



INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

*Team leo*

*Distributed and heterogeneous data and  
knowledge*

*Saclay - Île-de-France*

Theme : Knowledge and Data Representation and Management

*Activity*  
*R* *eport*

2010



## Table of contents

<b>1. Team</b>	<b>1</b>
<b>2. Overall Objectives</b>	<b>2</b>
<b>3. Scientific Foundations</b>	<b>2</b>
3.1. Efficient XML and RDF processing	2
3.2. Distributed query and inference systems	3
3.3. Web data management	3
3.4. Ontology-based Information Retrieval	3
3.5. Ontologies for heterogeneous data and knowledge management	3
3.6. Data-intensive visual analytics	3
<b>4. Application Domains</b>	<b>4</b>
<b>5. Software</b>	<b>4</b>
5.1. AlignViz	4
5.2. AnnoViP	4
5.3. EAP Framework	4
5.4. EdiFlow	4
5.5. LiquidXML	4
5.6. LN2R	5
5.7. MESAM	5
5.8. RDFViewS	5
5.9. SomeWhere	5
5.10. SpyWhere	5
5.11. TaxoMap	5
5.12. TaxoMap Framework	6
5.13. ViP2P	6
5.14. XUpOp	6
5.15. XUpIn	6
5.16. BPEL2CPN	6
<b>6. New Results</b>	<b>6</b>
6.1. Efficient XML and RDF data management	6
6.1.1. Access path selection for XML queries	7
6.1.2. Type-based Update Optimization for XML	7
6.1.3. XML query-update independence	7
6.1.4. Precision and complexity of XQuery type inference	7
6.1.5. Document and schema XML updates	8
6.1.6. Managing temporal XML documents	8
6.1.7. Materialized view selection for RDF	8
6.1.8. Hybrid models for XML and RDF	8
6.2. Web Data Management	8
6.2.1. Social networks	8
6.2.2. Data-centric workflow specification	9
6.3. Ontology-based Information Retrieval	9
6.3.1. Adaptive Ontologies for Web Information Retrieval	9
6.3.2. Semantic Annotation	9
6.4. Data and Knowledge Integration	9
6.4.1. Reference Reconciliation	9
6.4.2. Reference Fusion	10
6.4.3. Mapping between ontologies	10
6.4.4. Integration of Web resources	11
6.5. Distributed systems	12

6.5.1.	Distributed XML and RDF querying in ViP2P	12
6.5.2.	Module-based OWL data management	12
6.5.3.	Consequence Finding	13
6.5.4.	P2P consistency checking and query answering	13
6.5.5.	Mapping distributed ontologies	13
6.5.6.	Diagnosis of large distributed software systems	14
6.5.7.	Logical-based P2P diagnosis	14
6.5.8.	Distributed diagnosability	14
6.5.9.	Embedded diagnosis and diagnosability of engineered systems	14
6.5.10.	Satisfiability Checking with Powerful Proof Systems	15
6.6.	Data-intensive visual analytics	15
<b>7.</b>	<b>Contracts and Grants with Industry</b>	<b>15</b>
<b>8.</b>	<b>Other Grants and Activities</b>	<b>15</b>
8.1.	Regional Initiatives	15
8.1.1.	Digiteo EDIFlow project	15
8.1.2.	Digiteo SHIRI project	15
8.2.	National Initiatives	16
8.2.1.	ANR CODEX	16
8.2.2.	ANR Dataring	16
8.2.3.	ANR GEONTO project	16
8.2.4.	ANR UNLOC project	16
8.2.5.	FRAE SIRASAS project	16
8.2.6.	ANR VERSO PIMI project	16
8.2.7.	Participation to evaluation committees	17
8.3.	European Initiatives	17
8.3.1.	Webdam ERC Grant	17
8.3.2.	HYCON 2	17
8.4.	International Initiatives	17
8.4.1.	Cooperation within Europe	17
8.4.2.	Cooperation with Senegal	18
8.4.3.	Cooperation with North America	18
8.4.4.	Cooperation with Japan	18
<b>9.</b>	<b>Dissemination</b>	<b>18</b>
9.1.	Thesis	18
9.2.	Participation in Conferences	19
9.3.	Invited Presentations	21
9.4.	Animation of the scientific community	21
9.5.	Teaching	21
<b>10.</b>	<b>Bibliography</b>	<b>21</b>

*Leo has been created on January the 1<sup>st</sup>, 2010. Leo is a joint team with Laboratoire de Recherche en Informatique (UMR 8623 CNRS & University Paris-Sud), located in Orsay. See <http://leo.saclay.inria.fr>.*

# 1. Team

## Research Scientists

Serge Abiteboul [Senior Researcher INRIA, HdR]  
Ioana Manolescu [Senior Researcher INRIA, Team leader, HdR]  
Meghyn Bienvenu [Junior Researcher CNRS]

## Faculty Members

Nicole Bidoit [Professor, Univ. Paris 11, HdR]  
Philippe Chatalic [Associate Professor, Univ. Paris 11]  
Dario Colazzo [Associate Professor, Univ. Paris 11]  
Philippe Dague [Professor, Univ. Paris 11, HdR]  
François Goasdoué [Associate Professor, Univ. Paris 11]  
Nathalie Pernelle [Associate Professor, Univ. Paris 11]  
Chantal Reynaud [Professor, Univ. Paris 11, co-Team leader, HdR]  
Brigitte Safar [Associate Professor, Univ. Paris 11]  
Fatiha Saïs [Associate Professor, Univ. Paris 11]  
Laurent Simon [Associate Professor, Univ. Paris 11, HdR]

## External Collaborators

David Gross-Amblard [Associate Professor, Univ. Dijon]  
Tarek Melliti [Associate Professor, Univ. Evry Val d'Essonne]  
Philippe Rigaux [Professor, CNAM, HdR]  
Marie-Christine Rousset [Professor, Univ. Grenoble, HdR]  
Virginie Thion-Goasdoué [Associate Professor, CNAM]

## Technical Staff

Jesùs Camacho\_Rodriguez  
André Fonseca [since December]  
Nobal Niraula [until July]  
Alin Tilea [until May]

## PhD Students

Vincent Armant [ATER, Paris 11]  
Emilien Antoine [ERC grant, Paris 11]  
Mohamed Amine Baazizi [Allocataire MENRT, Paris 11]  
Michel Batteux [CIFRE with Sherpa Engineering, Paris 11]  
Pierre Bourhis [ENS Cachan]  
François Calvier [ATER until August, Paris 11]  
Alban Galland [X-Telecom, Paris 11]  
Fayçal Hamdi [ANR grant, Paris 11]  
Konstantinos Karanasos [ANR grant, Paris 11]  
Asterios Katsifodimos [Allocataire MENRT, Paris 11]  
Rania Khefifi [ANR grant, Paris 11]  
Julien Leblay [Allocataire MENRT, Paris 11]  
Wael Khemiri [Allocataire MENRT, Paris 11]  
Yingmin Li [ANR grant, Paris 11]  
Noor Malla [grant of Syrian government, Paris 11]  
Yassine Mrabet [Digiteo grant, Paris 11]  
Huu-Nghia Nguyen [Allocataire MENRT, Paris 11]  
Marina Sahakyan [Grant of French embassy in Armenia, co-supervised between Armenia and Paris 11]

Mouhamadou Thiam [Co-supervised between Sénégal and Paris 11]  
Federico Ulliana [Allocataire MENRT, Paris 11]  
Lina Ye [Allocataire MENRT, Paris 11]  
Nadjet Zemirline [Contract, Paris 11]

#### **Post-Doctoral Fellows**

Nada Abdallah [ATER, Paris 11]  
George Katsirelos [ANR grant]  
Ioannis Katsis [ERC grant]  
Hexuan Hu [Post Doc fellowship, until August]  
Othman Nasri [Post Doc fellowship, until September]  
Fabian Suchanek [ERC grant]

#### **Visiting Scientists**

Angela Bonifati [Researcher, CNR, Italy]  
Carlo Sartiani [Associate Professor, Università della Basilicata, Italy]  
Vasilis Vassalos [Associate Professor, AUEB, Greece]  
Yannis Papakonstantinou [Professor, UCSD, USA]

#### **Administrative Assistant**

Céline Halter [ITA]

## **2. Overall Objectives**

### **2.1. Highlights**

Significant efforts were invested this year in prototyping, leading to strong improvements in the ViP2P, TaxoMap and EdiFlow platforms, as well as to many new research prototypes.

Concerning the life of the team, Meghyn Bienvenu (CR2, CNRS, co-recipient of the AFIA PhD prize) joined the team in October. Four PhDs and an HDR were defended.

In our national and international scientific community, Leo has been very active with many participations to PCs and many events (co-)chaired. Leo has also proposed a Data Integration activity (coordinated by I. Manolescu) within the EIT ICT Labs “Knowledge and Innovation Community”, funded in 2011, and participates to a second activity “Cloud Computing for the Connected City”.

## **3. Scientific Foundations**

### **3.1. Efficient XML and RDF processing**

The development of Web technologies has led to a strong increase in the number and complexity of the applications which represent their data in Web formats, among which XML (for structured documents) and RDF (for Semantic Web data) are the most prominent. Leo has carried on research on algorithms and systems for efficiently processing expressive queries on such Web data formats.

We have considered the efficient management of XML and RDF data, both for query evaluation and for efficiently applying updates, possibly in concurrence with queries. The optimization techniques we considered are based on static analysis, and in particular on the usage of types, as well as on optimizing the access path to the XML data within the query processor.

## 3.2. Distributed query and inference systems

In many application domains, the available data and knowledge is naturally spread among independent sources/agents. Some of them can collaborate with others, but none of them has a global view of either the total available knowledge or the topology of the network of interacting agents. In this context, *peer-to-peer systems* are of particular interest, because in such settings there is no centralized control or hierarchical organization. This lack of global knowledge and global control, however, significantly complicates data and knowledge management tasks, since efficient processing (or optimization) typically requires some information about the system configuration and/or the users' needs. Another challenge arising in the P2P context is to propose distributed reasoning techniques, able to exploit all the global knowledge, despite the lack of a global view on this knowledge.

Our research has focused on consequence finding in distributed propositional systems, and consistency checking and query answering in distributed description logic systems.

## 3.3. Web data management

Use of the Web to share personal data is increasing rapidly with the emergence of Web 2.0 and social networks applications. However, users have yet to trust all the different hosts of their data and face difficulty with updates. We have worked on devising expressive models for handling, sharing and updating data in a trustworthy manner in distributed and in particular in social network systems.

We have also carried work on developing suitable specification mechanisms for data-centric workflows, frequently enacted through Web applications, understanding the expressive power of the different specification formalisms, and performing static analysis on workflow specifications.

## 3.4. Ontology-based Information Retrieval

The research advances on automating ontology population and document annotation are promising but even for named entity-based approaches or pattern-based approaches, it remains difficult to locate precisely concept instances since some of them may be blended in heterogeneous semi-structured documents. In a HTML or XML document, the granularity of the annotation could be precise, at the term level, or imprecise, at the node level. We have defined an annotation model expressive enough to represent and query semantic annotations made on heterogeneous semi-structured documents. In this context, Ontology-based user queries have to be reformulated to gather document nodes from documents that were annotated in an imprecise or incomplete manner by semantic annotation tools.

## 3.5. Ontologies for heterogeneous data and knowledge management

A main theme of the group is integration of heterogeneous information. The current explosion of data sources available and the availability of simple technologies for exchanging data has created an important demand for interoperability among resources. It increases the need for techniques which allow their integration and thus also the integration of schemas or ontologies used to describe the data and give it meaning.

Ontologies, which define domain concepts, are essential elements in integration systems. Leo works on ontology-based environments providing access to multiple heterogeneous sources and on reconciliation techniques, and on providing full life-cycle support for ontologies.

## 3.6. Data-intensive visual analytics

The increasing complexity and sophistication of information sources bring the need for powerful tools enabling users to interact with the data, possibly based on mechanisms simpler and more intuitive than through a query language. Visual analytics applications have been developed by experts in data visualisation, and enable users to interact in a rich and intuitive way with the data. However, most visual analytics applications are constrained in the maximum sizes of the data sets they can handle by their main-memory implementation. This raises the need for efficient persistent database support for visual analytics applications.

## 4. Application Domains

### 4.1. Application Domains

- Geographic data integration (GEONTO contract with IGN)
- Customer reference resolution (HEDI contract with Thales)
- Resolution of duplicates in INRIA HAL bibliographic data (EdiFlow contract)

## 5. Software

### 5.1. AlignViz

Name: AlignViz

Contact: Fayçal Hamdi (hamdi@lri.fr)

Other contacts: Brigitte Safar (safar@lri.fr) and Chantal Reynaud (chantal.reynaud@lri.fr)

Presentation: a visualization tool for alignments between ontologies

### 5.2. AnnoViP

Name: AnnoViP

Contact: Konstantinos Karanasos (konstantinos.karanasos@inria.fr)

Other contacts: Ioana Manolescu (ioana.manolescu@inria.fr) and Jesús Camacho\_Rodriguez (jesus.camacho-rodriguez@inria.fr)

Presentation: a tool for editing and exploiting XML documents with annotations in a distributed P2P setting

### 5.3. EAP Framework

Name: EAP Framework

Contact: Nadjat Zémirline (nadjat.zemirline@supelec.fr)

Other contacts: Chantal Reynaud (chantal.reynaud@lri.fr)

Presentation: a prototype helping to express adaptation strategies based on the use and a semi-automatic combination of patterns

### 5.4. EdiFlow

Name: EdiFlow (<http://scidam.gforge.inria.fr>)

Contact: Wael Khemiri (wael.khemiri@inria.fr)

Other contacts: Ioana Manolescu (ioana.manolescu@inria.fr), Jean-Daniel Fekete (jean-daniel.fekete@inria.fr), Pierre-Luc Hémerly (pierre-luc.hemery@inria.fr), Véronique Benzaken (veronique.benzaken@lri.fr)

Presentation: A platform for data-intensive visual analytics

### 5.5. LiquidXML

Name: Liquid XML (<http://vip2p.saclay.inria.fr/?page=liquidxml>)

Contact: Asterios Katsifodimos (asterios.katsifodimos@inria.fr)



Other contacts: Jesús Camacho\_Rodriguez (jesus.camacho-rodriguez@inria.fr), Ioana Manolescu (ioana.manolescu@inria.fr)

Presentation: a self-tuning P2P system for distributing XML content according to the query needs of the distributed peers

## 5.6. LN2R

Name: LN2R

Contact: Fatiha Saïs (saïs@lri.fr)

Other contacts: Nathalie Pernelle (pernelle@lri.fr)

Presentation: a logical and numerical tool for reference reconciliation.

## 5.7. MESAM

Name: MESAM

Contact: Nadjat Zémirline (nadjat.zemirline@supelec.fr)

Other contacts: Chantal Reynaud (chantal.reynaud@lri.fr)

Presentation: a plug-in for Protege 2000 to merge generic and specific models

## 5.8. RDFViewS

Name: RDFViewS (<http://rdfvs.saclay.inria.fr>)

Contact: Konstantinos Karanasos (konstantinos.karanasos@inria.fr)

Other contacts: François Goasdoué (fg@lri.fr), Julien Leblay (julien.leblay@inria.fr), and Ioana Manolescu (ioana.manolescu@inria.fr)

Presentation: a storage tuning wizard for RDF applications

## 5.9. SomeWhere

Name: SomeWhere

Contact: François Goasdoué (fg@lri.fr)

Other contacts: Philippe Chatalic (chatalic@lri.fr) and Laurent Simon (simon@lri.fr)

Presentation: a peer-to-peer infrastructure for propositional reasoning

## 5.10. SpyWhere

Name: SpyWhere

Contact: François-Elie Calvier (fcalvier@gmail.com)

Other contacts: Chantal Reynaud (chantal.reynaud@lri.fr)

Presentation: a generator of mapping candidates for enriching peer ontologies

## 5.11. TaxoMap

Name: TaxoMap

Contact: Fayçal Hamdi (hamdi@lri.fr)

Other contacts: Brigitte Safar (safar@lri.fr) and Chantal Reynaud (chantal.reynaud@lri.fr)

Presentation: a prototype to automate semantic mappings between taxonomies

## 5.12. TaxoMap Framework

Name: TaxoMap Framework

Contact: Fayçal Hamdi (hamdi@lri.fr)

Other contacts: Brigitte Safar (safar@lri.fr) and Chantal Reynaud (chantal.reynaud@lri.fr)

Presentation: an environment to specify treatments to refine mappings and to enrich ontologies

## 5.13. ViP2P

Name: ViP2P (views in peer-to-peer, <http://vip2p.saclay.inria.fr>)

Contact: Ioana Manolescu (ioana.manolescu@inria.fr)

Other contacts: Jesús Camacho\_Rodriguez (jesus.camacho-rodriguez@inria.fr), Asterios Katsifodimos (asterios.katsifodimos@inria.fr), Konstantinos Karanasos (konstantinos.karanasos@inria.fr)

Presentation: a P2P platform for disseminating and querying XML and RDF data in large-scale distributed networks.

## 5.14. XUpOp

Name: XUpOp (XML Update Optimization)

Contact: Dario Colazzo (colazzo@lri.fr)

Other contacts: Nicole Bidoit (bidoit@lri.fr), Marina Sahakian (Marina.Sahakyan@lri.fr), and Mohamed Amine Baazizi (baazizi@lri.fr)

Presentation: a general purpose type based optimizer for XML updates

## 5.15. XUpIn

Name: XUpIn (XML Update Independence)

Contact: Federico Ulliana (Federico.Ulliana@lri.fr)

Other contacts: Dario Colazzo (colazzo@lri.fr), Nicole Bidoit (bidoit@lri.fr)

Presentation: an XML query-update independence tester

## 5.16. BPEL2CPN

Name: BPEL2CPN

Contact: Yingmin Li (yingmin.li@lri.fr)

Other contacts: Philippe Dague (philippe.dague@lri.fr) and Tarek Melliti (tmelliti@ibisc.fr)

Presentation: a tool to convert a composite Web service written in BPEL into a Colored Petri Net model used to achieve decentralized diagnosis.

# 6. New Results

## 6.1. Efficient XML and RDF data management

**Participants:** Mohamed Amine Baazizi, Nicole Bidoit, Dario Colazzo, François Goasdoué, Konstantinos Karanasos, Julien Leblay, Noor Malla, Ioana Manolescu, Marina Sahakyan, Federico Ulliana.

### 6.1.1. Access path selection for XML queries

We have continued our work on optimizing XML queries through materialized view-based rewriting, implemented within the ViP2P system, jointly with V. Vassalos from AUEB, Greece. We have proposed an algorithm (to appear in the ICDE 2011 conference) for rewriting XQuery queries using materialized XQuery view, which improves the state of the art in terms of expressive power of the supported XQuery subset. A tutorial on this topic has been presented by A. Deutsch (UCSD, USA) I. Manolescu, and V. Vassalos (AUEB, Greece) at the French BDA 2010 conference.

### 6.1.2. Type-based Update Optimization for XML

XML projection is a well-known optimization technique for reducing memory consumption for XQuery in-memory engines in order to overcome the main-memory limitations of these systems (Galax, Saxon, QizX, and eXist). For instance, for eXist, QizX/open and Saxon it is not possible to update documents whose size is greater than 150 MB. We are currently investigating this technique for update optimization. The update language considered is XQuery Update Facility (XUF).

The main idea behind this technique is: given a query  $q$  over an XML document  $t$ , instead of evaluating  $q$  over  $t$ , the query  $q$  is evaluated on a smaller document  $t'$  obtained from  $t$  by pruning out, at loading time, parts of  $t$  that are irrelevant for  $q$ . The queried document  $t'$ , a projection of the original one, is often much smaller than  $t$  due to selectivity of queries.

The scenario and type-based projection proposed for XML queries, cannot be applied directly for updates. We have proposed a new scenario which is composed of four steps:

1. from the update  $U$  and the DTD  $D$ , a type projector  $P$  is inferred;
2. the document  $t$ , valid wrt  $D$ , is projected following  $P$  in a streaming manner, at loading time;
3.  $U$  is evaluated over the projection  $P(t)$  and produces a partial result  $U(P(t))$ ;
4. the initial document  $t$  is merged with  $U(P(t))$ , in a streaming manner, at writing (serializing) time in order to produce the final result  $U(t)$ .

The scenario has been first tested for a rather simple type projector. We are currently improving the technique by designing type projectors as efficient as possible. The goal is to refine the analysis of XUF in order to extract a projector as precise as possible. This analysis is made more intricate by the fact that projection should also prepare data for the fourth step of the scenario. The methods are validated by prototyping.

### 6.1.3. XML query-update independence

A query and an update are independent when the query result is not affected by update execution, on any possible input database. Detecting query-update independence is of crucial importance in many contexts: view maintenance, concurrency, access control policies etc. Benefits are amplified when query-update independence can be checked statically. We propose a novel schema-based approach for detecting XML query-update independence. Differently from traditional schema-based analysis for XQuery, our system infers sequence of labels, called *chains*, that are vertically navigated in each schema instance by query and update paths. More precisely, for each node that can be selected by a query/update path in a schema instance, the system infers a chain recording: a) all labels that are encountered from the root to the selected node, and b) the order in which these labels are traversed. The contextual and ordering information provided these chains is at the basis of and extremely precise static independence analysis.

### 6.1.4. Precision and complexity of XQuery type inference

A key feature of XQuery is its type system. Any language expression is statically typed and its type is used during program type-checking. In XQuery, types of input data and functions are defined in terms of regular expression types, but it is quite easy to write queries that generate non-regular languages. As a consequence, any type system for XQuery has to rely on a *type inference* process that approximates the (possibly non-regular) output type of a query with a regular type. This approximation process, while mandatory and unavoidable, may significantly decrease the precision of the inferred types. In [46] we study and compare in terms of precision and computational complexity two main existing XQuery type systems.

### 6.1.5. Document and schema XML updates

Starting from the update primitives supported by current language proposals, [51] deals with the data management issues that arise when documents and schema are updated and new versions created. Finally, it describes the current research proposals and the facilities of main enabled and native XML databases to handle updates at document and schema level.

### 6.1.6. Managing temporal XML documents

The management of temporal data is a crucial issue in many database applications. We are currently investigating efficient storage and update methods for temporal XML documents, with a focus on compactness of the representation. One of the method developed relies on the type based optimization method developed for updates.

### 6.1.7. Materialized view selection for RDF

Syntactically, RDF, the data format of the Semantic Web, resembles relational data. However, RDF query processing is significantly complicated by the irregular nature of RDF data and by its simplistic data model, which leads to syntactically complex queries (involving many joins over the whole triple data set). When a query workload is known, the performance of the workload can be significantly improved by materializing access support data structures such as materialized views. We have proposed and validated efficient algorithms for selecting a set of views to materialize in order to speed up the processing of a set of RDF queries, and developed a system, named RDFViewS, which implements and validates these ideas. Our algorithms appeared in [47] and RDFViewS was demonstrated in [19].

### 6.1.8. Hybrid models for XML and RDF

We have obtained interesting results in the area of jointly managing XML and RDF data. A first direction of work in this area was to support *annotated documents*, that is, XML documents where individual nodes or fragments could be annotated with RDF triples. This model allows to capture, for instance, blog comments, user ratings on social sites etc. We have formalized a language to query annotated documents, and implemented it based on materialized views in the AnnoViP system, which has been demonstrated in [45], [24]. More recently, we have started work on a more general model which fully subsumes XML and RDF and allows capturing the expressive power strengths of both the data and the semantic worlds.

## 6.2. Web Data Management

**Participants:** Serge Abiteboul, Emilien Antoine, Pierre Bourhis, Alban Galland, David Gross-Amblard, Ioana Manolescu, Fabian Suchanek.

Significant efforts were invested this year towards writing a new book on Web Data Management and Distribution [50], to appear in 2011.

### 6.2.1. Social networks

To allow for efficient, flexible sharing of data in social networks, we devised a model of distributed knowledge base with access control and cryptographic functionality. The model allows exchanging documents, access control statements, keys and instructions in a distributed setting. We are considering different implementations of this model that can be used to leverage technologies such as DHT or Gossiping. The model allows us to reason on the knowledge base, in particular on the access control and the distribution policies. The reasoning is based on trusted knowledge, the statements, whose authenticity can be easily checked.

In such a social network, participants may bring conflicting opinions. We have studied the problem of trying to corroborate information coming from a very large number of participants. We have proposed and evaluated various algorithms towards this goal [18].

### 6.2.2. Data-centric workflow specification

There has recently been a proliferation of workflow specification formalisms, notably data-centric, in response to the need to support increasingly ubiquitous processes centered around databases. Prominent examples include e-commerce systems, enterprise business processes, health-care and scientific workflows. In this context, we have developed a model for data-centric workflow specification. The model has been implemented within a prototype, named AXART, demonstrated at the VLDB 2010 conference [14].

## 6.3. Ontology-based Information Retrieval

**Participants:** Nathalie Pernelle, Chantal Reynaud, Mouhamadou Thiam, Yassine Mrabet.

### 6.3.1. Adaptive Ontologies for Web Information Retrieval

We published the experimental study of the Ontology-based Web search Framework made last year [20].

### 6.3.2. Semantic Annotation

SHIRI is an ontology-based system for integrating semi-structured documents related to a specific domain. The system purpose is to allow users to access to relevant parts of documents as answers to their queries. It relies on an automatic, unsupervised and ontology-driven approach for extraction, alignment and semantic annotation of tagged elements of documents.

- We have improved the order relation defined in the *SHIRI*-querying component [27], [42]. This component reformulates an ontology-based user query in order to obtain answers that come from more or less structured tagged elements.
- We currently define an approach which allows to integrate knowledge bases and semantic annotations that are provided by several annotators. In this setting, data and annotations are described using the vocabulary of different ontologies. Furthermore, the annotated data of the documents have to be reconciled or semantically linked. The new Querying component exploits uncertain ontology element alignments and uncertain data reconciliations to improve the recall. A tool is under implementation and the evaluation is done using the ontologies and the knowledge bases provided by DBPEDIA and KIM. The semantic annotations are made on WIKIPEDIA pages.

## 6.4. Data and Knowledge Integration

**Participants:** Fayçal Hamdi, Nibal Niraula, Nathalie Pernelle, Chantal Reynaud, Fatiha Saïs, Brigitte Safar, Fabian Suchanek.

### 6.4.1. Reference Reconciliation

The reference reconciliation problem consists in deciding whether different data descriptions refer to the same real world entity (same person, same conference, ...). In previous work, we have developed a knowledge-based approach named LN2R which combine two methods: a logical method and a numerical method. This approach allows computing a set of reference pairs that (1) refers to the same entity, (2) does not refer to the same entity and (3) may refer to the same entity. The logical method (L2R) translates knowledge that is declared in the ontology into a set of logical rules in order to infer certain (non) reconciliation decisions. The numerical method (N2R) computes a similarity score for each pair of references by taking into account knowledge that is declared in the ontology.

We are working on the extension and the adaptation of LN2R in several directions:

- design a Reference Reconciliation approach by considering different schemas (ontologies). In the setting of GeOnto project we have designed a combined approach of data and ontology reconciliation where (i) mappings (or similarity scores) between concepts are taken into account in the similarity computation between instances and (ii) reconciliation decisions (or similarity scores) are taken into account in the similarity computation between concepts. The influences between all these similarities are modeled by two non linear equation systems. A tool is under implementation and detailed description of this approach can be found in *GeOnto Report 2.2 of T0+30*.
- develop a method which can provide the user comprehensive and readable explanations for a set of reference reconciliation decisions. We have developed a graphical approach which is based on Coloured Petri Nets formalism. The obtained results will be published in the AKDM book chapter [52].
- in addition to the reconciliation and non reconciliation decisions, the logical method L2R allows inferring the semantic equivalence of heterogeneous basic values that are stored in a dictionary. We have studied how this dictionary can be automatically refined in order to improve the reconciliation results in the settings of collaboration with THALES (in the HEDI project). We have also studied the reduction of data comparisons by discovering association rules between significant substrings and n-grams that are extracted from electronic product descriptions.
- develop a reference reconciliation method for detecting redundant data in case of web data tables that are semantically annotated by an ontology. Each table cell values consists in numerical fuzzy set (NFS) or in symbolic fuzzy set (SFS). We have developed a method which uses ontology knowledge and computes similarity scores to decide the data redundancy. We have also proposed two similarity measures for numerical fuzzy set as well as symbolic fuzzy set. The proposed measures are more flexible than existing ones. First results will be published in the 2011 EGC conference.
- we have participated in the second Instance Matching track of Ontology Alignment Evaluation Initiative 2010 Campaign by using LN2R system. We participated in OWL data track and we obtained good results (third-top results). The obtained results will be published in [29].

#### 6.4.2. Reference Fusion

The issue of data fusion arises once reconciliations have been determined. The objective of the fusion is to obtain a unique representation of the real world entity. In previous work, we have proposed a fusion approach which deals with the uncertainty in the values associated with the attributes thanks to a formalism based on belief functions whose shapes are based on a set of criteria, using evidence theory formalism has been developed.

In [30], we propose a fusion method based on possibility theory, able to cope with uncertainty and with ontological knowledge. An implementation using W3C standards is provided. Rising from the fusion process, an ontology enrichment procedure is proposed to complete the global ontology. The evaluation of the proposed approach on Cora data set of scientific publication domain, obtained promising results.

The aim now is to build a flexible querying approach of the fused data where the user preferences are taken into account.

#### 6.4.3. Mapping between ontologies

We pursue our work on TaxoMap in the setting of the GeOnto project. We focused on two main points: mapping refinement and ontology enrichment.

Mapping refinement

- We designed and implemented an environment, called Taxomap framework, based on TaxoMap, helping an engineer to refine mappings to take into account specific conventions used in the ontologies.

In [21], we focus on the conception of the environment, on the definition of a first set of primitives to support the specification of treatments, and on a presentation of a use of the environment for the mapping refinement in the topographic domain.

In [22], we introduce the mapping refinement workflow and MPRL, the Mapping Refinement Pattern Language used in the environment to define mapping refinements. We present a set of mapping refinement patterns built in the setting of the ANR project GeOnto and experiments using these patterns.

Demonstrations of TaxoMap Framework have been made in RFIA, in the international symposium on Matching and Meaning and in the OntoGeo workshop [41].

- TaxoMap alignment and refinement modules have been used to take part to two tests in the international alignment contest OAEI [23] for the fifth consecutive time. It allows us to test the robustness of our tool and the new implemented techniques. We observed that the new version of Taxomap significantly improved over the previous version in terms of execution time and precision of generated mappings. As a participant of this contest, we have had to extend a web service interface and deploy our alignment tool as a web service accessed during the evaluation process. We used the SEALS evaluation service to test our system and launch our own experiments.

#### Ontology enrichment

- TaxoMap framework has been designed to allow the specification of other treatments based on alignment results. Currently, we study its use for enriching ontologies. In the setting of the ANR GeOnto project, we use it to enrich the IGN topographic ontologies with concepts extracted from the Rameau thesaurus. This led us to consider two different enrichment scenarios depending on whether the ontology used to enrich data is specific to the domain or not. This led us also to define new patterns. The first results will be published in the 2011 EGC conference.
- In parallel, we also study the enrichment of the same IGN ontology with Yago being developed at the Max Planck Institute for Computer Science. Yago is a huge semantic knowledge base knowing over two millions entities and about twenty million facts about these entities. We investigate the problem of identifying parts of Yago which are semantically close to the IGN ontology, a preliminary work on enrichment, based on our previous work on partitioning.

#### 6.4.4. Integration of Web resources

We proposed an approach to integrate resources available on the Web in Adaptive Hypermedia Systems (AHS). This approach allows creators to define their own adaptation strategies based on their own domain models. The expression of the strategies is independent of any adaptation engine. However, they can be automatically integrated into existing AHSs. This approach using adaptation patterns and based on the Nuggets language to define adaptation has been designed and implemented in the EAP framework. We published the approach in the conferences [43] and [36].

We also investigated how to integrate the EAP framework on other AHSs. We developed a collaboration with A. Cristea from the University of Warwick (UK) in order to study the expressivity and interoperability of our own flexible adaptation languages, LAG and Nuggets. We have implemented a translation module to generate adaptation strategies expressing using the EAP Framework to LAG. We evaluated the expression of adaptation using the EAP framework versus a rule-based language. We evaluated also the ease of use of LAG versus Nuggets. We are currently working to publish all these results in international journals.

On related topics, together with researchers from the Max-Planck institute, we have worked on enriching Semantic Web data repositories. More specifically, we have focused on extending the very large Yago ontology with temporal and spatial dimensions. A collaboration with Politecnico di Milano aims at bridging the gap between the Semantic Web and the Web service technologies they developed within the Search Computing ERC project.



## 6.5. Distributed systems

**Participants:** Nada Abdallah, Vincent Armant, Michel Batteux, François Calvier, Jesús Camacho\_Rodriguez, Meghyn Bienvenu, Philippe Chatalic, Philippe Dague, François Goasdoué, Konstantinos Karanasos, Asterios Katsifodimos, Hexuan Hu, Yingmin Li, Ioana Manolescu, Tarek Melliti, Othman Nasri, Huu-Nghia Nguyen, Chantal Reynaud, Alin Tilea, Laurent Simon, Lina Ye.

Leo research work has considered two main application fields: distributed dissemination and querying of Web data (in particular XML, RDF, and OWL data), and distributed inference.

### 6.5.1. Distributed XML and RDF querying in ViP2P

Within the area of distributed dissemination and querying, we have pursued our work consolidating and optimizing the ViP2P platform for peer-to-peer XML data management. In ViP2P, independent peers can publish XML and/or RDF data. Publishing means making the data available to other peers which may query it. ViP2P supports two modes for querying: long-running queries are treated as subscriptions, i.e., results are returned when they become present in the system, and stored at the subscribing/query peer; snapshot queries are answered instantly based on the contents of the existing subscriptions.

Based on a detailed performance analysis, we have brought numerous optimizations to ViP2P publication and query processing, leading to impressive scalability results. Thus, the platform is capable of disseminating hundreds of GBs of XML data based on fine-granularity subscriber requests in a WAN of 250 computers, in less than 15 minutes. These performance results are the best known to us so far in the area of P2P XML data management. Our work on ViP2P has been the focus of the invited presentation and demo by I. Manolescu and S. Zoupanos at the BDA 2010 summer school. Moreover, we have considered the problem of automatically deploying XML content in a large-scale distributed network according to the needs formulated through user subscriptions. We have devised and deployed the LiquidXML platform for XML data redistribution, and demonstrated it in [16]. Finally, we have deployed the AnnoViP model (XML documents with RDF annotations) also in the peer-to-peer framework of ViP2P. The interest of this deployment is to allow distributed, independent users to add to each other's knowledge by annotating the same documents or by building on each other's information.

### 6.5.2. Module-based OWL data management

The current trend for building an ontology-based data management system (DMS) is to capitalize on efforts made to design a preexisting well-established DMS (a *reference* system). The method amounts to extract from the reference DMS a piece of schema relevant to the new application needs – a module –, possibly to personalize it with extra-constraints w.r.t. the application under construction, and then to manage a dataset using the resulting schema.

We have revisited the reuse of a reference ontology-based DMS in order to build a new DMS with specific needs [48]. We go one step further by not only considering the *design* of a module-based DMS (i.e., how to extract a module from an ontological schema): we also study how a module-based DMS can benefit from the reference DMS to enhance its own data management skills. We carry out our investigations in the setting of the  $DL - lite_A$  dialect of  $DL - lite$ , which encompasses the foundations of the QL profile of OWL2 (i.e.,  $DL - lite_R$ ): the W3C recommendation for managing efficiently large datasets. Our contribution is to introduce and study novel properties of *robustness* for modules that provide means for checking easily that a robust module-based DMS evolves *safely* w.r.t. both the schema and the data of the reference DMS. From a module *robust to consistency checking*, for any data update in a corresponding module-based DMS, we show how to query the reference DMS for checking whether the local update does not bring any inconsistency with the data and the constraints of the reference DMS. From a module *robust to query answering*, for any query asked to a module-based DMS, we show how to query the reference DMS for obtaining additional answers by also exploiting the data stored in the reference DMS.



It is worth noticing that our investigations are sustained by real use-cases. For instance, the MyCF DMS (MyCorporisFabrica, <http://www.mycorporisfabrica.org>) has been built *by hand* from the FMA DMS (Foundational Model of Anatomy, <http://sig.biostr.washington.edu/projects/fm>). The extraction step has focused on particular parts of the human body (e.g., hand, foot, and knee), while the personalization step has enriched the descriptions of these parts with both 3D geometrical and bio-mechanical information. Notably, careful attention was paid so that MyCF still conforms with FMA at the end of the process.

### 6.5.3. Consequence Finding

In a P2P inference system (P2PIS) each peer models its application domain using a knowledge base (KB) and peers having similar interests can establish semantic correspondences, called mappings, between their respective KBs. Mappings not only characterize how the peers integrate with each other but also give rise to a semantic network, the decentralized KB of the P2PIS, in which it becomes possible to reason collectively.

SomeWhere is a platform that has been developed during the last years by Leo members, the scalability of which (up to a thousand of peers) has been demonstrated on synthetic data. It allows to perform truly decentralized consequence finding in networked peers with propositional logic theories. Consequence finding is one corner stone of many AI tasks such as common sense reasoning, diagnosis, or knowledge compilation.

- Bases of a new version of this platform have been set up this year and a complete rewriting and refactoring is underway.

### 6.5.4. P2P consistency checking and query answering

We have investigated a decentralized data model and associated algorithms for peer data management systems (PDMSs) based on the DL-LITE<sub>R</sub> description logic [37] :

- That logic corresponds to the new W3C recommendation for the Semantic Web: OWL2. Our approach relies on reducing query reformulation and consistency checking for DL-LITE<sub>R</sub> into reasoning in propositional logic. This enables a straightforward deployment of DL-LITE<sub>R</sub> PDMSs on top of SomeWhere, our scalable peer-to-peer inference system for the propositional logic.
- We have also shown how to answer queries using views – predefined queries – in DL-LITE<sub>R</sub> in the centralized and decentralized cases, by combining the query reformulation algorithm of DL-LITE<sub>R</sub> and the state-of-the-art query rewriting algorithm: MiniCon.

### 6.5.5. Mapping distributed ontologies

We have continued our experiments validating the approach implemented in SpyWhere:

- Test of the SpyWhere technique to select pairs of ontological elements to align
  - Ontologies in the publication domain have been used.
  - Two kinds of experiments: first scenarios with peers connected by a small number of mappings and then scenarios based on a network composed of a large number of peers.
  - Objectives: Evaluation of the process to identify elementary target elements and pairs of elements which are relevant to align - Evaluation of the various strategies related to the peers with which mappings are looking for.
  - The evaluation takes into account the number of proposed pairs of elements and also their relevance.
- Test of the alignment method, more particularly the inductive technique and the score measure to compute the confidence of a mapping.

The complete research work corresponding to mapping discovery in a peer-to-peer semantic system and its application to SomeRDFS is the core of the PhD of François-Elie Calvier defended on September 17, 2010 [12].

### **6.5.6. Diagnosis of large distributed software systems**

The work devoted to monitoring and diagnosis of composite Web services, and more generally of large distributed software systems, saw a first achievement with Yingmin Li's PhD defense on December, 2010. A CPN (Colored Petri Net) model is used to represent correct and faulty behaviors, for faults in the control process as well as in the data, and in particular faults propagation. This model gives birth to a set of algebraic linear equations describing its behavior through the dynamic evolution of the markings. It is turned into a set of linear symbolic inequalities the solutions of which, in terms of color variables and from initial symptoms (exception rise), express the minimal diagnoses and are computed by a backward propagation algorithm. This work has been extended to the decentralized case for cooperating choreographed BPEL Web services, where global diagnosis is achieved by a coordinator dialoguing with each local diagnoser. A complete implementation has been realized, where the CPN is automatically generated from the BPEL code and both local diagnosers and global coordinator are implemented as Web services too. It has been validated on the foodshop benchmark defined in the European project WS-DIAMOND.

Part of this work will be used by Huu-Nghia Nguyen in his thesis work devoted to exploration of the complementarities between model-based testing and diagnosis, with application to composition and coordination of services, such as Web services. This is a joint work with Fatiha Zaïdi and Pascal Poizat from the LRI testing team ForTesSE, in the context of a co-supervised thesis.

### **6.5.7. Logical-based P2P diagnosis**

Research on consistency-based distributed diagnosis, set up in the framework of propositional P2PISs, pursued in 2010. The algorithms developed are based on the distributed computation of prime implicants of the (unknown) global theory. Most of the work this year was devoted to implementation and experimentation with different jointree structures and different social networks topologies to have a deep understanding of the bottlenecks in scalability issues.

### **6.5.8. Distributed diagnosability**

The aim of diagnosability is to ensure that a given partially observable system has the property that any fault (taken from a set of faults given a priori) will be detectable and identifiable without ambiguity in a finite time after its occurrence. Work on diagnosability is led in the framework of discrete-event systems.

The work conducted along 2010 was focused on component-based systems modeled as synchronized labeled automata. Algorithms were designed for cooperative diagnosability analysis and for optimized distributed diagnosability verification, where components exchange only the minimum of abstracted information needed [35], [34]. Work about pattern diagnosability, where single fault represented by a transition label is replaced by a surveillance pattern represented by the language defined by an automaton, has also been led by using distributed simulation [33].

More recently, a generalization of diagnosability has been proposed for open systems, i.e. systems that are partially controllable from the environment (the classical diagnosability being the case where the system itself has total control on all its events). Such a system can happen to be diagnosable though it would be not diagnosable with the classical definition. Game structure is used to generalize the definition of diagnosability (called active diagnosability) and Alternating-time Temporal Logic is used to check it [26]. This framework is suitable not only for one system and its environment but also for a set of interacting systems. Next step will be precisely to extend the game metaphor for a distributed checking of active diagnosability.

### **6.5.9. Embedded diagnosis and diagnosability of engineered systems**

More applied research was led concerning embedded diagnosis of distributed functions in hardware/software architecture inside vehicles. This was done up to September 2009 in the framework of the DIAFORE ("DIAGnostic de FONctions REparties") ADEME/ANR/PREDIT project which results were published this year, about modeling and decentralized diagnosis of the Smart Distance Keeping function, used as a case study [31], [28]. The embedded (sensors faults) diagnostic architecture with a global diagnostic algorithm implemented in a dedicated Electronic Control Unit acting as diagnostic supervisor between the local

diagnostic algorithms is the first experiment of an embedded diagnostic function decentralized on several ECUs and decision making and reconfiguration action achieved at the supervisor level.

A new research has begun with the objective of making a bridge between off-line diagnosability analysis at design stage and on-line embedded diagnosis. This work is led in the framework of a CIFRE thesis with Sherpa Engineering and the case study is a fuel cell system [17].

#### **6.5.10. Satisfiability Checking with Powerful Proof Systems**

One of the limits of nowadays SAT solvers is related to the underlying proof system they are based on. In our recent work [32], [49] L. Simon added the extended resolution rule to our CDCL solver, glucose. This work was done with G. Audemard and G. Katsirelos with the support of the ANR Grant "UNLOC".

### **6.6. Data-intensive visual analytics**

**Participants:** Ioana Manolescu, Wael Khemiri.

We have continued the development of a database-backed platform for visual analytics, called EdiFlow. This platform is developed by W. Khemiri in collaboration with the AVIZ team of INRIA Saclay, and in particular with the engineer Pierre-Luc Hémerly. We have finalized the development of visual analytics applications where the data resides in a persistent database, users interact with the data, and changes to the database are immediately propagated to the visual analytics application, even in the case where it is remote (runs on a distinct machine). This approach has been published in [44] and will appear at ICDE 2011. It has allowed EdiFlow to be successfully deployed on WILD, the wall-sized interactive display of LRI. We are now working on propagating changes in the opposite directions, i.e. changes issued from user interaction with the visual analytics application and pushed back to the persistent database. This work is carried on in connection with the Digiteo DIM project EdiFlow.

## **7. Contracts and Grants with Industry**

### **7.1. HEDI project**

**Participants:** Fatiha Saïs, Nathalie Pernelle, Ioana Manolescu.

HEDI–Heterogeneous Electronic Data Integration project is a new collaboration agreement between Leo team of INRIA Saclay and Thales Corporate Services started in September 2009. This project aims at designing a data reconciliation tool for electronic component descriptions.

## **8. Other Grants and Activities**

### **8.1. Regional Initiatives**

#### **8.1.1. Digiteo EDIFlow project**

This project, led by I. Manolescu, is a collaboration between data management and information visualisation researchers (with V. Benzaken from LRI and J.-D. Fekete from INRIA AVIZ). The purpose is to study models and build a corresponding platform for efficient data-intensive workflow systems. In 2010, important development efforts have led to the integration of our first prototype with the InfoViz toolkit developed in AVIZ.

#### **8.1.2. Digiteo SHIRI project**

This research project, led by N. Pernelle is funded by the Ile de France region as a Digiteo project. It started on Oct. 1st 2007 and will last until Sept. 30th, 2011. It involves two partners of Digiteo, Supelec and the University of Paris-Sud. The aim of SHIRI is to design an annotation system to improve the relevance of the search on the Web when resources contain both semi-structured and textual data.

## 8.2. National Initiatives

In France, close links exist with groups at Orsay (proofs and programs, V. Benzaken; bio-informatics, C. Froidevaux; machine learning, M. Sebag; information visualization, J.-D. Fekete), with the Cedric Group at CNAM-Paris; some INRIA groups (Dahu, L. Segoufin, at INRIA-Saclay, Atlas, P. Valduriez at Montpellier; Exmo, J. Euzenat, at INRIA Rhone-Alpes; Mostrare at INRIA-Nord-Europe); the BIA group at INRA (P. Buche, C. Dervin), the GRIMM of the University of Toulouse Le Mirail (O. Haemmerlé), the LIRIS of the University of Lyon 1 (M. Hacid), the LIRMM of the University of Montpellier (M. Chein, M-L. Mugnier), INRA/CIRAD Montpellier (P. Buche, S. Destercke and R. Thomopoulos), the LI of the University of Tours (G. Venturini), and the UMPA at l'école normale supérieure de Lyon (Y. Ollivier).

### 8.2.1. ANR CODEX

Codex is a research project supported by the ANR Domaines Emergents call (2009-2012), coordinated by I. Manolescu (<http://codex.saclay.inria.fr>). The project involves INRIA Lille (Mostrare) and Grenoble (WAM) as well as the Universities of Blois, Paris-Sud, Paris 7, and a start-up. Leo work carried on within CODEX in 2010 was related to: the efficient management of XML document carrying RDF annotations, scaling up the ViP2P XML data management platform, maintenance of materialized XML views and efficient evaluation of XML queries and updates.

### 8.2.2. ANR Dataring

The DataRing project (2008-2011) lead by INRIA-Sophia (Patrick Valduriez) includes also U. Grenoble (Marie-Christine Rousset) and Telecom-ParisTech (Pierre Senellart). The project addresses the problem of P2P data sharing for online communities, by offering a high-level network ring across distributed data source owners. The DataRing work carried within Leo encompasses the LiquidXML work on adaptive P2P content redistribution.

### 8.2.3. ANR GEONTO project

The objective of this ANR MDCO project (2008-2011), is to make data in the geographic domain inter-operate. We focus on two main goals. On one hand, we aim at integrating heterogeneous geographic databases using schema matching techniques. On the other hand, we aim at querying a large collection of textual documents which are more various and for a larger readership than databases just mentioned before. This project is a collaboration between COGIT-IGN (Sébastien Mustière), the IC3 group at IRIT - Université Paul Sabatier (Nathalie Aussenac) and the DESI group at LIUPPA - Université de Pau et des Pays de l'Adour (Mauro Gaio). The home page of the project could be found at: <http://geonto.lri.fr>.

### 8.2.4. ANR UNLOC project

L. Simon is the coordinator of the ANR BLANC project about "incomplete search for UNSAT", which is an important theoretical and practical question. This projects started in early 2009 and includes physicists from Orsay and researchers from Univ. Artois (CRIL, Lens), Univ. Amiens and Univ. Marseille.

### 8.2.5. FRAE SIRASAS project

The SIRASAS ("Stratégies Innovantes et Robustes pour l'Autonomie des Systèmes Aéronautiques et Spatiaux") project, funded by the FRAE ("Fondation de Recherche pour l'Aéronautique et l'Espace"), ended on September 2010. Work on space applications was led along two directions: on one side on the case study of a rendezvous mission between two spacecrafts in Mars orbit with the task of detecting and isolating nozzles failures; on the other side on a more prospective topic which is studying reconfiguration task and its coupling with diagnosis in an autonomy context.

### 8.2.6. ANR VERSO PIMI project

The objectives of PIMI (Personal Information Management Through Internet) ANR project (2010-2013) are the definition of a design environment and a deployment platform for Personal Information Management System (PIMS). The future PIMS must provide the end-user personal data access with services that are

relevant to his needs. Ontologies will be used to describe semantically the services and the user needs. Ontology matching techniques will be defined to compare the services and the user needs during the automatic service composition. In order to take mobility into account, the PIMS will be accessed both by mobile devices (smartphone) and Internet-connected Personal Computers. This project is a collaboration between Leo team, ForTesSE team from LRI, IRIT, IT (Institut Telecom), GENIGRAPH, Montimage, Région Midi-Pyrénées and CTIE from Luxembourg.

### **8.2.7. Participation to evaluation committees**

- N. Bidoit has served as VP of the evaluation committee of the ANR SIMI2 programs (Blanc, JCJC). She is President of the evaluation committee for the Gilles Kahn PhD Prize.
- P. Dague has participated to the evaluation committees of the thematic ANR program ARPEGE (“Systèmes Embarqués et Grandes Infrastructures”) and of the non thematic ANR programs “Blanc” and “Jeunes Chercheuses et Jeunes Chercheurs”.
- C. Reynaud has participated to the evaluation committee of the non thematic ANR program “Blanc” and of the thematic ANR program CONTINT. She has also participated to the evaluation work carried out by AERES of the LIRIS laboratory (Univ. Lyon) and of the research unit U936 “Modélisation conceptuelle des connaissances biomédicales” (Univ. Rennes 1).

## **8.3. European Initiatives**

### **8.3.1. Webdam ERC Grant**

The Webdam grant (S. Abiteboul) started in December 2008. The goal is to develop a formal model for Web data management. This model will open new horizons for the development of the Web in a well-principled way, enhancing its functionality, performance, and reliability. Specifically, the goal is to develop a universally accepted formal framework for describing complex and flexible interacting Web applications featuring notably data exchange, sharing, integration, querying and updating. We also propose to develop formal foundations that will enable peers to concurrently reason about global data management activities, cooperate in solving specific tasks and support services with desired quality of service.

The Webdam project is shared between the Dahu and Leo teams, both from INRIA Saclay.

### **8.3.2. HYCON 2**

Philippe Dague participates to the FP7 NoE (Network of Excellence) HYCON2 (“Highly-complex and networked control systems”) that started on September 2010 for a 4 years duration. He will be involved more particularly in the following work packages: Analysis of complex systems, System-wide coordination and control, Self-organizing systems and control, for the tasks of distributed diagnosis and distributed diagnosability analysis.

## **8.4. International Initiatives**

### **8.4.1. Cooperation within Europe**

Several members of the team (I. Manolescu, C. Reynaud, B. Safar, F. Goasdoué, N. Pernelle, F. Saïs) have proposed a Data Integration activity within the EIT ICT Labs “Knowledge and Innovation Community”. The activity will very likely be funded in 2011, and is coordinated by Leo. Other partners are: Alcatel-Lucent from France, the DFKI and Fraunhofer institutes from Germany, Nokia and Aalto University from Finland, and the Twente University from the Netherlands. Team members (I. Manolescu, N. Bidoit, F. Goasdoué and P. Chatalic) also participate in a “Cloud Data Management” activity within the EIT ICT Labs.

I. Manolescu is a member of the Advisory Board of the SeCo (Search Computing) ERC project, headed by S. Ceri from Politecnico di Milano.

I. Manolescu is a member of the core working group “Data and knowledge management” of the ongoing FET-Flagship proposal “The Social Computer - Internet-Scale Human problem solving”.

Close links exist with University of Madrid (A. Gomez-Perez), University of Manchester (I. Horrocks), University of Rome (M. Lenzerini), University of Pisa (G. Ghelli and C. Sartiani), CRP Henri Tudor in Luxembourg (C. Pruski), Ahtens University of Management and Business (V. Vassalos), CNR Italy (A. Bonifati).

Closed links exist in the domain of adaptive hypermedia with A. Cristea from the University of Warwick in U.K. Nadjat Zemirline, whose thesis focuses on this field, has made two visits of one month each this year in the IAS group coordinated by A. Cristea.

#### **8.4.2. Cooperation with Senegal**

Leo started a cooperation with the Gaston Berger University in December 2006 which leads to a PhD in co-tutelle with Paris-Sud university. The subject of the thesis is the integration of semi-structured data for information retrieval. The PhD student is Mouhamadou Thiam.

#### **8.4.3. Cooperation with North America**

The Gemo group, predecessor of Leo, had for many years an Associated Team with the data management group at the University of California at San Diego (V. Vianu, A. Deutch, Y. Papakonstantinou). Leo keeps collaborating with this group. Y. Papakonstantinou visited in July, and a joint tutorial was written by I. Manolescu, A. Deutsch from UCSD and V. Vassalos from the Athens University of Economics and Business (AUEB), Greece.

#### **8.4.4. Cooperation with Japan**

Katsumi Inoué (from NII) and two other researchers from his group have visited Leo in September. We are considering possible collaboration actions on the topics of consequence finding and hypothesis finding, SAT, distributed reasoning handling knowledge sharing.

## **9. Dissemination**

### **9.1. Thesis**

The following HDR (*Habilitation à Diriger les Recherches*) was defended in 2010:

- Laurent Simon, *Logique propositionnelle: un langage simple, des problèmes complexes*, defended on December 13, 2010.

The following PhD thesis were defended in 2010:

- Nada Abdallah, *Raisonnements standard et non-standard pour les systèmes décentralisés de gestion de données et de connaissances* [11], defended on July 13, 2010
- François-Elie Calvier, *Découverte de mappings dans un système pair-à-pair sémantique : application à SomeRDFS* [12], defended on September 17, 2010
- Yingmin Li, *Diagnosis of Large Software Systems Based on Colored Petri Nets*, defended on December 9, 2010
- Mouhamadou Thiam, *Annotation Sémantique de Documents Semi-structurés pour la Recherche d'Information*, defended on December 9, 2010.

The following members of the group participated to HDR and PhD committees in 2010:

- Nicole Bidoit : President of the HDR defense committee of Julia Kempe.
- Philippe Dague: member of the PhD committee of Sara Boutouhami-Nouioua (Univ. Paris-Nord), of Johan Oudinet (Univ. Paris-Sud) and of Manel Khlif (Univ. de Technologie de Compiègne); member of the HDR committee of Fatiha Zaïdi (Univ. Paris-Sud) and of Laurent Simon (Univ. Paris-Sud).; member of the Senior Researcher committee of Marc Porcheron.
- François Goasdoué: member of the PhD committee of Mamadou Nguer (Univ. Paris Sud-11) and of Pierre Colomb (Univ. Blaise Pascal, Clermont-Ferrand).
- Chantal Reynaud: member of the PhD committee of Ivan Bedini (Univ. Versailles) and Noureddine Mokhtari (Univ. Nice-Sophia Antipolis), reviewer of the PhD committee of Rémi Tournaire (Univ. Grenoble). Member of the HDR committee of Amel Bouzhegoub (Evry).
- Laurent Simon: member of the PhD Committee of Benoit Da Mota (Univ. Angers).

## 9.2. Participation in Conferences

Members of the project have been chairs of scientific events:

N. Bidoit

- Co-chair of the scientific committee of the "Masses de données" thematic school (Les Houches, Mai 2010).

I. Manolescu

- Program chair of the international EDBT 2010 conference (Lausanne, Switzerland, March 2010)
- Demo co-chair of the international ICDE 2010 conference (Long Beach, CA, USA)
- ACM SIGMOD 2010 repeatability and workability co-chair

C. Reynaud

- Chair of the OntoGeo workshop of the conference SAGEO 2010, Toulouse.

F. Sais

- Co-chair of EvalECD *Evaluation des méthodes d'Extraction des Connaissances dans les Données* workshop in collaboration with EGC 2010

Members of the project have participated in program committees:

N. Bidoit

- Workshop on Semantics in Data and Knowledge Bases (ICALP) 2010

P. Chatalic

- *Congrès francophone Reconnaissance des Formes et Intelligence Artificielle* (RFIA) 2010
- IEEE International Conference on Tools with Artificial Intelligence (ICTAI) 2010
- *Journées Francophones de Programmation par Contraintes* (JFPC) 2010

D. Colazzo

- International Conference on Extending Database Technologies (EDBT) 2010

P. Dague

- *Congrès francophone Reconnaissance des Formes et Intelligence Artificielle* (RFIA) 2010
- *Journées Nationales de l'IA Fondamentale* (IAF) 2010
- 24th International Workshop on Qualitative Reasoning (QR) 2010
- 21st International Workshop on Principles of Diagnosis (DX) 2010

## F. Goasdoué

- *Congrès francophone Reconnaissance des Formes et Intelligence Artificielle (RFIA)* 2010
- European Conference on Research and Advanced Technology for Digital Libraries (ECDL) 2010
- IEEE International Conference on Tools with Artificial Intelligence (ICTAI) 2010
- *Journées Bases de Données Avancées (BDA)* 2010

## F. Hamdi

- *OntoGeo Ontologies géographiques* workshop in collaboration with the conference SAGEO 2010

## I. Manolescu

- International PhD workshop in collaboration with the International Conference on Data Engineering (ICDE) 2010
- *Journées en Bases de Données Avancées (BDA)* 2010

## Y. Mrabet

- *OntoGeo Ontologies géographiques* workshop in collaboration with the conference SAGEO 2010

## N. Pernelle

- *EvalECD Evaluation des méthodes d'Extraction des Connaissances dans les Donnés* workshop, in collaboration with EGC 2010
- *QDC Qualité des Données et des Connaissances* workshop, in collaboration with EGC 2010
- *Web Social* workshop in collaboration with EGC 2010

## C. Reynaud

- 17th International Conference on Knowledge Engineering and Knowledge Management (EKAW) 2010
- International Conference on Web and Information Technologies (ICWIT) 2010
- *Congrès francophone Reconnaissance des Formes et Intelligence Artificielle (RFIA)*
- *Conférence Extraction et Gestion des Connaissances (EGC)* 2010
- *21èmes Journées Francophones d'Ingénierie des Connaissances (IC)* 2010
- Workshop Reuse and adaptation of ontologies and terminologies (EKAW) 2010
- Demonstration and Poster Session (ESWC) 2010
- 4th International Workshop on Ontology Dynamics (IWOD in conjunction with ESWC) 2010
- International Workshop on Resource Discovery (RED) 2010
- *EvalECD Evaluation des méthodes d'Extraction des Connaissances dans les Donnés* workshop in collaboration with EGC 2010
- *QDC Qualité des Données et des Connaissances* workshop in collaboration with EGC 2010

## F. Saïs

- *Colloque sur L'Optimisation et les Systèmes d'Informations (COSI)* 2010
- *Web Social* workshop in collaboration with EGC 2010
- EGC Maghreb Conference (EGC-M) 2010

## L. Simon

- *Congrès francophone Reconnaissance des Formes et Intelligence Artificielle (RFIA)* 2010
- International Conference on Theory and Applications of Satisfiability Testing (SAT'2010)



### 9.3. Invited Presentations

N. Bidoit had an invited presentation at Workshop on Semantics in Data and Knowledge Bases (ICALP) 2010.

D. Colazzo: Invited speaker at the "Masses de données" summer school (Les Houches, Mai 2010).

I. Manolescu: Invited speaker at the "Masses de données" summer school (Les Houches, Mai 2010).

C. Reynaud had an invited presentation titled "Interopérabilité sémantique" in Journée Interopérabilité des systèmes, GIS INTEROP Grande Région.

### 9.4. Animation of the scientific community

#### Editors

S. Abiteboul

- Member of the steering committee of Proceedings of the VLDB Endowment (PVLDB) Journal, a journal that started in 2008.

I. Manolescu

- Editor in chief of the ACM SIGMOD Record, quarterly publication of the ACM's SIGMOD chapter
- Associated editor of the ACM Transactions on Internet Technology
- Member of the program committee of Proceedings of the VLDB Endowment (PVLDB) Journal

C. Reynaud

- Member of the Editorial Advisory Board of the book "Handbook on Research on E-Business Standards and Protocols: Documents, Data and Advanced Web Technologies", IGI Global.
- Revue Information - Interaction - Intelligence (RI3).

L. Simon

- Member of the Editorial Board of JSAT (the Journal on Satisfiability, Boolean Modeling and Computation)
- Guest Editor of a Special Issue of JSAT on SAT 2006 Competitions and Evaluations.

### 9.5. Teaching

- N. Bidoit and P. Chatalic are Vice-Presidents of the computer science department of Université Paris Sud 11.
- N. Bidoit, M. Biennu, D. Colazzo, P. Chatalic, P. Dague, F. Goasdoué, I. Manolescu, N. Pernelle, C. Reynaud, B. Safar, L. Simon, and F. Saïs all teach in the *Information, Apprentissage, Connaissance* Master program of Université Paris Sud 11.
- F. Goasdoué, I. Manolescu, N. Pernelle, F. Saïs all teach in the *Informatique : Systèmes Intelligents* Master program of Université Paris Dauphine 9.
- F. Goasdoué teaches in the *Systèmes d'information* Master program of the university of Beirut, Lebanon.
- L. Simon teaches in the Master 2 program at ENS Lyon.

## 10. Bibliography

### Major publications by the team in recent years

- [1] S. ABITEBOUL, P. BOURHIS, B. MARINOIU, A. GALLAND. [Demo] AXART - Enabling Collaborative Work with AXML Artifacts, in "Very Large Database", Singapour Singapour, 2010, <http://hal.inria.fr/inria-00521726/en>.

- [2] J. CAMACHO-RODRÍGUEZ, A. KATSIFODIMOS, I. MANOLESCU, A. ROATIS. *LiquidXML: Adaptive XML Content Redistribution*, in "ACM International Conference on Information and Knowledge Management", Canada Toronto, Oct 2010, <http://hal.inria.fr/inria-00508388/en>.
- [3] D. COLAZZO, G. GUERRINI, M. MESITI, B. OLIBONI, E. WALLER. *Document and Schema XML Updates*, in "Advanced Applications and Structures in XML Processing: Label Streams, Semantics Utilization and Data Query Technologies", C. LI, T. W. LING (editors), IGI Global, Nov 2010, <http://hal.inria.fr/inria-00537168/en>.
- [4] F. GOASDOUÉ, K. KARANASOS, J. LEBLAY, I. MANOLESCU. *RDFViewS: A Storage Tuning Wizard for RDF Applications*, in "ACM International Conference on Information and Knowledge Management", Canada Toronto, Oct 2010, <http://hal.inria.fr/inria-00509125/en>.
- [5] F. HAMDI, C. REYNAUD, B. SAFAR. *Pattern-based Mapping Refinement*, in "International Conference on Knowledge Engineering and Knowledge Management (EKAW)", Lisbon Portugal, 10 2010.
- [6] F. HAMDI, B. SAFAR, C. REYNAUD, H. ZARGAYOUNA. *Alignment-based Partitioning of Large-scale Ontologies*, in "Advances in Knowledge Discovery And Management", F. GUILLET, G. RITSCHARD, D. ZIGHED, H. BRIAND (editors), Springer, 2010, vol. 292, p. 251-269 [DOI : 10.1007/978-3-642-00580-0].
- [7] A. MADALINSKI, F. NOUIOUA, P. DAGUE. *Diagnosability verication with Petri net unfoldings*, in "International Journal of Knowledge-Based and Intelligent Engineering Systems, KES", 2010, vol. 14, n<sup>o</sup> 2, p. 49-55, <http://hal.inria.fr/inria-00540650/en>.
- [8] Y. MRABET, N. BENNACER, N. PERNELLE, M. THIAM. *Supporting Semantic Search on Heterogeneous Semi-structured Documents*, in "22nd International Conference on Advanced Information Systems Engineering, CAISE 2010", Tunisie Hammamet, LNCS, Springer-Verlag, Jun 2010, p. 224-229, Taux d'acceptation (13%), <http://hal.inria.fr/hal-00502286/en>.
- [9] F. SAÏS, R. THOMOPOULOS, S. DESTERCKE. *Ontology-Driven Possibilistic Reference Fusion*, in "OTM/ODBASE 2010 - the 9th International Conference on Ontologies, DataBases, and Applications of Semantics", Grèce Crete, R. MEERSMAN (editor), Springer Berlin / Heidelberg, Oct 2010, p. 1081-1098, <http://hal.inria.fr/inria-00519677/en>.
- [10] L. YE, P. DAGUE. *Diagnosability Analysis of Discrete Event Systems with Autonomous Components*, in "European Conference on Artificial Intelligence ECAI", Portugal Lisbonne, Aug 2010, <http://hal.inria.fr/inria-00540649/en>.

## Publications of the year

### Doctoral Dissertations and Habilitation Theses

- [11] N. ABDALLAH. *Raisonnements standard et non-standard pour les systèmes décentralisés de gestion de données et de connaissances*, Université Paris Sud - Paris XI, Jul 2010, <http://hal.inria.fr/tel-00536926/en>.
- [12] F.-É. CALVIER. *Découverte de mappings dans un système pair-à-pair sémantique : application à SomeRDFS*, Université Paris Sud - Paris XI, Sep 2010, <http://hal.inria.fr/tel-00530075/en>.

### Articles in International Peer-Reviewed Journal

- [13] A. MADALINSKI, F. NOUIOUA, P. DAGUE. *Diagnosability verification with Petri net unfoldings*, in "International Journal of Knowledge-Based and Intelligent Engineering Systems, KES", 2010, vol. 14, n<sup>o</sup> 2, p. 49-55, <http://hal.inria.fr/inria-00540650/en>.

### International Peer-Reviewed Conference/Proceedings

- [14] S. ABITEBOUL, P. BOURHIS, B. MARINOIU, A. GALLAND. *[Demo] AXART - Enabling Collaborative Work with AXML Artifacts*, in "Very Large Database", Singapour Singapour, 2010, <http://hal.inria.fr/inria-00521726/en>.
- [15] N. BIDOIT, D. COLAZZO, F. ULLIANA. *Detecting XML query-update independence*, in "26emes Journees Bases de Donnees Avancees", France Toulouse, Oct 2010, <http://hal.inria.fr/inria-00542526/en>.
- [16] J. CAMACHO-RODRÍGUEZ, A. KATSIFODIMOS, I. MANOLESCU, A. ROATIS. *LiquidXML: Adaptive XML Content Redistribution*, in "ACM International Conference on Information and Knowledge Management", Canada Toronto, Oct 2010, <http://hal.inria.fr/inria-00508388/en>.
- [17] P. FIANI, M. BATTEUX, P. DAGUE, N. RAPIN. *Fuel Cell System Improvement for Model-Based Diagnosis Analysis*, in "Vehicle Power and Propulsion Conference IEEE VPPC", France Lille, Sep 2010, <http://hal.inria.fr/inria-00540836/en>.
- [18] A. GALLAND, S. ABITEBOUL, A. MARIAN, P. SENELLART. *Corroborating Information from Disagreeing Views*, in "International Conference on Web Search and Data Mining (WSDM)", États-Unis New York City, Feb 2010, <http://hal.inria.fr/inria-00429546/en>.
- [19] F. GOASDOUÉ, K. KARANASOS, J. LEBLAY, I. MANOLESCU. *RDFViewS: A Storage Tuning Wizard for RDF Applications*, in "ACM International Conference on Information and Knowledge Management", Canada Toronto, Oct 2010, <http://hal.inria.fr/inria-00509125/en>.
- [20] N. GUELFY, C. PRUSKI, C. REYNAUD. *Experimental Assessment of the TARGET Adaptive Ontology-based Web Search Framework*, in "Workshop Web2Touch - Living experience through Web, in conjunction with 10th International Conference on New technologies of Distributed Systems NOTERE'2010", Tunisie Tozeur, Jun 2010, <http://hal.inria.fr/inria-00491153/en>.
- [21] F. HAMDI, C. REYNAUD, B. SAFAR. *A framework for mapping refinement specification*, in "International Symposium on Matching and Meaning", Royaume-Uni Leicester, Jun 2010, <http://hal.inria.fr/inria-00491121/en>.
- [22] F. HAMDI, C. REYNAUD, B. SAFAR. *Pattern-based Mapping Refinement*, in "International Conference on Knowledge Engineering and Knowledge Management (EKAW)", Portugal Lisbon, Jun 2010, <http://hal.inria.fr/inria-00491131/en>.
- [23] F. HAMDI, B. SAFAR, C. REYNAUD, N. NIRLAULA. *TaxoMap alignment and refinement modules: Results for OAEI 2010*, in "The Fifth International Workshop on Ontology Matching", Chine Shanghai, Nov 2010, <http://hal.inria.fr/inria-00538806/en>.
- [24] K. KARANASOS, S. ZOUPANOS. *Viewing a World of Annotations through AnnoVIP*, in "International Conference on Data Engineering (ICDE)", États-Unis Long Beach, California, 2010, <http://hal.inria.fr/inria-00541114/en>.

- [25] H. LAUW, R. SCHENKEL, F. SUCHANEK, M. THEOBALD, G. WEIKUM. *Harvesting Knowledge from Web Data and Text*, in "CIKM", Canada Toronto, SIG, 2010, <http://hal.inria.fr/inria-00534905/en>.
- [26] T. MELLITI, P. DAGUE. *Generalizing diagnosability definition and checking for open systems: a Game structure approach*, in "21st International Workshop on Principles of Diagnosis DX'10", États-Unis Portland, OR, Oct 2010, <http://hal.inria.fr/inria-00540849/en>.
- [27] Y. MRABET, N. BENNACER, N. PERNELLE, M. THIAM. *Supporting Semantic Search on Heterogeneous Semi-structured Documents*, in "22nd International Conference on Advanced Information Systems Engineering, CAISE 2010", Tunisie Hammamet, LNCS, Springer-Verlag, Jun 2010, p. 224-229, Taux d'acceptation (13%), <http://hal.inria.fr/hal-00502286/en>.
- [28] O. NASRI, H. SHRAÏM, P. DAGUE, O. HÉRON, M. CARTRON. *Model-based Decentralized Embedded Diagnosis inside Vehicles: Application to Smart Distance Keeping Function*, in "Conference on Control and Fault-Tolerant Systems Systol'10", France Nice, Oct 2010, <http://hal.inria.fr/inria-00540829/en>.
- [29] F. SAÏS, N. NIRLAULA, N. PERNELLE, M.-C. ROUSSET. *LN2R – a knowledge based reference reconciliation system: OAEI 2010 Results*, in "The Fifth International Workshop on Ontology Matching", Chine Shanghai, Nov 2010, <http://hal.inria.fr/inria-00527287/en>.
- [30] F. SAÏS, R. THOMOPOULOS, S. DESTERCKE. *Ontology-Driven Possibilistic Reference Fusion*, in "OTM/ODBASE 2010 - the 9th International Conference on Ontologies, DataBases, and Applications of Semantics", Grèce Crete, R. MEERSMAN (editor), Springer Berlin / Heidelberg, Oct 2010, p. 1081–1098, <http://hal.inria.fr/inria-00519677/en>.
- [31] H. SHRAÏM, O. NASRI, P. DAGUE, O. HÉRON, M. CARTRON. *Smart distance keeping: modeling and perspectives for embedded diagnosis*, in "1rst International Conference on Intelligent Systems, Modelling and Simulation ISMS'10", Royaume-Uni Liverpool, Jan 2010, <http://hal.inria.fr/inria-00540819/en>.
- [32] L. SIMON, G. AUDEMARD, K. GEORGE. *A Restriction of Extended Resolution for Clause Learning SAT Solvers*, in "AAAI", Atlanta États-Unis, 07 2010, <http://hal.inria.fr/inria-00543638/en/>.
- [33] Y. YAN, L. YE, P. DAGUE. *Diagnosability for Patterns in Distributed Discrete Event Systems*, in "21st International Workshop on Principles of Diagnosis DX'10", États-Unis Portland, OR, Oct 2010, <http://hal.inria.fr/inria-00540844/en>.
- [34] L. YE, P. DAGUE. *An Optimized Algorithm for Diagnosability of Component-based Systems*, in "10th International Conference on Discrete Event Systems WODES'10", Allemagne Berlin, Aug 2010, <http://hal.inria.fr/inria-00540764/en>.
- [35] L. YE, P. DAGUE. *Diagnosability Analysis of Discrete Event Systems with Autonomous Components*, in "European Conference on Artificial Intelligence ECAI", Portugal Lisbonne, Aug 2010, <http://hal.inria.fr/inria-00540649/en>.
- [36] N. ZEMIRLINE, Y. BOURDA, C. REYNAUD. *A Pattern-based Framework for expressing Adaptation Strategies in Adaptive Systems*, in "The 10th International Conference on Intelligent Systems Design and Applications (ISDA'10)", Égypte Caire, IEEE, 2010, p. ??-??, <http://hal.inria.fr/hal-00518385/en>.

### National Peer-Reviewed Conference/Proceedings

- [37] N. ABDALLAH, F. GOASDOUÉ. *Gestion décentralisée de données en DL-LITE*, in "Reconnaissance des Formes et Intelligence Artificielle", France Caen, Jan 2010, <http://hal.inria.fr/inria-00491509/en>.
- [38] J. CAMACHO-RODRÍGUEZ, A. KATSIFODIMOS, I. MANOLESCU, A. ROATIS. *LiquidXML: Adaptive XML Content Redistribution*, in "Bases de Données Avancées (BDA2010)", France Toulouse, Institut de Recherche en Informatique de Toulouse, Oct 2010, <http://hal.inria.fr/inria-00542525/en>.
- [39] S. GAHBICHE, N. PERNELLE, F. SAÏS. *Explication de décisions de réconciliation de références : approche fondée sur les réseaux de Petri colorés*, in "EGC", Tunisie Hammamet, 2010, p. 327-338, <http://hal.inria.fr/inria-00491197/en>.
- [40] F. HAMDI, C. REYNAUD, B. SAFAR. *L'approche TaxoMap Framework et son application au raffinement de mappings*, in "Reconnaissance des Formes et Intelligence Artificielle", France Caen, Jan 2010, <http://hal.inria.fr/inria-00432994/en>.
- [41] F. HAMDI, C. REYNAUD, B. SAFAR. *TaxoMap Framework appliqué à l'alignement d'ontologies géographiques dans le projet GéOnto*, in "Atelier OntoGeo "Ontologies Géographiques" associé à la conférence SAGEO'10", France Toulouse, Nov 2010, p. 51-53, <http://ontogeo2010.lri.fr/>, <http://hal.inria.fr/inria-00538804/en>.
- [42] Y. MRABET, N. BENNACER, N. PERNELLE, M. THIAM. *Une approche pour la recherche sémantique de l'information dans les documents semi-structurés hétérogènes*, in "CONFérence en Recherche d'Informations et Applications - CORIA 2010.", Tunisie Sousse, Centre de Publication Universitaire 2010, Mar 2010, p. 195-210, <http://hal.inria.fr/hal-00502292/en>.
- [43] N. ZEMIRLINE, Y. BOURDA, C. REYNAUD. *Réutilisation de patrons d'adaptation – Application aux systèmes hypermédia adaptatifs*, in "Ingénierie des Connaissances. IC 2010", France Nimes, S. DESPRES (éditeur), Ecole des Mines d'Alès, 2010, p. 107-118, <http://hal.inria.fr/hal-00488228/en>.

### Workshops without Proceedings

- [44] V. BENZAKEN, J.-D. FEKETE, W. KHEMIRI, I. MANOLESCU. *EdiFlow: data-intensive interactive workflows for visual analytics*, in "Journées Bases de Données Avancées", France Toulouse, Oct 2010, <http://hal.inria.fr/inria-00539718/en>.
- [45] J. CAMACHO-RODRÍGUEZ, K. KARANASOS, I. MANOLESCU, A. TILEA, S. ZOUPANOS. *Viewing a World of Annotations through AnnoVIP*, in "Bases de Données Avancées", France Toulouse, 2010, <http://hal.inria.fr/inria-00540632/en>.
- [46] D. COLAZZO, C. SARTIANI. *Precision and Complexity of XQuery Type Inference*, in "12th Italian Conference on Theoretical Computer Science (ICTCS)", Italie Camerino, Sep 2010, <http://hal.inria.fr/inria-00537161/en>.
- [47] F. GOASDOUÉ, K. KARANASOS, J. LEBLAY, I. MANOLESCU. *Materialized View-Based Processing of RDF Queries*, in "Bases de Données Avancées", France Toulouse, Oct 2010, <http://hal.inria.fr/inria-00536995/en>.
- [48] F. GOASDOUÉ, M.-C. ROUSSET. *Modules sémantiques robustes pour une réutilisation saine en DL-lite*, in "Bases de données avancées", France Toulouse, Sep 2010, <http://hal.inria.fr/inria-00533086/en>.

- [49] L. SIMON, G. AUDEMARD, K. GEORGE. *Une restriction de la résolution étendue pour les démonstrateurs SAT modernes*, in "JFPC", Orléans France, 06 2010, <http://hal.inria.fr/inria-00543640/en/>.

### **Scientific Books (or Scientific Book chapters)**

- [50] S. ABITEBOUL, I. MANOLESCU, P. RIGAUX, M.-C. ROUSSET, P. SENELLART. *Web Data Management and Distribution*, Cambridge University Press, Dec 2010, <http://hal.inria.fr/inria-00542796/en>.
- [51] D. COLAZZO, G. GUERRINI, M. MESITI, B. OLIBONI, E. WALLER. *Document and Schema XML Updates*, in "Advanced Applications and Structures in XML Processing: Label Streams, Semantics Utilization and Data Query Technologies", C. LI, T. W. LING (editors), IGI Global, Nov 2010, <http://hal.inria.fr/inria-00537168/en>.
- [52] S. GAHBICHE, N. PERNELLE, F. SAÏS. *Explaining Reference Reconciliation Decisions: a Coloured Petri Nets based Approach*, in "Advances in Knowledge Discovery and Management", F. GUILLET, G. RITSCHARD, D. ZIGHED, H. BRIAND (editors), Springer Berlin / Heidelberg, Dec 2010, <http://hal.inria.fr/inria-00519667/en>.