MEDINFO 2013 C.U. Lehmann et al. (Eds.) © 2013 IMIA and IOS Press. This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License. doi:10.3233/978-1-61499-289-9-970

# Fostering Ontology Alignment Sharing: A General-Purpose RDF Mapping Format

### Alberto Anguita<sup>a</sup>, Ana Escrich<sup>a</sup>, Victor Maojo<sup>a</sup>

<sup>a</sup> Biomedical Informatics Group, Universidad Politécnica de Madrid, Campus de Montegancedo, 28660 Madrid, Spain

#### Abstract and Objective

RDF has established in the last years as the language for describing, publishing and sharing biomedical resources. Following this trend, a great amount of RDF-based data sources, as well as ontologies, have appeared. Using a common language as RDF has provided a unified syntactic for sharing resources, but the semantics remain as the main cause of heterogeneity, hampering data integration and homogenization efforts. To overcome this issue, ontology alignment based solutions have been typically used. However, alignment information is usually codified using ad-hoc formats. In this paper, we present a general purpose ontology mapping format, totally independent from the homogenization approach to be applied. The format is accompanied with a Java API that offers mapping construction and parsing features, as well as some basic algorithms for applying it to data translation solutions.

Keywords: RDF, database integration, ontology alignment.

### Introduction

RDF (Resource Description Framework) based data integration relies on aligning pairs of RDF schemas. However, there is a lack of well-established formats for specifying alignments, and researchers are forced to adopt *ad hoc* solutions. This practice hampers the sharing and distribution of alignments, as developers cannot take advantage of previous efforts. In this sense, few works can be found, as in [1], but they have proven insufficient for the present problem. We present a generic RDF alignment format based on XML and a corresponding Java API for handling it. The proposed mapping format offers a solution specifically targeted at data translation/integration problems, and is designed to be compatible with any type of schema alignment approach.

## Methods

The mapping format is based on the alignment of pairs of RDF views of arbitrary complexity (each belonging to one of the mapped schemas). Each view can contain one or more paths, each of them formed by one or more triples (a statement of the type *subject, predicate, object*). Paths within the same view are bounded with *internal bounds*, specified in the format as XML attributes of the bounded elements. Elements from separate views can be bound together in a similar way, with *external bounds*. The format defines the tags for specifying the elements of each pair of semantically equivalent views, called *entries*. The format also supports the specification of additional metadata. Figure 1 shows an example code. Apart from the XML format, we provide a complete Java API for creating, parsing and editing mappings in our format.



*Figure 1 – Example of the mapping format code* 

## **Results and Discussion**

The presented format has been tested in two different research projects. First, it was used in the ACGT project [2] to support a federated data integration layer. More recently, it was applied in the p-medicine project (http://www.p-medicine.eu/), to enable the centralized integration of heterogeneous data sources. Although mappings are manually defined, the Java API enabled the use of end-user tools to facilitate this task. The resulting mappings allowed the automatic translation of queries in the ACGT project, and RDF data in the p-medicine project. In the former case, the decomposition of SPARQL queries in atomic paths allowed generating equivalent queries for the target schema. In the latter, data was queried according the mapping information and translated into the data warehouse schema. The Java API facilitates these operations. The presented XML format for codifying mappings between RDF schemas seeks to foster collaboration in the definition of alignment information. We believe this can be achieved by establishing generic standards that easily adapt to any data integration approach. The format and API are free to download from http://www.bioinformatics.org/mappingapi/wiki/.

#### Acknowledgments

This work was funded by the European Commission, through the p-medicine integrated project (FP7-ICT-2009-270089).

#### References

- Euzenat J. An API for Ontology Alignment. In: The Semantic Web – ISWC. 2004:698-712.
- [2] Martin L, Anguita A, Graf N, Tsiknakis M, Brochhausen M, Rüping S, Bucur A, Sfakianakis S, Sengstag T, Buffa F, Stenzhorn H. ACGT: advancing clinico-genomic trials on cancer four years of experience. In: *Stud Health Technol Inform.* 2011;169:734-8.