Linking data in digital libraries: the case of Puglia Digital Library

Tommaso Di Noia[1], Azzurra Ragone[2], Andrea Maurino[2], Marina Mongiello[1], Maria P. Marzocca[3], Giuseppe Cultrera[3], Mauro P. Bruno[4]

[1] Polytechnic University of Bari, Bari, Italy firstname.lastname@poliba.it
[2] University of Milano-Bicocca, Milano, Italy lastname@disco.unimib.it
[3] Innovapuglia SpA, Valenzano (BA), Italy
{mp.marzocca,g.cultrera}@innova.puglia.it
[4] Regione Puglia, Bari, Italy mp.bruno@regione.puglia.it

Abstract. The digital revolution has been a big shift in the creation, publication and storage of our digital heritage. New file formats and supports have followed over the years to produce and reproduce audio, photos and video. Based on these observations, the Puglia region started the *Puglia Digital Library* project with the aim to collect in a single public collection all the digital contents related to Puglia.
Due to its public nature, all the items available in the collection have been described using also RDF-based annotations and the final dataset has been exposed via a SPARQL endpoint. During the data-engineering process, attention has been paid in the selection of shared vocabularies thus allowing a plain integration with other projects such as *Cultura Italia*, *Europeana*, *Musei D'Italia*, *Internet Culturale* and *Sistema Archivistico Nazionale*. Thanks also to its links to DBpedia, Schema.org, FOAF and GeoNames, Puglia Digital Library can be considered as a new player in the Linked Open Data cloud.

1 Introduction

A Digital Library (DL) is a collection of digital resources (text, visual and audio material, etc.) which are stored in knowledge bases or databases. It has to provide means to make easy the storage and retrieval of media in the collection, as well as to link resources among various digital libraries thus allowing the sharing of knowledge among different providers. As of today, we have many examples of Digital Libraries maintained by institutions or organizations (e.g. Europeana,[5] World Digital Library[6]), by libraries (National Science Digital Library[7]) or academic institutions (Harvard University Library[8]).

In order to be as effective as possible in the whole document management chain, DLs are designed as very complex information systems which include

[5] http://europeana.eu/portal/
[6] https://www.wdl.org
[7] https://nsdl.org/
[8] http://library.harvard.edu/

digital document preservation, distributed database management, information filtering, information retrieval, intellectual property right management, query answering, resource discovery and selective dissemination of information [8]. The research efforts of the last years have been focused on ways to associate metadata to resources stored in DLs with the aim to provide an easy cataloging and browsing of the collections themselves.

More recently, the Linked Data initiative [10] has gained momentum as a set of best practices for publishing and connecting structured (open) reusable data on the Web [11]. In this respect, the Linked Open Data (LOD) initiative meets the need for a broader cooperation among different DLs, supporting the sharing of knowledge and resources among them and then a conceptual shift from document-centric to data-centric and metadata-based approaches [1]. Indeed, the reuse of knowledge coming from other repositories can be effective only if data are provided with common and shared metadata. Metadata are a key element in the digital library domain [13]. Indeed, cultural heritage institutions (museums, archives, libraries) use metadata, as well as thesauri to describe objects in their collections. Linking these metadata with datasets in the Linked Data cloud (GeoNames, DBpedia, FOAF, ecc.) greatly improves reusability and integration of diverse Digital Libraries [3]. The effectiveness in the use of LOD datasets to annotate digital resources in a DL is also witnessed by some successful use cases such as the one of the German National Library[9] or the British National Library[10] as well as the case of the National Library of Spain[11] and of the National Library of France.[12] Moreover, many cultural-heritage institutions have started to explore the benefits of LOD as a means for resource discovery for their hidden treasures [12].

There are many benefits in using linked metadata for DLs, among these: metadata openness and sharing, easiness in information discovery, identification of resource usage patterns, facet-based navigation and metadata enriched with links [1]. This latter has a very important role as makes the user able to navigate among DLs and external information providers. If all the DLs adopted the Linked Data principles they would play a dominant role in the Linked Data cloud as they store a great amount of legacy bibliographic and authority-list data [1]. In order to support the widely adoption of LOD in DLs, some general approaches to data management and transition from metadata to triples needs to be explored; this results in a fundamental shift in metadata design and development that has important implications for controlled vocabularies in terms of data cleanup and preparation [12].

In this paper we present the *Puglia Digital Library*[13] project (*PugliaDL*), whose digital resources are described according to standard vocabularies and can be exposed in the LOD cloud as they follow the well-known Linked Data

[9] `http://www.dnb.de/EN`
[10] `http://www.bl.uk/bibliographic/datafree.html`
[11] `http://datos.bne.es/`
[12] `http://data.bnf.fr/`
[13] `http://www.pugliadigitallibrary.it/`

principles.[14] The remainder of this paper proceeds as follow. Section 2 reviews some relevant work in the field of DLs focused on linking, sharing and reuse of data. Section 3 describes all the aspects related to *PugliaDL* including its Linked Data model. Section 4 is devoted to show some examples of resources available in *PugliaDL*. Conclusion closes the paper.

2 Related Work

In the last decades DLs have become very popular and several scientific investigations and practical efforts were put in place on them: there has been a considerable amount of effort in designing vocabularies, metadata standards, and thesauri to annotate DLs objects, so this section cannot be exhaustive.

Thousands of DLs are emerging around the world, crossing all disciplines and media, some are small community-based initiatives, while others are managed by e.g. public institutions, as national libraries offering a wide range of cultural treasures in multiple media [8]. Finding a unique definition of DL is almost impossible, as, during the years, different definitions has been proposed in the literature. The Digital Library Federation (DLF) puts more emphasis on the organizational aspects, stating that DLs are organizations that provide the resources to select, offer intellectual access to, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are available to a community or set thereof.[15] Borgman [4] states: "*DLs are a set of electronic resources and associated technical capabilities for creating, searching and using information. The content of DLs includes data and metadata*". The latter definition is important as it puts the accent on the presence of metadata, which are fundamental to classify and retrieve information in DLs.

At the beginning of the 2000's in the effort of managing and organizing DLs, several initiatives emerged as OCLC,[16] a global library cooperative with thousands of library members in more than 100 countries. Several DL Directories exist, among the others, the Library of Congress,[17] the Digital Library Directory[18] and the Alexandria Digital Research Library.[19] Europeana [9] is a DL linking more than 40 million digital items from the cultural and heritage domain (artworks, books, video, artefacts, sounds) that have been digitized throughout Europe. Europeana is a large-scale aggregator where the original data is abstracted to a common format and schema [3]. Europeana is a collector of digital objects, it does not store the digital content, but it just collects metadata about the items. Indeed, Europeana collects metadata about digitized content of over 3300 Europes galleries, libraries, museums, archives, etc. When the users find a content of interest, they are linked to the original site (content provider) that

[14] `https://www.w3.org/DesignIssues/LinkedData.html`
[15] DLF, April 21, 1999
[16] `https://www.oclc.org/`
[17] `https://www.loc.gov/collections`
[18] `http://www.digitallibrarydirectory.com/`
[19] `http://www.alexandria.ucsb.edu/`

holds the content itself, e.g. a museum, a library, a regional archive. To make the information searchable using metadata, initially, an extended Dublin Core model was used, named Europeana Semantic Elements. Now this has been superseded by a richer metadata standard named Europeana Data Model[20] (EDM), based on Semantic Web languages (OWL, RDF). EDM incorporate community standard such as LIDO[21] for museum, EAD[22] for archives or METS[23] for digital libraries. The EDM has also been used by other DLs. The Europeana approach, on the one side ensures great consistency and interoperability. Unfortunately, on the other side it may lose the richness of the original data [3].

The effort for producing standardized metadata to catalogue DLs started in the nineteenth century with regional and international consortia that tried to institute rigorous cataloguing principles and rules, just to cite a few AACR,[24] MARC,[25] ISBD,[26] FRBR,[27] RDA[28] [1]. Metadata standards such as FRBR and RDA are more devoted to human consumption rather than machine processing, indeed when these metadata are implemented using a technical format like MARC they show problems of metadata duplication, data inconsistency, lack of granularity and complexity [6, 1]. The solution to shift from a document-centric view (typical of the previous standards) to a data-centric view lies down in the adoption of LOD for metadata modelling, encoding, representation and sharing. The use of LOD is also justified by the need to make DLs freely and openly accessible, other than in a shareable, extensible and re-usable format[14]. In this direction, the Library of Congress and the Stanford University Libraries[29] have paved the way since 2011, followed by the Europeana initiative and the British Library, that have both developed a semantic metadata model compliant with LOD specifications [5, 9].

The task of search and discovery in DLs has often been faced using metadata describing information objects (e.g., documents). As the one of the Dublin Core Metadata initiative (DCMI) [7] focused on developing small usable set of vocabulary terms that can be used to describe the essential features of web resources (video, images, web pages, etc.), as well as physical resources like artworks, books, ecc. The Dublin Core Metadata can be used for both resource descriptions as well as to combine various metadata standards with the aim to provide interoperability among the metadata vocabularies in the LOD cloud. The current set of the Dublin Core vocabulary is composed by the DCMI Metadata Terms, all these terms are defined as RDF properties.

---

[20] `http://pro.europeana.eu/page/edm-documentation`
[21] `www.lido-schema.org/`
[22] `http://www.loc.gov/ead/`
[23] `http://www.loc.gov/mets/`
[24] Anglo-American Cataloguing Rules, 1967
[25] MAchine-Readable Cataloguing, 1960
[26] International Standard Bibliographic Description for Monographic Publications,1971
[27] Functional Requirements for Bibliographic Records, 1996
[28] Resource Description and Access, 2010
[29] `http://library.stanford.edu/`

3 Puglia Digital Library

Puglia Digital Library (*PugliaDL*) was conceived with the aim to preserve the memory of the regional heritage and to enable its sharing and reuse. For this reason *PugliaDL* wants to become a producer and supplier of LOD related to the regional heritage, making available data that usually are hard to find. The *PugliaDL* is a multimedia archive of books, magazines, newspapers, photographs, sounds and audiovisual materials, museum objects, historical and artistic sites, etc. Having in mind a DL were the information is open and shared, a key role is played by the services that allow the sharing of knowledge from and to national and international aggregators using standard metadata and LOD as access point to the DL. At the moment *PugliaDL* collects about 40 digital collections and 1700 resources organized with respect to three main levels: *topic*, *collection*, *digital resource*. *PugliaDL* mainly exposes its data on the Web in an ad-hoc Web portal where each digital resource has a preview (sound or picture), a brief description, a summary sheet, a map with the resource location, and can be downloaded. For what concerns books, it is possible to virtually browse them page by page; in this way it is possible to consult any book, even fragile and precious ones. In describing resources, a set of standard metadata and controlled vocabularies is used in order to favor interoperability. The *PugliaDL* is the only digital library in Italy that can interact with other systems, as Cultura Italia[30] (and through this with Europeana), Musei d'Italia,[31] Internet Culturale,[32] SAN-Sistema Archivistico Nazionale.[33] The sets of metadata used by the *PugliaDL* are depicted in Table 1.

The controlled vocabularies are:

- DCMIType (DCMI Type Vocabulary)
- PICO Thesaurus (that allow interoperability with Cultura Italia)
- Vocabularies ICCD (Italian standard for cataloging)
- AAT (Art & Architecture Thesaurus)
- TGN (Thesaurus of Geographic Names)

3.1 *Puglia Digital Library*: data modeling

The Puglia Digital Library is the pilot project in the Puglia region in the field of LOD. With reference to the 5-stars model proposed by Tim Berners-Lee[34], the *PugliaDL* publishes its data both in 3- and 5-stars level. The choice is justified by the fact that data published as CSV or XML can be exploited also by users that do not have knowledge about RDF and SPARQL. The data of *PugliaDL* are in the 5-stars level (LOD) as data are linked in the Semantic Web with external sources (DBpedia, GeoNames, etc.). Before publishing any data, the first step

[30] `http://www.culturaitalia.it`
[31] `http://www.culturaitalia.it/opencms/museid/index_museid.jsp?language=en`
[32] `http://www.internetculturale.it/opencms/opencms/it/`
[33] `http://san.beniculturali.it`
[34] `https://www.w3.org/DesignIssues/LinkedData.html`

DC	Dublin Core Metadata Initiative
DDI	Data Documentation Initiative
EAC-CPF	Encoded Archival Context - Corporate Bodies, Persons, and Families
EAD	Encoded Archival Description finding aid
FGDC	Federal Geographic Data Committee metadata
ISO 19115 2003 NAP	North American Profile of ISO 19115:2003 descriptive metadata
LC-AV	Technical metadata specified in the Library of Congress A/V prototyping project
LOM	Learning Object Model
MARC	MAchine-Readable Cataloging
MAG	Administrative and Management Metadata
METSRIGHTS	TSRights Declaration Schema
MODS	Library of Congress Metadata Object Description Schema
NISOIMG	NISO Technical Metadata for Digital Still Images (MIX)
PREMIS	PREservation Metadata: Implementation Strategies
TEIHDR	Text Encoding Initiative Header
TEXTMD	textMD Technical metadata for text
VRA	Visual Resources Association Core

Table 1: Sets of metadata used by the *Puglia Digital Library*

is their analysis with the aim to identify the data that should be published by following the criteria of integrity, coherence and completeness. The editorial staff of *PugliaDL* has defined some publishing rules, among these the fact that each *collection* should contain at least 5 digital resources (***dr***) and that to each ***dr*** the following metadata should be valued and published in an open format:

- **Identifying data**: title, ID, description, collection, subject, etc.
- **Category data**: genre, topic, etc.
- **Resource data**: title, author, chronological information, etc.
- **Geographical data**: location, geographic coordinates, curator, etc.
- **Technical data**: digital format, quality, etc.
- **Data on accessibility**: the copyright holder, license, etc.

All the data used to describe a ***dr*** are released using a CC0 license.[35] In the publishing phase the set of metadata is chosen with reference to the Italian legislation[36] rules. The *collections* of the *PugliaDL* are also published on the Open data portal of the Puglia Region.[37] One of the added values of using Linked Data technologies in a DL is the possibility to place a particular cultural asset in a proper context. Indeed, it is not possible to describe a cultural asset without reference to its historical period and style, as well as, its geographic location.

[35] `https://creativecommons.org/publicdomain/zero/1.0/`

[36] *Italian National guidelines for the valorization of public sector information* published by the Agency for Digital Italy (AgID 2014)

[37] `http://dati.puglia.it/`

Consider for instance Medieval art. There are different styles of Romanesque in different Italian regions: the Apulian Romanesque as well as the Lombard and Pisan Romanesque, and each one has its own characteristics. Thanks to Linked Data we can capture all these geographical and historical relations.

3.2 Resource annotation in *Puglia Digital Library*

Each resource in the system is described in accordance with existing ontologies, to make the system itself highly interoperable. The ontology chosen to describe such resources is CIDOC-CRM,[38] developed by ICOM,[39] which is the leading conceptual model for the heritage sector. Specifically its OWL implementation (Erlangen CRM/OWL[40]) has been adopted as more suitable for DLs. Such an ontology is able to model the interweaving of semantic relations among temporal and geographical dimensions, people, material and immaterial object descriptions. As a way of example, a video about the *Divine Comedy* can be linked to the video resource, or to its theatrical opera staged at the *Nuovo Teatro Verdi* located in *Brindisi* and to its director *Eimuntas Nekrošius*.

Furthermore, the CIDOC-CRM ontology is highly interoperable as it is mapped to several other ontologies.[41] *PugliaDL* itself uses standards that are all mapped to such ontologies: Dublin Core [7], EAD,[42] VRA,[43] ICCD,[44] EDM[45] and PICO.[46]

The PICO Application profile mapping[47] has been the main reference document to create the RDF annotations in *PugliaDL*. Indeed, this allows to exchange data with *Cultura Italia* and, through that, with *Europeana*. However, CIDOC-CRM has some limitations in describing multimedia resources, *e.g.* it does not have properties to describe technical details, like data rate mode, mimeType, channel configuration, etc. For this reason, *PugliaDL* also adopted other ontologies to overcome such limits, including Dublin Core, DBpedia,[48] Schema.org,[49] FOAF[50] and SKOS.[51]

In the following example one of the limitations of CIDOC is highlighted, in the representation of the asset location:

```
<crm:P53_has_former_or_current_location>
```

[38] http://www.cidoc-crm.org/
[39] http://icom.museum/
[40] http://erlangen-crm.org/
[41] http://www.cidoc-crm.org/crm_mappings.html
[42] https://www.loc.gov/ead/
[43] https://www.loc.gov/standards/vracore/
[44] http://www.iccd.beniculturali.it
[45] http://pro.europeana.eu/page/edm-documentation
[46] http://www.culturaitalia.it/opencms/export/sites/culturaitalia/attachments/documenti/picoap/picoap1.0.xml
[47] http://www.culturaitalia.it/opencms/documentazione_tecnica_en.jsp
[48] http://dbpedia.org
[49] http://schema.org
[50] http://www.foaf-project.org/
[51] https://www.w3.org/2004/02/skos/

```
      <crm:E53_Place>
         <crm:P87_is_identified_by>
            <crm:E48_Place_Name>
              <rdf:value>Bari</rdf:value>
            </crm:E48_Place_Name>
         </crm:P87_is_identified_by>
      </crm:E53_Place>
</crm:P53_has_former_or_current_location>
```

In such a description it is impossible to distinguish among *Municipality*, *District* or *Region*. While by integrating the DBpedia ontology it is possible to specify a property for each geographic area: `dbp:locationCity`, `dbo:province`, `dbo:region`, `dbo:address`. At the same time, Schema.org provides specific vocabularies for data with reference to audio, video, images, texts. FOAF is useful to model data about authors of resources. Finally, geospatial information are described through **LinkedGeoData**[52] and **Geonames**.[53]

3.3 Linking *Puglia Digital Library* to the Linked Data cloud

The digital resources of *PugliaDL* have been linked to external vocabularies and datasets of the Linked Data cloud to enrich the information provided with each resource. **DBpedia** is our main reference point. *PugliaDL* URIs are linked to DBpedia resources, especially for what concerns *category data* and *geographical data*, other than resource name, whenever present. Often, the author's name was not present in DBpedia, especially for local artists; for this reason we created an Authority File to be linked to VIAF[54] (Virtual International Authority File). **VIAF** is an international standard, accessible in RDF, that collects records coming from several authority files and gives them URIs, supporting in this way the search for author names independently of the language or of the alphabet. For what concerns geographical data, other than DBpedia, we linked our resources also with LinkedGeoData a broad geographical RDF knowledge base, based on data from Open Street Map, and interconnected with DBpedia and GeoNames. In the near future, we want to also integrate the service **Linked Open Street Map** (LOSM) [2] in *PugliaDL*.

With the aim to enhance the semantics of the data, we use controlled vocabularies to disambiguate terms in dependence of the context and to resolve synonymy and homonymy. For this reason we link to vocabularies of the **Getty Research Institute**[55] that, recently, have been released as Linked Open Data.[56] The Getty Research Institute vocabularies are constantly updated and compliant with international cataloging standards (e.g., VRA, CIDOC CRM, CCO,[57] etc.). They use a specific terminology for cultural and bibliographic fields and

[52] `http://linkedgeodata.org/`
[53] `http://www.geonames.org/ontology`
[54] `https://viaf.org/`
[55] `http://www.getty.edu/research/`
[56] `http://www.getty.edu/research/tools/vocabularies/lod/`
[57] `http://cco.vrafoundation.org/`

are accessible via SPARQL endpoints.[58] At the moment the vocabularies available as LOD are: **Art & Architecture Thesaurus** (AAT), **Union List of Artists Names** (ULAN) and **Getty Thesaurus of Geographic Names** (TGN). These vocabularies have been of great use as they are inter-linked, share the same data structure, are multilingual, and each term is identified by an ID[59].

Another vocabulary is the **PICO Thesaurus**, created by Cultura Italia. It is multilingual, compliant with RDF and SKOS, released in LOD format and it allows one to classify each resource with respect to its specific cultural domain. Each term in the vocabulary is identified by an URI. In the context of *PugliaDL* it has been mainly used to link keywords that identify the digital resource from a contextual and temporal point of view.[60] The interoperability is obtained thanks to the *Open Archives Initiative Protocol for Metadata Harvesting*[61] (OAI-PMH) that supports repository interoperability. Data Provider repositories expose structured metadata via OAI-PMH, while Service Providers make OAI-PMH service requests to harvest that metadata. Thanks to OAI-PMH protocol *PugliaDL* is linked to Europeana and to the LOD resources of the aforementioned LOD portals.

From a technological point of view, *PugliaDL* uses *OpenLink Virtuoso* ad triple-store and *LodView* as RDF viewer and IRI dereferencer. In Figure 1, the resource *"Masseria Aprile - Spazio Interno"* belonging to *PugliaDL* is depicted in its HTML rendering by LodView.

4 *PugliaDL* Linked Data in action

In this section we highlight how the *PugliaDL* is highly inter-operable thanks to properties that allow the linking with external dataset.

Content type. In order to model the type of content of digital resources (e.g. manuscript, film poster, book, farm etc.) we can use the following properties:[62] `crm:P2_has_type`, `dcterms:type`, `schema:category`, `dbo:category`, `dbp:category`. The values of such properties are linked both to the terms of ***Art & Architecture Thesaurus (AAT)*** and to ***DBpedia***. As a way of example, let us look at the *PugliaDL* resource *"Il Gattopardo"*; the digital resource is the film poster of the famous 1963 movie *"The Leopard"*.

```
<http://dati.puglia.it/resource/DigitalLibrary/Il_Gattopardo> a crm:E38_Image,
rdfs:Resource , schema:ImageObject , dbo:Image , dbp:Image ;
    rdfs:label "Il Gattopardo" ;
     crm:P2_has_type <http://it.dbpedia.org/resource/Poster> ,
<http://vocab.getty.edu/aat/300027221> ;
```

[58] `http://vocab.getty.edu/sparql`

[59] As a way of example, see the resource *poster* at `http://vocab.getty.edu/aat/300027221`

[60] As a way of example the *10th cent. A.D.* is linked with the URI `http://www.culturaitalia.it/pico/thesaurus/4.1#http://culturaitalia.it/pico/thesaurus/4.1#sec_x_d_c`

[61] `https://www.openarchives.org/OAI/openarchivesprotocol.html`

[62] All the prefixes we use are from `http://prefix.cc`.

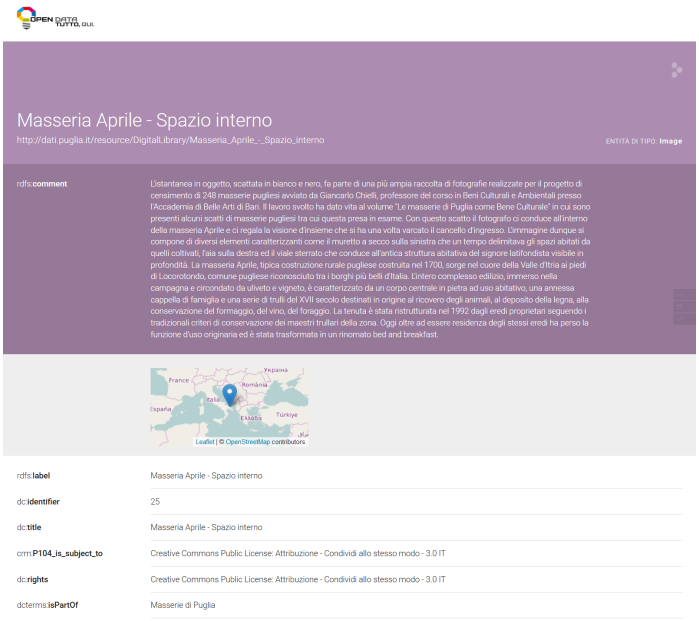

Fig. 1: An example of resource available in *PugliaDL* and displayed via LodView

```
    dbo:category <http://it.dbpedia.org/resource/Poster> ,
<http://vocab.getty.edu/aat/300027221> ;
    schema:category <http://it.dbpedia.org/resource/Poster> ,
<http://vocab.getty.edu/aat/300027221> .
```

The above example clearly shows that we preferred to add redundant properties in order to ease the reuse and a zero-effort interoperability with other datasets. Whenever possible we added both redundant properties and redundant resources (even when they could be connected via a `owl:sameAs` relation).

Resource author. As for the modeling of *authors* and their data we may use the following properties: `crm:P14_carried_out_by`, `dc:creator`, `dcterms:creator`, `schema:author`, `dbo:author`, `dbp:author`, `foaf:maker`.

Geographical data. Data about cities are modeled via properties linked to ***DBpedia*** and ***LinkedGeoData*** as well as to ***Getty Thesaurus of Geographic Names*** (TGN), among these: `dcterms:spatial`, `dbo:locationCity`, `dbo:city`, `dbp:locationCity`, `dbp:city`. The play *Macbeth* by *Francesco Maria Piave* will be described together with geographical information as:

```
<http://dati.puglia.it/resource/Macbeth> a crm:E33_Linguistic_Object,
 rdfs:Resource , schema:CreativeWork , dbo:WrittenWork ;
   rdfs:label "Macbeth" ;
   crm:P14_carried_out_by "Francesco Maria Piave" ;
   dc:creator "Francesco Maria Piave" ;
   schema:author "Francesco Maria Piave" ;
   dbo:author "Francesco Maria Piave" ;
   dbp:author "Francesco Maria Piave" ;
```

```
   foaf:maker "Francesco Maria Piave" ;
   dcterms:spatial
    <http://linkedgeodata.org/page/triplify/node1699232800> ,
    <http://vocab.getty.edu/tgn/7004105> ,
    <http://dbpedia.org/page/Barletta> , <http://it.dbpedia.org/resource/Barletta> ;
   dbo:locationCity
    <http://linkedgeodata.org/page/triplify/node1699232800> ,
    <http://vocab.getty.edu/tgn/7004105> ,
    <http://dbpedia.org/page/Barletta> ,  <http://it.dbpedia.org/resource/Barletta> ;
   dbp:city <http://linkedgeodata.org/page/triplify/node1699232800> ,
    <http://vocab.getty.edu/tgn/7004105> ,
    <http://dbpedia.org/page/Barletta> ,
    <http://it.dbpedia.org/resource/Barletta> .
```

Data about the *Region* is modeled by using the following properties: `dbo:region`, `dbp:region`, `dcterms:spatial`. Then, with reference to the digital resource *"Il Gattopardo"* we can add geographical information as in the following:

```
<http://dati.puglia.it/resource/DigitalLibrary/Il_Gattopardo> a crm:E38_Image,
rdfs:Resource , schema:ImageObject , dbo:Image , dbp:Image ;
rdfs:label "Il Gattopardo" ;
crm:P2_has_type <http://it.dbpedia.org/resource/Poster> ,
<http://vocab.getty.edu/aat/300027221> ;
dbo:category <http://it.dbpedia.org/resource/Poster> ,
<http://vocab.getty.edu/aat/300027221> ;
schema:category <http://it.dbpedia.org/resource/Poster> ,
<http://vocab.getty.edu/aat/300027221> ;
dcterms:spatial
    <http://it.dbpedia.org/resource/Bari> , <http://dbpedia.org/page/Bari> ,
    <http://it.dbpedia.org/resource/Puglia> , <http://dbpedia.org/page/Apulia> ,
    <http://vocab.getty.edu/tgn/7010380> ,
    <http://linkedgeodata.org/page/triplify/node2315669726> ;
dbo:region
    <http://it.dbpedia.org/resource/Puglia> , <http://dbpedia.org/page/Apulia> ,
    <http://vocab.getty.edu/tgn/7010380> ,
    <http://linkedgeodata.org/page/triplify/node2315669726> ;
dbp:region
    <http://it.dbpedia.org/resource/Puglia> , <http://dbpedia.org/page/Apulia> ,
    <http://vocab.getty.edu/tgn/7010380> ,
    <http://linkedgeodata.org/page/triplify/node2315669726> .
```

Finally, with reference to the properties to model keywords that identify the resource we use the following properties that are linked to ***Pico*** Thesaurus: `crm:P2_has_type`, `dc:subject`, `dcterms:subject`, `schema:category`, `dbo:category`, `dbp:category`.

```
<http://dati.puglia.it/resource/DigitalLibrary/Il_Gattopardo> a crm:E38_Image ,
 rdfs:Resource , schema:ImageObject , dbo:Image , dbp:Image ;
  rdfs:label "Il Gattopardo" ;
  crm:P2_has_type <http://culturaitalia.it/pico/thesaurus/4.1#cinema> ;
  dc:subject <http://culturaitalia.it/pico/thesaurus/4.1#cinema> ;
  schema:category <http://culturaitalia.it/pico/thesaurus/4.1#cinema> ;
  dbo:category <http://culturaitalia.it/pico/thesaurus/4.1#cinema> ;
  dbp:category <http://culturaitalia.it/pico/thesaurus/4.1#cinema> .
```

5 Conclusion

We presented *Puglia Digital Library* project, whose main aim is to preserve the memory and the beauty of the Puglia region heritage as well as to promote its sharing and reuse. All the items in the collections available in *PugliaDL*

have been described using also RDF annotations in order to be part of the LOD cloud. There are many benefits in using linked metadata for DLs; among these, the easiness of the process of information discovery and sharing. Indeed, the *PugliaDL* collections, thanks to their semantic annotations, can be easily integrated in other DLs projects such as Europeana. The *PugliaDL* dataset has been designed having interoperability in mind and this is the main reason why more than one standard vocabulary has been adopted to model properties and resources of the RDF triples.

Acknowledgements The authors acknowledge partial support of PON03PE_00136_1 Digital Services Ecosystem: DSE and Progetto Corvallis.

References

1. Alemu, G., Stevens, B., Ross, P., Chandler, J.: Linked Data for libraries: Benefits of a conceptual shift from library-specific record structures to RDF-based data models. New Library World 113(11/12), 549–570 (2012)
2. Anelli, V.W., Cali, A., Di Noia, T., Palmonari, M., Ragone, A.: Exposing Open Street Map in the Linked Data cloud. In: The 29th International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems (2016)
3. de Boer, V., Wielemaker, J., van Gent, J., Hildebrand, M., Isaac, A., Van Ossenbruggen, J., Schreiber, G.: Supporting Linked Data production for cultural heritage institutes: the Amsterdam Museum case study. In: The Semantic Web: Research and Applications, pp. 733–747. Springer (2012)
4. Borgman, C.L.: What are digital libraries? Competing visions. Inf. Process. Manage. 35(3), 227–243 (1999)
5. Free data services. Tech. rep., British Library (2011), `http://www.bl.uk/bibliographic/datafree.html`
6. Coyle, K., Hillmann, D.: Resource description and access (RDA): Cataloging rules for the 20th century. D-Lib Magazine 13(1/2) (2007)
7. Dublin Core metadata element set, version 1.1. DCMI recommendation, Dublin Core Metadata Initiative (2012), `http://dublincore.org/documents/dces/`
8. Fox, E.A., Marchionini, G.: Toward a worldwide digital library. Communications of the ACM 41(4), 29–32 (1998)
9. Gradmann, S.: Knowledge = information in context: on the importance of semantic contextualisation. In: Europeana White Paper. pp. 1–19. No. 1, Berlin School of Library und Information/Humboldt Universität zu Berlin (2010)
10. Heath, T., Bizer, C.: Linked Data: Evolving the Web into a global data space. Synthesis lectures on the Semantic Web: theory and technology 1(1), 1–136 (2011)
11. Lampert, C.K., Southwick, S.B.: Leading to linking: Introducing Linked Data to academic library digital collections. Journal of Library Metadata 13(2-3), 230–253 (2013)
12. Southwick, S.B., Lampert, C.K., Southwick, R.: Preparing controlled vocabularies for Linked Data: Benefits and challenges. Journal of Library Metadata 15(3-4), 177–190 (2015)
13. Tani, A., Candela, L., Castelli, D.: Dealing with metadata quality: The legacy of digital library efforts. Information Processing & Management 49(6), 1194–1205 (2013)
14. Library Linked Data incubator group final report. Tech. rep., W3C (2011), `http://www.w3.org/2005/Incubator/lld/XGR-lld/`