Journal Subline

LNCS 5530

# Journal on Data Semantics XIII

Stefano Spaccapietra
Editor-in-Chief

Springer

Lecture Notes in Computer Science 5530

*Commenced Publication in 1973*
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Stefano Spaccapietra Esteban Zimányi
Il-Yeol Song (Eds.)

# Journal on Data Semantics XIII

Springer

Volume Editors

Stefano Spaccapietra
École Polytechnique Fédérale de Lausanne
EPFL-IC
Database Laboratory
1015 Lausanne, Switzerland
E-mail: stefano.spaccapietra@epfl.ch

Esteban Zimányi
Université Libre de Bruxelles
Department of Computer and Decision Engineering
50 av. F.D. Roosevelt, 1050 Bruxelles, Belgium
E-mail: ezimanyi@ulb.ac.be

Il-Yeol Song
Drexel University
College of Information Science and Technology
Philadelphia, PA 19104, USA
E-mail: song@drexel.edu

CR Subject Classification (1998): H.3, H.4, H.2, C.2, D.3, F.3, D.2

ISSN 0302-9743 (Lecture Notes in Computer Science)
ISSN 1861-2032 (Journal on Data Semantics)
ISBN-10 3-642-03097-1 Springer Berlin Heidelberg New York
ISBN-13 978-3-642-03097-0 Springer Berlin Heidelberg New York



Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12660797 06/3180 5 4 3 2 1 0

# The LNCS Journal on Semantics of Data

Computerized information handling has changed its focus from centralized data management systems to decentralized data exchange facilities. Modern distribution channels, such as high-speed Internet networks and wireless communication infrastructure, provide reliable technical support for data distribution and data access, materializing the new, popular idea that data may be available to anybody, anywhere, anytime. However, providing huge amounts of data on request often turns into a counterproductive service, making the data useless because of poor relevance or an inappropriate level of detail. Semantic knowledge is the essential missing piece that allows the delivery of information that matches user requirements. Semantic agreement, in particular, is essential to meaningful data exchange.

Semantic issues have long been open issues in data and knowledge management. However, the boom in semantically poor technologies, such as the Web and XML, has boosted renewed interest in semantics. Conferences on the Semantic Web, for instance, attract big crowds of participants, while ontologies on their own have become a hot and popular topic in the database and artificial intelligence communities.

Springer's *LNCS Journal on Data Semantics* aims at providing a highly visible dissemination channel for remarkable work that in one way or another addresses research and development on issues related to the semantics of data. The target domain ranges from theories supporting the formal definition of semantic content to innovative domain-specific application of semantic knowledge. This publication channel should be of the highest interest to researchers and advanced practitioners working on the Semantic Web, interoperability, mobile information services, data warehousing, knowledge representation and reasoning, conceptual database modeling, ontologies, and artificial intelligence.

Topics of relevance to this journal include:

- Semantic interoperability, semantic mediators
- Ontologies
- Ontology, schema and data integration, reconciliation and alignment
- Multiple representations, alternative representations
- Knowledge representation and reasoning
- Conceptualization and representation
- Multimodel and multiparadigm approaches
- Mappings, transformations, reverse engineering
- Metadata
- Conceptual data modeling
- Integrity description and handling
- Evolution and change
- Web semantics and semi-structured data

- Semantic caching
- Data warehousing and semantic data mining
- Spatial, temporal, multimedia, and multimodal semantics
- Semantics in data visualization
- Semantic services for mobile users
- Supporting tools
- Applications of semantic-driven approaches

These topics are to be understood as specifically related to semantic issues. Contributions submitted to the journal and dealing with semantics of data will be considered even if they are not from the topics in the list.

While the physical appearance of the journal issues resembles the books from the well-known Springer LNCS series, the mode of operation is that of a journal. Contributions can be freely submitted by authors and are reviewed by the Editorial Board. Contributions may also be invited, and nevertheless carefully reviewed, as in the case for issues that contain extended versions of the best papers from major conferences addressing data semantics issues. Special issues, focusing on a specific topic, are coordinated by guest editors once the proposal for a special issue is accepted by the Editorial Board. Finally, it is also possible that a journal issue be devoted to a single text.

The Editorial Board comprises an Editor-in-Chief (with overall responsibility), a Co-editor-in-Chief, and several members. The Editor-in-Chief has a four-year mandate. Members of the board have a three-year mandate. Mandates are renewable and new members may be elected anytime.

We are happy to welcome you to our readership and authorship, and hope we will share this privileged contact for a long time.

Stefano Spaccapietra
Editor-in-Chief
`http://lbd.epfl.ch/e/Springer/`

# JoDS Volume XIII – Special Issue on Semantic Data Warehouses

Data warehouses have been established as a fundamental and essential component of current decision-support systems. Many organizations have successfully used data warehouses to collect essential indicators that help them improve their business processes. Furthermore, the combination of data warehouses and data mining has allowed these organizations to extract strategic knowledge from raw data, allowing them to design new ways to perform their operations.

In recent years, research in data warehouses has addressed many topics ranging from physical-level issues, aiming at increasing the performance of data warehouses in order to deal with vast amounts of data, to conceptual-level and methodological issues, which help designers build effective data warehouse applications that address the needs of decision makers better.

Nevertheless, globalization and increased competition pose new challenges to organizations, which need to dynamically and promptly adapt themselves to new situations. This brings new requirements to their data warehouse and decision-support systems, particularly with respect to (1) heterogeneity, autonomy, distribution, and evolution of data sources, (2) integration of data from these data sources while ensuring consistency and data quality, (3) adaptability of the data warehouse to multiple users with multiple and conflicting requirements, (4) integration of the data warehouse with the business processes of the organization, and (5) providing innovative ways to interact with the data warehouse, including advanced visualization mechanisms that help to reveal strategic knowledge. In addition, data warehouses are increasingly being used in non-traditional application domains, such as biological, multimedia, and spatio-temporal applications, which demand new requirements for dealing with the particular semantics of these application domains.

Therefore, building next-generation data warehouse systems and applications requires enriching the overall data warehouse lifecycle with semantics in order to support a wide variety of tasks including interoperability, knowledge reuse, knowledge acquisition, knowledge management, reasoning, etc.

The papers in this special issue address several of the topics mentioned above. They all provide different insights into the multiple benefits that can be obtained by envisioning data warehouses from a new semantic perspective. As this is a relatively new domain, these papers open many new research directions that need to be addressed in future work. This research will definitely have a huge impact on the next generation of data warehouse applications and tools.

January 2009

Esteban Zimányi
Il Yeol Song

## Referees for the Special Issue

We would like to thank all the reviewers for their excellent work in evaluating the papers. Without their committment the publication of this special issue of JODS would not have been possible.

Alberto Abelló, Universitat Politécnica de Catalunia, Spain

Omar Boussaïd, Université du Lyon 2, France

Matteo Golfarelli, University of Bologna, Italy

Panagiotis Kalnis, National University of Singapore, Singapore

Jens Lechtenbörger, University of Münster, Germany

Wolfgang Lehner, Dresden University of Technology, Germany

Tok Wang Ling, National University of Singapore, Singapore

Sergio Luján Mora, University of Alicante, Spain

Elzbieta Malinowski, Universidad de Costa Rica, Costa Rica

Svetlana Mansmann, University of Konstanz, Germany

Rokia Missaoui, Université du Québec en Outaouais, Canada

Ullas Nambiar, IBM India Research Lab, India

Torben Bach Pedersen, Aalborg University, Denmark

Mario Piattini, Universidad de Castilla La Mancha, Spain

Stefano Rizzi, University of Bologna, Italy

Markus Schneider, University of Florida, USA

Alkis Simitsis, Stanford University, USA

Dimitri Theodoratos, New Jersey Institute of Technology, USA

Juan-Carlos Trujillo Mondéjar, Universidad de Alicante, Spain

Panos Vassiliadis, University of Ioannina, Greece

Robert Wrembel, Poznan University of Technology, Poland

# Previous Issues of the Journal

| | |
|---|---|
| JoDS I | Special Issue on Extended Papers from 2002 Conferences, LNCS 2800, December 2003<br>Co-editors: Sal March and Karl Aberer |
| JoDS II | Special Issue on Extended Papers from 2003 Conferences, LNCS 3360, December 2004<br>Co-editors: Roger (Buzz) King, Maria Orlowska, Elisa Bertino, Dennis McLeod, Sushil Jajodia, and Leon Strous |
| JoDS III | Special Issue on Semantic-Based Geographical Information Systems, LNCS 3534, August 2005<br>Guest Editor: Esteban Zimányi |
| JoDS IV | Normal Issue, LNCS 3730, December 2005 |
| JoDS V | Special Issue on Extended Papers from 2004 Conferences, LNCS 3870, February 2006<br>Co-editors: Paolo Atzeni, Wesley W. Chu, Tiziana Catarci, and Katia P. Sycara |
| JoDS VI | Special Issue on Emergent Semantics, LNCS 4090, September 2006<br>Guest Editors: Karl Aberer and Philippe Cudre-Mauroux |
| JoDS VII | Normal Issue, LNCS 4244, November 2006 |
| JoDS VIII | Special Issue on Extended Papers from 2005 Conferences, LNCS 4830, February 2007<br>Co-editors: Pavel Shvaiko, Mohand-Saïd Hacid, John Mylopoulos, Barbara Pernici, Juan Trujillo, Paolo Atzeni, Michael Kifer, François Fages, and Ilya Zaihrayeu |
| JoDS IX | Special Issue on Extended Papers from 2005 Conferences (continued), LNCS 4601, September 2007<br>Co-editors: Pavel Shvaiko, Mohand-Saïd Hacid, John Mylopoulos, Barbara Pernici, Juan Trujillo, Paolo Atzeni, Michael Kifer, François Fages, and Ilya Zaihrayeu |
| JoDS X | Normal Issue, LNCS 4900, February 2008 |

JoDS XI Special Issue on Extended Papers from 2006 Conferences, LNCS 5383, December 2008
Co-editors: Jeff Z. Pan, Philippe Thiran, Terry Halpin, Steffen Staab, Vojtech Svatek, Pavel Shvaiko, and John Roddick

JoDS XII Normal Issue, in press, March 2009

# JoDS Editorial Board

# Table of Contents

# Multidimensional Integrated Ontologies: A Framework for Designing Semantic Data Warehouses

Victoria Nebot[1], Rafael Berlanga[1], Juan Manuel Pérez[1], María José Aramburu[1], and Torben Bach Pedersen[2]

[1] Universitat Jaume I, Av. Vicent Sos Baynat, s/n
E-12071 Castelló, Spain
{romerom,berlanga,juanma.perez,aramburu}@uji.es
[2] Aalborg University, Selma Lagerløfs Vej 300,
DK-9220 Aalborg Ø, Denmark
tbp@cs.aau.dk

**Abstract.** The Semantic Web enables organizations to attach semantic annotations taken from domain and application ontologies to the information they generate. The concepts in these ontologies could describe the facts, dimensions and categories implied in the analysis subjects of a data warehouse. In this paper we propose the Semantic Data Warehouse to be a repository of ontologies and semantically annotated data resources. We also propose an ontology-driven framework to design multidimensional analysis models for Semantic Data Warehouses. This framework provides means for building a Multidimensional Integrated Ontology (MIO) including the classes, relationships and instances that represent interesting analysis dimensions, and it can be also used to check the properties required by current multidimensional databases (e.g., dimension orthogonality, category satisfiability, etc.) In this paper we also sketch how the instance data of a MIO can be translated into OLAP cubes for analysis purposes. Finally, some implementation issues of the overall framework are discussed.

**Keywords:** Data warehouses, Semantic Web, Multi-ontology integration.

## 1 Introduction

The Semantic Web is a rich source of knowledge whose exploitation will open new opportunities to the academic and business communities. One of these opportunities is the analysis of information resources for decision support tasks such as the identification of trends, and the discovery of new decision variables. Semantic annotations are formal descriptions of information resources which usually rely on widely accepted domain ontologies. The main reason for using domain ontologies is to set up a common terminology and logic for the concepts involved in a particular domain. Semantic annotations are especially useful for describing unstructured, semi-structured and text data, which cannot be managed properly by current database systems. Nowadays many applications (e.g., medical applications) attach metadata and semantic annotations to the information they produce, for example medical image, laboratory tests, etc. In the near future, large repositories of semantically annotated data will be available, opening new opportunities for enhancing current decision support systems.

S. Spaccapietra et al. (Eds.): Journal on Data Semantics XIII, LNCS 5530, pp. 1–36, 2009.

Data warehouse systems are stores of information aimed at analysis tasks. This information is extracted from existing databases and is pre-processed to harmonize its syntax and semantics. Thus, one of the main purposes of data warehouse systems is the integration of information coming from several sources. Afterwards, OLAP systems can be applied to efficiently exploit the stored information. Both types of systems rely on multidimensional data models, which distinguish the stored measures from the analysis dimensions that characterize them.

In this paper we tackle the problem of combining data warehouse and Semantic Web technologies. Our proposal is a framework for designing multidimensional analysis models over the semantic annotations stored in a Semantic Data Warehouse (SDW). In our approach, an SDW is conceived as a XML repository that includes web resources, domain ontologies and the semantic annotations made with them. Being a data warehouse, this repository is subject oriented, and therefore it is aimed at recording only data that is relevant for specific analysis tasks.

Our work is being carried out in the context of a larger research project about the integration and exploitation of biomedical data provided by clinicians for research tasks. The framework presented here is based on the specification of a Multidimensional Integrated Ontology (MIO) over the SDW ontologies in order to retrieve the ontology classes and instances that will later be used in the multidimensional analysis. To our best knowledge, our approach is the first one on addressing the following requirements:

- *Multi-ontology design*. Much semantic data is generated in the context of very complex scenarios involving several domain ontologies. The framework proposed in the paper allows the selection of the concepts needed for the analysis through different ontologies.
- *Scalability*. As domain ontologies usually have a considerably large size, the method for building MIOs must be scalable. We will achieve these scalability requirements by extracting only those modules or fragments that are necessary from the source ontologies.
- *Formally well-founded approach.* In order to keep the semantics and inference mechanisms of the source ontologies, the proposed design process relies on formalisms that have been widely accepted for the Semantic Web (e.g., Description Logics).

The main contributions of the paper can be summarized as follows:

1. A framework for designing and building Semantic Data Warehouses.
2. An application scenario and a running use case to establish the requirements and to illustrate the usefulness of our techniques.
3. A methodology for the design, automatic generation and validation of Multidimensional Integrated Ontologies. By integrating the concepts and properties of several ontologies coming from the same application domain, a MIO establishes the topics, measures, dimensions and hierarchies required by a specific data analysis application.
4. The automatic construction of a multidimensional cube, according to the specifications of a MIO, starting from the annotated data stored in the SDW, in order to allow the analysis of this data by using traditional OLAP operators.
5. The study of several alternatives for implementing the proposed SDW.

The rest of the paper is organized as follows. Section 2 describes an application scenario that motivates our approach. Section 3 reviews the related work including: Description Logics, OWL and OLAP; the existing approaches to annotate biomedical data; the combination of Semantic Web and data warehouse technologies; and different alternatives for exploiting knowledge from multiple ontologies. Section 4 introduces our approach to a Semantic Data Warehouse. Section 5 explains the methodology proposed for designing Multidimensional Integrated Ontologies and Section 6 gives some implementation guidelines. Finally, Section 7 presents some conclusions and future work.

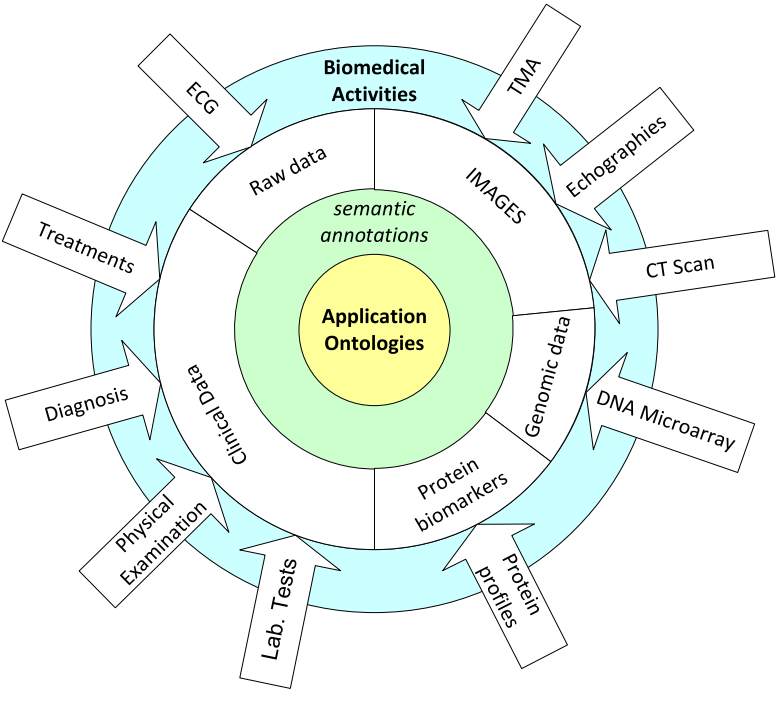


**Fig. 1.** Generation of semantic annotations in the biomedical domain

## 2 Application Scenario and Use Case

In this section we describe an application scenario for an SDW along with a use case that will serve to define the examples of the rest of the paper. By defining this application scenario, we will identify a list of requirements that can be considered common to many applications of SDWs, and that, therefore, can be applied to prove the usefulness of the framework proposed in this paper.

Our application scenario is Biomedicine in which, at the moment, vast amounts of semantically annotated data are being generated by many different types of data management systems (see section 3.2). In order to guide the process of semantically annotating the data, current data management systems adopt specific application ontologies relying on one or more widely accepted domain ontologies. A domain

ontology is a very large corpus of semantically related data that describe the knowledge and vocabularies agreed by the relevant biomedical community. The reader can find a good review of the main biomedical ontologies in (Rubin et al., 2007).

Figure 1 shows the usual process of generating semantic annotations for the data elements that biomedical activities produce. The application ontologies that rule the structure of the semantic annotations are located in the core of the data management system. At the cortex part, we find the different types of complex data elements, coming from very different biomedical activities and departments, that need to be annotated before being exploited in the context of an SDW. Typically, semantic annotations are expressed in XML or RDF formats.

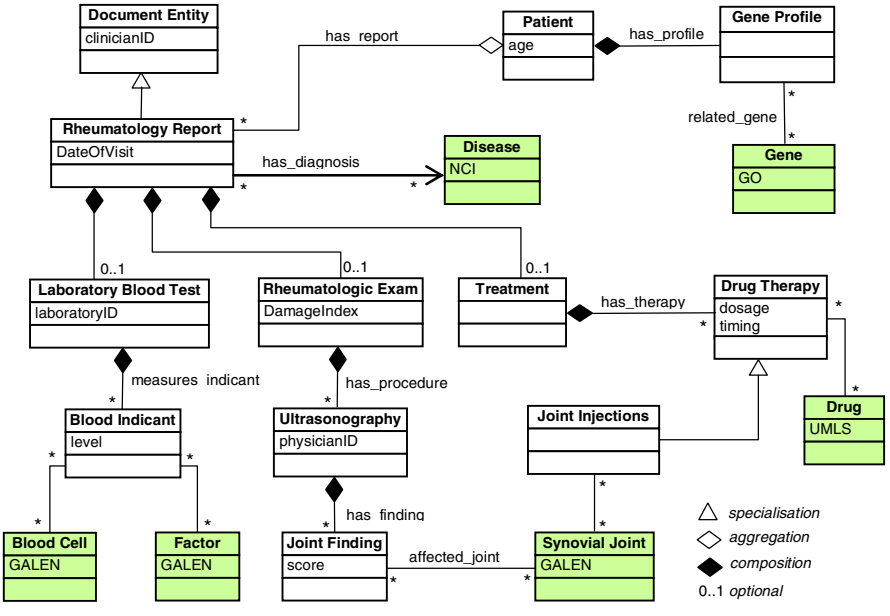


**Fig. 2.** A fragment of an application ontology for Rheumatology

In the biomedical scenario, semantically annotated data consists of many different types of data (e.g. lab test reports, ultrasound scans, images, etc.) originating from heterogeneous data sources. This data also presents complex relationships that evolve rapidly as new biomedical research methods are applied. As a consequence, this data cannot be properly managed by current data warehouse technology, mainly because it is complex, semi-structured, dynamic and highly heterogeneous.

Figure 2 illustrates an ontology fragment for the Rheumatology domain. As the figure shows, a patient may have different rheumatology reports, authored by some clinicians, consisting of the results of some blood tests and rheumatologic exams, the diagnosis of a disease (defined in the domain NCI ontology) and the proposed treatment. The objective of these examinations is to estimate an overall damage index by performing some ultrasonography tests. The treatment is modelled as a collection of drug therapies, sometimes applied in the affected joints. The joint set is compiled

from the GALEN domain ontology. The patient has a genetic profile. The cells and genes involved in the genetic profiles are described by the GALEN and GO domain ontologies, respectively.

Although in Figure 2 we have used UML to graphically represent the ontology fragment, the actual representation formalism will in practice rely on standard languages such as RDF/S and OWL. External concepts coming from domain ontologies are represented in the UML diagram with shaded boxes, indicating the source ontology within the attribute section (e.g. NCI, GO, etc.). Domain ontologies can be used to control the vocabulary and to bring further semantics to the annotated data. Table 1 shows an example of semantically annotated data generated from the application ontology of Figure 2 and stored as RDF triples.

**Table 1.** Application ontology instances stored as RDF triples

| Subject | Predicate | Object |
|---|---|---|
| Patient8991u | type | Patient |
| Patient8991u | age | 15 |
| Patient8991u | sex | Male |
| Patient8991u | has_report | RR001u |
| RR001u | type | Rheumatology |
| RR001u | dateOfVisit | 2008/02/22 |
| RR001u | clinicianID | Clinician2293u |
| RR001u | has_Diagnosis | RA1 |
| RA1 | label | Rheumatoid_Arthritis$^{NCI}$ |
| RA1 | type | Disease$^{NCI}$ |
| RR001u | has_Section | LBT1234u |
| … | … | … |

In the context of this application scenario, our aim is to build a warehouse where semantically annotated data can be analysed with OLAP-based techniques. As use case, we propose to analyse the efficacy of different drugs in the treatment of several types of inflammatory diseases, mainly rheumatic ones. The analysts of this use case should define the dimensions, measures and facts that will allow the analysis of the semantic annotations, gathered from several hospitals and, therefore, expressed with different application ontologies. Notice that at this point, the analyst does neither know the values nor the roll-up relationships that will eventually be used in the resulting cube. As we will show, the framework presented in this paper will capture this information from the application and the domain ontologies involved in the analysis.

Figure 3 shows the seven dimensions that we have selected in order to study this use case from different points of view, including: the patient's age and gender, the subtype of disease (diagnosis), the biomarkers taken from the patient, the damage index of patient´s joints and the drugs administered during the follow-up visits of the patient. Since we consider that the relation that exists between disease symptoms and affected body parts is very relevant for the analysis, we have introduced the category Anatomy in the disease dimension. The biomarkers of interest include blood cells, blood factors and genes. The category Tissue has been similarly introduced in the biomarkers dimension in order to relate biomarkers with their associated tissues.

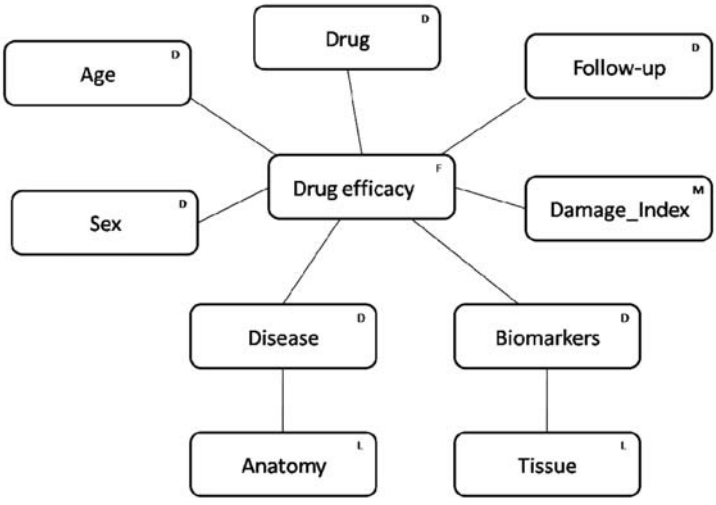


**Fig. 3.** Dimensions defined for analyzing rheumatology patients. We use the letter *D* for dimensions, *F* for facts, *M* for measures and *L* for dimension levels.

In this use case, OLAP technologies can be applied to perform useful analysis operations over the gathered data, as for example:

- By applying roll-up operations, we can aggregate data into coarser granularities such as drug families, active principles, types of diseases, and so on. On the contrary, by means of the available drill-down operations, we can refine each of the analysis dimensions to obtain data with a finer granularity. This kind of operations can give useful information to the clinicians about the relation between diagnosis and treatment efficacy.
- By applying selection and projection operations, we can restrict the analysis to patient subsets according to criteria based on age, sex, affected body parts, etc.

In this section we have defined an application scenario and a use case for the SDWs we want to achieve. In this scenario we identify the following set of application requirements:

1. Integration of biomedical data, information and knowledge to gain a comprehensive view of patients.
2. Scalable data storage functionalities to store the collected semantic information as well as the relevant application and domain ontologies.
3. Flexible ways of specifying analysis dimensions, measures and facts based on medical criteria.
4. Easy exploration of large domain ontologies considering their implicit semantics, and the possible overlapping in their concepts (e.g. mappings).

In the context of other application scenarios these requirements should not be much different, so from our point of view, they can be considered as a basic set of requirements for a generic analysis application of an SDW. It is worth mentioning that the contributions of this paper described in the introduction are aimed at covering all these requirements.

# 3 Background and Related Work

In this section we review the basic concepts involved in the representation, generation and storage of semantic annotations of data, as well as some related work about the analysis of semantic data.

## 3.1 OWL, Description Logics and OLAP

The Ontology Web Language (OWL) is a language for the specification of ontologies, whose definition by the W3C Consortium has empowered the biomedical community to develop large and complex ontologies like the NCI thesaurus, GALEN, etc. OWL provides a powerful knowledge representation language that has a clean and well defined semantics based on Description Logics (DL). Description Logics are a family of knowledge representation formalisms devised to capture most of the requirements of conceptual modelling. These formalisms are decidable subsets of First Order Logic that are expressive enough to capture interesting conceptual modelling properties. The main purpose of DLs is to provide a formal theory that can be used to validate conceptual schemata (Franconi & Ng, 2000) of heterogeneous databases (Mena et al., 2000), data warehousing design and multidimensional aggregation modelling (Baader & Sattler, 2003). It is worth mentioning that Baader & Sattler (2003) and Franconi & Ng (2000) apply DLs in the context of a traditional warehouse. Our proposal is different; we propose to design the warehouse starting from a collection of semantically annotated data. We use DLs for helping the warehouse designer to transform ontology fragments into analysis dimensions, by testing if these dimensions satisfy a set of properties desirable for OLAP applications.

Let us briefly introduce the basic constructors of Description Logics through the basic language $\mathcal{ALC}$ (Schmidt-Schauss & Smolka, 1991), which is summarised as follows:

$$\mathcal{ALC} ::= \bot \mid A \mid C \mid \neg C \mid C \sqcap D \mid C \sqcup D \mid \exists R.C \mid \forall R.C$$

The basic elements of $\mathcal{ALC}$ are concepts (classes in OWL notation), which can be either atomic ($A$) or derived from other concepts (expressions $C$ and $D$). Complex concepts are built by using the classical Boolean operators over concepts, namely: *and* ($\sqcap$), *or* ($\sqcup$) and *not* ($\neg$). Value restrictions on the concept individuals (instances in OWL notation) are represented through roles (object properties in OWL notation), which can be either existential ($\exists R.C$) or universal ($\forall R.C$). The universal concept is denoted with $\top$, whereas the empty concept is denoted with $\bot$. The empty concept is usually associated with inconsistencies and contradictions in the ontology.

Currently there exist several reasoners that deal with some Description Logic languages[1], although most of them do not fully support the retrieval of large sets of asserted instances. Indeed, the complexity of these reasoners is PSpace-complete, which does not guarantee scalability for large domains.

[1] See http://www.cs.man.ac.uk/~sattler/reasoners.html for an exhaustive list. More information about DLs can be found at http://dl.kr.org/

Additionally, several DL constructors have been proposed to capture the main elements of conceptual modelling for databases. For example, *concrete domains* were introduced to account for the usual data types in a conceptual database schema. It has been demonstrated that domains like the integers and strings can be easily introduced into a DL without losing decidability[2] (Lutz et al., 2005). Furthermore, users can state features (i.e., relations between instances and values from these domains) with predicates expressing value comparisons. OWL languages support these constructors via the so-called data type properties. Another interesting constructor for OLAP applications is that of role composition, $R \circ P$, which recently has been introduced in OWL. Role composition allows us to express joined relationships making the intermediate involved concepts implicit. Reasoning over role compositions has been shown to be decidable (Horroks & Sattler 2003), but it is not fully supported by current reasoners yet.

Concerning data warehouse operations, Baader & Sattler (2003) introduced aggregates over concrete domains. The resulting language is called $\mathcal{ALC}$ ($\sum$), and extends the basic language $\mathcal{ALC}$ with concrete domains and a limited set of aggregation functions, namely: sum, min, max and count. Aggregates are introduced through complex features of the form $\Gamma(R \circ u)$, which relate each instance with the aggregate $\Gamma$ over all the values reachable from R followed by the feature u. For example, we can define the following complex feature *sum*(*month* ∘ *income*) to relate instances with their annual incomes. With this complex feature we can ask for employees having annual incomes greater than 100,000 Euros by means of the concept:

$$\mathsf{Employee} \sqcap \exists \mathsf{year}.>(\mathsf{sum}(\mathsf{month} \circ \mathsf{income}), 100000)$$

However, DLs formalisms present important limitations for representing complex measures and aggregations. Baader & Sattler (2003) also demonstrate that handling aggregates in DLs usually leads to undecidability problems, even for very simple aggregates such as sum and count. Moreover, decidable cases present a level of computational complexity too high for practical real-world applications. Baader and Sattler indicate that some interesting inference problems for multidimensional models, such as summarizability, have not been treated by the proposed DLs. Finally, there are no reasoners able to deal with the advanced features required by these new constructors.

Because of these reasons, we propose a new framework to define an integrated ontology that will be used to build a multidimensional data schema over which to apply the OLAP operations required by the analysis tasks. In this way, summarizability will be ensured by building a valid cube from this multidimensional schema so that aggregations are performed over it, out of the DL formalism.

### 3.2 Annotating Biomedical Data

In the biomedical scenario there exist a large number of initiatives for annotating biomedical databases for the Semantic Web. For example, in the SEMEDA project, Köhler et al. (2003) use a controlled vocabulary and an RDF-like ontology to annotate

[2] This occurs whenever the introduced domain satisfies the so-called *admissibility* property.