheda F and Scheda OA), entities and properties described in the ontologies are of widespread application and are, to our knowledge, the vastest and most complete model for describing photographs, works of art, and related documentation. Archives planning to describe collections of photos in CIDOC-CRM could immediately find our experience in working with Zeri's catalog crucial for their own representation purposes.

This paper extends and details works already introduced in [Ciro Mattia Gonano et al. 2014]. In addition to what described in that paper, we provide a thorough revision of the F Entry Ontology triggered by the OA Entry Ontology development. This includes the addition of topics that had not been tackled previously: a full description of works of art depicted in catalogued photographs, the development of aspects that were deemed secondary when first developing the F Entry in isolation, and the uniformity of choices for RDF representation of both work types (photos and works of art). Then we proceeded with the mapping of terms belonging to Scheda F and Scheda OA used in Zeri Photo Archive into RDF, according to aforementioned ontologies. Finally, we realized the Linked Open Data publication of the first subset of the Zeri catalog, including photos of artworks of 15^{th}-16^{th} centuries. All the steps of the process, from the mapping to the ontologies development and the LOD publication, have been shared with the ICCD and are still under discussion in order to define a unique general model that could be proposed to all the Italian institutions that have to manage these kind of data.

The rest of the paper is organised as follows. In Section 2 we introduce some related works on these topics. In Section 3 we present the methodology we adopted for the development of the ontologies, which are introduced in Section 4. In Section 5 we discuss the alignments among the two content standards (i.e., Scheda F and Scheda OA) into RDF. In Section 6 we describe the LOD dataset we have created that includes all the RDF data obtained from converting the Zeri Photo Catalog. Finally, in Section 7 we conclude the paper sketching out some future works.

2. RELATED WORKS

As said in the introduction, the digital cultural heritage scenario is facing a strong transformation. In particular, publishing strategies of cultural heritage data are changing the traditional methodologies. Cleaning, reconciliation, enrichment and linking [Seth Van Hooland and Ruben Verborgh 2014] are the new keywords in the domain of libraries, archives and museums. The Semantic Web, and the Linked Open Data theory and practice, have determined a revolution in data production and access, and both machines and final users are experimenting new ways for acquiring knowledge from URIs and typed RDF links (http://linkeddata.org).

Many projects, e.g. LODLAM (http://lodlam.net) and OpenGlam (http://openglam.org), demonstrate how much the community is growing around these themes (for an updated survey see the Task force activity on LOD: https://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData). Datasets are increasing in number (see the datahub.io classification) and the LOD cloud is larger day after day. Together with the publication of LOD collections, new vocabularies are modelled (see LOV, Linked Open Vocabularies, http://lov.okfn.org/dataset/lov/), in order to address the needs of the cultural heritage domain. The scientific literature about methods in producing LOD shows the directions of the research around these themes: authorities reconciliation and linking, but also ontologies reuse and development, are the basis for realizing the idea of LOD.

The need to convert flat traditional element schemes, such as the above-mentioned Scheda F and Scheda OA, into ontology terms forced us to develop new models as a result of a reflection on existing conceptualizations and their shortcomings. The adoption of models that were widely appreciated in the specific domain of cultural heritage is the first obvious step, but we considered also other models already capable to address additional important issues.

As described in the next sections, many ontologies represent important facets in our ontology development process. Each chosen ontology is used to cover one aspect of our heterogeneous scenario and were therefore first reused, but then also extended. The SPAR ontologies (http://www.sparontologies.net), mostly used in academic literature, were necessary in order to describe the documentation used in the cataloguing process. But, by describing even the catalog entries using SPAR, we aimed at making them available also as bibliographic sources. In particular, FaBiO [Silvio Peroni and David Shotton 2012] was chosen for managing bibliographies according to the FRBR approach [International Federation of Library Associations and Institutions 2009]; CiTO [Silvio Peroni and David Shotton 2012] was reused for describing the different citations that cataloguers provide to support their attributions, and PRO [Silvio Peroni et al. 2012] was fundamental for documenting people's role in photographic, arts, publishing and cataloguing domains. HiCO [Marilena Daquino and Francesca Tomasi 2015], was created by extending PROV-O [Timothy Lebo et al. 2013], in order to describe the interpretation process relative to subjective attributions, a fundamental aspect of our conceptualization. And of course CIDOC-CRM [Patrick Le Boeuf et al. 2015] was finally used in order to describe all the features directly related to catalogued objects, so as to guarantee semantic interoperability between stakeholders.

The need of ontological interconnection represents a trend: as Europeana demonstrates, in particular within the EDM (Europeana Data Model) [Europeana 2016], the variety of existing vocabularies in the cultural heritage domain brings the need of reusing and aligning models [Martin Doerr 2010]. The mapping to CIDOC-CRM forced on us an analysis of methodologies in merging and mapping, as technical processes,

theoretical activities and common, shared, approaches also in the cultural heritage domain [Achille Felicetti et al. 2013].

In addition to the ontologies, we made use of authorities, controlled vocabularies and other datasets (e.g. VIAF, DBpedia, Getty AAT and ULAN, Wikidata, Geonames), in order to define appropriate links for both mutual connections and alignment (a list in [Antoine Isaac et al. 2011]).

Other existing datasets in the artworks domain were finally analysed (e.g. Smithsonian Art Museum, http://americanart.si.edu/collections/search/lod/about/; Yale Center for British Art, http://britishart.yale.edu/collections/using-collections/technology/linked-open-data).

3. METHODOLOGY AND ONTOLOGY DEVELOPMENT

Modelling complex information provided by Scheda F/OA content standards required to manage different knowledge domains, i.e. bibliographic, archival, photography, arts and cultural heritage domains. Thus, our priority was to enable the coexistence of heterogeneous and existing models, already available and developed for describing one of the aforementioned domains, as big modular consistent ontologies for describing all the entities involved in the Zeri's Catalog.

Usually, cultural institutions adopt the most suitable vocabulary for their purpose, without taking care about the way similar concepts are actually modelled in other relevant domains, leaving the possibility of aligning them as possible future work. The same object may be described in different ways according to the data creator's cultural background. For example, Arts and Library Sciences conceive and apply differently the Functional Requirements for Bibliographic Records (FRBR) conceptual model [Murtha Baca 2006] [Murtha Baca and Sherman Clarke 2007] [Karen Coyle 2016] [Richard P. Smiraglia et al. 2013] into their object description. Cataloguing artworks generally entails aggregating under the same label (i.e. type of work) functional, formal, and morphological features of the object, by considering these issues inseparable from the conception of the work itself, i.e. the FRBR Work – which is a peculiarity of the arts domain. In the bibliographic domain instead, such definition may be formalized by considering such features at other levels than the Work, e.g. the Expression and the Manifestation levels. Thus the RDF representation of a painting may significantly vary when defined in one of the two aforementioned domains, which will affect the way how data can be integrated.

In addition to pure theoretical and designing consideration, the way different data providers label their data should be taken into account as well. For instance, cataloguers use specific and, often, customised criteria for recording authorship attributions, e.g. by means of their own classification system, that are not immediately sharable to other stakeholders without a reconciliation into a controlled vocabulary. Thus, one of the main aspect we had to address was to provide flexible mechanisms [Michele Pasin and John Bradley 2015] to record the particular criteria adopted by a cataloguer for cataloguing a work, still allowing their evolution over time if new catalog guidelines would be proposed and substituted to the old one, without any loss of precious information. In addition, keeping track the multiple and even inconsistent interpretations in cataloguing was something that had to be handled properly as well.

In order to provide a model to describe the aforementioned cases, we also wanted to rely on existing standards and models, instead of reinventing everything from scratch. The use of FRBR-compliant ontologies (i.e., SPAR Ontologies) for modelling the bibliographic and documental aspects of our domain, the HiCo Ontology to provide attributions of provenance information, and CIDOC-CRM, i.e. the international standard for describing cultural artefacts, appeared to be quite useful for providing a first, even if still partial, description of works of art and photographs and their related data.

We initially tried to figure it out if an alignment of Scheda F/OA elements to entities of CIDOC-CRM model could be possible. Usually, the alignment to existing models is not a straightforward operation and could require several iterations and integrations to be prepared properly. We conducted a brief analysis of the

CIDOC-CRM data model, which revealed that several concepts we needed in order to describe our domain in a proper way were not available in such reference model.

Therefore, our development first proceeded in defining terms and relations so as to accomplish the full description of our domain, and then we refactored and aligned them into CIDOC-CRM. We have also developed two OWL 2 DL ontologies [Boris Motik et al. 2012], released with Creative Commons licenses to foster broad reusability, for expressing concepts and relations of the main descriptive areas of the Italian content standards Scheda F/OA in RDF that were lacking in CIDOC-CRM. The approach we followed was mainly data-centric. In fact, such ontologies have to properly describe all the data represented in a real and representative dataset compliant with F/OA standards, i.e. the Zeri Photo Archive catalog. Thus, the use of such actual dataset allowed us to develop ontologies where their entities have meaningful names, and to provide real examples of their usage at every stage of the development process.

The methodology we adopted was the Simplified Agile Methodology for Ontology Development (*SAMOD*) [Silvio Peroni 2016]. SAMOD is an agile methodology, developed starting from guidelines proposed in well-known and existing ontology development methodologies such as [Mariano Fernández et al. 1997] [Mike Uschold 1995], which allowed us to develop the model by means of several small and iterative steps, and to create documentation by using exemplars of data. In particular, this methodology required us to consider small issues iteratively, and provided us a way to test our under-development model immediately on a real use case in any phase of the development.

Good practices in ontology development were respected to ensure semantic interoperability and facilitate the model reuse: concepts and relations defined were refactored, directly reusing most appropriate external entities from established and well known ontologies, or aligning them to the ones of other models.

Documentation of both the F Entry and OA Entry ontology development processes is available online, respectively at http://www.essepuntato.it/2014/03/fentry/samod and http://oaentry-ontology.sourceforge.net/samod/OAdevelopment.zip.

In the next section we provide a complete description of covered aspects and their formalization as ontological entities, accompanied by exemplar of their usage.

4. THE F ENTRY ONTOLOGY AND THE OA ENTRY ONTOLOGY

As above said, the main objective of our project is to expose Federico Zeri Foundation data to the world according to the standard CIDOC-CRM, the model chosen as *lingua franca* for sharing data between stakeholders. However, CIDOC-CRM lacks some peculiar entities of photography and arts domains, as well as several relations that in our opinion would truly make possible an integration between museums, and between museums and other domains.

Therefore, we extended the CIDOC-CRM model by creating two complementary ontologies, the F Entry Ontology and the OA Entry Ontology, that are described in the following subsections and that allow one to address the modelling scenarios that are not addressable properly by CIDOC-CRM. Thus, the actual mapping between CIDOC-CRM and the F/OA content standards is described in Section 5.

The first developed ontology, by using the above described methodology, is the F Entry Ontology, which deals with issues related to photography domain and several aspects that are in common with the OA Entry content standard for describing works of art. The latter ontology, the OA Entry Ontology, is thus very similar to the former one, and basically provides additional entities and relations proper of a comprehensive representation of the arts domain. The development of the OA Entry Ontology also resulted in a revision of the F Entry Ontology, so as to provide a specular description of same issues when appropriate. For the sake of brevity, common aspects of the two ontologies are detailed once in Section 4.1.

4.1 The F Entry Ontology (FEO)

The current version of F Entry Ontology (FEO) revises the previous version introduced in [Ciro Mattia Gonano et al. 2014], and it is available online at http://www.essepuntato.it/2014/03/fentry. The Graffoo diagram [Riccardo Falco et al. 2014] of FEO in Fig. 1 provides an overview of its classes – i.e., the yellow rectangles –, object properties – i.e., the blue dotted lines beginning with a solid circle and ending with a solid arrow –, and assertions among classes – i.e., the black lines ending with a solid arrow (the full specification of Graffoo graphical elements is available at http://www.essepuntato.it/graffoo/specification/current.html).

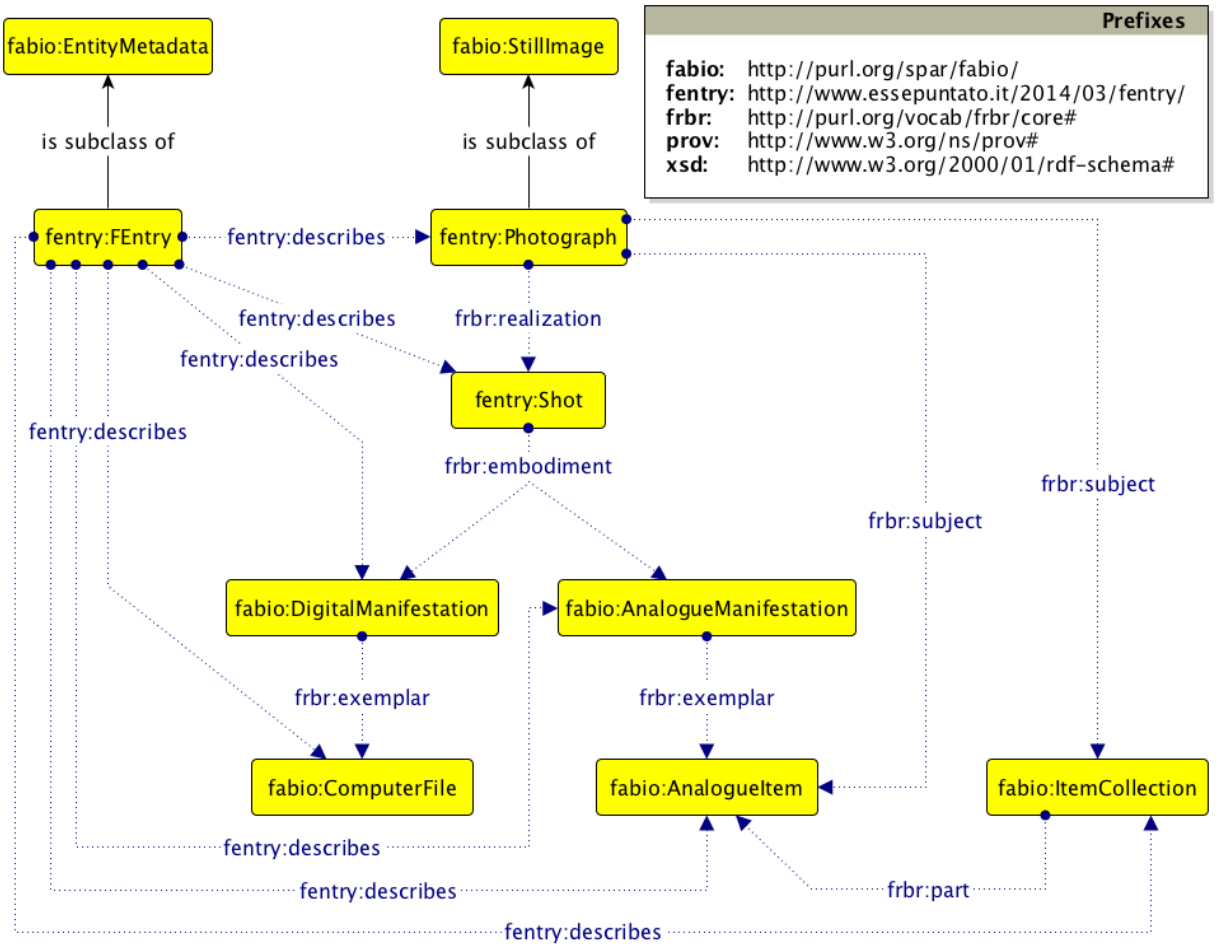

Fig. 1. The Graffoo diagram of the F Entry Ontology main classes and properties.

FEO introduces classes and properties needed to characterize three main concepts: the photograph, the subject portrayed in the photograph and the F Entry describing the photograph and its subjects. Each of the aforementioned entities is characterised in terms of FRBR. In particular:

— the photograph is represented as a FRBR Work when describing its essence, as a FRBR Expression when dealing with information about its realization (i.e. the shot), as a FRBR Manifestation when describing each tangible form of the photograph, and as FRBR Item for each individual copy with different features;

— in the Zeri Photo Archive the portrayed subject of a photograph is always a concrete work of art. Therefore, it can be represented as a FRBR Work for defining its essence, and as FRBR Item for describing the concrete object. The entities representing this sort of subject have been refined when the OA Entry Ontology was created, explained in details in Section 4.2. When the subject is not identified as a concrete object, a place or an agent, it can be simply considered as a term from an open/controlled vocabulary or thesaurus;

— the F Entry is a work containing metadata about a photograph and its cataloguing; it is subject of several revisions, each of them related to responsible entities (i.e. cataloguers and supervisors); we are not interested in how it is preserved, in which formats and in how many copies. Thus a F Entry is represented as a FRBR Work when describing its creation and as a FRBR Expression when describing its contents and revisions.

Existing ontologies have been imported in FEO so as to provide a precise description of specific aspects of the domain in consideration. In particular, we imported the FRBR-aligned Bibliographic Ontology (FaBiO, prefix *fabio*) [Silvio Peroni and David Shotton 2012], the Publishing Roles Ontology (PRO, prefix *pro*) [Silvio Peroni et al. 2012], the Historical Context Ontology (HiCO, prefix *hico*) [Marilena Daquino and Francesca Tomasi 2015], and the Citation Typing Ontology (CiTO, prefix *cito*) [Silvio Peroni and David Shotton 2012]. In addition to the terms from these ontologies, terms from an OWL 2 DL version of FRBR (prefix *frbr*, http://purl.org/spar/frbr, imported by FaBiO) are used, so as to represent hierarchical and associative relations between the main entities, as well as terms defined in the Provenance Ontology (PROV-O, prefix *prov*) [Timothy Lebo et al. 2013].

In the next sections we describe, with more details, how we use and extend the imported ontologies – i.e., FaBiO, PRO, CiTO, HiCO – for describing the different aspects of the domain. The complete RDF example of the excerpts provided in the following sub-sections is available at http://dx.doi.org/10.6084/m9.figshare.3175252.

4.1.1 *Extending FaBiO for defining the cultural object .* FaBiO has been originally developed for describing bibliographic entities according to FRBR conceptual model. It mainly addresses issues related to published texts, by introducing wide taxonomies of possible kind of works, expressions, manifestation and items, also defining them by means of several object properties domain and range constrains. Therefore it was a good starting point that we have refined for modelling our main entities (i.e. F/OA entries, photographs, and works of art), so as to enable a deeper characterization of them according to well-known and used models.

An F Entry can describe a photograph in each phase of its lifecycle: from its creation (i.e. the essence of the photograph) to its realization (i.e. the shot), from its development into a visible image (i.e. negative, positive, slide, digital image) to its publishing and reproduction. Each phase of the lifecycle of the photograph corresponds to an instance of a class defined in terms of FRBR as defined in FaBiO.

An F Entry is a document containing metadata about a photograph and its cataloguing. Therefore it is defined in terms of FRBR Work (i.e. an instance of the class *fentry:FEntry*, subclass of *fabio:EntityMetadata*). Subjects of the entry are considered as FRBR Works as well: the photograph (i.e. an instance of the class *fentry:Photograph*, subclass of *fabio:StillImage*) and, in the Zeri Photo Archive context, the work of art portrayed in the photograph (i.e. an instance of the class *fabio:ArtisticWork*).

The object property *fentry:describes* links an instance of the class *fentry:FEntry* to instances of the proper class defining its subject, among which there are classes *fentry:Photograph* and *fabio:ArtisticWork*. For ex-

ample, the natural language scenario "a F Entry describes the photograph portraying the *Jesus's baptism* painting", may be expressed, in Turtle syntax, as follows:

```
:fentry-72486 a fentry:FEntry ;
  fentry:describes
    :jesus-baptism-photo-work , :jesus-baptism-photo-item ,
    :jesus-baptism-work , :jesus-baptism-item .
:jesus-baptism-photo-work a fentry:Photograph .
:jesus-baptism-photo-item a fabio:AnalogItem .
:jesus-baptism-work a fabio:ArtisticWork .
:jesus-baptism-item a fabio:AnalogItem .
```

Thus, the shot (i.e. an instance of the class *fentry:Shot*, subclass of *fabio:Expression*) is a realization of a photograph, which can take several forms when developed, principally distinguished by means of their analog/digital formats (i.e. instances of the classes *fabio:AnalogManifestation* or *fabio:DigitalManifestation*) and be individually described (i.e. an instance of the classes *fabio:AnalogItem* or *fabio:DigitalItem*).

For example, the natural language scenario "The shot of the photograph portraying the *Jesus's baptism* painting that had been taken by Brogi before 1940 was published by himself in 1940", can be expressed, in Turtle syntax, as follows:

```
:jesus-baptism-photo-work
  frbr:realization :jesus-baptism-photo-shot .
:jesus-baptism-photo-shot a fentry:Shot ;
  frbr:embodiment :jesus-baptism-photo-positive .
:jesus-baptism-photo-positive a fabio:AnalogManifestation .
```

4.1.2 *Using PRO for describing the lifecycle of the object.* PRO allows one to describe scenarios where agents hold roles with respect to a particular time and context, e.g. the fact that a person has been (property *pro:holdsRoleInTime*) the owner (the role in consideration, specified through the property *pro:withRole*) of a certain photograph (the context, linked with the property *pro:relatesTo*) from the 27th of November 2002 to the 16th of June 2015 (the time interval, specified by means of the property *tvc:atTime*). An instance of the class *pro:RoleInTime* is created every time we need to specify such kinds of roles.

For instance, in our context, the role of Brogi as photographer is held on the shot, i.e. a FRBR Expression, and the publisher role is held on the positive of the photograph, i.e. a FRBR Manifestation. These relations can be represented as follows:

```
:brogi a foaf:Agent ;
  pro:holdsRoleInTime :brogi-photographer-jesus-baptism-photo-shot ;
  pro:holdsRoleInTime :brogi-publisher-jesus-baptism-photo-positive .
:brogi-photographer-jesus-baptism-photo-work a pro:RoleInTime ;
  pro:withRole scoro:photographer ;
  pro:relatesTo :jesus-baptism-photo-shot ;
  tvc:atTime :jesus-baptism-photo-shot-date .
:brogi-publisher-jesus-baptism-photo-positive a pro:RoleInTime ;
  pro:withRole pro:publisher ;
  pro:relatesTo :jesus-baptism-photo-positive ;
  tvc:atTime :jesus-baptism-photo-publishing-date .
```

A particular situation is when the creator and the realizer of the shot may not be the same person. Therefore the main photographer is described by means of the provided CIDOC-CRM terms for representing a creation of a work (explained in next Section 5) and the terms belonging to PRO (or to ScoRO, http://purl.org/spar/scoro, which extends PRO with additional roles) for describing other roles than the *creator* role.

4.1.3 *Using HiCO for providing provenance of assertions.* HiCO has been developed for describing the interpretative process underlying questionable information, by means of a provenance statement characterizing the entity that may be contradicted by another data provider. Each questionable information is, thus, defined by an individual of the class *hico:InterpretationAct*, which allows one to specify the scope (property

hico:hasInterpretationType) for the current interpretation, and the criteria (*hico:hasInterpretationCriterion*) underlying such interpretative choice.

In addition, each instance of *hico:InterpretationAct* is also linked (property *hico:isExtractedFrom*) to the text source where such questionable information is stated in natural language. In our domain, such source is always the content of the F Entry, subclass of *fabio:MetadataDocument* – i.e., a FRBR Expression. Thus, it is worth mentioning that the creator (property *prov:wasAssociatedWith*, defined in PROV-O [Timothy Lebo et al. 2013]) of the RDF statements describing an interpretation act and the author of the original text they were extracted from are not necessarily the same person. Finally, the relation between the agent holding a certain role (an individual of the class *pro:RoleInTime*, described in the previous sub-section) and such RDF-defined interpretation act is introduced by means of the object property *prov:wasGeneratedBy*.

For instance, consider the following natural language text derived from an existing F Entry and describing some information about the photograph portraying the *Jesus's baptism* painting by Leonardo da Vinci:

> the attribution of Brogi as the publisher of the photograph portraying the *Jesus's baptism* painting, was motivated by a formal analysis of the photograph itself, which revealed on its verso an inscription naming Brogi as publisher.

Such natural language scenario can be expressed in RDF, by using the ontological entities introduced above, as follows (in Turtle syntax):

```
:brogi-publisher-jesus-baptism-photo-positive
  prov:wasGeneratedBy :jesus-baptism-photo-publisher-attribution .
:jesus-baptism-photo-publisher-attribution a hico:InterpretationAct ;
  hico:hasInterpretationType :role-attribution ;
  hico:hasInterpretationType :zeri-preferred-attribution ;
  hico:hasInterpretationCriterion :formal-analysis ;
  hico:hasInterpretationCriterion :inscription ;
  hico:isExtractedFrom :fentry-72486-expression ;
  prov:wasAssociatedWith :crr-mm .
:role-attribution a hico:InterpretationType .
:zeri-preferred-attribution a hico:InterpretationType .
:formal-analysis a hico:InterpretationCriterion .
:inscription a hico:InterpretationCriterion .
:fentry-72486-expression a fabio:MetadataDocument .
:crr-mm a foaf:Agent .
```

In the excerpt, the instance `:zeri-preferred-attribution` is provided in order to further distinguish the current interpretation as the one chosen by the cataloguing institution – so as to distinguish it from other possible alternative interpretations specified.

4.1.4 *Using CiTO for relating documents and attributions.* The relation between an interpretation and a heterogeneous source supporting a certain cataloguer's interpretation can be defined as proper (even implicit) citation. CiTO allows one to mark citation links between authors, and to specify the intent of such citation by providing a wide set of object properties relating citing and cited entities. To this end, we can use the object properties provided in CiTO for linking an individual of the class *hico:InterpretationAct* to the original textual interpretation from which such interpretation act has been derived.

For instance, the fact the cataloguer is citing as evidence an inscription recognized on the verso of the photograph, can be represented by means of the object property *cito:citesAsEvidence* as follows:

```
:jesus-baptism-photo-publisher-attribution
  cito:citesAsEvidence :jesus-baptism-photo-verso .
```

4.2 The OA Entry Ontology

The Scheda F content standard provides just few elements regarding the work of art that may be portrayed in a photograph. However, there exists another content standard, i.e. the Scheda OA, that aims at giving an

exhaustive reference document for the complete description of any work of art – not only those portrayed in photographs. Therefore, all the peculiar aspects of the work of art portrayed in a photograph have been modelled in the OA Entry Ontology (available at http://purl.org/emmedi/oaentry). The Graffoo diagram in Fig. 2 provides an overview of its main classes and properties.

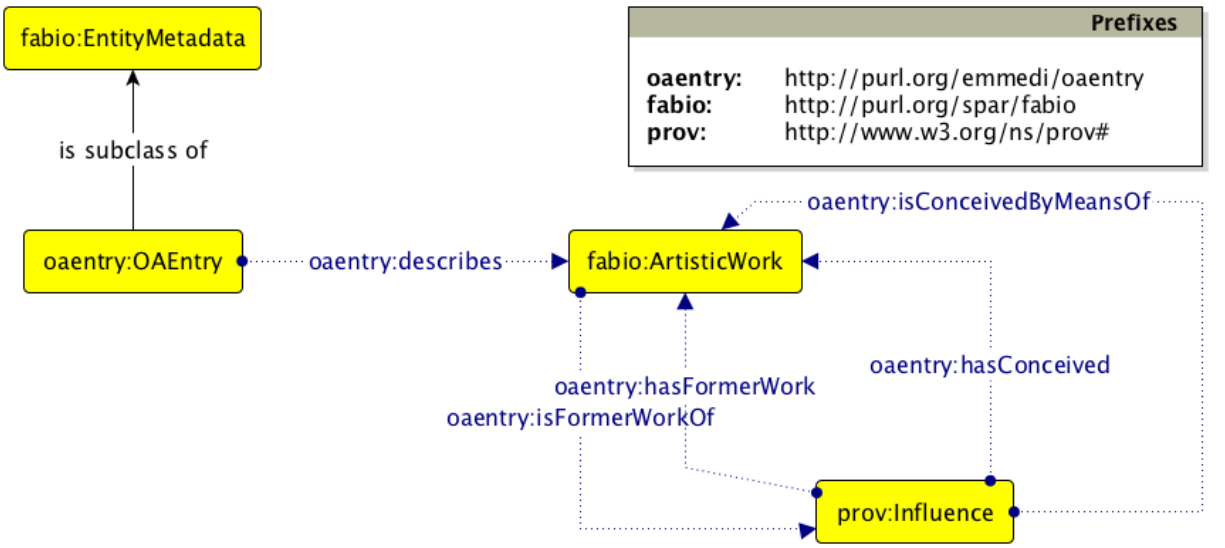

Fig. 2. The Graffoo diagram of the OA Entry Ontology main classes and properties.

The OA Entry Ontology introduces three main concepts: the work of art, the OA Entry that contains metadata about the work of art, and the influence between works of art. In particular:

— a work of art that can be described in different phases of its lifecycle (creation, restoration, location, ownership, custody etc.). Thus, we decided to represent a work of art as a FRBR Work to describe its essence, as a FRBR Manifestation to provide information about its physical features that may vary over time, and as a FRBR Item when dealing with information about legal aspects and its location. Modelling works of art in these terms required to take into account established considerations in the arts domain [Murtha Baca 2006] [Murtha Baca and Sherman Clarke 2007];

— the OA entry is a document containing metadata about a work of art and its cataloguing: in the Zeri Photo Archive an OA entry is always created when a F entry describes a work of art portrayed in a photograph. This means that several metadata of such entries are shared, e.g., same authors of the entries, updating and described entities. In our ontology, an OA entry is defined as a FRBR Work when describing its creation and with an FRBR Expression when addressing issues related to its content;

— the influence between two works of art concerns the way one affects (the conception of) the other. Such influence has an effect mainly between conceptual entities (i.e., considering the FRBR Work level) rather than their concrete objects realization (i.e., the FRBR Expression level). For instance a sinopia or a preliminary drawing of a work of art is considered a proper work of art *per se*, even if it is strictly related to the final work created by an artist.

As before, in the OA Entry Ontology we reuse the models introduced for the F Entry Ontology – i.e. FaBiO, PRO, CiTO, HiCO, FRBR and PROV-O – so as to provide a precise description of specific aspects of the domain in consideration.

An OA Entry is a document containing metadata about a work of art and its cataloguing. Therefore, the OA Entry Ontology defines it in terms of a FRBR Work (i.e. an instance of the class *oaentry:OAEntry*, subclass of *fabio:EntityMetadata*). The work of art (class *fabio:ArtisticWork*) described by the entry is considered as a FRBR Work as well.

The object property *oaentry:describes* allows one to link an OA Entry to its the work of art it describes. Its range involves all the four FRBR levels, since, as introduced above, several metadata in the Zeri Photo Archive actually refers to peculiar aspects related with the material used for embodying the work of art (i.e., FRBR Manifestation level) as well as its current physical location (i.e., FRBR Item level).

4.2.1 *Extending PROV-O and HiCO for describing the influence between works.* Original-to-derivative relations between two works of art can be represented by means of the class *prov:Influence*. In order to enable the description of different kinds of influence that can exist between works of art, the OA Entry Ontology extends PROV-O by adding appropriate subclasses to *prov:Influence* (*oaentry:Cartoon*, *oaentry:Copy*, *oaentry:Derivation*, etc.) – derived by analysing the field "ROFF" (i.e., the *status of work*) in all the concrete OA entries available in the Zeri Photo Catalogue and considering terms of the controlled vocabulary provided by the Scheda OA content standard.

The object property *oaentry:hasFormerWork* (i.e. a sub-property of *prov:entity*) allows one to link an individual of any of the influence classes to the original work of art. The object property *oaentry:hasConceived* enables one to link an individual of any of the influence classes to the derivative work of art in consideration.

By means of the aforementioned ontological entities, it is possible to model properly a scenario described in the sentence "the anonymous drawing of Sistine Chapel is conceived as a drawing of Michelangelo Buonarroti's frescos in Sistine Chapel" as shown in the following excerpt in Turtle syntax (the complete example is available at http://dx.doi.org/10.6084/m9.figshare.3175048):

```
:anonymous-drawing-sistine-chapel-work
  oaentry:isConveivedByMeansOf :michelangelo-fresco-sistine-chapel-drawing .
:michelangelo-fresco-sistine-chapel-drawing a oaentry:Drawing ;
  oaentry:hasFormerWork :michelangelo-fresco-sistine-chapel-work ;
  oaentry:hasConceived :anonymous-drawing-sistine-chapel-work .
```

It is worth noticing that such influence assertion can be even defined as a questionable information. In this case, an instance of the class *hico:InterpretationAct* can be created to specify explicitly that the claimed influence was actually derived from a cataloguer's subjective choice. To this end, the OA Entry Ontology extends HiCO by providing terms for defining types of interpretation strictly related to the arts domain – e.g. the attribution of an influence between works may be represented as the individual *oaentry:influence-between-works-attribution*. This can be defined in RDF as follows:

```
:michelangelo-fresco-sistine-chapel-drawing a oaentry:Drawing ;
  prov:wasGeneratedBy :michelangelo-fresco-sistine-chapel-drawing-attribution .
:michelangelo-fresco-sistine-chapel-drawing-attribution a hico:InterpretationAct ;
  hico:hasInterpretationType oaentry:influence-between-works-attribution ;hico:hasInterpretationType :zeri-
       preferred-attribution ;
  hico:hasInterpretationCriterion :cataloguer-choice ;
  hico:isExtractedFrom :oaentry-15429-expression ;
  prov:wasAssociatedWith :crr-mm .
oaentry:influence-between-works-attribution a hico:InterpretationType .
:zeri-preferred-attribution a hico:InterpretationType .
:cataloguer-choice a hico:InterpretationCriterion .
:oaentry-15429-expression a fabio:Expression .
:crr-mm a foaf:Agent .
```

4.2.2 *Extending PRO for providing a controlled vocabulary of roles in the Arts domain.* In addition to information about a work of art, the OA Entry generally requires the cataloguer to provide several information also about the artist responsible for the creation of the work, including the actual role he/she had in that creative process. Moreover, when cataloguing the work of art, several responsibilities are attributed to cataloguers as well. Therefore, in the OA Entry Ontology we added appropriate individuals to the class *pro:Role* so as to describe additional roles proper to the arts and cataloguing domains. In particular, we have created two new subclasses of *pro:Role*, i.e. *oaentry:ArtisticRole* and *oaentry:CataloguingRole*, each including specific individuals:

— those defined as instances of *oaentry:ArtisticRole* have been recognized by means of the open vocabulary adopted by the Zeri Foundation for describing roles of artists and the controlled vocabulary provided by Scheda OA, such as *oaentry:antiquarian*, *oaentry:architect*,*oaentry:art-dealer*, etc.;

— those defined as instances of *oaentry:CataloguingRole* have been recognized as the main roles involved into the cataloguing process: *oaentry:catalogue*r, *oaentry:cataloguing-institution*, *oaentry:cataloguing-supervisor*, and *oaentry:competent-institution*.

5. MAPPING F/OA ENTRY TO RDF

Each Scheda F/OA content standard is organized in sections. Each section contains data about a work (a photograph, a work of art) that can be described in terms of one of the four FRBR entities (Work, Expression, Manifestation, Item).

In this section we briefly introduce the mapping we propose between the fields in the Scheda OA and Scheda F content standards and RDF according to the ontologies introduced in the previous section, and by enriching such description with terms coming from the CIDOC-CRM specification. It is worth mentioning that a complete mapping is not in the scope of this paper. In particular, in this work we focus on the mapping between all the fields in Scheda F/OA that are actually used in the Zeri Photo Archive Catalogue – which counts about 118 fields out of more than 300 provided by Scheda F, and about 97 fields out of 280 provided by Scheda OA. Thus, the mappings defined include all the mandatory elements prescribed by the standard creators (called *inventory level of description*), and a selection of other significant fields.

Operatively, one author (i.e. Marilena Daquino) performed a first round of the mapping by looking at all the aforementioned fields in the Scheda F and Scheda OA content standards one by one. For each field she provided a first mapping to RDF accompanied by a meaningful example of use. After this initial process, another author (i.e. Silvio Peroni) checked again the whole mapping document produced in order to correct mistakes, to suggest modifications, and to harmonize it where ambiguous mappings were introduced. After another iteration by the first author, so as to address all the suggestions introduced by the second one, the resulting mapping document was analysed by the other three authors – i.e. a member of the Zeri Foundation responsible for the Zeri's catalog (Francesca Mambelli), a digital humanist (Francesca Tomasi), and a computer scientist with strong background in markup languages (Fabio Vitali) – so as to gather additional feedback. A new version of the mapping document was, thus, finally released by the first author.

The mapping process resulted in the creation of two distinct documents, i.e., *F Entry to RDF* (https://dx.doi.org/10.6084/m9.figshare.3175273) and *OAEntry to RDF* (https://dx.doi.org/10.6084/m9.figshare.3175057), accompanied by exemplar data that represent an F Entry (http://dx.doi.org/10.6084/m9.figshare.3175252) and an OA Entry (https://dx.doi.org/10.6084/m9.figshare.3175048) in RDF created according to such mappings. All the mapping documents contain tables structured as shown in Fig. 3. Such tables reproduce the structure of related content standard they refer to, and are organized in three columns. The first and the second column contain the name of the field in the Scheda F/OA and a brief description. The third column details the mapping with RDF terms and accompanies it with examples of usage.

OG – OBJECT AND SUBJECT DESCRIPTION *		
OGT – OBJECT *		

OGTD *	DEFINITION
	
	A term identifying the main type of a described work of art. It may belong to a local open thesaurus and/or to an established one, e.g. the *Art and architecture Thesaurus*.
		CRM:E55_TYPE (CLASS)
		
		According to the *Cataloging Cultural Objects (CCO) project* of the *Visual Resources Association Foundation* (http://cco.vrafoundation.org/), which suggests to consider only the FRBR Work level when describing type of works of art, terms belonging to the open vocabulary identified in this field are considered specializations of a work of art at the FRBR Work level of description, i.e. individuals of the classes `fabio:ArtisticWork` and `crm:E28_Conceptual_Object`.
		By means of the object property `crm:P2_has_type` a work of art may be associated to an individual defining the type of work.
		
		EXEMPLAR USAGE:
		`:oa-47172`
		`a crm:E28_Conceptual_Object , fabio:ArtisticWork ;`
		`crm:P2_has_type :polyptych .`
		
		Terms of the `crm:E55_Type` hierarchy shall be aligned to an established controlled vocabulary or thesaurus, e.g. the AAT Getty Thesaurus (http://www.getty.edu/research/tools/vocabularies/aat/).
		
		EXEMPLAR USAGE:
		`:polyptych a crm:E55_Type ;`
		`rdfs:seeAlso`
		`<http://vocab.getty.edu/aat/300178235> .`

OGTT	OBJECT TYPE
	
	A term specializing the main type of the described work of art, excluding its functional and morphological features.
		CRM:E55_TYPE (CLASS)
		
		Here, as in the OGTD field, an heterogeneous open or controlled vocabulary is used to further describe formal features of a work of art: the value of this field shall be a complementary definition of the previous one or may be considered as another specification, through the use of the property `crm:P2_has_type`.
		E.g. OGTD: 'fountain' ; OGTT: 'basin' .
		
		EXEMPLAR USAGE:
		`:oa-75147`
		`crm:P2_has_type :basin ;`
		`crm:P2_has_type :fountain .`
		`:fountain a crm:E55_Type .`
		`:basin a crm:E55_Type .`
		
		OR
		`:oa-75147`
		`crm:P2_has_type :basin-fountain .`
		`:basin-fountain a crm:E55_Type .`

Fig. 3. An excerpt of the mapping document "Mapping OA Entry to RDF".

The following sections are organized on the basis of the four FRBR level, with an introduction about the representation of top-level relations between entries and subjects therein described. The ontological entities related with the F Entry Ontology and the OA Entry Ontology (presented in Section 4) are directly used without further explanation, while the use of CIDOC-CRM in the RDF excerpts is detailed.

5.1 The entry and its subject

Any F/OA Entry can be defined in CIDOC-CRM as instance of the class *E31 Document*, and it is related to their respective subjects by means of the object property *P70 documents*. An explicit relation between the F Entry describing a photograph and an OA Entry describing the work of art portrayed in the mentioned photograph can be represented by using *P67 refers to*.

Identifiers of the entries are represented by means of an instance of the class *E42 Identifier*. An entry can have several identifiers, and all the identifiers can be represented in the same way and then characterized with the property *P2 has type*, so as to specify which type of identifier is associated (levels of cataloguing, regional codes, general catalog numbers, etc.).

We decided to use terms belonging to PRO as well as the individuals of the class *oaentry:CataloguingRole* for describing the cataloguing process, which involves different roles at different times.

The following Turtle excerpt below introduces an example of all the aforementioned aspects:

```
:fentry-72486 a fentry:FEntry , crm:E31_Document ;
  fentry:describes :photo-72486 , :oa-47172 ;
  crm:P67_refers_to :oaentry-43677 .
:fentry-72486-creation a crm:E65_Creation ;
  crm:P14_carried_out_by :cataloguer ;
  crm:P4_has_time_span :2016-
:oaentry-43677 a oaentry:OAEntry , fabio:Work , crm:E31_Document ;
  oaentry:describes :oa-47172 ;
  crm:P140i_was_attributed_by
    :oaentry-43677-catalog-level-assignment ,
    :oaentry-43677-nctr-assignment , :oaentry-43677-nctn-assignment ;
:md-cataloguer-oaentry-43677 a pro:RoleInTime ;
  pro:relatesTo :oaentry-43677 ;
  pro:isHeldBy :md ;
  pro:withRole oaentry:cataloguer ;
  tvc:atTime :2012-11-04 .
```

5.2 The Work level

Both the photograph (described in a F Entry) and the work of art (described in an OA Entry) are defined in CIDOC-CRM terms as instances of the class *E28 Conceptual Object* when describing their essence and their creation.

The class *E65 Creation* is required to define the authorship of photographs and works of art, i.e. to identify the main photographer of the photograph and the main artist or group of artists of the work of art. The actors (*E39 Actor*, or one of its subclasses, i.e., *E21 Person* and *E74 Group*) involved in this authorship attribution are specified by using the object property *P14 carried out by*. The object property *P4 has time span* is used for specifying the duration of the creation event.

Moreover, the shot can be associated to a place (the class *E53 Place*) by using the object property *P8 took place at*, and can be associated to a specific occasion, i.e. an instance of the class *E4 Period* (further specified in *E5 Event*) by means of the object property *P10 falls within*.

Both the creators of the two aforementioned works can be associated to a cultural context (for instance a school of painters or a workshop) by using the property *P107i is current of former member of*, which relates he/she to an individual of the class *E74 Group*.

According to Bountouri and Gergatsoulis [Lina Bountouri and Manolis Gergatsoulis 2011a] [Lina Bountouri and Manolis Gergatsoulis 2011b], the archival description of the photograph, i.e. the hierarchical organization of the containers that include the catalogued object, can be described as nested relations by using the object property *P106 is composed of* for relating conceptual entities.

One or more titles can be attributed to the entities, by means of the object property *P102 has title* associated to an instance of the class *E35 Title,* which can be further specialized by using *P2 has type* to define if the title is an attributed, traditional or an alternate one.

When bibliography or other sources are provided to support the cataloguing, e.g. letters, audio-recorded works, catalogs, entries etc., a generic relation can be represented with *P70i is documented in*, by linking to an individual of the broader class *E31 Document*.

A direct relation between the photograph and the depicted work of art can be established relating the photograph (the FRBR Work) at the concrete object of art (an FRBR Item).

The following Turtle excerpt introduces an example of part of the aforementioned aspects:

```
:photo-72486 a fentry:Photograph , crm:E28_Conceptual_Object ;
  crm:P94i_was_created_by :photo-72486-creation ;
  crm:P106i_forms_part_of :folder-leonardo ;
  crm:P102_has_title :jesus-baptism-verrocchio ;
  frbr:subject :oa-47172-item ;
  crm:P70i_is_documented_in :document-f2336 .
:photo-72486-creation a crm:E65_Creation ;
  crm:P94_has_created :photo-72486 , :photo-72486-expression ;
  crm:P7_took_place_at :florence ;
  crm:P10_falls_within :exhibition-of-paintings ;
  crm:P4_has_time_span :1926-1932 ;
  crm:P14_carried_out_by :brogi-studio .
:folder-leonardo a fabio:Work , crm:E90_Symbolic_Object ;
  crm:P106i_forms_part_of :subseries-leonardo .
:subseries-leonardo a fabio:Work , crm:E90_Symbolic_Object ;
  crm:P106i_forms_part_of :series-leonardo .
:series-leonardo a fabio:Work , crm:E90_Symbolic_Object ;
  crm:P106i_forms_part_of :zeri-photo-archive .
:zeri-photo-archive a fabio:WorkCollection .
:oa-47172 a fabio:ArtisticWork , crm:E28_Conceptual_Object ;
  crm:P94i_was_created_by :oa-47172-creation ;
  fabio:hasPortrayal :oa-47172-item .
```

5.3 The Expression level

While in arts domain we are not interested in representing contents of a work of art separately from their conception, in the photography domain we can describe the photograph with regards to its content, i.e. its FRBR Expression, which is realized at the same time of the creation of the work. None of such information are precisely covered by the CIDOC-CRM, then terms from FaBiO and F Entry Ontology are used.

The following Turtle excerpt introduces an example of the aforementioned aspects:

```
:photo-72486-creation a crm:E65_Creation ;
  crm:P94_has_created :photo-72486 , :photo-72486-expression .
:photo-72486-expression a fentry:Shot ;
  crm:P94i_was_created_by :photo-72486-creation ;
  frbr:realizationOf :photo-72486 .
```

5.4 The Manifestation level

The description about the format used for embodying an object applies to the Manifestation level.

On the one hand, in photographs different formats identify several manifestations, e.g. digital images, slides, negatives, and positives. Each manifestation belongs also to the class *E22 Man-Made Object* and represents a specific form that the work may have, which can be further characterised by means of the object property *P2 has type*.

On the other hand, in works of art the FRBR Manifestation level should be applied any time a relevant change affected the object, e.g. any time a restoration intervention is recorded.

Both the photograph and the work of art may be described in terms of the material they are made, i.e. an instance of the class *E57 Material*. The class *E16 Measurement* is used to define the various dimensions (*E54 Dimension*) related with such manifestations (e.g. weight and height). Other specific features characterizing the manifestation, e.g. colour, are linked with the object property *P56 bears feature*.

The following Turtle excerpt introduces an example of all the aforementioned aspects:

```
:photo-72486-positive
  a crm:E22_Man-Made_Object , fabio:AnalogManifestation ;
  crm:P45_consists_of :gelatin-silver ;
  crm:P56_bears_feature :black-and-white ;
  crm:P39i_was_measured_by :photo-72486-positive-measurement .
:photo-72486-positive-measurement a crm:E16_Measurement ;
  crm:P40_observed_dimension :height-194mm ;
  crm:P40_observed_dimension :width-250mm .
```

5.5 The Item level

The photograph is considered as a concrete object that can be defined as an individual of the class *E22 Man-Made Object* (also inferred as *E84 Information Carrier*) linked to the portrayed work of art by means of the object property *P62 depicts*.

Specific features regarding the object reported by cataloguers during an assessment (represented with the class *E14 Condition Assessment*) are defined as instances of the class *E3 Condition State*. It is worth mentioning that any instance of *E3 Condition State* can be further specialized by using the *P2 has type* property, e.g. for defining the status of the object recorded during the assessment.

Several situations may be related to the production of the concrete object representing the photograph or the work of art, where several actors with a specific role may be involved. Therefore terms belonging to PRO Ontology are here preferred.

Different sorts of location (the class *E53 Place*) can be attributed to a concrete object by using the object property *P55 has current location*. When filling the section regarding the geographical and administrative location of an object (either a photograph, a work of art or a related work of one of them), a cataloguer can refer to a place, i.e. class *E53 Place,* and/or to a current keeper (*P50 has current keeper*) which resides in such place. In the latter case, the current keeper (*E39 Actor*) is further related to a place by using the object property*P74 has current or former residence*, and to an address by using the property *P76 has contact point*. The place in which the keeper resides is defined by means of a chain of nested places (e.g. village, town, district, region, country), each other related through the object property *P89 falls within*.

Such complete description of the current location of an object, i.e. in terms of a place and an agent keeping it, enables the description of its transfers of custody (*E10 Transfer of Custody*). If a particular keeper is not identified, the change of location can be simply defined in terms of *E9 Move*, linked the location the object has been moved (*P26 moved*) from another one (*P27 moved from*).

The ownership of such concrete object can be defined by using the property *P52 has current owner*. Each owner (*E39 Actor*) can have acquired such work (property *P22i acquired title through*) as consequence of an acquisition event *E8 Acquisition*, that can be further specialized by using *P2 has type* to describe ownership specifications.

The last considered section regarding the concrete object addresses its participation in exhibitions (*E5 Event*). Such events can, thus, be linked to the object included in the event (*P12 occurred in the presence of*) to the location (*P7 took place at*) and date (*P4 has time span*) of the event, which can be have a formal appellation (*E41 Appellation*) specified through the property *P1 is identified by*.

The following Turtle excerpt introduces a partial example of the aforementioned aspects regarding only a photograph, being a work of art described similarly:

```
:photo-72486-positive-item a fabio:AnalogItem , crm:E22_Man-Made_Object ;
  crm:P62_depicts :oa-47172-item ;
  crm:P57_has_number_of_parts "1" ;
  crm:P34i_was_assessed_by :photo-72486-positive-item-condition ;
  crm:P55_has_current_location :large-formats-room ;
  crm:P140i_was_attributed_by :photo-72486-invn-assignment ;
  crm:P52_has_current_owner :university-of-bologna ;
  crm:P12i_was_present_at :exhibition-london-1987 ;
  crm:P30i_custody_transferred_through :photo-72486-item-provenance-1 .
:photo-72486-positive-item-condition a crm:E14_Condition_Assessment ;
```

```
  crm:P35_has_identified :photo-72486-positive-item-condition-state .
:photo-72486-positive-item-condition-state a crm:E3_Condition_State ;
  crm:P2_has_type :discrete ;
  crm:P3_has_note "silver mirror" .
:large-formats-room a crm:E53_Place ;
  crm:P89_falls_within :ex-convent-santa-cristina .
:photo-72486-item-provenance-1 a crm:E10_Transfer_of_Custody ;
  crm:P28_custody_surrendered_by :villa-i-tatti ;
  crm:P29_custody_received_by :zeri-foundation ;
  crm:P30_transferred_custody_of :photo-72486-positive-item ;
  crm:P4_has_time_span :1989 .
```

6. THE ZERI PHOTO ARCHIVE RDF DATASET

The specific heterogeneous nature of the domains considered in the *Zeri and LODE* project (and in the PHAROS project) and the lack of similar Linked Open Data datasets available on the Web have been important issues to address for providing a comprehensive description of such different kind of information by means of Semantic Web technologies. Thus, from our perspective, publishing a dataset about these kinds of objects may be considered a pioneering attempt *per se*, and it would represent, in principle, an established practice for creating a network of LOD with other Italian cultural institutions. In addition, it can be considered an available source of precious information for linking existing international datasets in the field of Arts.

The Zeri Photo Archive RDF dataset we have realized (available online at https://w3id.org/zericatalog/), which contains data compliant with the ontologies and the mappings described in the previous sections, is the result of an analysis on a dataset originally created for testing purposes in the context of the PHAROS project. In particular, we gathered about 31.000 F Entries and 19.000 OA Entries from Zeri Photo Archive catalog, stored as XML documents (compliant with no particular schema) that contained the data prescribed by the Scheda F and Scheda OA content standards. In particular, XML contents were organized in *paragraphs*, which correspond to Scheda F/OA descriptive sections. The data gathered have been chosen according to a thematic organisation. The first subset of Zeri catalog considered was based on a collection of OA Entries describing works of XV-XVI centuries, and included all the F Entries describing photographs portraying the aforementioned works.

The resulting RDF dataset has been created by means of an XSL transformation of the provided XML documents into RDF statements – where one RDF/XML file has been created for each of the considered entries. RDF statements mainly refer to photographs and works of art, and also include relevant information describing about 4,500 bibliographic entities, 6,000 artists, 2,000 photographers and 2,000 catalogs. Such additional information have been provided by the Zeri Foundation by means of other XML documents representing a local authority files. In particular, such XML files reproduce the fields required by the ICCD guidelines for creating people authority files (http://www.iccd.beniculturali.it/index.php?it/473/standard-catalografici/Standard/55[1]) and bibliographic references authority files (http://www.iccd.beniculturali.it/index.php?it/473/standard-catalografici/Standard/58). The aforementioned authority files have been converted into RDF (by creating one RDF/XML file for each record in an authority file) by means of an XSL transformation as well.

All the RDF resources have been labelled in Italian and, when possible, in English, in order to facilitate their understanding to a larger audience. In addition, the IRIs of such resources have been created in English, in order to ensure their easy reuse in other non-Italian datasets. The document describing the IRI design choices is available at the homepage of the project, (https://w3id.org/zericatalog, section *Data*).

[1] http://www.iccd.beniculturali.it/index.php?it/473/standard-catalografici/Standard/55

All the open vocabularies adopted by cataloguers have been implemented as local controlled vocabularies, which will have to be enhanced when the dataset will be enriched with all the data coming from the full Zeri catalog.

All the RDF resources have been further enhanced with direct links to external authority files and datasets, in particular:

— the Getty Art & Architecture Thesaurus (http://www.getty.edu/research/tools/vocabularies/aat/) was considered for object types, materials and techniques;

— VIAF (http://viaf.org), Getty Union List of Artist Names (http://www.getty.edu/research/tools/vocabularies/ulan), Wikidata (https://www.wikidata.org), and Dbpedia (http://dbpedia.org) have been chosen to identify artists and photographers (the latter still in progress);

— geoNames (http://www.geonames.org) and, again, Wikidata and Dbpedia were considered for identifying places.

Reconciliation of entities to established authority files like VIAF and geoNames was mainly obtained through a semi-automatic process, by using ad hoc plugins implemented in OpenRefine (http://openrefine.org/). Moreover, a PERL script (http://search.cpan.org/dist/App-wdq/lib/App/wdq.pm) was useful to access the Wikidata Query Service, which also enabled us to directly link founded entities to Wikipedia pages, Dbpedia and Getty ULAN entities.

All the RDF statements have been stored in an Apache Jena Fuseki2 triplestore (https://jena.apache.org/documentation/fuseki2). We chose Fuseki2 since it is easy to deploy and manage even by non-expert users, but it will be probably changed until the end of the PHAROS project, due to its limits in managing huge amounts of data. All the data stored in the triplestore are distributed under the license CC-BY-NC, Creative Commons Attribution-NonCommercial 4.0 International (http://creativecommons.org/licenses/by-nc/4.0/).

As of September 11, 2016, the dataset includes about 11,400,000 RDF statements relating 1,600,000 unique typed entities. Among these, about 3,000 have been linked with external resources already available in the Linked Open Data. In particular, we create links to 2,200 different VIAF resources, 1,200 to Getty ULAN resources, 1,500 different GeoNames resources, and 2,260 different Dbpedia and Wikidata resources. We estimate to have 100 million RDF statements in the dataset once the full Zeri catalog will be processed.

All the RDF data can be queried in SPARQL by making appropriate REST requests to the related SPARQL endpoint made available by the triplestore at https://w3id.org/zericatalog/sparql. We have also prepared a web interface (available at http://data.fondazionezeri.unibo.it/query/) for allowing users to query the triplestore directly on the Web by means of SPARQL queries. Finally, we have used LODView (http://lodview.it/) for allowing direct browsing of all the RDF data included in the triplestore – for instance, Fig. 4 shows the HTML visualisation of the resource https://w3id.org/zericatalog/photo/59972, i.e. the photograph "Alinari, Fratelli , Perugia, Collegio del Cambio: Cappella di S. Giovanni Battista, volta - S. Giovanni e S. Luca (Giannicola di Paolo)".

All F Entries and OA Entries defined in the dataset include links to the current photographic and arts catalog of the Zeri Foundation (available at http://catalogo.fondazionezeri.unibo.it), enabling users to go from the LOD-based representation of the catalog to the traditional Web pages presenting the same entities. The inverse link from HTML pages to the related RDF resources has been implemented and it is currently in test phase (and not officially released yet).

7. CONCLUSIONS

In this article we have introduced the main current outcomes of the *Zeri and LODE* project, i.e. the F Entry Ontology, the OA Entry Ontology, the mapping documents between two Italian content standards (i.e. the Scheda F and the Scheda OA) and CIDOC-CRM accompanied by the aforementioned ontologies, and the creation of a real LOD dataset listing about 50,000 entries from the Zeri Photo Archive. Our project is set

Fig. 4. An example of a resource of the Zeri Photo Archive visualized in LODView.

within the PHAROS project, which groups an international consortium of several institutions for creating a common platform for research on images and metadata of Western and non-Western works.

As required by the PHAROS Consortium, the *Zeri and LODE* project has already produced a first stable outcome that allows the description of a large set of metadata about photographs and works of art. In the next phase, we will drive the adoption of these ontologies by other institutions (a dialog with ICCD has been set up already so as to study a strategy for the adoption of our models at a national level) and we will extend the current LOD dataset with all the data expressed in the other F/OA entries in the Zeri's catalog, so as to make available openly and in machine-readable form the full richness of Federico Zeri's knowledge base. Of course, this extension will require a new hardware architecture – e.g. a new powerful server, the use of a more performant triplestore such as Blazegraph (https://www.blazegraph.com/), as well as additional disk space and RAM for storing and managing the new data – which is something we will deal with in the near future.

At the same time, the ontologies we developed and the corresponding mapping documents will be revised and extended in order to complete the analysis of the F/OA contents standards, also considering relevant topic introduced in the earliest version of the Scheda F (version 4.00 at http://www.iccd.beniculturali.it/index.php?it/473/standard-catalografici/Standard/62). Therefore a complete mapping will be provided to allow the most comprehensive description of the Photography and Arts domain as addressed in ICCD content standards.

It is worth mentioning that, even if the mapping we provided was based on national content standards and required to compare and merge two slightly different perspectives (i.e., the official cataloguing standard requirements and its implementation in a real use case where some customizations and a few new fields were defined by the Zeri Foundation itself), we claim it can be considered an exhaustive resource for any international cultural heritage institution to publish its data as LOD. The large number of models used to map such standards into RDF was useful to define a coherent and comprehensive descriptive scenario, based on state-of-the-art ontologies for the cultural heritage domain.

In addition, it is possible that future revisions of the ontologies and of the mapping documents will take into account additional models that were not considered at this stage, such as FRBRoo. While FRBRoo matches well with CIDOC-CRM by design, its adoption in our project was not part of the requirements exposed by the PHAROS Consortium. In addition, at the current stage of our project and according to the domain we have to describe, FRBRoo would not add any meaningful additional detail to the current data, and thus we decided to postpone its use in future phases of the project – when at least a formal alignment to it will be provided. In addition to FRBRoo, we are planning to include proper models for archival description, e.g. the Reload Ontologies (http://labs.regesta.com/progettoReload/), extensively adopted in Italian archives, in order to enable an easier integration of the Zeri catalog to LOD datasets made available by other archival institutions.

Finally, once other partners of the PHAROS Consortium publish their data according to the defined models, the second phase of the project will allow us to integrate all the data in a common environment, together with a searchable repository for images (at the monent, a beta version of this application to search on images is available at http://en.images.pharosartresearch.org/).

ACKNOWLEDGEMENTS

We would like to thank Ciro Mattia Gonano, who firstly addressed relevant issues in mapping Scheda F to CIDOC-CRM and whose work was the basis for the current evolution of the project, and Raffaele Messuti, who helped us in managing the current RDF dataset, deploying and maintaining the server, and in reconciling data to external vocabularies and datasets. We would also like to thank the editor and the reviewers of the paper, who have provided us with valuable comments and suggestions for improving the article.

REFERENCES

Murtha Baca. 2006. *Cataloguing Cultural Objects: Cco; a Guide to Describing Cultural Works and Their Images*. American Library Association, Chicago, IL.

Murtha Baca and Sherman Clarke. 2007. FRBR and Works of Art, Architecture, and Material Culture. *Understanding FRBR: What It Is and How It Will Affect Our Retrieval Tools*. ed. Arlene Taylor (Westport CT: Libraries Unlimited), 227–242.

Lina Bountouri and Manolis Gergatsoulis. 2011. Mapping Encoded Archival Description to CIDOC-CRM. In 1st Workshop on Digital Information Management, Corfu, Greece, March 30-31, Ionian University, 8–25.

Lina Bountouri and Manolis Gergatsoulis. 2011. The Semantic Mapping of Archival Metadata to the CIDOC-CRM Ontology. *Journal of Archival Organization*, 9, 3-4, 174–207.

Karen Coyle. 2016. FRBR, Before and After: A Look at Our Bibliographic Models. American Library Association, Chicago.

Marilena Daquino and Francesca Tomasi. 2015. Historical Context (HiCO): a conceptual model for describing context information of cultural heritage objects. In *Metadata and Semantics Research, Communication in Computer and Information Science 544*, Eds. Emmanouel Garoufallou, Richard J. Hartley and Panorea Gaitanou, 424–436. Springer, Cham. DOI: http://dx.doi.org/10.1007/978-3-319-24129-6_37

Victor De Boer, Jan Wielemaker, Judith van Gent, Michiel Hildebrand, Antoine Isaac, Jacco van Ossenbruggen and Guus Schreiber. 2012. Supporting Linked Data Production for Cultural Heritage Institutes: The Amsterdam Museum Case Study. In *The Semantic Web: Research and Applications, Lecture Notes in Computer Science 7295*, Eds. Elena Simperl, Philipp Cimiano, Axel Polleres, Oscar Corcho and Valentina Presutti, 733–747. Springer, Heidelberg. DOI: http://dx.doi.org/10.1007/978-3-642-30284-8_56

Martin Doerr. 2010. The Europeana Data Model (EDM) mapping to CIDOC-CRM. http://www.cidoc-crm.org/docs/CRM-EDM_FRBR.ppt

Europeana. 2016. Definition of the Europeana Data Model v5.2.7. http://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation//EDM_Definition_v5.2.7_042016.pdf

Riccardo Falco, Aldo Gangemi, Silvio Peroni and Fabio Vitali. 2014. Modelling OWL ontologies with Graffoo. In *The Semantic Web: ESWC 2014 Satellite Events, Lecture Notes in Computer Science 8798*. Eds. Valentina Presutti, Eva Blomqvist, Raphael Troncy, Harald Sack, Ioannis Papadakis and Anna Tordai, 320–325. Springer, Berlin. DOI: http://dx.doi.org/10.1007/978-3-319-11955-7_42

Achille Felicetti, Tiziana Scarselli, Maria Letizia Mancinelli and Franco Niccolucci. 2013. Mapping ICCD Archaeological Data to CIDOC-CRM: the RA Schema. In *Proceedings of the Workshop Practical Experiences with CIDOC-CRM and its Extensions*. CEUR-WS.org, Aachen. http://ceur-ws.org/Vol-1117/paper2.pdf

Mariano Fernández, Asunción Gómez-Pérez, Natalia Juristo. 1997. METHONTOLOGY: from ontological art towards ontological engineering. In *Proceedings of the AAAI97 Spring Symposium*. Stanford, 33–40.

Ciro Mattia Gonano, Francesca Mambelli, Silvio Peroni, Francesca Tomasi and Fabio Vitali. 2014. Zeri e LODE: Extracting the Zeri photo archive to linked open data: formalizing the conceptual model. In *Proceedings of the 2014 IEEE/ACM Joint Conference on Digital Libraries (JCDL 2014)*. IEEE, Washington, 289–298. DOI: http://dx.doi.org/10.1109/JCDL.2014.6970182

International Federation of Library Associations and Institutions. 2009. Functional Requirements for Bibliographic Records. *Final Report, February 2009*. http://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf

Antoine Isaac, William Waites, Jeff Young and Marcia Zeng. 2011. Datasets, Value Vocabularies, and Metadata Element Sets. *W3C Incubator Group Report, 25 October 2011*. http://www.w3.org/2005/Incubator/lld/XGR-lld-vocabdataset/

Timothy Lebo, Satya Sahoo and Deborah McGuinness. 2013. PROV-O: The PROV Ontology. *W3C Recommendation 30 April 2013*. http://www.w3.org/TR/prov-o/

Patrick Le Boeuf, Martin Doerr, Christian Emil Ore and Stephen Stead. 2015. Definition of the CIDOC Conceptual Reference Model. http://www.cidoc-crm.org/docs/cidoc_crm_version_6.2.1.pdf

Julia Marden, Carolyn Li-Madeo, Noreen Whysel and Jeffrey Edelstein. 2013. Linked open data for cultural heritage: evolution of an information technology. In *Proceedings of the 31st ACM International Conference of Design of Communication*, 107–112. ACN, New York. DOI: http://dx.doi.org/10.1145/2507065.2507103

Boris Motik, Peter F. Patel-Schneider and Bijan Parsia. 2012. OWL 2 Web Ontology Language: Structural Specification and Functional-Style Syntax (Second Edition). W3C Recommendation, 11 December 2012. http://www.w3.org/TR/owl2-syntax/

Michele Pasin and John Bradley. 2015. Factoid-based prosopography and computer ontologies: towards an integrated approach. In *Digital Scholarship in the Humanities,* 30.1, 86-97. DOI: http://dx.doi.org/10.1093/llc/fqt037

Silvio Peroni. 2016. SAMOD: an agile methodology for the development of ontologies. *figshare*. http://dx.doi.org/10.6084/m9.figshare.3189769

Silvio Peroni and David Shotton. 2012. FaBiO and CiTO: ontologies for describing bibliographic resources and citations. *Journal of Web Semantics: Science, Services and Agents on the World Wide Web*, 17 (December 2012), 33–43. DOI: http://dx.doi.org/10.1016/j.websem.2012.08.001

Silvio Peroni, David Shotton and Fabio Vitali. 2012. Scholarly publishing and the Linked Data: describing roles, statuses, temporal and contextual extents. In *Proceedings of the 8th International Conference on Semantic Systems (i-Semantics 2012)*. Eds. Harald Sack and Tassilio Pellegrini, 9–16. ACM, New York. DOI: http://dx.doi.org/10.1145/2362499.2362502

Inge Reist, David Farneth, R. Samuel Stein and Remigius Weda. 2015. An Introduction to PHAROS: Aggregating Free Access to 31 Million Digitized Images and Counting. *Speech at CIDOC 2015*. New Dehli, September 5–9, 2015. http://network.icom.museum/fileadmin/user_upload/minisites/cidoc/BoardMeetings/CIDOC_PHAROS_Farneth-Stein-Weda_1.pdf

Max Schmachtenberg, Chris Bizer and Heiko Paulheim. 2014. State of the LOD Cloud 2014 (version 0.4). http://linkeddatacatalog.dws.informatik.uni-mannheim.de/state/

Richard P. Smiraglia, Pat Riva, Maja Žumer. 2013. *The FRBR Family of Conceptual Models: Toward a Linked Bibliographic Future* (1st. ed.), Routledge, New York, NY.

Pedro Szekely, Craig A. Knoblock, Fengyu Yang, Xuming Zhu, Eleanor E. Fink, Rachel Allen and Georgina Goodlander. 2013. Connecting the Smithsonian American Art Museum to the Linked Data Cloud. In *The Semantic Web: Semantics and Big Data, Lecture Notes in Computer Science 7882*, Eds. Philipp Cimiano, Oscar Corcho, Valentina Presutti, Laura Hollink and Sebastian Rudolph, 593–607. Springer, Heidelberg. DOI: http://dx.doi.org/10.1007/978-3-642-38288-8_40

Mike Uschold. 1995. Towards a Methodology for Building Ontologies. In *Proceedings of IJCAI-95's Workshop on Basic Ontological Issues in Knowledge Sharing*, 82, 74–82.

Seth Van Hooland and Ruben Verborgh. 2014. Linked Data for Libraries, Archives and Museums. How to clean, link and publish your metadata. Amer Library Assn Editions. ISBN: 978-0838912515