# A Linked Open Data service for performing arts

Pierfrancesco Bellini, Paolo Nesi

Distributed Systems and Internet Technology Lab, DISIT, http://www.disit.dsi.unifi.it
Dipartimento Ingegneria dell'Informazione, University of Florence, Italy
pbellini@dsi.unifi.it, paolo.nesi@unifi.it

**Abstract.** Linked Open Data (LOD) is a new way of sharing information about digital/physical resources allowing connected computers to better use information related with the resources. LOD allows to enrich information about resources possibly improving the user experience when using the resources or on finding them. This paper presents the experience in publishing as linked data the information which is present on ECLAP portal about multimedia content on performing arts. The system provides access to information about content, the terms of the taxonomy used to classify the content and also structural information like connections with groups managing the content, use in playlists and collections. Moreover information about annotations on audio/visual content is provided and also information about users is available (e.g., the 'friends' graph). The enrichments made on geographical information present in the content metadata (e.g., city/country of the performance) allowed to link content with the GeoNames database that is available as liked data.

**Keywords:** Linked open data, performing arts, metadata enrichment

## 1      Introduction

The Web is a  big source of potentially relevant background information, for example Wikipedia or WordNet provide detailed information about terms, concepts and relations among them, that can be used to enhance existing content metadata descriptions and improve the user experience in the access or while searching content [1][2]. However in the Web of Documents the information is not so easy to be extracted from HTML pages, for this reason in recent years the Web of Data is born to provide a machine friendly representation of the information on the Web.

Linked Data is a technique for data publishing that uses common Web technologies to connect related data and make them accessible on the Web. It is based on identifying resources with HTTP Uniform  Resource Identifiers (URI), and, using standards like the Resource Description Framework (RDF) to provide data about these resources and to connect them to other resources on the Web [3].

Regarding the description of resources the best practices for publishing linked data suggest to reuse vocabularies already available. Reuse can be done by using classes and properties as they are or by creating a specific vocabulary and defining sub clas-

ses and sub properties of the ones already defined. Some well known basic vocabularies are present:

- *Dublin Core* (http://purl.org/dc/terms/) for the description of human-created artifacts,
- *Friend of a Friend* (http://xmlns.com/foaf/1.0/) for the description of persons, organization and relations among them,
- *Creative Commons* (http://creativecommons.org/ns#) for the representation of legal information about works,
- *Basic Geo Vocabulary* (http://www.w3.org/2003/01/geo/wgs84_pos#) for basic properties for the representation of geographical coordinates.

In the field of performing arts, there are some contributions, not all the aspects are covered by a single vocabulary. The *Music Ontology* [4] used by BBC programmes and music [5] and DBtune covers only the music related information, the *Linked Movie Database* has a vocabulary specific for the film domain and other ontologies like DBpedia [6] and Freebase are quite generic.

Another relevant aspect is the description of annotations of multimedia content, the Annotea project [7] was one of the first to adopt semantic web technologies for annotations but it was originally designed for annotations of web sites and therefore offers limited capabilities for annotating multimedia objects. The LEMO annotation framework [8] built on top of Annotea model supports annotations of media fragments [9]. Recently the Open Annotation Collaboration [10] model has been proposed and it is designed for the use as linked data.

The paper is organized as follows: Section 2 presents and overview of the ECLAP portal and project, Section 3 presents the ECLAP Semantic model describing the entities and the relations supported. Section 4 describes the linked data implementation, in Section 5 a tool used for the display and navigation of the semantic information is presented. Conclusions are drawn in Section 6.


## 2    ECLAP

ECLAP is both a Best Practice Network and provider of Content and User Services. As **Best Practice Network**, ECLAP consists of working groups that analyze the state of the art and identify the best practices and guidelines to cope with technical and strategic problems of the performing arts sector. To this end, three main ECLAP Working Groups (with corresponding blogs and forums) have been set up to cover the areas of: digital libraries and models for performing arts content, intellectual property management and tools, and digital content based tools for teaching and learning of performing arts in the new era. For facilitating the networking and discussions, ECLAP is also a repository of technical documents, demonstrators, best practices and standards that can be used to better understand the sector problems and find corresponding guidelines, state of the art solutions as well as future activities and project proposals.

The **ECLAP Content and User Service** exploits the use of advanced social media and semantic computing technologies and solutions for the content enrichment, aggregation and distribution of rich multilingual performing art content towards PCs and mobiles. Presently, ECLAP distributes more than 110,000 distinct objects (video, audio, images, texts, 3D), from 31 content providers, in up to 13 major metadata languages, towards a community of about 2000 registered users, world-wide distributed.

## 3   ECLAP Semantic model

In Figure 1, the general ECLAP semantic model is shown where almost all the mayor entities managed by the system are reported.
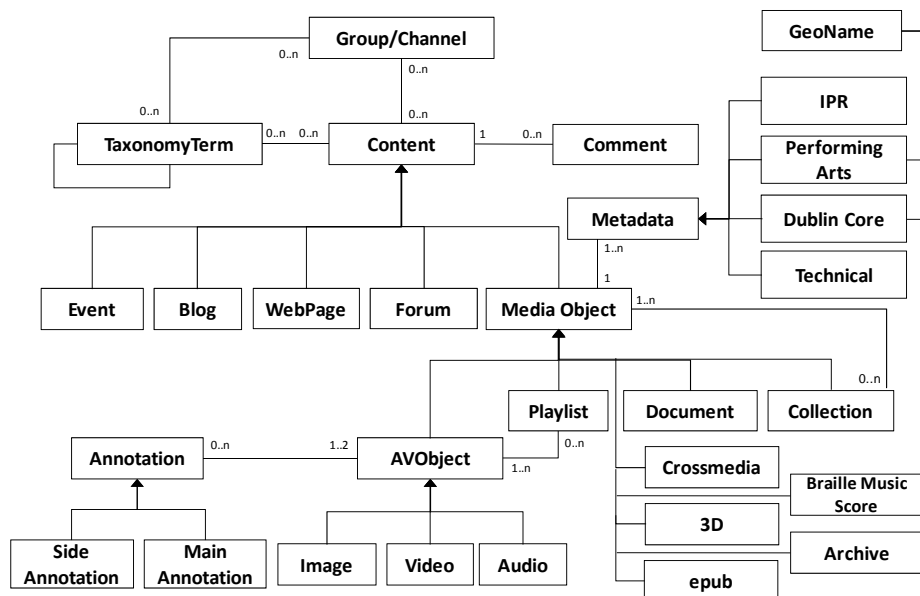


**Fig. 1.** ECLAP Semantic model

The *Content* elements that represent the content managed by the portal are associated with *Groups/Channels* providing the content (each ECLAP content provider has a group). Moreover each content can have *Comment*s and can be associated with terms taken from a taxonomy. *Content* is specialized in *Events*, *Blogs*, *WebPage*, *Forum* and *Media Objects*. Blogs, WebPages and Forums are used for providing news, general unstructured information and for stimulating users discussions on specific topics while *Media Objects* represent the multimedia content and their aggregations that are accessible from the ECLAP portal. The *Media Objects* are specialized in *AVObjects* (Image, Video, Audio) that can be used in *Annotations* and in *Playlists*. *Playlists* aggregate *AVObjects* in a sequence allowing to use even a fragment of the Audio/Video. *Collections* aggregate a set of *Media Objects* and in this case can in-

clude also *Documents*, *Playlists* and also other *Collections*. *Annotations* are used to associate a textual description with an audio visual object or to its fragment, moreover it is possible to associate with another audio visual object (or its fragment) allowing to link two audio visual contents or even different parts of the same content. This kind of annotations are presented to the user via the MyStoryPlayer tool [11][12].

In Fig. 2, the relations among *Users* and other entities are depicted. A user may be a member of one or more groups and he/she can be the group administrator. Each content is provided by a user and each *Media Object* may be marked as favorite (similar to the facebook "Like") by an user, moreover a group administrator can mark the content as to be featured on the featured object list on the portal. *Comments* and *Annotations* are linked to the user that created them. Finally users are linked with other users with the 'knows' relation that builds the classical 'social graph' and each user can specify the topic of interests among the taxonomy terms.

Each *Media Object* is associated with different sets of metadata, the general *Dublin Core* metadata (e.g. title, subject, type, description), the *Technical* metadata related to the content and its management (e.g. audo/video duration, workflow type), the metadata per IPR management and specific metadata for performing arts information (e.g., performance place, performance date, performing arts type, etc.). To be noted that some of the Dublin core and performing arts metadata elements (e.g. coverage, spatial, performance place, performance city and country) can be associated with GeoNames entities to allow to link to the GeoNames linked data service.
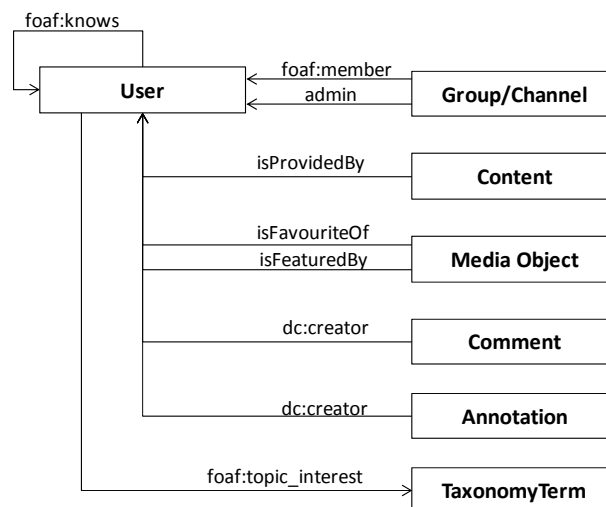


**Fig. 2.** Relations of users with other entities

In Table 1, some of the properties specifically defined for performing arts are reported. These properties are defined as specialization of Dublin core properties and they were identified analyzing the metadata schemas used by ECLAP partners as well as schemas used by other projects and metadata standards. Among the properties identified information about the performance depicted in the resource (place, city,

country and date), the premiere of the performance (place, city, country and date), the contributors to the creation of the performance each with the specific role (actor, dancer, light designer, hairdresser) is present.

**Table 1.** Some performing arts metadata properties

| property | sub property of | description |
| --- | --- | --- |
| performancePlace | dcterms:spatial | theatre or venue of the performance |
| performanceCity | dcterms:spatial | city of the performance |
| performanceCountry | dcterms:spatial | country of the performance |
| performanceDate | dcterms:issued | date of the performance |
| firstPerformancePlace | dcterms:spatial | venue of the premier of the performance |
| firstPerformanceCity | dcterms:spatial | city of the premier |
| firstPerformanceCountry | dcterms:spatial | country of the premier |
| firstPerformanceDate | dcterms:issued | date of the premier |
| plotSummary | dcterms:abstract | summary of the plot |
| performingArtsProfessional | dcterms:contributor | person involved in the performance realization |
| dancer | performingArtsProfessional | |
| actor | performingArtsProfessional | |
| director | performingArtsProfessional | |
| … | | |

A complete description of the ECLAP metadata fields is reported in [13] while indexing is described in [14].

### 3.1 Linking to Geonames dataset

In order to link content record with Geonames, the data fields containing geographical information were analyzed to find matches with names that are present in the Geonames dataset. The fields considered are the (first)performance place, city and country and dublin core spatial and coverage. Since exact matching does not produce sufficient results the matching was done using full text search of the metadata field over the geographical names, the results have been filtered requiring that the words of the matched name have to be present in the metadata field. Moreover, when the country field is indentified for the identification of city or place the search is limited to names of that country. The matching is not perfect and it can be improved and compared with other techniques [15][16][17].

# 4    ECLAP LOD service

The ECLAP portal allows to access RDF descriptions of digital resources that are available on it using specific URIs, the RDF description of the resource is provided in case of a LOD enabled browser otherwise the standard web browsers are redirected to the usual HTML page with a human readable description. Also the taxonomy terms used to classify content are accessible using LOD as well as the content annotations that relate them, the groups to which the content is bound (e.g., the group of the provider) and the ECLAP users.

The URIs currently supported are in the forms:

- http://www.eclap.eu/resource/object/<axoid>
- http://www.eclap.eu/resource/term/<tid>
- http://www.eclap.eu/resource/annotation/<aid>
- http://www.eclap.eu/resource/group/<gid>
- http://www.eclap.eu/resource/user/<uid>

where *<axoid>* is the identifier assigned to the content when uploaded (e.g., urn:axmedis:00000:obj:04e0caef-b33b-4f4a-ba50-a80d96766192), *<tid>* is the identifier of the vocabulary term (e.g., 501 for Dance), *<aid>* is the identifier assigned to the annotation, *<gid>* is the identifier of the group (e.g., 3160 for the Development group) and *<uid>* is the identifier of the user (e.g., 1 is the portal administrator).

The following links are present among:
- the content and the vocabulary terms,
- a content and the aggregated content (e.g., collection, playlist) containing it,
- the content and the groups that are used to provide the content (each ECLAP content provider has a group),
- a content and the annotations that describe it,
- the users and content, groups and annotations,
- the content and the GeoNames vocabulary for the places where performances were held, they are provided as a result of an enrichment made on the metadata.

In Figure 3, an example of how a content is related with vocabulary/taxonomy terms, collections and annotations is reported. For the description of the entities a specific ontology has been designed, this ontology is available as a linked data. All URIs used for properties and classes are dereferenceable  and point to the ontology description (e.g., http://www.eclap.eu/schema/eclap/performancePlace) both as RDF and human readable documentation in HTML.
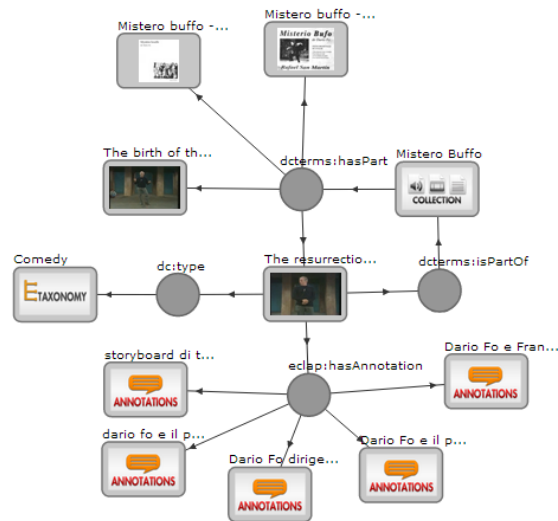
**Fig. 3.** Example relation among a content with collections, taxonomy terms and annotations

## 4.1 Content description

Each content is described using RDF, the fields that are already Dublin Core terms in the ECLAP model are provided as they are, while the specific fields for ECLAP are provided by using specific properties (e.g., *eclap:performancePlace*) that are declared refinements of more generic properties taken from standard schemas (e.g., dcterms:spatial).

The relations with the vocabulary are provided by using specific properties (e.g., *eclap:genre* for the terms of the genre hierarchy) linking the LOD URIs to the terms. Also these properties are declared as sub properties of Dublin Core terms.

The relations with other aggregated content like collections are provided using *dcterms:isPartOf* and *dcterms:hasPart* properties. Relations with the group of the content provider that is giving the content are offered by specific properties, *eclap:isProvidedBy* and *eclap:provides* (both sub properties of *dc:relation*). These relations allows to link all the content, in particular it can be useful for crawlers allowing them to harvest all the content of a provider. The following is an example:

```
<rdf:RDF …>
  <rdf:Description
rdf:about="http://www.eclap.eu/resource/object/urn:axmedis:00000:obj:04..">
    <dc:title xml:lang="en">you PARA | DISO</dc:title>
    <dc:description xml:lang="en">In July 2010 Emio Greco and Pieter C. Scholten
presented their perormance "you PARA | DISO" at Salle Garnier de l'Opéra de Monte-
Carlo. You PARA | DISO is the last performance around Dante's Divina Commedia. …
</dc:description>
    <dc:publisher xml:lang="en">iTheatre</dc:publisher>
    <dc:subject xml:lang="en">dance</dc:subject>
```

```
<dc:format>video</dc:format>
<dc:rights xml:lang="en">erik lint</dc:rights>
<dc:rights xml:lang="en">emo greco &amp; pc</dc:rights>
<dc:creator xml:lang="en">emio greco &amp; pc</dc:creator>
<dc:creator xml:lang="en">erik lint</dc:creator>
<eclap:performancePlace>Salle Garnier de l'Opéra de Monte-
Carlo</eclap:performancePlace>
<eclap:performanceCity>Monte-Carlo</eclap:performanceCity>
<eclap:performanceCountry>Monaco</eclap:performanceCountry>
<eclap:performanceCountry rdf:resource="http://sws.geonames.org/2993457/"/>
<eclap:performanceDate>July 2010</eclap:performanceDate>
<eclap:choreographer>Emio Greco</eclap:choreographer>
<eclap:choreographer>Pieter C. Scholten</eclap:choreographer>
<eclap:historicalPeriod
   rdf:resource="http://www.eclap.eu/resource/term/567"/>
<eclap:performingArtsType
   rdf:resource="http://www.eclap.eu/resource/term/501"/>
<eclap:managementAndOrganization
   rdf:resource="http://www.eclap.eu/resource/term/514"/>
<dcterms:isPartOf
   rdf:resource="http://www.eclap.eu/resource/object/urn:axmedis:00000..."/>
<eclap:hasAnnotation
rdf:resource="http://www.eclap.eu/resource/annotation/SideAnnotation_130..." />
   …
  </rdf:Description>
</rdf:RDF>
```

## 4.2    Taxonomy description

ECLAP provides six thesauri of terms for the classification of content (for a total of 231 terms):

- Subject (e.g., Teaching, Philosophy, Multiculture)
- Genre (e.g., Comedy, Comic, Drama)
- Historical period (e.g., Contemporary, Classical, XX Century)
- Movement and style (e.g., Experimental, Theatre of the absurd)
- Performing arts type (e.g., Dance, Ballet, Music, Rock, Theatre, Noh)
- Management and organization (e.g., Performance, Choreography )

Each term in the thesauri is described using SKOS [18], the relations among the concepts are provided using the *broader/narrower* properties, and each term is described with multilingual labels in 13 different languages. Moreover, each term is linked with all the content items that use that term using a specific *isSubjectOf*  property. The following is an example:

```
<rdf:RDF>
  <skos:Concept rdf:about="http://www.eclap.eu/resource/term/501">
    <skos:prefLabel xml:lang="it">Danza</skos:prefLabel>
    <skos:prefLabel xml:lang="en">Dance</skos:prefLabel>
    …
    <skos:broader rdf:resource="http://www.eclap.eu/resource/term/664"/>
```

```
      <skos:narrower rdf:resource="http://www.eclap.eu/resource/term/540"/>
      <skos:narrower rdf:resource="http://www.eclap.eu/resource/term/539"/>
      <skos:narrower rdf:resource="http://www.eclap.eu/resource/term/507"/>
      <skos:narrower rdf:resource="http://www.eclap.eu/resource/term/506"/>
      <eclap:isSubjectOf
         rdf:resource="http://www.eclap.eu/resource/object/urn:axmedis:000... "/>
      <eclap:isSubjectOf
         rdf:resource="http://www.eclap.eu/resource/object/urn:axmedis:000... "/>
      …
   </skos:Concept>
</rdf:RDF>
```

## 4.3    Annotations description

The annotations are used to describe the whole content or its fragment associating it with it a textual description and with another content or fragment. The annotations can be also associated with a classification term (e.g. scene, gesture, character).

The annotations are described using the Open Annotation Collaboration ontology (http://www.openannotation.org/) that is currently in working draft, the *hasTarget* property refers to the object being annotated, the *FragmentSelector* class is used to specify the temporal fragment of the annotated resource that is subject to the annotation and the *hasBody* property refers to the annotation body that can be the reference to another content or a text description. The *annotatedBy* property is used to relate the annotation to the user that created it and the *annotatedAt* indicates when the annotation was created. The following is an example:

```
<rdf:RDF xmlns:oa="http://www.w3.org/ns/oa#"
         xmlns:cnt="http://www.w3.org/2011/content#">
   <oa:Annotation
     rdf:about="http://www.eclap.eu/resource/annotation/SideAnnotation_13010...">
     <oa:hasTarget>
       <oa:SpecificResource>
         <oa:hasSource
            rdf:resource="http://www.eclap.eu/resource/object/urn:axmedis:0.."/>
         <oa:hasSelector>
            <oa:FragmentSelector>
             <rdf:value>t=npt:10,60</rdf:value>
             <dcterm:conformsTo rdf:resource="http://www.w3.org/TR/media-frags/"/>
            </oa:FragmentSelector>
         </oa:hasSelector>
       </oa:SpecificResource>
     </oa:hasTarget>
     <oa:hasBody
        rdf:resource="http://www.eclap.eu/resource/object/urn:axmedis:..."/>
     <oa:hasBody>
       <cnt:ContentAsText>
         <cnt:chars>this is an annotation</cnr:chars>
```

```
          <dc:format>text/plain</dc:format>
        </cnt:ContentAsText>
      </oa:hasBody>
      <oa:annotatedBy rdf:resource="http://www.eclap.eu/resource/user/1"/>
      <oa:annotatedAt>2013-02-28T20:00:00</oa:annotatedAt>
    </oa:Annotation>
</rdf:RDF>
```

## 4.4    User description

Considering the privacy implication of publishing personal information about the user, minimal personal user information is provided, namely only the nickname is provided. However other relations are available such as:

- the 'knows' relation that connect with 'friends' users,
- the featured content,
- the favorite content,
- the uploaded content,
- the annotations created,
- the subscribed groups,
- the taxonomy terms of interest to the user.

The following is an example of the description of an user:

```
<rdf:RDF …>
  <foaf:Person rdf:about="http://www.eclap.eu/resource/user/45">
    <foaf:nick>bellini</foaf:nick>
    <foaf:knows rdf:recource="http://www.eclap.eu/resource/user/1" />
    <foaf:topic_interest rdf:resource="http://www.eclap.eu/resource/term/501" />
    <eclap:isMemberOf rdf:resource="http://www.eclap.eu/resource/group/3160" />
    <eclap:isAdminOf rdf:resource="http://www.eclap.eu/resource/group/3160" />
    <eclap:createdAnnotation
      rdf:resource="http://www.eclap.eu/resource/annotation/SideAnnotation..." />
    <eclap:hasFavorite
      rdf:resource="http://www.eclap.eu/resource/object/urn:axmedis:000..." />
    <eclap:hasFeatured
      rdf:resource="http://www.eclap.eu/resource/object/urn:axmedis:000..." />
  </foaf:Person>
</rdf:RDF>
```

The property *isMemberOf* is the inverse of the *foaf:member* property and the *createdAnnotation* property is the inverse of *oa:annotatedBy*. The *hasFavourite* property is defined as a sub property of *foaf:interest*.

# 5 Relations display and navigation

The ECLAP portal allows to display and to navigate the relations present among the entities managed by the portal. The 'Social Graph' of a Media Object is shown when a content is played or at user login. This graph is a simplification of the information that is available via Linked Data and the terminology used for relations is not always the same used in LOD.

The graph is made of two kind of nodes, rectangular shaped nodes represent entities (content, terms, users, etc.) while circular shaped nodes represent relations. Directed edges connect an entity node to a relation node and a relation node to an entity node. Examples of relations are shown in Figure 3. Regarding the user interactions the user is able to:

- **Expand** an entity node with its relations adding them to the graph
- **Focus** on an entity, in this case the graph is cleared and only the focused node is shown with its relations
- **Open** the page associated with the node
- **use the Back button** to go back to previous states of the graph (e.g. after a focus)
- **Zoom/Pan** the view
- **Hide/show** types of relations to reduce the complexity of the graph

A special node is the 'More' node that is presented when in a relation are present many nodes (e.g., the content associated with a group). In this case, providing all nodes could be infeasible thus a limited number of nodes is provided and a 'more' node is added to the relation. Clicking on it other nodes are added to the relation in a way similar to classical pagination used to present long lists in HTML.

In Figure 4, an example of ECLAP social graph of a content is shown after expanding some nodes.

# 6 Conclusions

The ECLAP portal is now publishing as linked open the description of more than 110,000 content items, the taxonomy used to classify them, as well as the annotations defined over them. This data is available and it can be used by semantic crawlers to find information about performing arts and other linked data systems can reuse our ontology or taxonomy terms or link to our content. Moreover the relations among the information that is present on the portal can be visualized using the ECLAP social graph that allows to navigate these relations allowing to find new content or some unexpected relation.

However the work is still in progress for linking the dataset to other source of information like DBpedia but also to identify person names that are present in the descriptions to create an authority file of people in the performing arts.
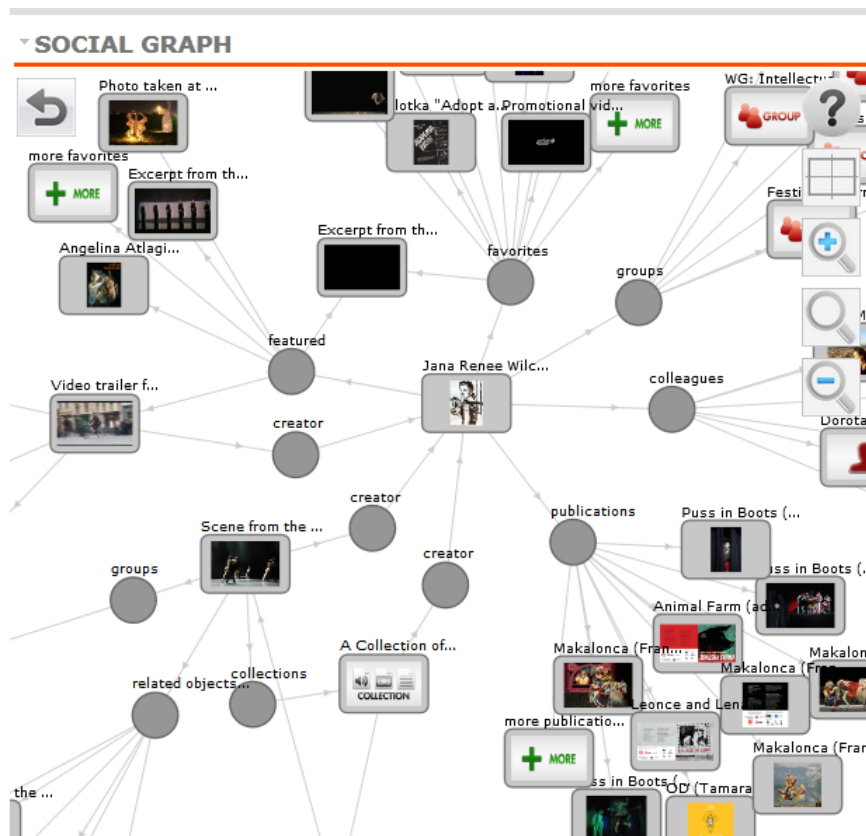
**Fig. 4.** The ECLAP Social Graph

# References

1. Brunetti J. M., Rosa G., Gimeno J. M.,Garcia R.: Improved linked data interaction through an automatic information architecture. In: Int. J. of Software Engineering and Knowledge Engineering, 22(03), 325-343 (2012)..
2. Waitelonis J., Sack H.: Towards Exploratory Video Search Using Linked Data. In: Proc. 11th IEEE Int. Symposium on Multimedia (ISM '09). IEEE Computer Society, Washington, DC, USA, pp 540-545 (2009)

3. Bizer C., Heath T., Berners-Lee T.: Linked Data—the story so far. In: Int. J Semantic Web Inf Syst 5(3):1–22 (2009)

4. Raimond Y., Sandler M.B.: A Web of musical information. In: Bello JP, Chew E, Turnbull D (eds) ISMIR, pp 263–268 (2008)

5. Kobilarov G., Scott T., Raimond Y., Oliver S., Sizemore C., Smethurst M., Bizer C., Lee R.: Media meets Semantic Web—how the BBC uses DBpedia and Linked Data to make connections. In: Proceedings of the 6th European Semantic Web conference. Springer, Berlin, pp 723–737 (2009)

6. Bizer C., Lehmann J., Kobilarov G., Auer S., Becker C., Cyganiak R., Hellmann S.: DBpedia — a crystallization point for the Web of Data. J Web Sem 7(3):154–165 (2009)

7. Kahan J., Koivunen M.R.: Annotea: an open RDF infrastructure for shared Web annotations. In: WWW '01: proceedings of the 10th international conference on World Wide Web. ACM Press, New York, pp 623–632 (2001)

8. Haslhofer B., Jochum W., King R., Sadilek C., Schellner K.: The LEMO annotation framework: weaving multimedia annotations with the Web. Int J Digit Libr 10(1):15–32 (2009)

9. W3C: Media fragments URI 1.0. W3C Media Fragments Working Group URL http://www.w3.org/TR/media-frags/ (2009)

10. Haslhofer B., Simon R., Sanderson R., Van de Sompel H.: The Open Annotation Collaboration (OAC) Model. In: Proc. of the 2011 Workshop on Multimedia on the Web (MMWEB '11). IEEE Computer Society, Washington, DC, USA, pp 5-9 (2011)

11. Bellini P., Nesi P., Paolucci M., Serena M.: Models and tools for content aggregation and audiovisual cross annotation synchronization. In: Proc. of 2011 IEEE Int. Symp. on Multimedia, pp 210-215 (2011)

12. Bellini P., Nesi P., Serena M.: Mystoryplayer: Semantic Audio Visual Annotation And Navigation Tool. In: Proc. of the 17th Int. Conf. on Distributed Multimedia Systems (2011)

13. ECLAP DE4.3 Metadata Descriptors Interoperability, http://www.eclap.eu/urn:axmedis:00000: obj: 6d1e8e22-55d4 -4082-9486-f37a3752704a

14. Bellini P., Cenni D., Nesi P.: On the Effectiveness and Optimization of Infor-mation Retrieval for Cross Media Content. In: Proc. of the KDIR 2012 part of IC3K 2012, Intern. Joint Conf. on Knowledge Discovery, Knowledge Engineering and Knowledge Management (2012)

15. Scharffe F., Euzenat J.: Alignments for data interlinking: analysed systems. http://melinda.inrialpes.fr/systems.html (2009)

16. Tordai A., van Ossenbruggen J., Schreiber G.: Combining vocabulary alignment techniques. In: Proc. of the fifth int. conf. on Knowledge capture (K-CAP '09). ACM, New York, NY, USA, pp 25-32 (2009)

17. Volz J., Bizer C., Gaedke M., Kobilarov G.: Silk—A link discovery framework for the Web of Data. In: Proceedings of the 2nd international workshop on Linked Data on the Web (LDOW). Madrid, Spain (2009)

18. W3C Semantic Web Deployment Group: SKOS simple knowledge organization system reference. URL http://www.w3.org/TR/2009/REC-skos-reference-20090818/ (2009)