Towards Standardized Integration of Images in the Cloud of Linked Data

Ruben Tous[1], Jaime Delgado[1], Frederik Temmermans[2,3], Bart Jansen[2,3], and Peter Schelkens[2,3]

[1] Universitat Politècnica de Catalunya - BarcelonaTech (UPC), Barcelona, Spain
{rtous,jaime.delgado}@ac.upc.edu
[2] Department of Electronics and Informatics - ETRO, Vrije Universiteit Brussel, Brussels, Belgium
[3] iMinds, Dept. of Future Media and Imaging (FMI), Ghent, Belgium
ftemmerm@etro.vub.ac.be

Abstract. Currently, there are several ways of describing and referring to images in RDF. This ambiguity results in a proliferation of vocabularies for image descriptions, complicating the cross-community data integration of information related to digital images. In addition, there are no standardized guidelines on how to integrate RDF data into the image metadata. Therefore, the JPEG standardization committee has recently initiated some activities to streamline the integration of images in the cloud of Linked Data. One effort is the standardization of an ontology for describing digital images. JPEG aims at providing a technology that enables the uniform description of photos and videos with technical, administrative and semantic metadata compliant with the RDF specification and the principles of Linked Data. Secondly, specifications to integrate RDF data into JPEG images are elaborated. Finally, since descriptions often only apply to a certain part of an image, a last effort is to formalize the specification of regions of interest. In this paper, members of the JPEG committee provide a detailed overview of these ongoing activities.

Keywords: Ontology, JPEG, JPSearch, image, photo, video, controlled vocabulary, semantic description, RDF (Resource Description Framework), linked data, computer vision.

1 Introduction

Semantic image annotation is nowadays a hot topic and hundreds of different metadata vocabularies are being developed in an ad-hoc way [10,3,2]. A metadata vocabulary is a specification of what terms to use in the metadata for a certain domain, and how these terms are defined. Modern metadata vocabularies have transitioned from shared natural-language definitions to shared machine-processable representations [4]. Commonly, adopting the shared formal model provided by the W3C's RDF, which is used to support Linked Data. Linked Data is a method of publishing metadata, where every single item is a resource.

Y.T. Demey and H. Panetto (Eds.): OTM 2013 Workshops, LNCS 8186, pp. 388–397, 2013.

Every resource can be linked to other resources, public Linked Data databases (e.g. DBpedia) or to formal vocabularies, i.e. ontologies. Since anyone can define ontologies, the development of new ontologies is unlimited. However, for the sake of interoperability and to prevent proliferation, someone should select, refine, harmonize, catalogue, register and disseminate those vocabularies. Therefore, the Joint Photographic Experts Group (JPEG, formally ISO/IEC JTC1/SC29/WG1) has recently initiated some new activities on next generation image metadata. The main goal of this activity is to provide a simple and uniform way of annotating JPEG images with metadata compliant to the Linked Data principles. To accomplish this objective, a first task is the specification of a JPEG Ontology for Still Image Description (JPOnto). JPOnto provides a set of classes, properties, and restrictions that can be used to represent and interchange information about still and moving images generated in different systems and under different contexts. It can also be specialized to create new classes and properties to model image information for different applications and domains. A second task is designing a mechanism which allows embedding RDF metadata annotations within JPEG (ISO/IEC 10918) or JPEG 2000 (ISO/IEC 15444) files. Finally, since descriptions often only apply to a certain part of an image, a last effort is to formalize the specification of regions of interest.

Section 2 describes the context of this work while Sections 3, 4 and 5 explain the new activities in more detail. It should be noted that the described work is work in progress and has to go to several revisions and procedures before becoming an International Standard. During the process some specifics might change. The authors of this article are members of the JPEG committee and participated in launching and elaborating this work.

2 Contextualization

2.1 The JPEG Committee

JPEG is a joint working group of the Joint Technical Committee 1 (JTC 1) of the International Standardization Organization (ISO) and the International Electrotechnical Commission (IEC). More specifically, the JPEG committee is Working Group 1 (WG1), Coding of Still Pictures, of JTC 1's subcommittee 29 (SC29), Coding of Audio, Picture, Multimedia and Hypermedia Information. The word "Joint" in JPEG does not refer to the joint efforts of ISO and IEC, but to the fact that the JPEG activities are the result of an additional collaboration with the International Telecommunication Union ITU [5]. The JPEG committee is well known for its image coding standards, including JPEG and JPEG2000. In addition, JPEG produced standards related to image security (JPSEC), transfer protocols (JPIP) and image metadata and search (JPSearch). The work described in this article is part of the activities of the JPSearch group.

2.2 JPSearch

JPSearch is an activity within JPEG that aims to address interoperability in image search and retrieval systems [6,7,1]. For this purpose, JPSearch puts

forward an abstract image search and retrieval framework. Interfaces and protocols for data exchange between the components of this architecture are standardized, with minimal restrictions on how these components perform their respective tasks. The use and reuse of metadata and associated metadata schemas is thus facilitated. A common query language enables search over distributed repositories. Finally, an interchange format allows users to easily import and export their data and metadata among different applications and devices. In the JPSearch framework, interoperability can be defined in different ways: between self-contained vertical image search systems providing federated search, between layers of an image search and retrieval system so that different modules can be supplied by distinct vendors, or at the metadata level such that different systems may add, update, or query metadata [6]. JPSearch currently consists of five Parts, all of them recently reached the state of International Standard:

Part 1. System framework and components
Part 2. Registration, identification and management of schema and ontology
Part 3. JPSearch Query Format
Part 4. File format for metadata embedded in image data
Part 5. Data interchange format between image repositories
Part 6. Reference Software

2.3 JPSearch RA

JPSearch compliant systems can manage multiple proprietary or community-specific metadata vocabularies formalized as XML schemas or RDF ontologies. The multiplicity of vocabularies is solved by allowing the publication of machine-readable translations between metadata terms belonging to proprietary vocabularies and metadata terms in the core metadata vocabulary provided by JPSearch. This core vocabulary is formalized as an XML Schema, but all its terms have equivalent counterparts within the JPOnto ontology, thus facilitating the linkage to them from Linked Data.

In order to rationalize the usage of vocabularies and translation rules across different JPSearch systems, a global authority for schemas and their translation rules has been established. JPSearch RA is the official body designated by ISO to serve as a registration authority for Part 2 of ISO/IEC 24800 (JPSearch). Via the authority, all JPSearch compliant applications can obtain the information they need. The JPSearch RA was assigned to the Distributed Multimedia Applications Group (DMAG) of the Universitat Politècnica de Catalunya - BarcelonaTech (UPC). The JPSearch RA entered into full operation early 2013 at `http://dmag.ac.upc.edu/jpsearch-ra`. It provides a registration procedure for the publication of metadata vocabularies and translation rules. The information is directly stored in the RA's internal server. Furthermore, it provides a registry look up service in which the RA displays a list of registered metadata vocabularies and translation rules on the website. The displayed vocabulary information includes links to the schema or ontology and also to its related ISO/IEC 24800-2 translation rules. Registration forms are available and any person or organization is eligible to apply.

2.4 Workplan

The work items discussed in this article will initially be amendments to the current Parts of JPSearch and as such, become part of International Standards. More specifically, ontologies will extend Part 2, while image integration will extend Part 4. The region of interest specification is shared with a new REST API activity that will amend Part 3. In order to officially complement a standard, amendments go trough the following stages:

- Proposed Draft Amendment (PDAM)
- Final Proposed Draft Amendment (FPDAM)
- Final Draft Amendment (FDAM)
- Amendment (ADM)

The committee meets three times a year, at each meeting cycle, a document can move from one stage to the next stage. At the time of writing, the work is in the first stage, i.e. initial drafts have been produced. As a consequence, it will take at least one year before this work can promote to International Standard. At that point, a new version of a particular Part can be created that integrates the amendment(s).

3 JPEG Ontology for Image Description (JPOnto)

The JPEG work on the JPSearch Ontology for Still Image Description (JPOnto) started in February 2012 [9] and the first working draft was released in April 2013 [8]. JPOnto provides a set of classes, properties, and restrictions that can be used to represent and interchange information about still images generated in different systems and under different contexts. It can also be specialized to create new classes and properties to model still image information for different applications and domains. JPOnto users may only need to use parts of the entire ontology, depending on their needs and according to how much detail they want to include in their still image information. For this, the JPOnto terms (classes and properties) are grouped into two sub-vocabularies to provide an incremental introduction to the ontology, **JPOnto-core** and **JPOnto-visual**.

On the one hand, a core vocabulary named JPOnto-core is the basic set of entities and properties which serves as a central component to interconnect all the other sub-vocabularies of JPOnto. Three classes provide a basis for the remaining part of the JPOnto-core:

- **jponto:Image** : Represents a digital still image.
- **jponto:RegionOfInterest** : A class specialized in representing a certain region within a still image.
- **jponto:Agent** : An agent (eg. person, group, software or physical artifact).

In addition to these primary classes JPOnto provides classes and properties for describing organizations (*jponto:Organization*) and people (*jponto:Person*). Figure 2 outlines the relationships among the JPOnto-core classes but also

includes some relevant relationships with some classes from JPOnto-visual (*jponto:PersonGroup* and *jponto:People*). JPOnto provides also many other datatype properties (that link individuals to data values) that are not depicted in the figure (e.g. *jponto:creationDate*, *jponto:boundingPolygon*, etc.).

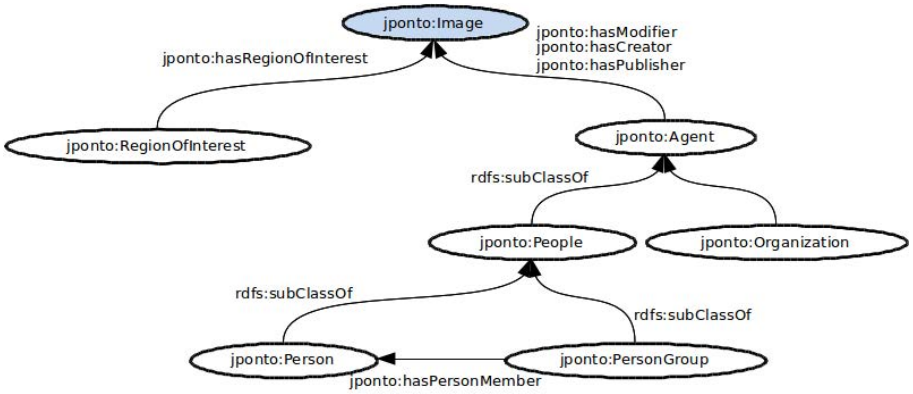

Fig. 1. Outline of JPOnto-core

JPOnto partially overlaps with well known vocabularies including, but not limited to, Dublin Core, the W3C Ontology for Media Resources and Friend of a Friend (FOAF). JPEG aims at releasing a self-included specification in order to ease the development of compliant tools. For this reason, terms in JPOnto with equivalent counterparts in external vocabularies have been redefined within the JPOnto namespace instead of directly reusing or subtyping the external terms. The proper links between related terms (e.g. *owl:equivalentClass, owl:sameAs, etc.* will be included in the final specification.

On the other hand, **JPOnto-visual** provides a rich set of constructs for semantic visual content description, including, but not limited to, a uniform description of identities, features, aspect, relationships, actions and emotional information of people appearing in the images, as well as description of events, locations and objects. JPOnto-visual will enable the description of the semantics of digital images in a simple, structured and uniform way. The resulting specification will have multiple applications, such as serving as a unified annotation format for evaluation forums about the annotation and retrieval of digital images (e.g. ImageCLEF), for improving Web accessibility for visually impaired Web users, and many others.

The **jponto:depicts** property is the core component of JPOnto-visual. The *jponto:depicts* property allows binding an image with the real world objects and events that it represents. Defining the *jponto:RegionOfInterest* as a subclass of *jponto:Image*, allows to also use the *jponto:depicts* predicate to bind ROIs with real world objects and events.

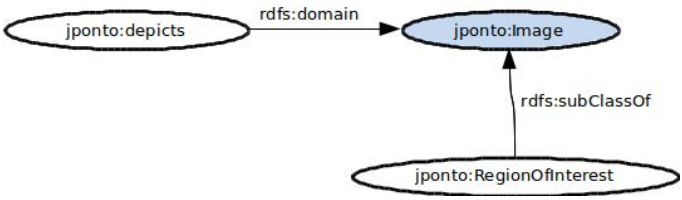

Fig. 2. The *jponto:depicts* property allows binding an image with the real world objects and events that it represents

The example in Code 1 shows the usage of the *jponto:depicts* predicate for the semantic description of the contents of an imaginary image.

Code 1. Semantic description of the contents of an imaginary image

```
@prefix jponto: <http://www.jpeg.org/ns/jponto#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix : <http://example.com/> .
:image1
  rdf:type jponto:Image;
  jponto:depicts [
      rdf:type  jponto::Person;
      jponto:givenName "Ruben Tous";
      jponto:mbox <mailto:rtous@ac.upc.edu> ];
  jponto:depicts [
    rdf:type  jponto::Building;
    jponto:givenName "Namdaemun gate";
    jponto:wikipedia <https://en.wikipedia.org/wiki/Sungnyemun> ].
```

To describe people appearing in an image, JPOnto-visual allows for describing their:

- identities and relationships
- actions and emotions
- type (gender, age, race, profession, etc.).
- aspects (make up, clothes, etc.)

In addition to the class *jponto:Person* defined in JPOnto-core, JPOnto-visual provides an alternative class for referring to persons appearing in the photo in an unspecific way (class *jponto:People*) and a class to refer to specific groups of 2 or more persons (class *jponto:PersonGroup*). Other classes related to the description of people are *jponto:Feeling* and *jponto:Action*.

JPOnto-visual provides also classes and properties for the description of events, locations, objects, atmospheric conditions, lighting and combustion effects, image views, etc.

4 Embedding JPOnto Instances within JPEG and JPEG 2000 Images

JPEG is currently specifying a compact and uniform way to embed JPOnto instances within JPEG (ISO/IEC 10918) or JPEG 2000 (ISO/IEC 15444) files. The work is focused on providing a canonical serialization syntax for the metadata statements. The location and signaling of the resulting metadata UTF-8 stream within the image file is specified by the JPSearch Part 4 standard (ISO/IEC 24800-4).

An embedded JPOnto instance, which is basically an RDF graph, must follow the following rules to be compliant with the standard:

- It makes use of character encoding UTF-8.
- It is serialized using the W3C's Terse RDF Triple Language (a.k.a. Turtle) and the canonical serialization syntax style defined by JPEG.

As the Turtle syntax is very flexible, JPEG wants to constraint its usage to a canonical serialization syntax style to facilitate the development of of conformant readers. The canonical serialization syntax specifies that the serialized metadata must include one predicate list referring to a single instance of the Class jponto-core:Image. Any triple with a subject different to IMAGE_INSTANCE_RELATIVE_IRI will be represented using nested unlabeled blank nodes as shown in the example of Code 2.

Code 2. Embedded JPOnto instance basic structure (Turtle syntax)

```
@base <IMAGE_INSTANCE_BASE_IRI>.
@prefix jponto: <http://www.jpeg.org/ns/jponto#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
:IMAGE_INSTANCE_RELATIVE_IRI
    rdf:type jponto:Image;
    jponto:hasCreator [
        jponto:givenName "Ruben Tous";
        jponto:mbox <mailto:rtous@ac.upc.edu> ].
```

More compact serializations are also enabled under certain circumstances. When the instance only makes use of namespaces endorsed by the standard (e.g. the JPOnto and RDF namespaces) it is possible to omit the namespaces declaration if the standard prefixes are used (e.g. jponto, rdf, etc.). It is also valid to omit the jponto prefix. The example in Code 3 shows a very compact serialization.

5 Region of Interest Specification

Often, descriptions only apply to a certain part of an image rather than to the whole image. Therefore, a proper Region Of Interest (ROI) specification is an

Code 3. Embedded JPOnto instance using a compact serialization style

```
:hasView  :Outdoor;
:hasQuality :MotionBlur;
:depicts [
    a  :Building;
    :givenName "Namdaemun gate";
    :wikipedia <https://en.wikipedia.org/wiki/Sungnyemun> ].
:depicts [
    a  :Vehicle;
    :color    :Green  ];
:depicts [
    a  :Person;
    :givenName "Ruben Tous";
    :mbox <mailto:rtous@ac.upc.edu> ].
:depicts :Rainy.
```

important part of the description framework. Currently, some ad-hoc methods of specifying ROI's in RDF have been proposed. The W3C Wiki[1] provides an example that adopts the SVG (Scalable Vector Graphics) Path specification to specify a polygonal ROI, analogously to the example given in Code 4.

Code 4. ROI specification using the SVG representation

```
@prefix foaf: <http://xmlns.com/foaf/0.1#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
<http://www.site.com/image.jpg>
    img:hasPart [
      a  img:Polygon;
      img:polypath "M20 160 L80 160 L80 20 L20 20 Z";
      foaf:regionDepicts [
        a  foaf:Person;
        foaf:name "Ruben Tous";
      ].
    ].
```

The path specification is a sequence of points, specified by the two coordinates of the point preceded by M or L. M stands for "move to" and is in this case used to move to the start position. L stands for "line to" and connects the current position with the associated point. The path specification ends with a Z, which is the "close path" command. The SVG path specification provides additional commands allowing to describe more complex shapes. Moreover, SVG specifies alternative representations for common shapes include rectangles, circles, ellipses and polygons. These could be used analogously. This representation is often used

[1] http://www.w3.org/wiki

because its SVG compatibility might allow specific applications to render the region. However, it is not at all guaranteed that clients can interpret the syntax.

Another approach is specifying a region with a specific resource identifier, allowing a much shorter and readable notation. One option is to adopt W3C's Media Fragments recommendation[2]. A fragment is the part of a URL preceded with a "#". The current specification only supports rectangular selections using the *xywh* parameter. The value is a comma separated sequence of four integers denoting respectively the upper left corner's x and y coordinates, width and height. The example in Code 5 describes a person in an area of 128 by 68 starting at x=66 and y=89.

Code 5. ROI specification using a custom resource identifier

```
@prefix foaf: <http://xmlns.com/foaf/0.1#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
<http://www.site.com/image.jpg#xywh=66,89,128,64>
    foaf:depicts [
      a  foaf:Person;
      foaf:name "Ruben Tous";
    ].
```

Currently, another ongoing activity within JPSearch is the definition of a standardized REST API for image search and retrieval. ROI specification is also part of it. Rather than fragments, the region is specified by some specific arguments in the query string of a URL. The query string of a URL is the part consisting of key value pair arguments and preceded by a "?". The query syntax implies that the specific region could be requested by a client application. For compatibility reasons, the preferred syntax in this context is taken over from JPIP (ISO/IEC 15444-9). JPIP allows specifying many types of regions, including rectangles, ellipses, quadrilaterals, oriented ellipses or arbitrary polygons. For example, a rectangle is specified by its width and height ($rsiz$) and coordinates of the upper left corner ($roff$). The following resource identifier specifies a rectangle region of 128 by 64 pixels with the upper left corner at position 66, 89: `http://www.site.com/image.jpg?rsiz=128,64&roff=66,89`.

Since the JPSearch API is still under development, it has not yet been decided which specification will be recommended for usage in RDF image descriptions. A main predicament is finding a good balance between flexibility and complexity.

6 Conclusions

Image metadata descriptions are evolving from natural-language definitions to shared machine-processable representations. In the open world of Linked Data anyone can define his own ontologies, leading to a proliferation of vocabularies.

[2] `http://www.w3.org/TR/media-frags/`

To obtain interoperability, someone should select, refine, harmonize, catalogue, register and disseminate those vocabularies. In addition, seamless integration of images in the cloud of Linked Data requires standardized guidelines on how to embed these descriptions into images. Finally, a formal region of interest specification allows more precise and interoperable descriptions. The Joint Photographic Experts Group aims to provide an answer to these predicaments with the initiatives described in this article.

Acknowledgments. This work has been partly supported by the Spanish government (TEC2011-22989) and iMinds. Bart Jansen is currently funded by Innoviris.

References

1. Doeller, M., Tous, R., Temmermans, F., Yoon, K., Park, J.-H., Kim, Y., Stegmaier, F., Delgado, J.: Jpeg's jpsearch standard: Harmonizing image management and search. IEEE Multimedia 99 (PrePrints), 1 (2012)
2. Hollink, L., Worring, M.: Building a visual ontology for video retrieval. In: Proceedings of the 13th Annual ACM International Conference on Multimedia, MULTIMEDIA 2005, pp. 479–482. ACM, New York (2005)
3. Hollink, L., Schreiber, G., Wielemaker, J., Wielinga, B.: Semantic annotation of image collections. Knowledge Capture, 41–48 (2003)
4. Miller, Y., Zhitomirsky-Geffet, M., Bar-Ilan, J.: Exploring the effectiveness of ontology based tagging versus free text tagging (2012)
5. Schelkens, P., Bruylants, T., Temmermans, F., Barbarien, J., Dooms, A., Munteanu, A.: The jpeg 2000 family of standards, 724802–724802-9 (2009)
6. Temmermans, F., Dufaux, F., Schelkens, P.: Jpsearch: Metadata interoperability during image exchange [standards in a nutshell]. IEEE Signal Processing Magazine 29(5), 134–139 (2012)
7. Temmermans, F., Doeller, M., Vanhamel, I., Jansen, B., Munteanu, A., Schelkens, P.: Jpsearch: An answer to the lack of standardization in mobile image retrieval. Signal Processing: Image Communication 28(4), 386–401 (2013)
8. Tous, R., Bailer, W., Delgado, J., Temmermans, F.: ISO/IEC JTC1/SC29/WG1/N6369 JPEG Ontology for Image Description (JPOnto) WD 1.0. Output Document of the 61st JPEG Meeting, Incheon, Korea (April 2013)
9. Tous, R., Delgado, J.: ISO/IEC JTC1/SC29/WG1/N5995 Proposal for Innovations in Metadata Management in JPEG. Input Document of the 57th JPEG Meeting, San Jose, US (February 2012)
10. Tousch, A.-M., Herbin, S., Audibert, J.-Y.: Semantic hierarchies for image annotation: A survey. Pattern Recognition 45(1), 333–345 (2012)