Thèse d'habilitation

Université Jean Monnet

École doctorale 488 Sciences, Ingénierie et Santé (SIS)

Interopérabilité pour le Web sémantique : une approche par médiation faiblement couplée

Soutenue publiquement par :

Antoine Zimmermann
MINES Saint-Étienne

Membres du jury :

Olivier Boissier	Professeur, MINES Saint-Étienne	Tuteur de recherche
Pierre-Antoine Champin	Maître de conférence HDR, Université Claude-Bernard	Examinateur
Jérôme Euzenat	Directeur de recherche, Inria	Examinateur
Fabien Gandon	Directeur de recherche, Inria	Rapporteur
Andreas Harth	Professeur, Friedrich-Alexander-Universität Erlangen-Nürnberg	Rapporteur
Axel Polleres	Professeur, Wirtschaftsuniversität Wien	Examinateur
Marie-Christine Rousset	Professeur, Université Grenoble Alpes	Examinatrice
Hala Skaf-Molli	Matre de conférence HDR, Université de Nantes	Rapportrice

le 23 février 2021

Habilitation Thesis

Université Jean Monnet

École doctorale 488 Sciences, Ingénierie et Santé (SIS)

Interoperability for the Semantic Web: A Loosely Coupled Mediation Approach

Presented and defended by:

Antoine Zimmermann
MINES Saint-Étienne

Jury members:

Olivier Boissier	Professor, MINES Saint-Étienne	Research tutor
Pierre-Antoine Champin	Associate professor, Université Claude-Bernard	Examiner
Jérôme Euzenat	Research director, Inria	Examiner
Fabien Gandon	Research director, Inria	Reviewer
Andreas Harth	Professor, Friedrich-Alexander-Universität Erlangen-Nürnberg	Reviewer
Axel Polleres	Professor, Wirtschaftsuniversität Wien	Examiner
Marie-Christine Rousset	Professor, Université Grenoble Alpes	Examiner
Hala Skaf-Molli	Associate professor, Université de Nantes	Reviewer

on the 23rd of February 2021

Interoperability for the Semantic Web: A Loosely Coupled Mediation Approach

Antoine ZIMMERMANN

March 2, 2021

Abstract

Dans cette thèse, je défends deux opinions : une opinion générale, à savoir que je mérite maintenant le diplôme d'habilitation, qui est le but principal de la rédaction de ce document ; et, dans le cadre de cette argumentation, une opinion spécifique, qui est la thèse scientifique que la plupart de mes travaux ont soutenue, et qui donne un sens au titre de la thèse. La thèse préconise une approche particulière pour permettre l'interopérabilité sur le Web (sémantique). Plus précisément, pour réaliser la vision du web sémantique, il faut rendre l'information sur le web plus facilement utilisable, compréhensible et traitable par les agents, qu'il s'agisse de personnes, d'applications logicielles ou de dispositifs matériels. Une caractéristique clé du Web est son ouverture qui permet à chacun de publier des données, des informations et des connaissances sur n'importe quoi. Bien qu'il s'agisse là d'une force considérable du Web, elle entraîne une grande hétérogénéité. Par conséquent, l'exploitation des ressources du Web nécessite souvent des processus *ad hoc* qui sont adaptés à la source d'information. Il est donc difficile d'exploiter le Web dans son ensemble, plutôt qu'une seule source à la fois. Mon travail de recherche a consisté à atténuer cette difficulté en définissant des moyens de combiner des sources d'information hétérogènes. Je l'ai surtout fait avec une approche de médiation : si l'on veut tirer profit d'une diversité de sources d'information qui ne peut pas être traitée facilement de manière uniforme, on peut utiliser une ressource ou un mécanisme séparé qui fournit une interface unique, tout en offrant des connexions, des liens, des correspondances, des métadonnées, *etc.* explicites entre les sources d'information ou au sommet de celles-ci. Il en résulte une plus grande interopérabilité des systèmes. J'illustre cette idée à différents niveaux d'abstraction et pour différents types d'interopérabilité, en décrivant et en me référant à mes travaux antérieurs : au niveau des données pour l'interopérabilité syntaxique ; au niveau des connaissances pour l'interopérabilité sémantique ; et au niveau de la décision, vers l'interopérabilité des processus. En plus de cet aperçu de mes contributions scientifiques, je termine cette thèse par un résumé de mes activités de travail jusqu'à présent, démontrant que j'ai mené des recherches et des études indépendantes, ainsi que supervisé des étudiants.

Résumé

Ma thèse d'habilitation préconise une approche particulière permettant l'interopérabilité sur le Web (sémantique). Plus précisément, pour réaliser la vision du web sémantique, il faut rendre l'information sur le web plus facilement utilisable, compréhensible et traitable par les agents, qu'il s'agisse de personnes, d'applications logicielles ou de dispositifs matériels. Une caractéristique clé du Web est son ouverture qui permet à chacun de publier des données, des informations et des connaissances sur n'importe quel sujet. Bien qu'il s'agisse là d'une force considérable du Web, elle entraîne une grande hétérogénéité. Par conséquent, l'exploitation des ressources du Web nécessite souvent des processus ad hoc qui sont adaptés à la source d'information. Il est donc difficile d'exploiter le Web dans son ensemble, plutôt qu'une seule source à la fois. Mon travail de recherche a consisté à atténuer cette difficulté en définissant des moyens de combiner des sources d'information hétérogènes. Je l'ai surtout fait avec une approche de médiation : si l'on veut tirer profit d'une diversité de sources d'information qui ne peut être traitée facilement de manière uniforme, on peut utiliser une ressource ou un mécanisme séparé qui fournit une interface unique, tout en offrant des connexions, des liens, des correspondances, des métadonnées explicites entre les sources d'information ou en complément de celles-ci. Il en résulte une plus grande interopérabilité des systèmes. J'illustre cette idée à différents niveaux d'abstraction et pour différents types d'interopérabilité, en décrivant et en me référant à mes travaux antérieurs : au niveau des données pour l'interopérabilité syntaxique ; au niveau des connaissances pour l'interopérabilité sémantique ; et au niveau de la décision, vers l'interopérabilité des processus. En plus de cet aperçu de mes contributions scientifiques, je fournis un résumé de mes activités de travail jusqu'à aujourd'hui, démontrant que j'ai mené et encadré des recherches indépendantes dignes de ce diplôme.

Acknowledgement

The work presented in this thesis is the result of many collaborations with people that deserve being acknowledged. For the most part, they are mentioned in Appendix E as co-authors of the paper we published together. However, I would like to specifically thank some people that deserve more credit than a mere mention in an author list.

First, I would like to thank the esteemed scientists that evaluated my work in order to deliver the diploma of the *habilitation à diriger des recherches*: in alphabetical order, Fabien Gandon, Andreas Harth, and Hala Skaf-Molli, the rapporteurs; Pierre-Antoine Champin, Jérôme Euzenat, Axel Polleres, and Marie-Christine Rousset, the examiners; and Olivier Boissier, my tutor for this habilitation. They not only accepted the task, but also made the defence an enjoyable moment of deep, and challenging, scientific discussions. Among them, special thanks go to:

- Marie-Chritine Rousset for accepting the responsibility to be the president of the jury... again! She had been the president of the jury during the defence of my PhD thesis, more than twelve years ago;

- Jérôme Euzenat for taking me on the doctoral journey in the framework of the Knowledge Web project, which gave me access to an excellent network of international collaborators, more than fifteen years ago. He managed to give me just enough guidance to lead me to the successful defence and yet leave me with almost unlimited freedom to investigate the topic in the way I wanted;

- Axel Polleres, for hiring me as a postdoc in the legendary DERI, Galway, where I spent twenty wonderful months. Axel became more than a boss but also a friend. He gave me responsibilities, and I felt trusted.

- Last but not least among the members of the jury, Olivier has been an exceptional colleague all along since 2011 when I joined his team, an excellent advisor for the completion of this thesis and the preparation of my defence, and, too, a dear and trusting, as well as trustworthy, friend. There seems to be no limit to how much and how often he is willing to help his colleagues and students, even in time of extreme business!

I am grateful to the PhD students I supervised, but especially to:

- Andrei Ciortea, who opened up a whole avenue of research for me and others – Andrei, you are my best student so far ;-)

- Noorani Bakerally, who never ceased to surprise me in unexpected way, and who hold the record for shortest time to complete a PhD among my students;

- José Miguel Giménez-García, whom I wish will overcome the difficulties and eventually defend his thesis.

Among the other people that I had the chance to work with, I would like to thank some more people, in chronological order:

- Sébastien Laborie, the PhD student (at the time in Grenoble) who worked on a very different topic from mine but with whom I collaborated just because we wanted to;
- Chan Le Duc, who really helped me shape an essential part of the my PhD thesis and who contacted me recently to work on sweet abstract nonsense;
- Aidan Hogan, a researcher for whom I have the deepest respect, with his impressive brightness, extraordinary writing skills, amazing intellectual honesty and kindness;
- Maxime Lefrançois, with whom I share so many views with respect to the Semantic Web and together with whom we are bringing our team forth nicely among the community of Semantic Web research.

More thanks go to my parents and brothers, and beyond everyone else to Cécile and Rémi.

Contents

Introduction

Modern information systems are nowadays typically accessible remotely by commodity hardware and software. In most cases now, a simple Web browser accessing an authentication page is sufficient to grant access to the inside of an organisation's system. All interactions are conveniently performed by HTTP calls via the Web navigator. Many people can connect at the same time, using software and hardware from many different vendors who did not have to agree beforehand on the supporting functionalities. The interoperability between the individual clients and the organisation's servers happens by design of the Web standards and architectures.

Sometimes, the information system itself has to connect to remote resources provided by third parties in order to automatise some tasks (for instance, update itself with the latest bug fixes). However, while human interaction with Web resources has become extremely portable, where people can use almost any type of computerised devices – laptop, phone, tablet, car, fridge, etc. – as an interface between them and remote resources of any kind – text, pictures, sounds, films, games, apps –, the level of interoperability required for automating the interactions – that is, letting an autonomous device act upon the remote resources without human intervention – is not yet realised. My work in the past fifteen years has been towards empowering systems with more interoperability for more automation capacities. This dissertation is an account of my contributions towards this end, as well as a thesis defending **loosely coupled mediation** as an approach to achieve this goal.

The type of problems we want to address assumes that we have an information need that requires combining multiple information resources. This is greatly facilitated by the uniformity of the network protocol (the Internet) and the trend towards using the Web as the application layer of choice. This allows interoperable access to information resources, as one just needs a URL to get a resource. Further, resources can provide more URLs to navigate from resources to resources via hypermedia.

But interoperability does not stop at the bottom of the application layer. To ensure strong (semantic) interoperability, the initial plan for the Semantic Web was to have a uniform data layer with XML as the surface syntax for all data exchanges, RDF as the data model for resource descriptions and interlinking, a standard ontology language (that eventually would be named OWL) to allow description of the background knowledge to more easily understand RDF graphs and make inferences, etc. This vision was schematised by the so-called "Semantic Web layer cake" of Figure 1, which was revised many times.

Unfortunately, while pieces of the Semantic Web stack became standardised little by little, they did not all become *de facto* standards. All layers upon the Web have a diversity of technologies that coexist. It may be surprising that, while the Web was quickly adopted as a uniform application layer, the rest of the technology stack that should have allowed semantic interoperability did not spread so well. In my

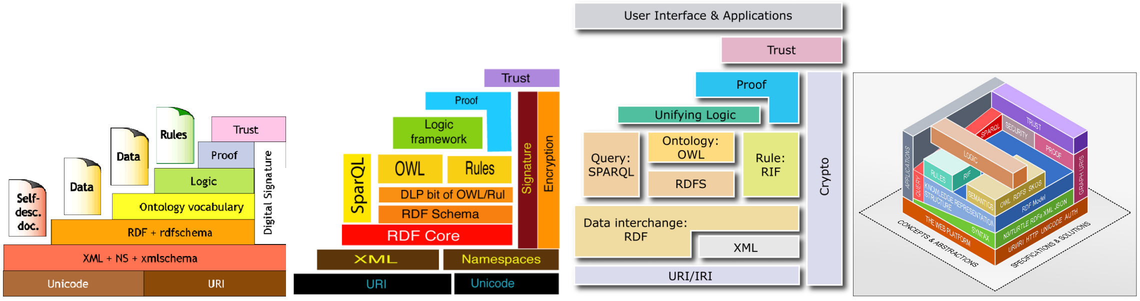


Figure 1: The Semantic Web layer cake and its evolution. From left to right: from [Berners-Lee et al., 2001] in 2001; 2004 version from W3C; 2007 version from Wikipedia; 2009 version posted on Flickr

opinion, one important reason is that the Web filled a gap that was not well filled with existing technologies. It made access to online resources so easy, as opposed to, for instance, FTP where the IP addresses must be known, the directory structure must be explored in order to find a relevant resource. Then nothing in the FTP ecosystem allows one to navigate easily from one server to another. On the contrary, when XML, RDF and so on were designed as layers for the Semantic Web on top of core Web standards, information systems already had existing ways of consuming and exporting data. The dominant data model in use was mostly relational, which is not easily transformed or migrated to RDF. RDF requires special modelling that database specialists are not familiar with. Ontology engineering is a complex and not so widespread skill.

As a result, we currently have much heterogeneity in the layers on top of the basic Web technologies that hinders strong interoperability. However, my claim is that instead of fighting for uniformity at all layers of operations, we can **embrace the diversity and still allow interoperability by way of a loosely coupled mediation approach**. The fact that it is loosely coupled is important because it is what distinguishes the approach from how the Web itself serves as a mediation layer between vastly heterogeneous information systems and web clients that themselves run on very heterogeneous software and hardware architectures. Web servers mediating between local infrastructures and remote web clients are implemented *at the information source*, but a mediation layer for data and knowledge and decision could be provided by third parties at very different places. This vision is illustrated in Figure 2, where classical mediation is compared to loosely coupled mediation.

In classical mediation, wrappers are attached to sources, and a mediator provides the interface between the user application and the sources through the wrappers. In a loosely coupled mediation approach, mediators, wrappers, and information sources are resources on the Web. Mediators may or may not be attached to wrappers. Instead, clients may decide to dynamically connect a mediator (such as M2) to difference wrappers, and similarly decide to connect a wrapper (such as W2) to different information sources. This selection by the client is visualised as the dashed ellipses in Figure 2.

The majority of my work instantiates this vision in different ways. In many cases, the mediator and wrappers in my contributions are not described as software components: the mediator is a model or language that the wrappers instantiate, while a generic engine takes care of the operational mediation. In order to provide a coherent

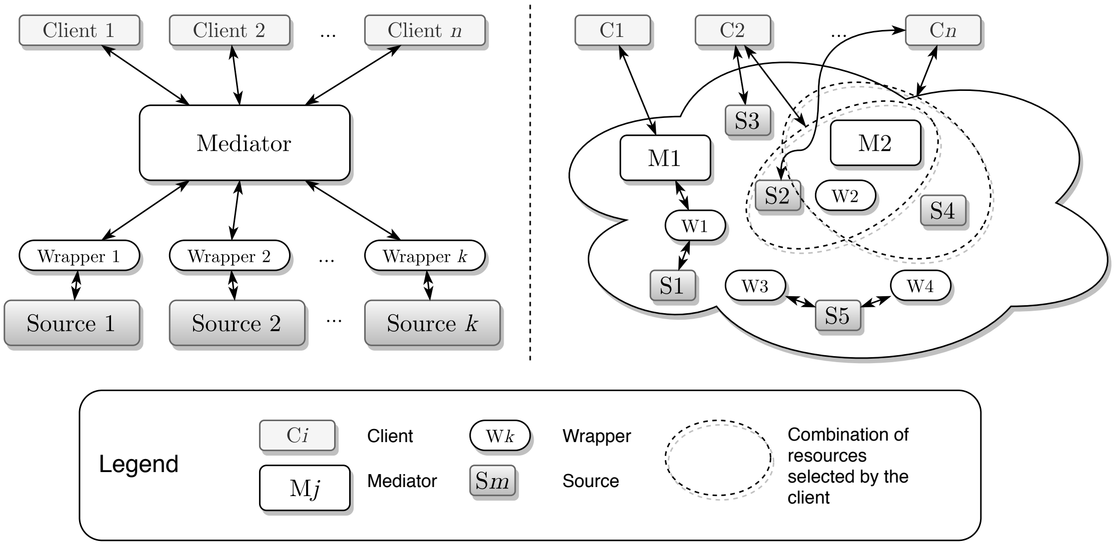


Figure 2: On the left, classical mediation. On the right, loosely coupled mediation on top of the Web.

view of my contributions according to this vision, I define layers of interoperability as in Figure 3 that start at the bottom with the network layer which allows communication between the participants of distributed or decentralised systems. I assume that all communications are mediated through the Web and therefore I do not address this layer. On top of it, the data layer is concerned with providing a uniform data model and uniform data access. Currently, no such model nor access method prevails, in spite of the existence of a standard data model for the Web, namely RDF. My work at this layer consists of two approaches: provide a flexible mediator that translates all types of data into RDF, or provide a uniform access mechanism to heterogeneous data, via a uniform RESTful interface or a single query language. As a result of this work, we can assume that the upper layer can be entirely defined on top of the RDF data model. Even with a common data model, it is possible to have heterogeneous knowledge models, depending on viewpoints, objectives, contexts. In my work, mediation at this layer is enabled by adding meta-information about knowledge models, either in the form of explicit correspondences between terminologies, or as representation of the context of knowledge. Leveraging this meta-information allows reasoning across contexts and viewpoints. Finally, at the top layer, processes that utilise one common knowledge model may still diverge. Specifically, two agents that are given exactly the same knowledge may take different actions, make different decisions, thus fail to coordinate. My work at this layer consists in facilitating the cooperation process through a Web platform, enabled by explicit knowledge.

In the next chapters of this dissertation, I will show how my scientific contributions fit in this general vision. My contributions are structured according to the three upper layers of Figure 3. Each layer is associated with a part in this thesis, ordered from bottom to top layers, divided into chapters that each correspond to a contribution made in the context of a collaborative project or a PhD student supervision. I outline the structure of this document by highlighting the collaborations and supervision, as well as the main challenge addressed.

- Part I concentrates on the data layer of Figure 3, with three approaches to

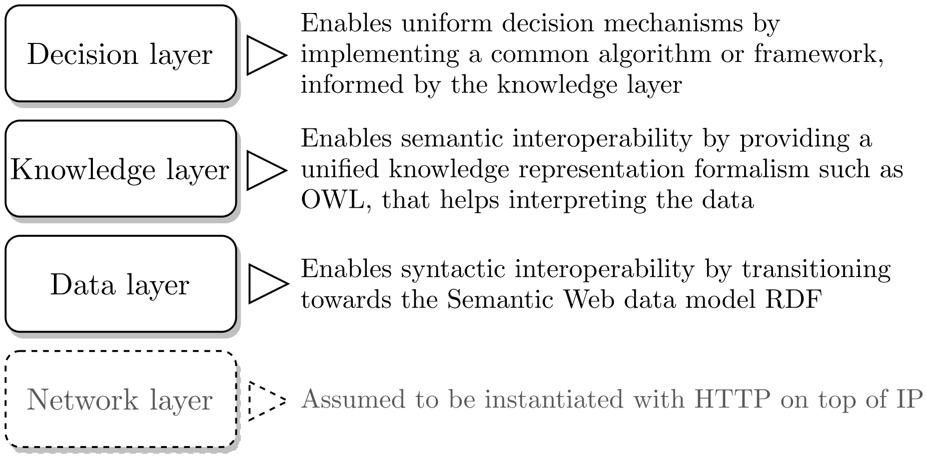


Figure 3: A general layered model for interoperability

reconcile data formats and data interoperability:

- In Chapter 1, I present a language for data transformation – SPARQL-Generate – that can be used to give a uniform view on heterogeneous data in the form of RDF graphs. The challenge consists in providing a flexible mechanism for data transformation from any data format to the Semantic Web data model RDF. The transformation is the mediating mechanism between heterogeneous data sources and the user of RDF data. This was made possible by my coordination of ANR project OpenSensingCity and my involvement in ITEA project SEAS. Relevant publications: [Lefrançois et al., 2016b, Lefrançois et al., 2017c, Lefrançois et al., 2017b].
- In Chapter 2, I present a federated query engine – PolyWeb – that relies on data transformations and metadata to query multiple data sources using their native data model query language. The challenge consists in defining how an original query can be split and translated to different data base engines that use different languages. The federation engine mediates between the user with his single query language and the heterogeneous and distributed data base systems. This is the result of a collaboration with colleagues of Insight Galway research center. Relevant publications: [Khan et al., 2017b, Khan et al., 2019].
- In Chapter 3, I present a workflow, language and engine for current RDF data in a more accessible and navigable way, following the Linked Data Platform 1.0 standard. The challenge consists in facilitating the deployment of data to platforms that are easy to exploit by open data developers. The tools of the workflow mediates between the raw data and their exposition as structured linked data on a Web platform. This was the result of my supervision of PhD student Noorany Bakerally made possible by getting project OpenSensingCity funded under my coordination. Relevant publications: [Bakerally et al., 2016, Bakerally and Zimmermann, 2017, Bakerally et al., 2018a, Bakerally et al., 2018b] and Noorani's thesis [Bakerally, 2018].

• Part II concentrates on the knowledge layer of Figure 3, with a particular focus on context representation and reasoning:

- In Chapter 4, I present a summary of my PhD thesis contributions, followed by a more recent continuation of it that consists in formalising multi-level networks of aligned ontologies. The challenge is to be able to treat networks of aligned ontologies as a single ontology that may be aligned with other networks of aligned ontologies, thus creating levels of alignments. This way, a set of heterogeneous ontologies can be used a unified knowledge model. The alignments help mediate between the user of networked knowledge and individual ontologies that compose it. This was done as a result of the supervision of Sihem Klai from University of Annaba. Relevant publications: [Klai et al., 2016a, Klai et al., 2016b] and Sihem's thesis [Klai, 2016].
- In Chapter 5, I present a formalism – Annotated RDFS – for making the context of a logical statement explicit as a value from an algebraic structure, with associated semantics, reasoning method and query language. The challenge is to have a uniform treatment of many types of contextual annotation (temporal, provenance, fuzzy, etc.). This was made possible by a collaboration started within the context of COST Action Agreement Technologies, of which I was a work package leader. Relevant publications: [Lopes et al., 2010a, Lopes et al., 2010b, Zimmermann et al., 2012, Lopes et al., 2012].
- In Chapter 6, I present a model – NdFluents – for reifying contextual information into standard semantic web formalisms such that reasoning can be performed over the reified model in the same way as it is done in a non contextual setting. The challenge is to provide a representation of context that is fully within the standard knowledge representation formalism, yet allows some form of contextual reasoning. This contrasts with the previous chapter where a non-standard formalism is introduced. Multi-contextual knowledge may mediate through a transformation mechanism that assembles the sources into a unified model. This work was made possible by my participation in H2020 project WDAqua and my supervision of PhD student José Giménez-García. Relevant publications: [Giménez-García et al., 2016b, Giménez-García et al., 2017, Zimmermann and Giménez-García, 2017a, Zimmermann and Giménez-García, 2017b].

- Part III makes the assumption that we have a more uniform data and knowledge layer, and takes advantage of it for two specific cases:
  - In Chapter 7, interacting social agents and things on the Web, leveraging social networking platforms that are semantically described to allow autonomous decision making. The challenge is to make interactions between things and agents easier and more systematic in the context of the Internet of Things. Social platforms are semantically described, providing a mediation point through which agents can communicate and coordinate. This work resulted from my supervision of PhD student Andrei Ciortea, who was in a *cotutelle* between Mines Saint-Étienne and Polytechnic University of Bucharest. Relevant publications: [Ciortea et al., 2013, Ciortea et al., 2014, Ciortea et al., 2015, Ciortea et al., 2016b, Ciortea et al.,

2016a, Ciortea et al., 2018, Ciortea et al., 2017, Ciortea et al., 2019] and Andrei's thesis [Ciortea, 2016].

- In Chapter 8, multi-goal pathfinding in ubiquitous environments, where ontologies and knowledge graphs are used to decide itineraries that also satisfy goals in passing. The challenge is to have a pathfinding algorithm that can leverage formal description of a dynamic environment, with possible issues of latency and recomputation at travel time. The communications between the actors of this type of system may take advantage of the social Web of things described in the previous chapter. This work resulted from my supervision of PhD student Oudom Kem. Relevant publications: [Kem et al., 2016, Kem et al., 2017d, Kem et al., 2017a, Kem et al., 2017b, Kem et al., 2017c] and Oudom's thesis [Kem, 2018].

- Part IV summarises the contributions and present the research directions I want to pursue in the future:
  - In Chapter 9, I provide a synthesis that highlights the relations and dependencies between the different parts of this thesis, focusing on the two main aspects, namely *interoperability* and *loosely coupled mediation*.
  - In Chapter 10, I give several research directions that I would like to investigate in order to go beyond my present contributions. While I focus on a few building blocks, the overall long term plan consists in developing interoperable socio-technical systems on the Web.

- Lastly, the appendices provide a summary of my past career, in terms of teaching (Appendix A), research duties (Appendix B), supervision (Appendix C), projects (Appendix D), and publications (Appendix E). This final, mostly quantitative synthesis of my professional activities should testify of my abilities to conduct independent research, communicate it appropriately, collaborate towards its development, train others to further it, and participate in the communities by sharing thoughts and burden.

Part I

Data Interoperability: Transformation, Organisation, Interrogation

Outline of the contributions

A first step towards the realisation of the vision presented in the introduction is to enable a Web of Data. Although the Web is used to expose all kinds of data formats, an application that exploits various data sets must have a cohesive model to which it can relate information it consumes.

This part presents my contributions about enabling interoperability at the data layer. There are several approaches to homogenise access to data on the Web, each leading to different challenges:

- First, all data may be transformed into a unique data model that can in turn be integrated in the client's system or application. The challenge is to provide tools that make the transformations easy, flexible, and language-neutral. To this aim, I first describe my contribution to a declarative transformation language called SPARQL-Generate (Chapter 1);

- Second, precise data elements may be retrieved according to a query language, but it poses research challenges when the data sources are database systems that use different data models and query languages. My second contribution addresses federated query answering over multiple data models, with an approach called PolyWeb, building on the idea of Polystores introduced by Stonebraker (Chapter 2);

- Third, one may homogenise the way data can be navigated through. Linked Data Platform 1.0 is a standard that allows this, but the complexity of putting such platforms in place makes their use challenging. My third contribution consists in mainly a declarative language that describes how raw data must be organised in a Linked Data Platform that is deployed automatically. The engine for the language is thus mediating the data flow between the data sources and the Web server that hosts the platform (Chapter 3).

Resulting from these contributions, we can assume that all data on the Web can be presented to the client or user as RDF through a transformation, a platform, or a query endpoint. At the end of this part, I summarise the contribution putting them in perspective of future research that they enable.

The work presented in this part results from: national project OpenSensingCity that I coordinated (Chapter 1 and 3); my supervision of PhD student Noorani Bakerally in the context of project, jointly with Prof. Olivier Boissier (Chapter 1 and 3); European project SEAS (Chapter 1); and collaboration with Insight Galway, particularly the resaerch team of Ratnesh Sahay, whom I unofficially supervised in his PhD during my time in Galway and until his defense in 2012 (Chapter 2).

Chapter 1

Transforming Heterogeneous Data

Enabling an homogeneous view on the data can be achieved by transforming all available data into a single, easily integrated data model such as RDF. Migrating existing data towards RDF was part of the early efforts of the Semantic Web community. Tools called *triplifiers*[1] were developed to transform application-specific formats to RDF (such as BibTeX, ID3Tag, iCal). A large collaboration of mostly academic actors worked towards the publication of open data in RDF in order to connect distributed data across the Web.

The result of the effort is visualised in the Linked Open Data Cloud.[2] At first sight, there seems to be a vast amount of data of all kinds available in this form. However, with a closer look, we must admit that this data cloud is but a tiny drop in an immeasurable ocean of Web data in very heterogeneous forms as CSV, JSON, XML, PDF, Excel files, etc. If there must be a common data model for data integration across the Web, then there may be a need for a mediation approach that enables homogenising data formats for the consumer.

Our vision of loosely coupled mediation can be achieved at the data layer by relying on the following components:

- A language for declaring data transformation from all kinds of formats to a *lingua franca*;
- A mediator being the implementation of an engine for the said language;
- Wrappers that take the form of files instantiating the language for specific transformations.

In our case, the *lingua franca* of the Web is RDF. There exist several languages for expressing transformations to RDF from heterogeneous data, as shown, for instance, by Dimou et al. [Dimou et al., 2018]. One of them is SPARQL-Generate, that we introduced with my colleague Maxime Lefrançois, for the reasons explained in the next section. A description of how SPARQL-Generate works is then given in Section 1.2.

[1] ConverterToRDF. W3C wiki. `https://www.w3.org/wiki/ConverterToRdf`, retrieved 31 August 2020.

[2] The Linked Open Data Cloud. `https://lod-cloud.net/` retrieved 24 July 2019.

1.1 Rationale for introducing SPARQL-Generate

The design of SPARQL-Generate stems from an analysis of the needs of several use cases where RDF is the interchange data model of choice. In the field of open data, developers often limit their use of open data sets to one at a time because of the heterogeneity of the data formats, schemas, lack of documentation or ambiguous terminology. As pointed out in a study based on interviews with open data reusers (data scientists and Web developers) [Dymytrova and Paquienséguy, 2019], conducted during project OpenSensingCity, data transformation, structuring and filtering is particularly time consuming.[3] In the field of the Internet of Things, while there are reference vocabularies in the form of Web ontologies, strong constraints prevent things to directly communicate in RDF. Interoperability between things on the IoT based on Web ontologies requires a translation step from raw data to RDF conforming to these ontologies. This second use case was particularly relevant to ITEA project Smart Energy Aware Systems (or SEAS for short),[4] where a unified knowledge model for the energy domain was proposed [Lefrançois et al., 2017a] to which all data interchanged among energy-related systems should conform.

As a result of analysing these cases, we identified the following requirements for our transformation framework:

R1: transform several sources having heterogeneous formats;

R2: contextualize the transformation with an RDF Dataset;

R3: be extensible to new data formats;

R4: be easy to use by Semantic Web experts;

R5: integrate in a typical semantic web engineering workflow;

R6: be flexible and easily maintainable;

R7: transform binary formats as well as textual formats.

The result took the form of a language that extends SPARQL, the standard query language for RDF, with the capability to extract data items from any data source and generate RDF from them, thus the name *SPARQL-Generate*. Building on top of SPARQL allows semantic web developer to use a single language for all data manipulation in their workflow. The SPARQL-Generate engine is a mediator that can either be made available as a Web service or integrated in an application; a specific SPARQL-Generate transformation acts as a wrapper to a data source. A transformation can hardcode the data source that it transforms, or can be generic, applied to a data set chosen at execution time by the engine. Other resources such as a repository of transformations could be made available to help decide what transformations to use and where to find them in function of the data sets to be integrated.

[3] "*La transformation des données et leur mise en forme (filtrage en fonction des éléments recherchés, élimination de redondances, structuration et transformation dans les formats utilisés) sont des étapes particulièrement chronophages*".

[4] SEAS project overview at ITEA `https://itea3.org/project/seas.html` retrieved 31 August 2020.

1.2 The SPARQL-Generate language

The principles behind the writing of a data transformation in SPARQL-Generate is the following:

1. define a graph shape for the output of the transformation, in the form of a graph pattern where some parts of the graph (nodes or arcs) are replaced with variables;
2. identify the data sources that must be transformed by providing their URLs;
3. describe the portions of the source data files that must be extracted to instantiate the variables of the graph shape. This can be done with a selection pattern expressed in a language that is specific to the source data format (such as XPath, JSONPath, CSS selector). This is typically done in two parts:
   (a) first, identify a recurring structure in the source file (such as a line in a CSV file) over which the data generation must iterate;
   (b) second, select within each substructure the specific data value that must be extracted (for instance, the value of a column in a CSV line).

Figure 1.1 shows an example of a simple SPARQL-Generate query.

Default graph (Turtle)

```
<s25> a :TempSensor ;
     geo:lat 38.677220 ;
     geo:long -27.212627 .
<s26> a :TempSensor ;
     geo:lat 37.790498 ;
     geo:long -25.501970 .
<s27> a :TempSensor ;
     geo:lat 37.780768;
     geo:long -25.496294 .
```

Document position.txt

```
37.780496,-25.495157
```

Document measures.json

```
{ "s25": 14.24,
  "s26": 18.18 }
```

Output (Turtle)

```
  <s26> a :NearbySensor ;
       :temp 18.18 .
  <s27> a :NearbySensor .
```

SPARQL-Generate query

```
GENERATE {
  ?sensor a :NearbySensor .

  GENERATE {
    ?sensorIRI :temp ?temp .
  }
  ITERATOR sgiter:JSONListKeys(?measures) AS ?sensorId
  WHERE {
    BIND( IRI( ?sensorId ) AS ?sensorIRI )
    FILTER( ?sensor = ?sensorIRI )
    BIND( CONCAT( "$." , ?sensorId ) AS ?jsonPath )
    BIND( sgfn:JSONPath( ?measures , ?jsonPath ) AS ?temp )
  } .
}
SOURCE <position.txt> AS ?pos
SOURCE <measures.json> AS ?measures
WHERE {
  BIND( sgfn:SplitAtPosition(?pos,"(.*),(.*)",1) AS ?long )
  BIND( sgfn:SplitAtPosition(?pos,"(.*),(.*)",2) AS ?lat )
  ?sensor a :TempSensor .
  ?sensor geo:lat ?slat .
  ?sensor geo:long ?slong .
  FILTER( ex:distance(?lat, ?long, ?slat, ?slong) < 10 )
}
```

Figure 1.1: Example of a SPARQL-Generate query execution on a default graph and two documents. From [Lefrançois et al., 2017b]

Item 1 is written in the transformation as a kind of SPARQL graph pattern that allows the use of convenient syntactic sugar, for conciseness. This pattern is put in a special clause at the beginning of the query (`GENERATE`) to comply with how `CONSTRUCT` queries are written in SPARQL. Typically, the graph pattern is instantiated multiple times from a given data source, and it is possible to nest other `GENERATE` patterns in order to generate multiple subgraphs.

Item 2 is given with a `SOURCE` clause which provides the URL to be fetched, and assigns the content of the file at this URL to a variable, as a literal value.

Item 3a is specified in the clause `ITERATOR` which uses format-specific functions that produce a set of solution mappings based on a selection pattern, such as an XPath query or a regular expression with parenthesised patterns. The iteration functions work in a way similar to SPARQL extension functions except that they provide a set of mappings rather than a single value. More iteration functions can be added in the same way as extension functions are added to a SPARQL engine. This makes the language easily extensible.

Item 3b is done by binding certain variables to an expression that can rely on SPARQL operators as well as extension functions that implement each selection languages (XPath, JSONPath, etc.). These functions are regular SPARQL extension functions, such that the part given in the `WHERE` clause is a valid SPARQL `WHERE` (modulo syntactic sugar that we added for convenience). A SPARQL-Generate transformation can make use of a local triplestore in addition to its data extraction from source files. This is given "for free" by choosing to base our language on SPARQL.

SPARQL-Generate semantics The formal semantics for SPARQL-Generate is given in [Lefrançois et al., 2017b]. I will not reproduce it here in full, but give the main idea.

A SPARQL-Generate query works on a data structure that combines an RDF dataset and what we call a *documentset*, being a set of pairs (i, d) where i is an IRI and d is a document that we model as a literal whose datatype IRI identifies the file format of the document, and such that the IRIs do not appear in 2 or more pairs.

A `SOURCE` clause that mentions an IRI and a variable is executed by assigning the document associated with the IRI in the documentset to the variable. Several `SOURCE` clauses may exist, so multiple variables can be bound in this way.

An `ITERATOR` uses an iteration function, which produces as output a sequence of substructures (literals which are typed depending on the structure, such as a DOM tree for an XPath iterator). This sequence is assigned to a variable in different SPARQL solution mappings.

The `WHERE` clause is evaluated as in SPARQL and joined with the solution mappings from the `ITERATOR` and `SOURCE` clauses. This results in a SPARQL solution set that in turn is used to instantiate the graph pattern in the `GENERATE` clause. In case there are nested `GENERATE`, the same is done multiple times for each solution mapping of the enclosing `GENERATE`.

The current version of SPARQL-Generate allows iterators to bind multiple variables at the same time, and provide syntactic sugar that simplifies the writing of transformations very much. This does not affect the formal semantics.

1.3 SPARQL-Generate implementation

Our reference SPARQL-Generate engine has been implemented by Maxime Lefrançois, with some contributions by Noorani Bakerally, El Mehdi Khalfi, Omar Qawasmeh and myself. It is based on Apache Jena, an open source framework for Semantic Web development. At the time of writing this dissertation, it is the only implementation of the language. The source code is available at `https://github.com/sparql-generate/`.

Our tests have shown that we have a competitive implementation in terms of efficiency, but it is an active domain where changes happen fast: indeed, our initial paper [Lefrançois et al., 2017b] shows much better performance than the RML engine at the time. A later comparison of RDF generation engines showed honorable results from SPARQL-Generate.[5] But a recent test by a new version of the RML engine inverted the domination [Haesendonck et al., 2019]. Since then, version 2 of SPARQL-Generate was released, with improvements on the performances, but no comparison has been performed yet. However, the advantages of SPARQL-Generate do not reside in the throughput performance of the implementation, as explained in the rationale for the language (Section 1.1).

The engine is available in different forms:

- An online testing tool, that we call *the SPARQL-Generate playground*, can be used for new users to get familiar with the language, or to quickly test queries on small input, especially for debugging purposes. A screenshot of the playground can be seen in Figure 1.2;
- An executable jar that can be used as a command line tool.
- An extension of text editor Sublime Text with syntax highlight, execution and debugging information of SPARQL-Generate transformation from a keyboard shortcut.
- An open source API available at Maven Central.[6]

More recently, Maxime Lefrançois added the possibility to do `SELECT` queries in addition to `GENERATE` queries, thus making SPARQL-Generate a full-fledged query language for heterogeneous data. On top of this, he integrated Olivier Corby's STTL [Corby and Faron-Zucker, 2015], an extension of SPARQL for generating arbitrary text strings based on RDF data. STTL string templates can be used in combination with `GENERATE` transformations, and vice versa, thus making the language a complete data manipulation framework. All of this is available in the current implementation on Github. This shows the benefit of choosing SPARQL as the basis for our data transformation language, as it can be easily extended to cover much more data processing tasks.

1.4 Summary

This work was made in collaboration with Maxime Lefrançois (postdoc then associate professor at Mines Saint-Étienne) and Noorani Bakerally (Phd student at Mines Saint-Étienne) in the context of projects SEAS and OpenSensingCity. My publications that relates to this work are: [Lefrançois et al., 2016b, Lefrançois et al., 2017c, Lefrançois et al., 2017b]. Figure 1.3 describes how it instantiates the general idea of a loosely coupled mediation approach that I am defending in this thesis.

[5] Maxim Kolchin: A practical review of non-RDF to RDF converters. `https://medium.com/datafabric/a-practical-review-of-non-rdf-to-rdf-converters-51686338927f` retrieved 31 August 2020.

[6] SPARQL-Generate at Maven Central `https://search.maven.org/search?q=fr.mines-stetienne.ci.sparql-generate`

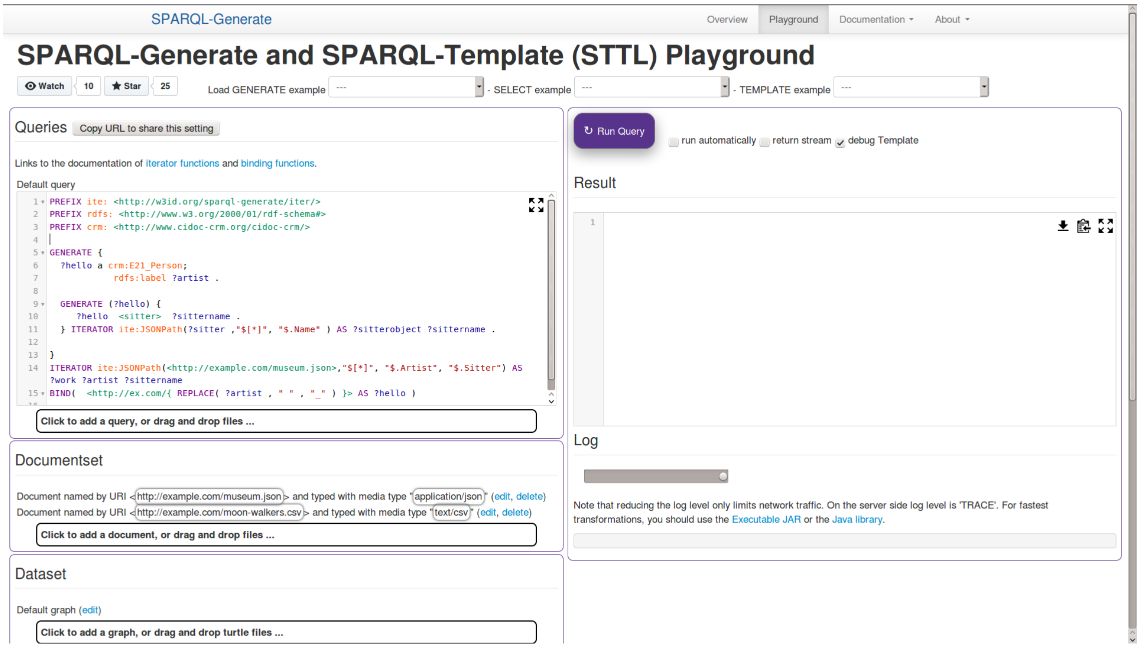


Figure 1.2: SPARQL-Generate online playground. A textarea contains the editable query, with syntax and error highlighting. An RDF dataset can be edited to simulate data on a triplestore. A documentset can be edited, but if an IRI is not present there, a `SOURCE` that mentions the IRI will dereference it from the Web. The result of the transformation is given in the text box on the right hand side, and below are log messages.

The information sources are data files, that could be static or dynamically obtained from Web services (grey boxes inside the cloud-like shape). The mediation makes all data appear to be RDF, by way of the SPARQL-Generate engine (the mediator, large white boxes) that makes use of SPARQL-Generate transformations, that act as wrapper specifications (Q1 to Q4). SPARQL-Generate transformations are usually referencing data sources, but they can also be generic, such that the client select the sources that it wants to transform with a given SPARQL-Generate query. It is also possible that a data source provides a link or reference to a SPARQL-Generate transformation, such that it can be interpreted as RDF if need be (e.g., S4 and S5 in the picture). The SPARQL-Generate engine can exist as a Web resource or be embedded in the client (as in C3). Thereby, there is a loose coupling between all the components of this mediation architecture, as opposed to, for instance, a Web server that provides an HTTP access to its internal data or application.

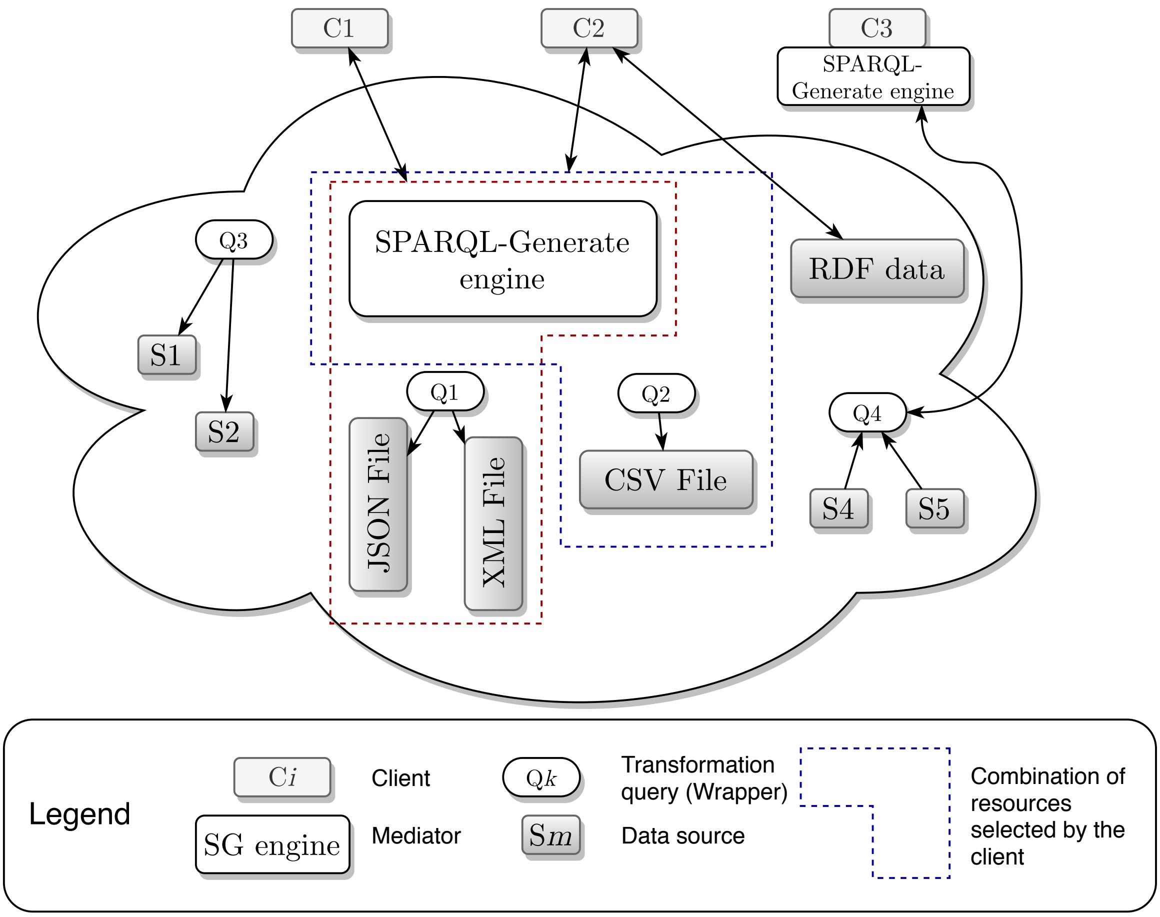


Figure 1.3: SPARQL-Generate as an instance of loosely coupled mediation

Chapter 2

Querying Heterogeneous Data Models

Sometimes, data is not made directly available in the form of files that must be processed by the client. Instead, an access point with a rich query language allows an application to directly request precise data values from the data source. This allows the server to decide and optimise how the data is processed.

However, there can be heterogeneity in the query languages used at different data sources. The difference in query languages usually comes from a difference in the data models used internally behind the access points. In 2015, Michael Stonebraker noted that translating all data to a single data model that can be queried with a single query language is very costly, and may not be efficient because local optimisations on a data model cannot easily be translated to the target data model [Stonebraker, 2015]. Thus there was a need for a new form of federated database systems where multiple data models could coexist, but that should be accessible with a uniform interface. This is the idea of Polystores, that was first implemented by the Big Data Working Group in the system known as BigDAWG [Duggan et al., 2015].

It is important to note that the concept of Polystore is an abstract idea of unifying querying over multiple data models which can be implemented using different technologies (such as relational, array, or graph models). For instance, a Polystore may provide an interface that only supports SQL queries, and the engine is in charge of selecting local storages, splitting the query into subqueries, translating the SQL into the local, native query language, issuing the transformed queries and joining the results into a table.

A federated query system is essentially a mediation systems that stands between a user making queries and multiple database management systems that need not be known by the user.

In the case of the work I am describing here, conducted in collaboration with researchers from the Insight Center for Data Analytics in Galway, Ireland, we chose to devise a Polystore entirely based on semantic technologies. It is called PolyWeb [Khan et al., 2017b, Khan et al., 2019] and I give more details on its principles (in Section 2.1) and its implementation and performances (in Section 2.2). My part in this work was mainly to provide a rigorous formalisation of the approach that my collaborators implemented.

2.1 The PolyWeb approach

PolyWeb can be considered as a federated SPARQL engine, in the sense that it processes a single SPARQL query that is used to select relevant query endpoints, split the query into subqueries, issue the subqueries to the right endpoint, get the results and join them into a single SPARQL solution set. We identified 12 federated SPARQL engines in [Khan et al., 2019], but none are able to query RDF and non-RDF databases at the same time. We fill this gap with PolyWeb by introducing extra steps to define the subqueries (not necessarily in SPARQL) and join the results (not necessarily following the SPARQL query results format).

Consequently, the queries must be translated into the native query language(s) of the data stores, and the returned results translated back into a SPARQL solution set. The strategies for data source selection are somewhat similar to those of other SPARQL federation engines, but the information used to apply these strategies is quite different.

2.1.1 Data models and mapping definitions

The PolyWeb approach can only deal with data models for which certain assumptions hold:

- the data model must provide a native query language (Definition 1);
- it must be possible to translate data sets of a data model to RDF, using a mapping definition (Definition 2).

This section clarifies these notions, starting from a description of a data model. We introduce Definition 1 in order to abstract away from any specific data model, generalising the principle of Polystores.

Definition 1 (Data model) *A data model* dm *defines a set* $\mathsf{DS}_{\mathsf{dm}}$ *of data sets that instantiate the data model, and a query language* $\mathsf{QL}_{\mathsf{dm}}$*. The query language allows one to pose a query* $q \in \mathsf{QL}_{\mathsf{dm}}$ *against a data set* $D \in \mathsf{DS}_{\mathsf{dm}}$ *to obtain a response in a result format* $\mathsf{RS}_{\mathsf{dm}}$*. A function* $\mathrm{Eval}_{\mathsf{dm}} : \mathsf{QL}_{\mathsf{dm}} \times \mathsf{DS}_{\mathsf{dm}} \to \mathsf{RS}_{\mathsf{dm}}$ *defines what response is obtained from issuing a query against a data set.*

As an example, in the relational data model, data sets are relational databases (sets of relational tables), the query language corresponds to the relational algebra (or SQL), responses are in the form of a relational table. In the RDF data model, data sets are RDF graphs, the query language for RDF is SPARQL, which provides responses in the form of SPARQL result sets [Harris and Seaborne, 2013]. We denote the RDF data model with rdf, the set of RDF graphs with $\mathcal{G}$, the set of SPARQL queries sparql, the set of SPARQL results as $\mathsf{RS}_{\mathsf{sparql}}$, and the evaluation of a SPARQL query q against an RDF graph G with $[\![q]\!]_G$. In order to be able to deal with many data models in a seamless way using SPARQL-only queries, we need a way to express the translation between a data set in native data model into RDF. This is done with the notion of mapping definition that serves to define a transformation from a non-RDF data set to RDF.

Definition 2 (Mapping definition) *A* mapping definition *for a data model* dm *is a function* $\mathrm{md} : \mathsf{DS}_{\mathsf{dm}} \to \mathcal{G}$*.*

A mapping definition can be used to query non-RDF data sets with the SPARQL query language. Indeed, if a SPARQL query q is posed on a data set $D \in \mathsf{DS}_{\mathsf{dm}}$ associated with a mapping definition md, then the response to the query can be defined as $[\![q]\!]_{\mathsf{md}(D)}$. However, this definition suggests that the data source should be completely translated to RDF according to the mapping definition, which would not be convenient when the data set is huge. Sometimes, it is not even possible to convert the source data because it is only made accessible via a query endpoint. Instead, the mapping definition should be leveraged to define a query translation, defined as follows:

Definition 3 (Query translation) *A query translation for a mapping definition* md *on a data model* dm *is a pair* $\langle \mathrm{qt}_{\mathsf{md}}, \mathrm{rt}_{\mathsf{md}} \rangle$ *where:*

- $\mathrm{qt}_{\mathsf{md}} : \mathsf{sparql} \to \mathsf{DS}_{\mathsf{dm}}$ *and*
- $\mathrm{rt}_{\mathsf{md}} : \mathsf{RS}_{\mathsf{dm}} \to \mathsf{RS}_{\mathsf{sparql}}$,

such that for all $q \in \mathsf{sparql}$ *and all* $D \in \mathsf{DS}_{\mathsf{dm}}$, $[\![q]\!]_{\mathsf{md}(D)} = \mathrm{rt}_{\mathsf{md}}(\mathrm{Eval}_{\mathsf{dm}}(\mathrm{qt}_{\mathsf{md}}(q), D))$.

Mapping definitions can be written in dedicated mapping languages such as R2RML [Das et al., 2012] (for mapping relational databases to RDF), RML [Dimou et al., 2014] (extending R2RML to other data models, such as XML, JSON, CSV, HTML), XSPARQL [Polleres et al., 2009] (initially for XML), SPARQL-Generate [Lefrançois et al., 2017c] (CSV, JSON, XML, HTML, CBOR, others), XSLT [Kay, 2017] (if limited to transformations that results in RDF/XML), defined as a JSON-LD [Sporny et al., 2014] `@context` for JSON, or using the CSVW vocabulary [Tennison and Kellogg, 2015a] for CSV.

2.1.2 Data summaries for source selection

In order to determine what data sources are relevant for a given query, PolyWeb relies on metadata that we call *data summaries*. In general, data summaries consist of a very small amount of data and are easy to compute. In the case of PolyWeb, data summaries are exclusively based on URIs that appear in predicate position of triples in the data sets. However, since data sets may not be in RDF, the notion of "predicate position", let alone "triple" is irrelevant. So PolyWeb makes use of mapping definitions that must be associated with every non-RDF data sources. Additionally, PolyWeb differentiate predicates that are exclusively occurring in a single data source; it also distinguishes certain predicates as being "unsafe". An unsafe predicate is one that is used on some instances of a class, but not all instances. Formally, an unsafe predicate in an RDF graph is defined as follows:

Definition 4 (Unsafe predicate) *For a given graph* $g \in \mathcal{G}$*, an IRI* p *is an* unsafe predicate *if and only if there exist* c, s, o, s', o' *such that* $s \neq s'$, $\{\langle s, p, o\rangle, \langle s, \mathtt{rdf:type}, c\rangle, \langle s', \mathtt{rdf:type}, c$ *g and* $\langle s', p, o'\rangle \notin g$.

The reason for these specific choices for data summaries is primarily due to the application domain that PolyWeb was developed for: biomedical big data, in the context of a project funded by Science Foundation Ireland. I was not involved in the project, nor the decisions on what to use as a data summary. I will criticise this choice in Section 2.3. The following section shows how the approach compares to other federated SPARQL query engines that require all the data to be translated to RDF.

2.2 Implementation and evaluation

PolyWeb is implemented in Java, makes use of the RML mapping language for conversion of non-RDF data sources to RDF, and for translation of SPARQL queries into native local queries. At the moment, it can query triplestores, relational databases, and Apache Drill data stores for CSV files. It can be noticed that the two non-RDF data models are both based on tabular data. Therefore, experimental results should be taken with a grain of salt. Yet, in spite of the strong limitation of the tested system, the results are instructive in several ways.

First, let us observe Table 2.1. This table shows statistics about data sets used for the PolyWeb experiments. The last column in particular is worthy of consideration, as it shows that converting all the data to RDF takes a significant amount of time. If the conversion is done only once, or only on rare occasions, then the investment pays off, as we will see later. However, if the data source changes frequently, then any solution that relies exclusively on RDF data has a serious disadvantage.

Table 2.1: Overview of the data sets used in the PolyWeb experiments in [Khan et al., 2019]

Data set	**format**	**size**	**records**	**RDFising**	**triples**	**sub.**	**pred.**	**obj.**	**RDF size**
COSMIC-CNV	CSV	3.2 GB	29 M	3 hours	37 M	1 M	18	0.3 M	6.54 GB
TCGA-OV-CNV	RDB	212 MB	2.6 M	2 min.	10 M	1.8 M	6	1.2 M	494 MB
CNVD-CNV	TSV	34 MB	0.2 M	3 sec.	1.7 M	0.2 M	12	1.7 M	128 MB
Total	-	3.5 GB	32 M	∼3 hours	49 M	2 M	36	3 M	7 GB

Second, we compare PolyWeb's performances with a version of PolyWeb that only queries RDF data (PolyWeb-RDF, that does not do query and result conversion) and two other federated SPARQL query systems (FedX [Schwarte et al., 2011] and HiBISCuS [Saleem and Ngonga Ngomo, 2014]). The two federation engines have been chosen because they showed best performances in recent benchmarks and, interestingly, FedX is an index-free system (in the sense that it does not rely on data summaries), while HiBISCuS computes a data summary.

Figure 2.1 and Figure 2.2 show the time for selecting the relevant data sources and the execution time of the queries, respectively. We can see that PolyWeb's source selection strategy is extremely efficient when compared to the other federated engines, but this advantage is completely overturned when considering query execution time. PolyWeb is nearly up to two orders of magnitude slower than the other systems on some queries. However, if the ability to query hetereogeneous data models is turned off (using PolyWeb-RDF), the system becomes competitive again. I discuss these results in the next section.

2.3 Discussion

Careless attention to the previous section may put too much focus on the alarming results of Figure 2.2. In this section, I will mitigate these results. To ensure fair comparison, I will only compare PolyWeb to PolyWeb-RDF, because they rely on the same source selection strategies. The difference in execution time then can only

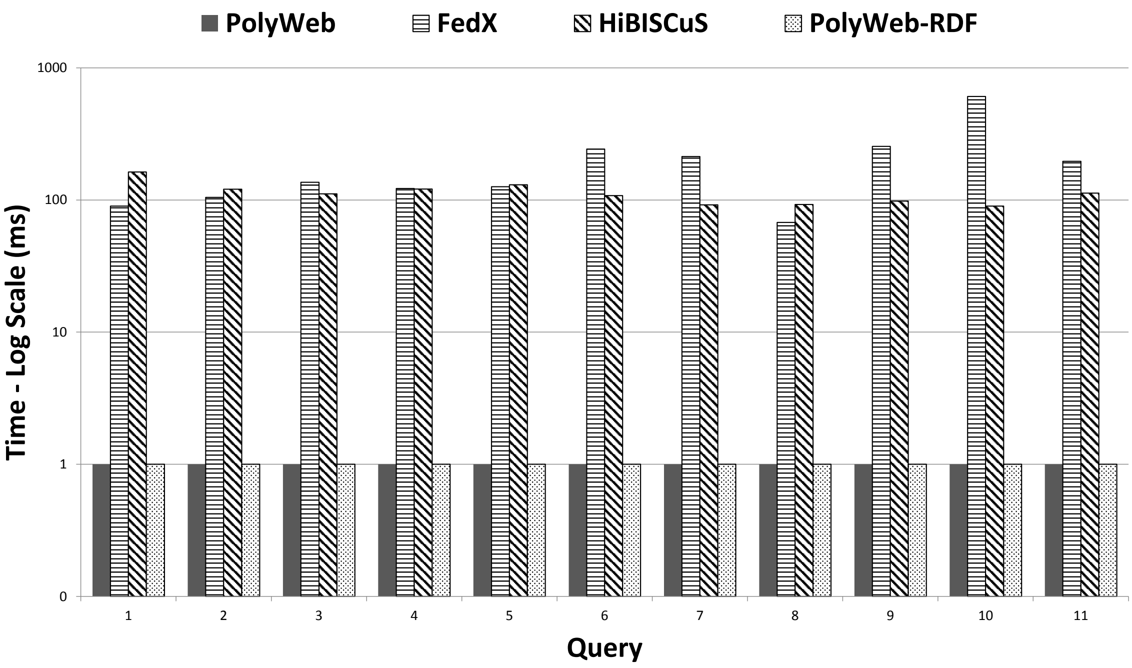


Figure 2.1: Comparison of source selection time

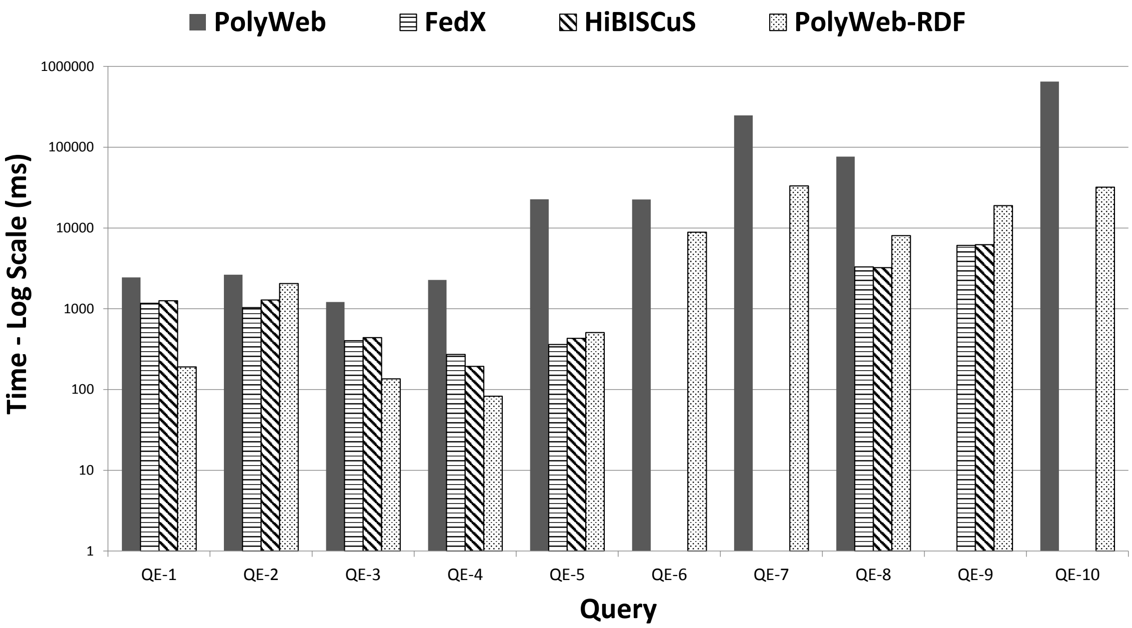


Figure 2.2: Comparison of query execution time. An absent bar means a timeout

be explained by the following factors:[1]

1. PolyWeb does query rewriting when PolyWeb-RDF does not;

2. With PolyWeb, the translated queries are executed against a non-RDF datastore, while PolyWeb-RDF executes the original SPARQL queries on a triplestore;

[1]Here I dismiss the ever possible human errors that introduce bugs in the programme or mistakes in the experimental setups, etc.

3. PolyWeb must translate the query results back to SPARQL solution sets.

It is doubtful that Item 1 and Item 3 are responsible for the important overhead. Query rewriting is applied to a rather small set of queries that are themselves small. The complexity of the rewriting task would not lead to a long lasting process. Translating results may have to be performed on a large result set, but it is a very straightforward, linear operation.

So the remaining issue must be the execution of the translated query using the native query system. It is very much possible that our benchmark has queries that are typical of a graph-based data model, but that are uncommon and awkward in other data models. Then this leaves us with a question: why not turn the data into a graph if the information need is that of a graph-based query?

The answer is simple: data transformation becomes unbearable as the pace of changes in the sources increase. This highlights a common problem with a mediation approach: its benefit usually cannot be shown at run time, when compared to a tightly coupled integrated system that is already in place. But quantitative comparison of execution time hide the true cost of tight integration. Loosely coupled components can be developed fast and can evolve independently from each others. Thus, the choice of a mediation approach must be balanced with the accumulated costs of design, migration, transformation and execution time.

Another aspect that was not so well developed in [Khan et al., 2019] is data summaries and selection strategy. PolyWeb, as a concept, does not enforce any specific data summarisation, nor does it impose a specific selection strategy. But the assumption that each data source must be accompanied with a mapping definition provides a strong advantage in defining what data sources to query. Indeed, a declarative mapping definition (such as a SPARQL-Generate or RML transformation) can be seen as a schema for the target data model (namely RDF). Said schema can accurately determine whether a subquery is likely to return results, thus possibly minimising the number of queries, and joins, to be performed on the sources. Moreover, when the data evolve, the mapping can stay the same, eliminating the need to recompute data summaries at each update.

2.4 Summary

This work was mainly made in collaboration with Yasar Khan (postdoc Insight Galway) and other contributors from Insight Galway. My publications that relates this work are: [Khan et al., 2017b, Khan et al., 2019]. Figure 2.3 describes how it instantiates the general idea of a loosely coupled mediation approach that I am defending in this thesis.

The information sources are query endpoints of different data models. The mediator, which is the federated query engine PolyWeb, offers a uniform SPARQL query endpoint, while each source is wrapped according to 1) a transformation from SPARQL to the local query language and from query results to SPARQL query results (T1, T2), and 2) a data summary for source selection (DS1, DS2, DS3). The description of PolyWeb here allows a loose coupling of the components but in our prototype, data summaries, transformations and data sources were hardcoded. However, distributing and retrieving at run time these components from anywhere on the Web is not likely to produce significant overhead since we expect the rest of

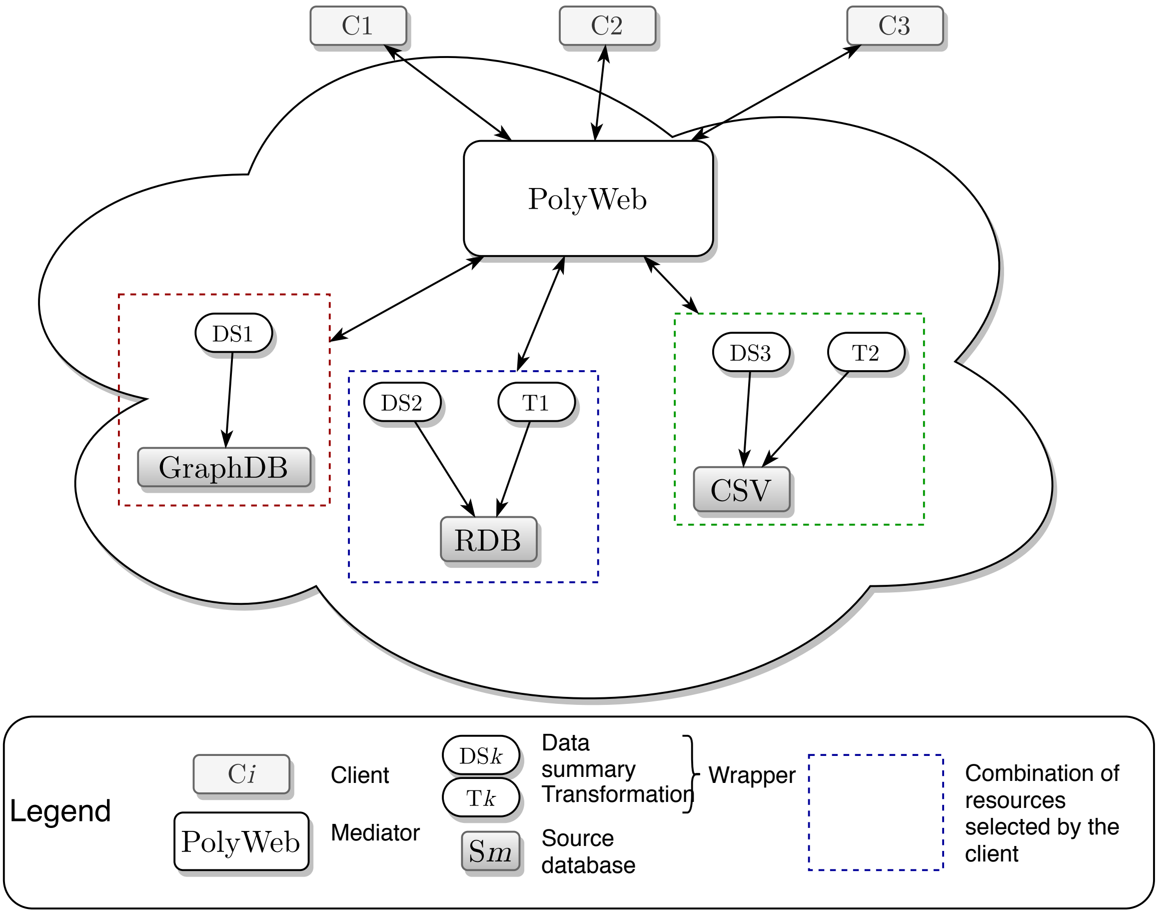


Figure 2.3: PolyWeb as an instance of loosely coupled mediation

the process (source selection, query transformation, query execution, result joins) to be the most time consuming tasks.

Chapter 3

Automatic Deployment of Linked Data Platforms

Once the data model has been uniformised, it is still possible to improve data interoperability by organising the data in more accessible way. Linked Data Platforms, a type of data server architecture standardised by the W3C, improve data access by relying on REST principles for data access and management. In this section, I first provide an overview of Linked Data Platform 1.0, the W3C standard, to show how it helps navigating through linked data (Section 3.1). Then I present the work done with PhD student Noorani Bakerally in order to automatise the deployment of such platforms, where a key component is a language that describes how to organise the data on the platform. This fits the vision of a loose mediation because it places a set of software components between raw, unorganised data (data dumps or distributed data files) and a platform that exposes the data in a uniformly navigable way. This contribution is meant to simplify and automate the deployment of Linked Data, a task that has been little investigated in the past. This work is split in two main contributions: a workflow for data deployment that decompose the task into subtasks that each can be implemented as software modules (Section 3.2); and a language to describe how to organise data on the final platform, such that data can be deployed automatically on any implementation of the standard (Section 3.3).

3.1 Linked Data Platform 1.0

The Linked Data Platform (or LDP for short) standard [Speicher et al., 2015] provides guidelines for implementing Web servers that expose data in compliance with the Linked Data principles, especially describing the use of HTTP for accessing, updating, creating and deleting resources.

The LDP standard allows fine-grained management of data items, called Linked Data Platform Resources (LDPR), with CRUD[1] operations over the HTTP protocol. A resource in this sense is anything that is deemed of interest to describe on the platform, such as people, sale products, places, events, documents. A special kind of resources is defined by the standard to organise LDPRs in a grouped, hierarchical way: Linked Data Platform Containers (LDPC), that are artifacts of the platform that facilitates navigation and management of similar resources.

[1] Create, Read, Update, Delete.

LDPs can host any kind of data, but it gives special status to data in RDF. A Linked Data Platform RDF Source is a resource represented in RDF on the platform, while a Linked Data Platform Non-RDF Source is not represented in RDF. Linked Data Platform Containers are always LDP RDF Source and group LDPRs together by linking from them to their members. Since LDPCs are themselves LDP resources, they can be organised in a hierarchy where some LDPCs are made members of other LDPCs.

The organisation of data on an LDP makes data discoverable and navigable easily through hypermedia. It also simplifies data management, as creating a resource in a container (with a new link from it) is as easy as a HTTP POST request to the URI of an LDPC. Linked Data Platforms and the LDP standards were identified as a result of the investigations of project OpenSensingCity, where open data professionals were interviewed to assses their needs, expectations, and requirements. From this study, we found that Semantic Web technologies are identified as crucial to enable interoperability and reusability of open data [Larroche and Dymytrova, 2017, Dymytrova, 2018] and more particularly, LDP satisfies most of the desiderata of open data users [Bakerally, 2018]. However, LDP is largely unknown by the actors of open data, and its use is currently reserved to those who are technical experts of Linked Data. Noorani's work highlighted the difficulties in using LDP implementations and paved the way towards much easier Linked Data deployment.

3.2 A workflow for data deployment

The deployment of a linked data platform requires deciding on a number of design choices, and can be split in different tasks. The workflow described here does not entirely fit in the mediation approach, in the sense that it is addressing other concerns such as methodological or architectural, but it helps understanding the purpose of the language and module described in Section 3.3. This work was published in [Bakerally and Zimmermann, 2017] and [Bakerally et al., 2018a]. We can see a depiction of the workflow in Figure 3.1.

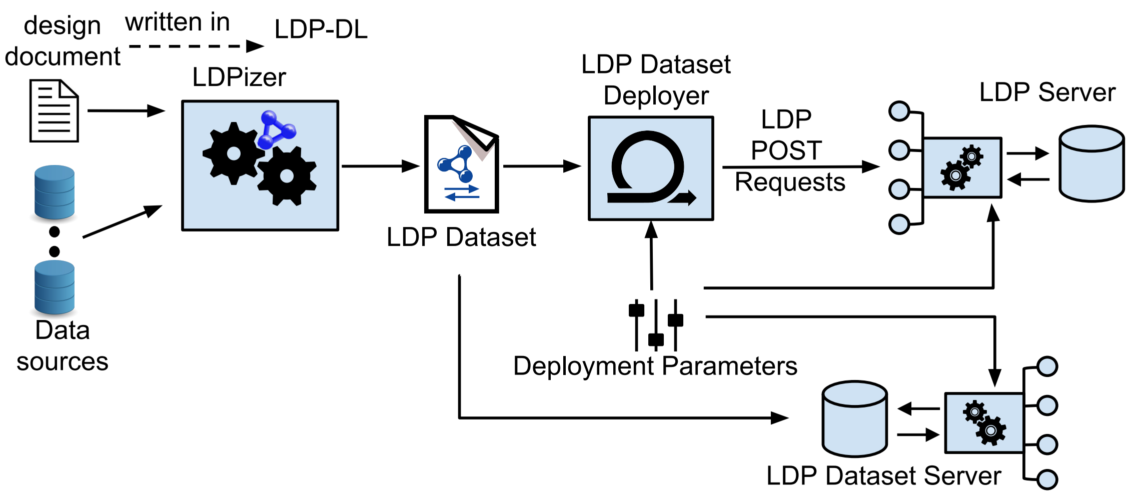


Figure 3.1: A workflow for the generation of linked data platforms

Our approach uses model-driven engineering where components consume and generate new models or transform them into running systems. It centrally relies on a domain-specific language, LDP-DL, that defines the core part of the platform. Specifically, a document written in LDP-DL (called a *design document*) indicates what data sources to use (see left side of Figure 3.1), and how data from the sources must be organised into containers, subcontainers, and non-container resources.

A software component called the LDPizer ingests a design document and source data sets to generate an intermediary model that describes how data will be organised on the deployed platform, but does so independently of (1) the location of the end platform (URLs are not fixed), (2) the access rights, (3) the implementation of the LDP standard. This model is called an LDP dataset, and can then be used in different ways, or just exchanged or published as is.

If an LDP server is in place, then the LDP dataset can be provided to a component in charge of its deployment on the platform. This essentially consists in performing POST requests, according to the LDP standard specification, with parameters suggesting what URIs to be used on newly created resources.[2]

Another option consists in using an implementation of LDP that directly consumes an LDP dataset. The advantage being that any change to the LDP dataset is immediately visible on the server, which can make the platform more dynamic and flexible.

The main components in this workflow are the design document and the LDPizer, which serve as a mediator between the data sources and the deployed platform. In the next section, I describe in more details the language for generating the LDP dataset.

3.3 Linked Data Platform Design Language

This work was published in [Bakerally et al., 2018b]. Our goal with the LDP design language is to declare how data available as large RDF datasets must be exposed on a linked data platform conforming with the LDP standard. Deploying an LDP is not as straightforward as putting a data dump for download, and quite different from providing a query endpoint. On an LDP, data is split into smaller information resources that should be navigable in a RESTful manner. For better navigability, resources are grouped into containers. This means that, given available data in RDF, the following decisions must be taken:

- what containers must be created?
- what resources must be put in which container?
- what RDF graph must be presented when retrieving the resource?
- what URIs should be used to identify the resources?

The difficulty with the first item is that the set of containers may not be fixed a priori. For instance, it is possible that a new container must be created for each instance of a certain class. Moreover, the containers can be nested. Regarding the third item, it is possible that part of the data in the sources is ignored, and that extra (meta)data is added in the "LDPization" process. Thus, there is no

[2] URIs can only be *suggested* to a Web server. The server always take the final decision on how to identify newly created resources.

"one size fits all" design for this. For the fourth item, it may seem that we could simply reuse the URIs in the source data, since we assume they are already in RDF. However, compliance with the linked data principles enforces dereferenceability of URIs. Therefore, when a URI is looked up, one should get a description of the resource from the platform, not from an arbitrary web space, even less a 404.

3.3.1 Illustrative example

I illustrate these design decisions with Figure 3.2, taken from [Bakerally et al., 2018b].

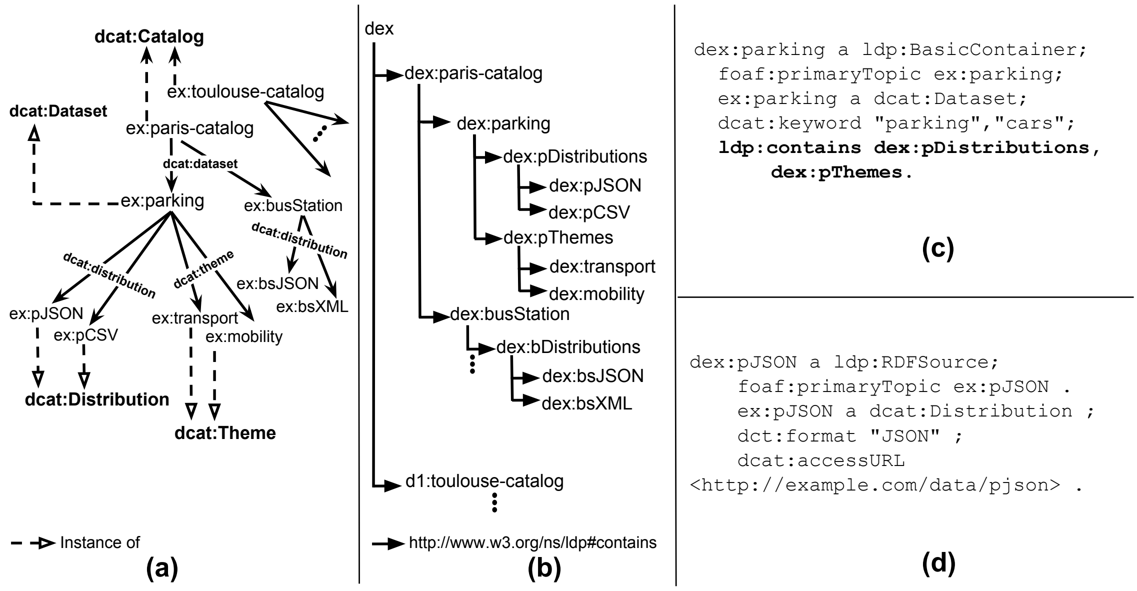


Figure 3.2: Example of an RDF Graph, a container hierarchy generated from it, and content of some LDP resources [Bakerally et al., 2018b]

On the left, this is an example data source, describing open data catalogs, following the W3C standard DCAT [Maali and Erickson, 2014]. These are catalogs from open data portals of French cities. Catalogs contain datasets (in this case, transportation data) and datasets may have one or more distributions that provide the data in a certain format, and themes. In the middle, we find the container hierarchy. There is a root container for all the data in the platform (`dex`) that contains as many sub-containers as there are catalogs in the source. Each catalog-container is then divided into multiple containers, one for each dataset of the catalog. A dataset-container has sub-containers for distributions and for themes. On the right, we have two examples of RDF graphs that we would like to get when dereferencing the URIs of the LDP resources. On the top, it corresponds to the parking container. We see that it describes the parking dataset from the source, and also contains other resources. At the bottom, this is a resource that describes a *distribution* in JSON for the parking dataset. It should be noticed that in order to describe the resource `ex:pJSON`, we introduce an extra resource `dex:pJSON` on the platform (different namespace). This means that we can navigate over all the descriptions of the resources from the source while "staying" on the platform.

3.3.2 Overview of the language

The abstract model of a design document is depicted in Figure 3.3, which provides the main components of the abstract syntax of LDP-DL.

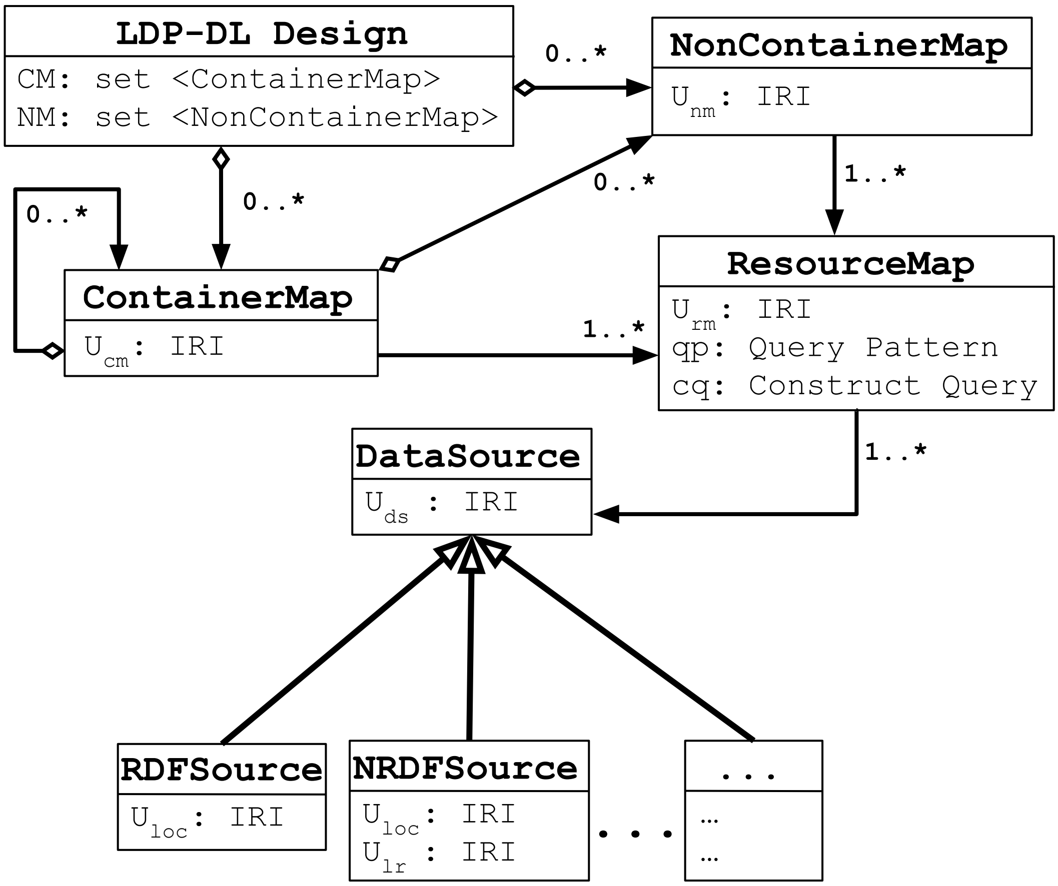


Figure 3.3: Abstract model of LDP-DL in UML notation. From Noorani Bakerally's thesis [Bakerally, 2018]

A design document contains an LDP-DL design that instantiates the upper left class of the diagram. The design has two components that are sets: a set of `ContainerMap`s and a set of `NonContainerMap`s. A `ContainerMap` gives the instructions to create containers and their contents, while a `NonContainerMap` describes resources that will not be identified as containers and thus will not contain other resources. Regardless of whether they are containers or not, resources created on the LPD are associated with resources existing in the data sources, and each LDPR will be associated with an RDF graph that will be served when dereferencing the URI of the resource. These are given by `ResourceMap`s that not only query the source data to identify relevant resources from which to create LDP containers or non containers (using query pattern `qp`) but also define what graph is associated with them (using construct query `cq`). These queries are issued against a `DataSource` that can be an RDF graph (`RDFSource`), or the result of a transformation applied to non RDF source (`NRDFSource`), or possibly other implementation-specific types of data sources, such as an authenticated call to a Web service. A `ContainerMap` may have a couple more components: a set of `ContainerMaps` and a set of `NonContainerMaps` that define the resources that must be created as members of the newly created containers.

While these explanations form the core of the language, a few more details are required to understand how design documents work. The query patterns and construct queries in `ResourceMaps` may contain special reserved variables. These are written $\{\rho, \nu, \pi_1, \pi_2, \ldots, \pi_n, \ldots\}$ and are infinite in number. As said, each resource created on the LDP is associated with a resource found in the data sources. We call the resource from the data sources associated with an LDPR the *related resource* of the LDPR. The reserved variable ρ serves to identify the related resource, and the query pattern `qp` is projected on this variable to determine what they are for a given `ContainerMap` or `NonContainerMap`. The variable ν is reserved to indicate, in a construct query, the URI of the newly created LDP resource for which we want to define an RDF graph. In some sense, ν resembles the special keyword `this` is Java or C++. Finally, π_1, π_2, etc. are references to the resources related to the "parent" container, the "grand-parent" container, and so on.

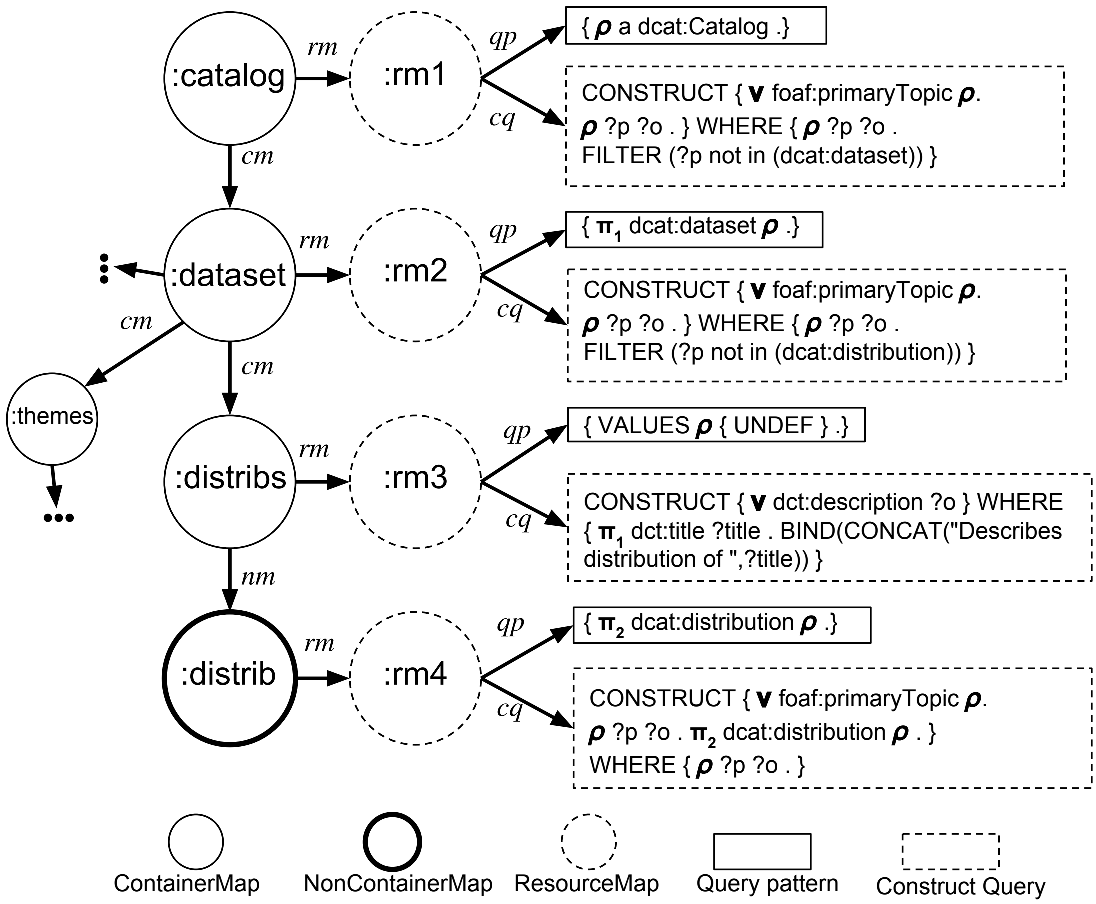


Figure 3.4: A graphical rendition of the abstract syntax of an example LDP-DL design document. From [Bakerally et al., 2018b]

Figure 3.4 shows an example design document following the abstract syntax. One of the query patterns is remarkable: `:rm3` makes use of the special SPARQL keywords `VALUES` and `UNDEF` that make the variable ρ unbound. This is different from having an empty solution set. This allows an LDP resource to exist without having a related resource in the data source. It is especially useful when one wants to group resources in containers that have no corresponding entity in the data sources. It is also useful to provide metadata about the platform that are unrelated to the data sources.

3.3.3 Formal semantics

The goal of the formal semantics is to determine, given data available on a deployed LDP, whether or not such deployment conforms to a given design document. This means that we need an abstract formal structure that describes the content of an LDP. To simplify things a bit, we assume that all data is publicly accessible to everyone, such that we do not have to model access rights and other such parameters. LDP datasets (Definition 5) formalises the state of an LDP in terms of its data content. The semantics allows us to define whether an LDP dataset satisfies an LDP-DL design. In turn, an LDP-DL implementation will be formally conformant if every time it deploys data to an LDP, the resulting LDP dataset satisfies the design document.

Definition 5 (LDP dataset) *An* LDP dataset *is a pair* $\langle \mathbf{NG}, \mathbf{NC} \rangle$ *where* $\mathbf{NG}$ *is a set of named graphs and* $\mathbf{NC}$ *is a set of* named container*, that is a set of triples* $\langle n, g, M \rangle$ *such that* $n \in \mathbf{IRI}$ *(called the container name),* g *is an RDF graph and* $M \in 2^{\mathbf{IRI}}$*. In addition to this, there are some constraints on an LDP dataset. In an LDP dataset* Σ*:*

- *no IRI appears more than once as a (graph or container) name;*
- *for all* $\langle n, g, M \rangle \in \mathbf{NC}$*, and for all* $u \in M$*, there exists a named graph or container having the name* u *in* Σ*.*

An LDP dataset D can be seen as the representation of the data on an LDP server by understanding it as follows: if a pair $\langle n, g \rangle$ or triple $\langle n, g, M \rangle$ belongs to D, then if an HTTP `GET` request for n is sent to the server, then it must respond with a representation of the RDF graph g, and an indication that the resource is supporting interactions according to the LDP standard (as recommended in [Speicher et al., 2015, Section 4.2.1.4]). Moreover, if a triple $\langle n, g, M \rangle$ belongs to D, the server must respond to a request for n with an indication that the resource is an LDP container (as recommended in [Speicher et al., 2015, Section 5.2.1.4]) and the resources that are members of this container are exactly those identified by the IRIs in M.

Having this notion of LDP dataset seems to suggest that we can build a specific instance of such dataset as a result of parsing and traversing the structure of an LDP-DL design document. However, this is not so straightforward as the LDP standard is quite permissive in terms of what and how information must be provided. Moreover, we want the validity of an LDP dataset to be independent of the specificities of the deployed platform, in terms of location (what URLs it uses) and implementation (any standard-compliant LDP implementation should be usable).

As a result, we defined a model theory for LDP-DL with a rather unorthodox model structure. However, the choices are not arbitrary. With this approach, it is possible to define a formal interpretation of parts of a design document independently of where the part occurs (for instance, a `ContainerMap`), because the same

Map could be reused in the context of different design documents. Moreover, the interpretation of Maps is relative to the hierarchy of containers in which the resources are created.

The details of the formal semantics are not much important for the present dissertation and can be found in the ESWC 2018 paper [Bakerally et al., 2018b] and with more details and examples in Noorani Bakerally's thesis [Bakerally, 2018].

3.4 Implementation

Noorani Bakeralli implemented the complete workflow for LDP generation and deployment. However, I will only insist on some aspects of the implementation. An overview of the complete set of implemented components is shown in Figure 3.5.

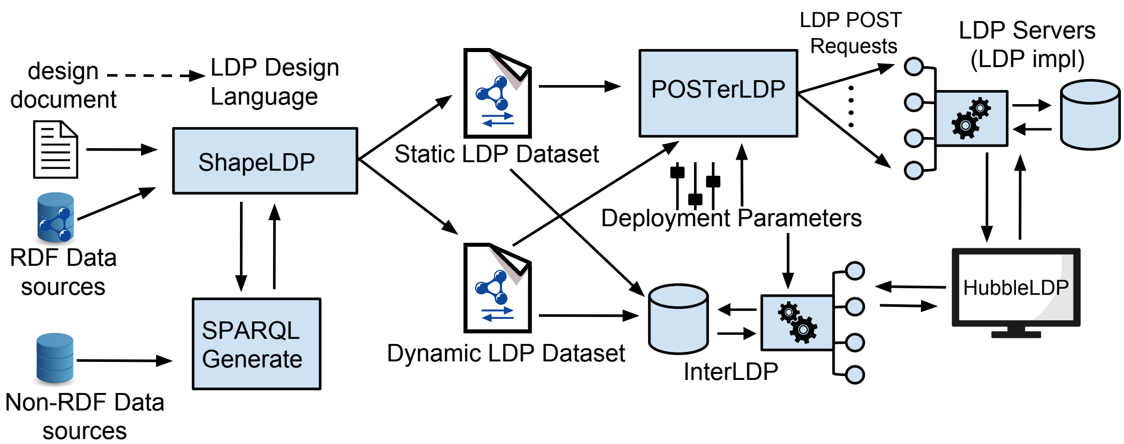


Figure 3.5: Overview of LDP Generation Toolkit. From Noorani Bakerally's thesis [Bakerally, 2018]

First, the concrete syntax for LDP-DL is RDF, based on a vocabulary.[3] As an RDF graph can easily be split into multiple RDF graphs and recombined by a standard merge operation, it is possible to modularise LDP-DL documents. The LDPizer takes a set of RDF files as input, merge them and interpret them as a single design document. One module can contribute to multiple designs.

Second, the data sources can include non-RDF data because a SPARQL-Generate transformation can be referenced as part of a DataSource. The transformation is given by its URI, so it can be hosted remotely. Therefore, LDP-DL can manage arbitrary heterogeneous data.

Third, the main component that serves as an engine for LDP-DL, ShapeLDP,[4] can generate two types of intermediary output. A *static LDP dataset* is very similar to an LDP dataset as defined in Definition 5 except that the IRIs of LDP resources are stored in relative form. At deployment phase, the relative IRIs are used as indicative names to the server, which can simply append a prefix (the base URI of the LDP) or create a different identifier. Either way, it does not impact the

[3]LDL Design Language vocabulary. `https://w3id.org/ldpdl/#`. Currently, this URL is not working due to changes in servers at école des mines. (**TODO**) The page is also available at `https://github.com/noorbakerally/LDPDesignLanguageSpecificationHTMLPage`, retrieved 25 July 2020.

[4]Source avilable at `https://github.com/noorbakerally/ShapeLDP`, retrieved 25 July 2020.

validity of the deployment with respect to the formal semantics. A *dynamic LDP dataset* differs from an LDP dataset by not storing any RDF graph representing the resources. Instead, it materialises a `CONSTRUCT` query from the template given in a `ResourceMap` (replacing the reserved variables with their values) and a reference to the data source that must be used to execute the query.

```
# description of LDP resources
<catalog> {
  <catalog> a ldp:BasicContainer;
    ldp:contains <animations> .

  # graph description provides compiled CONSTRUCT query
  # and data source on which query must be executed
  <catalog> ldl:graphDescription [
    ldl:graphQuery "CONSTRUCT {
        <https://bistrotdepays.opendatasoft.com/api/v2/catalog/exports/ttl>
        ?p ?o . } WHERE {
        <https://bistrotdepays.opendatasoft.com/api/v2/catalog/exports/ttl>
        ?p ?o FILTER ( ?p NOT IN (<http://www.w3.org/ns/dcat#dataset>) ) }
        " ;
    ldl:dataSource <DataSource1>
  ] .
  <DataSource1> a ldl:DataSource;
    ldl:location
        "http://bistrotdepays.opendatasoft.com/api/v2/catalog/exports/ttl"
        .
}
```

Figure 3.6: A sample code for a dynamic LDP dataset

If someone wants to use an existing, running implementation of the LDP standard, the LDP dataset can be deployed there using PosterLDP,[5] a tool in charge of creating the appropriate resources and containers using HTTP interactions as defined by the LDP standard. Instead, one can use Noorani's implementation of a read-only LDP, InterLDP,[6] that directly consumes (static or dynamic) LDP datasets. Finally, the data in an LDP can be navigated in interactively in a Web browser user HubbleLDP.[7]

3.5 Summary

This work was made in collaboration with Noorani Bakerally (Phd student at Mines Saint-Étienne) and Olivier Boissier (Professor at Mines Saint-Étienne), in the context of project OpenSensingCity. My publications that relates to this work are: [Bakerally et al., 2016, Bakerally and Zimmermann, 2017, Bakerally et al.,

[5]Source available at `https://github.com/noorbakerally/PosterLDP`, retrieved 25 July 2020.
[6]Source code available at `https://github.com/noorbakerally/InterLDP`, retrieved 25 July 2020.
[7]Source code available at `https://github.com/noorbakerally/HubbleLDP`, retrieved 31 August 2020.

2018b, Bakerally et al., 2018a]. Figure 3.7 describes how it instantiates the general idea of a loosely coupled mediation approach that I am defending in this thesis.

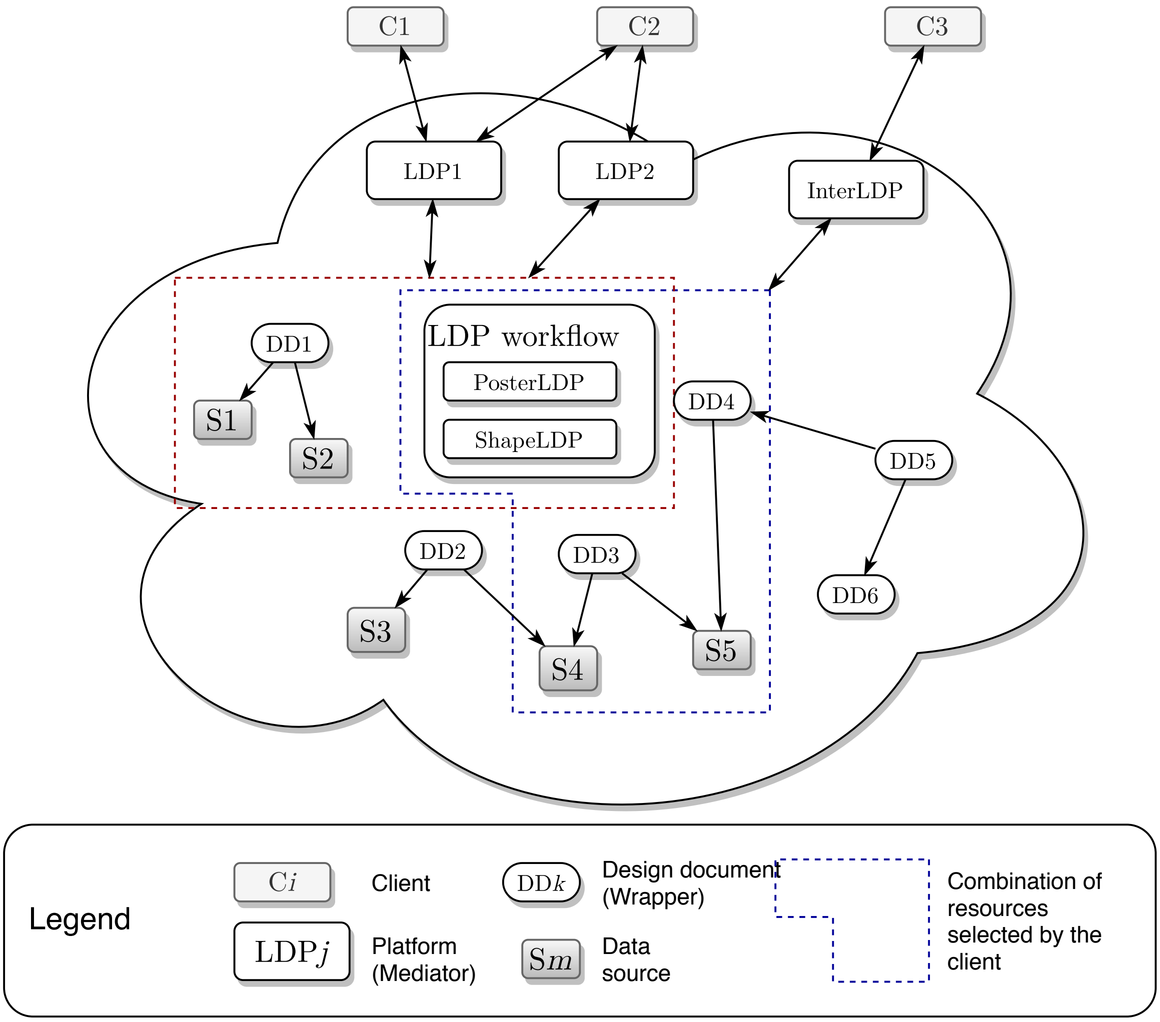


Figure 3.7: LDP-DL language and modules as an instance of loosely coupled mediation

The information sources (S1 to S5 in Figure 3.7) are existing data that a data provider wants to expose as Linked Data in a systematic way. LDP servers (LDP1, LDP2 or InterLDP) can be seen as mediators that enable homogeneous access. In this case, the LDP-DL workflow and components constitute the wrapper. However, if the data provider is the target user of this work, we can rather see the LDPizer and LDP dataset deployer as a mediator for an installed LDP server, that wraps the data according to an LDP-DL document (DDk). Design documents may also reference other design documents, modularly (see DD5). Loose coupling is ensured by having a generic engine for a declarative language, such that the wrapper (a specific design for specific data sources) can be stored independently of the mediator.

Summary and future work

The three contributions correspond to three ways an information system can make its data available to a client: the first corresponds to a data export where the client is free to process it, integrate it, index it, however it wants; the second corresponds to providing access to the data via a query language, allowing the client to flexibly extract precise pieces of data while relieving it from the burden of indexing, optimisation, storage, etc.; the third corresponds to exposing the data in an organised interface that is easily navigable, much like a web site may expose the content of a database in a readable, hyperlinked way.

In each case, there exist distributed and heterogeneous data sources that need to exported, exposed, or interrogated. Each contribution allows access to the data with a single interface, while the actual data or request flow go through a wrapping mechanism that relies on a declarative language, which enable the flexible loose coupling that I am advocating. This is illustrated in Figure 3.8.

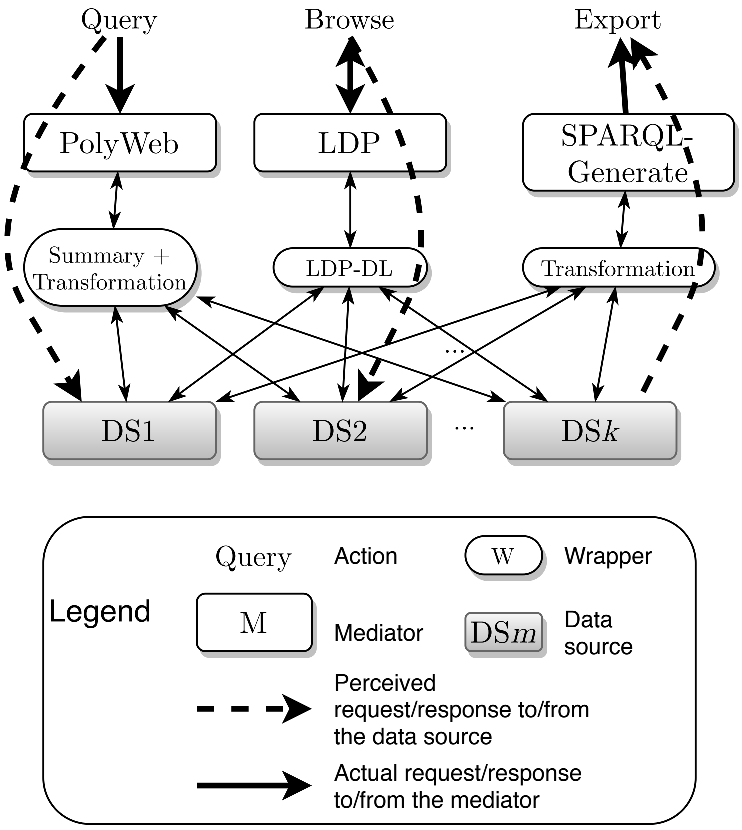


Figure 3.8: Three data access methods in a loosely coupled mediation architecture

The three contributions are interdependent: LDP-DL make use of SPARQL-

Generate in order to exploit both RDF and non-RDF sources; LDP-DL also rely on SPARQL to extract relevant data pieces, so it can exploit federate SPARQL systems as well, including PolyWeb. PolyWeb relies on transformations that could be written in SPARQL-Generate (although not for the moment).

In each case, there are limitations and possible improvements, sometimes challenging issues, that could be investigated further:

- In the case of SPARQL-Generate, a lot of improvements have been made to the language and its implementation since our original publication [Lefrançois et al., 2017c]. Some of the improvements to the language have repercussions on the semantics of the language, which should be formalised (handling of streams, assignment of multiple variables at once, generation of non-RDF data, etc.). Besides, a stronger user study should be made to validate our hypothesis that SPARQL-Generate is easier and more efficient to use by Semantic Web experts than other data transformation solutions. Beyond SPARQL-Generate, the problem of translating arbitrary data to RDF is offering other challenges: the bottleneck is mainly the selection of the right vocabulary and decisions on the shape of the target data. This later task is related to knowledge modelling and overlaps with what is discussed in the next part.

- In the case of PolyWeb, we observed strong limitations in terms of performance, and the current prototype is supporting very few data models. The translation of SPARQL queries into queries to tree-based structures (XML or JSON) is one of the challenges. Investigating how SPARQL-Generate could be used to transform the queries rather than the data is another open issue. In the general case of federated SPARQL queries, the choice of an adequate data summary, both precise enough to help deciding which data source to query and simple enough to make it easily computable, is an important research area.

- In the case of LDP-DL, we only scratched the surface of automating Linked Data deployment. For the moment, only basic Linked Data Platform containers are supported, while the LDP standard provides also Direct LDPC and Indirect LDPC, which are noticeably more complex. Additionally, the focus has been put on open data, such that access rights and authentication did not need to be tackled. Also, LDP-DL requires strong technical knowledge that could be made easier by introducing templates for common data deployment schemes. Extensions should support paging, which consists in providing a maximum number of resources in container, with a link to the next batch of resources if there are more than the maximum, as is commonly done in Web sites where lists of links are provided with a fixed number at a time. Finally, although the LDP standard formalises some of the Linked Data principles, it is not possible to host a container on an LDP which contain a resource not described in the same platform.

Part II

Knowledge Interoperability via Context

Outline of the contributions

Once data heterogeneity has been addressed, there is still interoperability problems due to divergence in how data is interpreted. Interpretation is what turns data into knowledge, but if the interpretation of those who emit data differ from those who receive them, then false conclusions may enfold. The definition of a formal semantics for a data model helps in this regard, as it gives a common ground on what can be concluded or not, independently of the intended interpretation. However, in an environment of multiple, independent information sources, this is not sufficient.

In order to interoperate, systems have to agree on the abstract concepts that must be represented, instantiated and related. This is the part related to ontology engineering, where an ontology is a formal description of a domain of knowledge, typically consisting of a vocabulary of terms that denote things, classes of things, or types of relations between things, and of logical axioms that constrain how to interpret the terms.

Ontologies may vary much in their vocabularies and axioms according to viewpoints, requirements, domains of applications, and so on. This is a barrier to interoperability. Additionally, the knowledge bases instantiating the concepts of ontologies may describe different "realities" according to their context. This part presents my contributions about enabling interoperability at the knowledge layer, that partly overcome these issues:

- First, I will briefly discuss my work related to my PhD thesis, formalising the structure and semantics of a network of aligned ontologies. This contribution allows one to consider a set of independent ontologies, connected together by (possibly third party's) ontology alignments, as a single unified knowledge model. This work allows me to introduce the research challenge addressed in Chapter 4, where we investigate how to exploit multiple networks of aligned ontologies, leading to multi-level networked knowledge supporting a hierarchical modularisation.

- Second, when knowledge is instantiated from within different contexts, it is sometimes possible to explicitly and exactly represent the context on top of the knowledge in the form of annotations. In this case, reasoning across contexts is made possible by the computation of algebraic operations over annotations, while retaining much of the reasoning procedure of the underlying knowledge formalism, as presented in Chapter 5.

- Third, I address the problem of context representation differently than in the second contribution: instead of relying on a new formalism of annotated knowledge, which is not standard, I present how context can be represented and reasoned with within the boundaries of standard knowledge representation formalisms RDF and OWL (Chapter 6).

As a result of these contributions, we can assume that clients have a unified view of all knowledge, whether at a general ontological level, or at particular factual knowledge, that they can reason with all together. As in the first part, at the end of this one, I summarise the contributions putting them in perspective of future research that they enable.

The work presented in this part results from: my initial research as a PhD student and its continuation in Sihem Klai's PhD thesis, whom I jointly supervised with Prof. Md. Tarek Khadir (Chapter 4); my work as a post doctorate researcher in Galway, thanks to a collaboration started in the context of COST Action Agreement Technologies, in which I was co-leading a work package on semantics (Chapter 5); the supervision of PhD student José M. Giménez García, in the context of H2020 ITN project WDAqua, jointly with Prof. Pierre Maret (Chapter 6).

Chapter 4

Ontologies and Alignments

Ontologies are ways to make explicit how states of affair relating to a particular domain must be described, by abstracting away from the *particulars* and stating knowledge about the *universals*. But ontologies are of little use if they are not *shared*, reused and agreed upon by many actors, as emphasised by multiple authors [Borst, 1997, Studer et al., 1998, Guarino et al., 2009].

In a Web that has reached common agreement at the data level – say, by adopting RDF as the data model of choice for every data exchange – the possibility to disagree at the knowledge level is still a barrier to interoperabililty. Consider Figure 4.1 that depicts two possible models for representing the notion of role. These diagrams should be understood as follows: on the left-hand side, it is specified that a person holds a role during an event (a relation `holdsRole` connects an instance of `Person` to an instance of `RoleDuringEvent`), while a *role during event* exists with a role (relation `withRole` connects a `RoleDuringEvent` to a `Role`) and for the duration of an event (relation `during` connects a `RoleDuringEvent` to an `OrganisedEvent`). This is the model defined for the ScholarlyData project,[1] that aims at providing semantic web compatible data about semantic web research, including the description of semantic web events and their organisation.

On the right-hand side, we describe the membership of a person to an organisation with a certain role at a certain time. More precisely, an instance of `Membership` is connected to:

- an instance of `Person` with relation `member`;
- an instance of `Role` with relation `role`;
- an instance of `Organization` with relation `organization`;
- an instance of a temporal entity, such as `time:Interval`, with relation `memberDuring`.

This second model is what the W3C recommends in its Organization Ontology [Reynolds, 2014]. Using these two models to represent the fact that Antoine Zimmermann was a programme committee member of the International Semantic Web Conference 2020 can be done with the graphs in Figure 4.2.

If one wants to use both ontologies, or bring together and reason with data conforming to each of the two models, it would be relevent to make explicit that the two concepts of `Role` correspond to each other. This is what ontology alignments

[1] `http://www.scholarlydata.org/`, retrieved 25 July 2020.

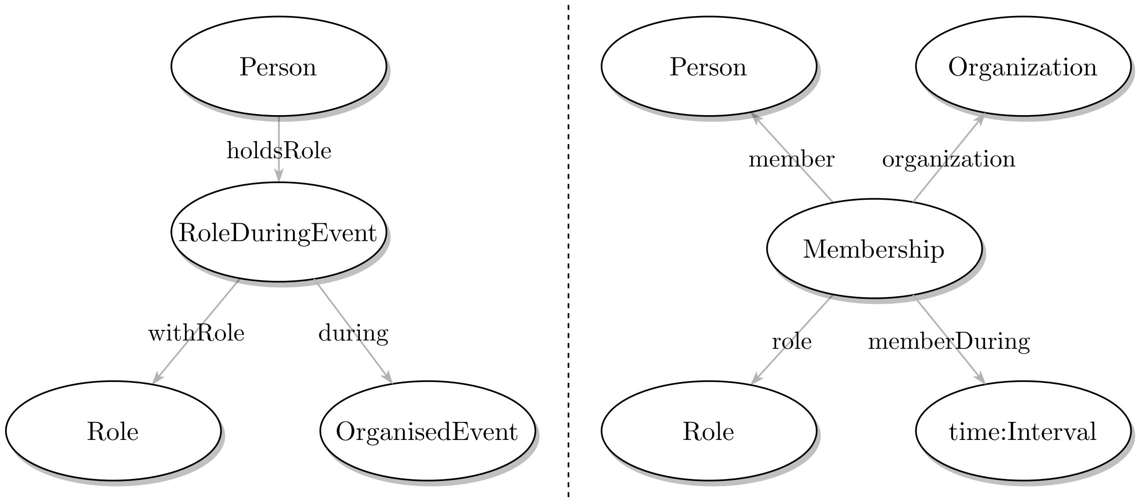


Figure 4.1: Comparison of two models for representing roles. On the left side, the model used by the Conference Ontology. On the right, the model followed by the Organization Ontology.

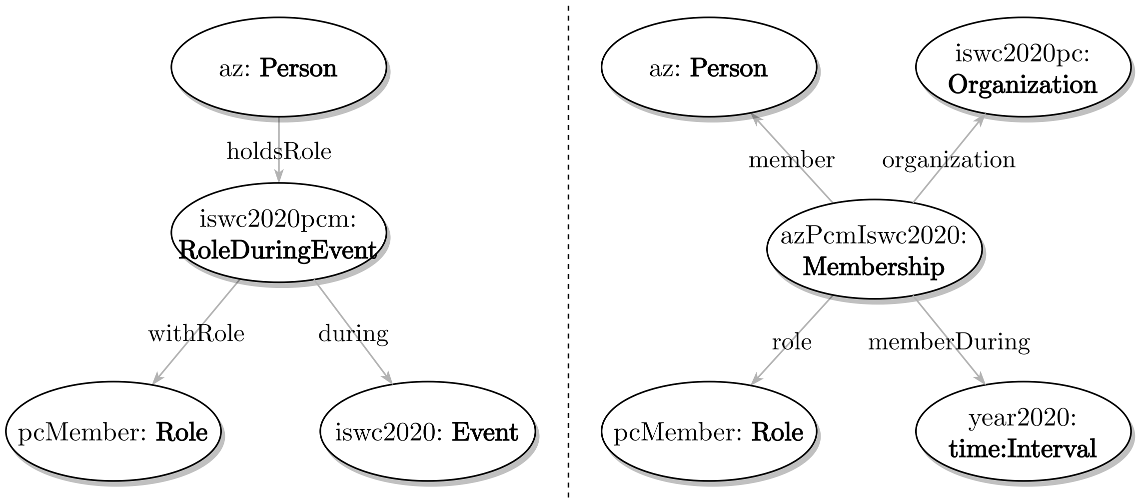


Figure 4.2: Instantiating the two models of Figure 4.1

are: a set of correspondences between pairs of ontologies. In this chapter, I describe my work related to the semantics of networks of aligned ontologies which highlight the idea that ontology alignments are the tools that mediate between local views of the world (namely, individual ontologies) and a global, unified knowledge (the ontology network). The first section is related to my PhD thesis and a few additional contributions as a postdoc. The second section describes the work done with Sihem Klai under my supervision on multi-level networks of ontologies. I keep these sections brief and concise because they are not part of the research directions that I want to pursue in the future.

4.1 Semantics of networks of ontologies

There are different ways of exploiting multiple ontologies: importing them, merging them, integrating them after updates. However, when two ontologies describe a common domain of knowledge, a typical task consists in finding correspondences between the terms of the ontoogies. This is called ontology matching and it leads to the construction of ontology alignments. In my PhD thesis, I considered ontology alignments as first class citizens in knowledge representation, giving them a complementary role to ontologies [Zimmermann, 2008]. I described operations on alignments in category-theoretical terms [Zimmermann et al., 2006], I presented an abstract categorisation of the semantics of alignments [Zimmermann and Euzenat, 2006], distinguishing between alignments as mere axioms (simple semantics), alignments as *bridge rules* in the sense of Giunchiglia's contextual reasoning [Giunchiglia, 1993] (distributed semantics, a category in which fall DFOL [Ghidini and Serafini, 1998], DDL and COWL [Borgida and Serafini, 2003, Bouquet et al., 2003a], E-connection [Kutz et al., 2004], DDL revisited [Homola, 2007], $E^{DDL}_{HQ^{+}}$ SHIQ [Vouros and Santipantakis, 2012]), and alignments as indicating a global viewpoint on the aligned ontologies, distinct from each local viewpoints (integrated semantics, of which IDDL [Zimmermann, 2007] is an instance).

The idea of the integrated semantics that I introduced is that each ontology in a network represents a viewpoint that must be interpreted separately. Additionally, there exists a supplementary viewpoint, which is the one of the network of aligned ontology. Formally, these means that a network of n ontologies is associated with $n+1$ interpretations, with n *local* interpretations, and 1 *global* interpretation. Alignments are constraining how the global interpretation must coincide with the local ones.

By using this type of semantics on description logic ontologies, I could define a reasoning procedure such that local reasoners, assigned to each ontology, could be used as black boxes by a global reasoner that is only aware of the alignments [Zimmermann and Le Duc, 2008b]. By calling the local reasoners with different inputs, the global reasoner is able to establish the global consistency. This way, one can see the global reasoner as a mediator between the user of the ontology network, and the individual ontologies.

What follows are contributions directly extending on my PhD thesis but made as a post doc in Galway:

- In [Le Duc et al., 2010], we described an API for enabling reasoning over a network of aligned ontologies, which allowed different forms of non-standard modular ontology reasoning. The API was implemented for a distributed reasoner described in [Le Duc et al., 2013], with an account of its performances over benchmark ontologies.

- In [Zimmermann et al., 2009] and later [Sahay et al., 2013], we presented a theroretical investigation of the logical formalisms for reasoning on heterogeneous ontologies, in relation to patient records data exchange and inferences. Those formalisms and their effectiveness is highly dependent on the quality of ontology alignments, for which generic automated methods of discovery are rather inefficient. As a result, we proposed a methodology to improve the construction of both ontologies and alignments in the health care domain [Sahay et al., 2011].

4.2 Multi-level network of ontologies

With the formal semantics of networks of aligned ontologies, one can see an ontology network as just another case of logical theory, describing the world with a distinct viewpoint (the global interpretation mentioned before). If we push the notion further, one may question what it means to exploit multiple networks of ontologies: this what we investigated with PhD student Sihem Klai. We proposed the concept of multi-level knowledge network, where ontologies are a first level of so called *knowledge nodes*, then some ontologies are matched, forming networks of aligned ontologies that are second level of knowledge nodes, then again different ontology networks may be grouped by way of higher level *links* (that can be thought of as alignments of ontology networks) that form third level knowledge nodes, and theoretically, higher level knowledge nodes may be aligned or linked to other knowledge nodes to form arbitrarily layered knowledge networks. The formalisation and some algorithms for multi-level knowledge network were published in [Klai et al., 2016a, Klai et al., 2016b], and additional results in [Klai et al., 2019]. More details are given in Sihem Klai's thesis [Klai, 2016].

Now, considering that there are measures existing for reconciling different ontologies, in the following, I will assume agreement was reached on how to model the domain of discourse. Yet, it is still possible to find knowledge disparity when following the same representation because different situations are described according to different viewpoints, or in different context. The notion of context is the main topic of the next two sections.

4.3 Summary

This work was mostly made in collaboration with Jérôme Euzenat (*directeur de recherche* at Inria), in part with Chan Le Duc (postdoc at Inria at the time, now associate professor at Université Paris 8) when I was a PhD student. Later contributions on multi-level networks of ontologies were made in collaboration with Sihem Klai (Annaba University, Algeria). My publications that relates to this work are: [Zimmermann and Euzenat, 2006, Zimmermann, 2007, Zimmermann and Le Duc, 2008b, Zimmermann, 2008, Zimmermann et al., 2009, Le Duc et al., 2010, Le Duc et al., 2013, Sahay et al., 2013, Klai et al., 2016a, Klai et al., 2016b, Klai et al., 2019]. Figure 4.3 describes how it instantiates the general idea of a loosely coupled mediation approach that I am defending in this thesis.

Information sources are ontologies (O1 to O7). Reasoning is made over multiple sources with a reasoning engine for networks of aligned ontologies. Such distributed reasoner constitutes the mediator. In this case, ontology alignments can be seen as wrapper specifications for pairs of ontologies (OA1 to OA6). Note that if alignments are seen as a global view on pairs of ontologies and networks of aligned ontologies, then it is possible to define “auto-alignments” that relate an ontology to itself, such as OA4 in the picture. These alignments provide a different view on the ontologies by making explicit correspondences between terms of the same ontology. Besides, the same pair of ontologies can be aligned in different ways (OA1 and OA2 between ontologies O1 and O2). A client may want to use certain types of alignments more than others, which is made possible by the loose coupling of the components of the mediator architecture.

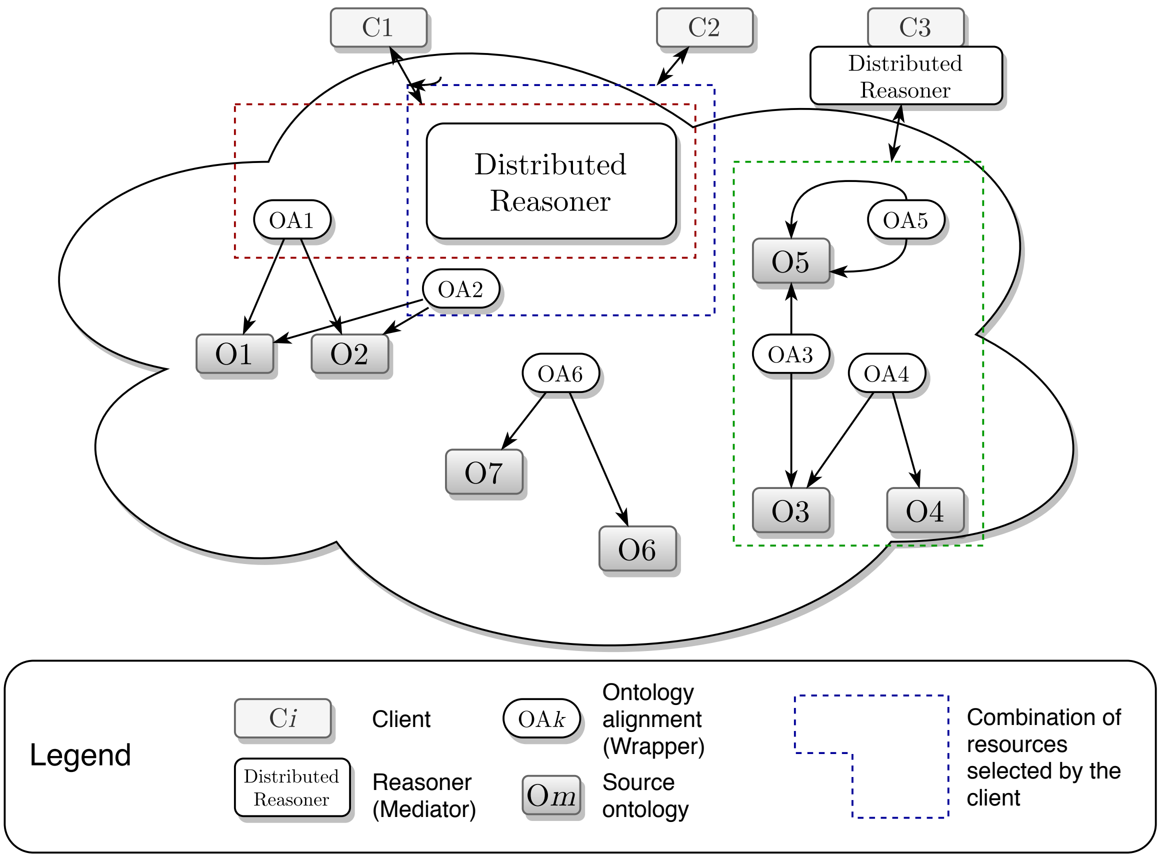


Figure 4.3: Networks of aligned ontologies as an instance of loosely coupled mediation

Chapter 5

Annotated RDFS

With the assumption that ontologies have been aligned, we can suppose that there exists a single unified knowledge model that can be used across applications. This means that the general terms, such as classes and relations, are agreed upon, or at least interrelated explicitly. However, there is still a possibility for knowledge to diverge for the particulars. The same instance can be described differently, depending on viewpoints, perspectives, opinions, epochs, places, in one word *context*. This chapter and the next one deal with this notion in different ways.

In some well defined cases, context can be quantified as a value of a certain algebraic structure. For instance, one can describe the temporal validity of a statement by indicating the set of time points over which it holds true. As time points can be encoded as decimal numbers, or even discretised to integers, and considering that validity holds over a continuous period of time between a beginning and an end point, one can in practice define temporal validity as a pair marking the extremities of a temporal interval. Similarly, quantifying the confidence in a statement can be done by assigning a numerical value to it. Asserting a degree of truth, for fuzzy statements, can follow a similar approach.

Several proposals were made to extend RDF in different ways to allow for reasoning that would include a quantified context for: time [Gutiérrez et al., 2007, Pugliese et al., 2008, Tappolet and Bernstein, 2009], fuzziness [Mazzieri and Dragoni, 2008, Straccia, 2009], probability [Udrea et al., 2006b], provenance [Flouris et al., 2009], trust [Schenk, 2008, Hartig, 2009], or access rights [Lopes et al., 2012] where the representation and reasoning process ultimately rely on the same principles.

Udrea et al. [Udrea et al., 2006a, Udrea et al., 2010] first recognised the similarity of these diverse approaches, pointing out the strong connection between RDF with annotation values (temporal, fuzzy, provenance, etc.) and the formalism of Annotated Logic Programming [Kifer and Subrahmanian, 1992]. Thus, they defined *Annotated RDF*, where annotations (such as timestamps) must belong to a finite partial order. The semantics of Annotated RDF does not consider the specificities of the RDFS vocabulary, with which subclass, subproperty, domain, and range can be expressed.

With DERI colleagues Nuno Lopes and Axel Polleres, and CNR Prof. Umberto Straccia, we developed an improved version of Annotated RDF that we called *Annotated RDFS* [Zimmermann et al., 2012], that provided the following additions:

- It relies on a richer, not necessarily finite structure and provide additional inference capabilities;

- We provide a query language for Annotated RDFS, based on SPARQL 1.1 and including rich features like optional, filters, assignments, aggregates and solution modifiers;
- We formally define *compound annotation domains* that allow to combine multiple annotations (such as temporal and provenance) in the same framework.

I explain each of the previous items in the next subsections.

5.1 RDFS with annotations

Here, I formally define the logic Annotated RDFS with its syntax and semantics. I slightly modify the definitions given in [Zimmermann et al., 2012] in order to make Annotated RDFS compatible with RDF 1.1 [Cyganiak et al., 2014, Hayes and Patel-Schneider, 2014], which was standardised after our paper got published. These updates do not change the formalism significantly.

5.1.1 Syntax

The formalism of Annotated RDFS assumes the same basic constructs as RDF: an infinite set $\mathbf{I}$ of IRIs, and infinite set $\mathbf{L}$ of literals, and infinite set $\mathbf{B}$ of blank nodes, each being disjoint from the two others. We define an RDF triple as an element of $\mathbf{B}\cup\mathbf{I}\cup\mathbf{L}\times\mathbf{I}\times\mathbf{B}\cup\mathbf{I}\cup\mathbf{L}$, therefore allowing literals in subject position, contrary to the RDF standard. In $\mathbf{I}$, we distinguish 5 IRIs that play a special role in the semantics: $\{\mathsf{type}, \mathsf{sc}, \mathsf{sp}, \mathsf{dom}, \mathsf{range}\}$. These IRIs are taken from the ρdf [Muñoz et al., 2007] fragment of the RDFS vocabulary, which is considered the "essential" part of the RDFS formalism. type is instance-of relation, sc is the subclass relation, sp is the subproperty relation, dom is the domain and range is range.

Additionally, we define the notion of an annotation domain:

Definition 6 (Annotation domain) *An* annotation domain *for RDFS is an idempotent, commutative semi-ring*

$$D = \langle L, \oplus, \otimes, \bot, \top \rangle \ ,$$

where $\oplus$ is $\top$-annihilating. That is, for $\lambda, \lambda_i \in L$

1. *$\oplus$ is idempotent, commutative, associative;*
2. *$\otimes$ is commutative and associative;*
3. *$\bot \oplus \lambda = \lambda$, $\top \otimes \lambda = \lambda$, $\bot \otimes \lambda = \bot$, and $\top \oplus \lambda = \top$;*
4. *$\otimes$ is distributive over $\oplus$, i.e., $\lambda_1 \otimes (\lambda_2 \oplus \lambda_3) = (\lambda_1 \otimes \lambda_2) \oplus (\lambda_1 \otimes \lambda_3)$;*

Given an annotation domain Δ, an *annotated triple* over Δ is an expression $\tau : \alpha$ where τ is an RDF triple and $\alpha \in \Delta$. An *annotated graph* is a finite set of annotated triples. A *ground* annotated graph is an annotated graph that does not contain any blank node. Finally, we call a *map* a function $\mu : \mathbf{B} \to \mathbf{I} \cup \mathbf{B} \cup \mathbf{L}$ and for an annotated graph G, we denote by $\mu(G)$ an annotated graph obtained from G by replacing blank nodes b in it by $\mu(b)$.

5.1.2 Semantics

The semantics of Annotated RDFS is defined following a classic model-theoretic approach: we start with a notion of interpretation, followed by the notion of satisfaction, which allows us to define models and entailment as usual.

Let us assume $D = \langle L, \oplus, \otimes, \bot, \top \rangle$ is an annotation domain. In a nutshell, an interpretation $\mathcal{I}$ in Annotated RDFS assigns to a triple τ an element $\lambda \in L$ of the annotation domain, dictating that under $\mathcal{I}$, the annotation of τ is greater or equal than (*i.e.*, $\succeq$) λ. Formally, an *annotated interpretation* is defined as follows:

Definition 7 (Annotated interpretation) *An* annotated interpretation *is a tuple* $\mathcal{I} = \langle \Delta_R, \Delta_P, \Delta_C, P[\![\cdot]\!], C[\![\cdot]\!], \cdot^{\mathcal{I}} \rangle$ *such that:*

1. Δ_R *is a nonempty finite set, called the domain or universe of* $\mathcal{I}$ *(the set of* resources*);*

2. Δ_P *is a nonempty finite set, not necessarily disjoint from* Δ_R *(the set of* properties*);*

3. $\Delta_C \subseteq \Delta_R$ *is a distinguished subset of* Δ_R *(the set of* classes*);*

4. $P[\![\cdot]\!]$ *maps each property* $p \in \Delta_P$ *into a function* $P[\![p]\!] : \Delta_R \times \Delta_R \to L$*, i.e., assigns an annotation value to each pair of resources;*

5. $C[\![\cdot]\!]$ *maps each class* $c \in \Delta_C$ *into a function* $C[\![c]\!] : \Delta_R \to L$*, i.e., assigns an annotation value representing class membership in* c *to every resource;*

6. $\cdot^{\mathcal{I}}$ *maps each* $t \in \mathbf{I}$ *into a value* $t^{\mathcal{I}} \in \Delta_R \cup \Delta_P$ *and each* $t \in \mathbf{L}$ *into a value* $t^{\mathcal{I}} \in \Delta_R$*.*

The main difference between this definition and the one from [Zimmermann et al., 2012] is that an interpretation here is not defined with respect to a vocabulary, that is, all IRIs and all literals are interpreted in the universe of $\mathcal{I}$.

An interpretation $\mathcal{I}$ is a *model* of an annotated ground graph G when it satisfies the following conditions:

Definition 8 (Satisfaction) *An annotated interpretation* $\mathcal{I}$ satisfies *an annotated ground graph* G*, denoted* $\mathcal{I} \models G$*, iff:*

Simple:

1. $(s, p, o)\colon \lambda \in G$ *implies* $p^{\mathcal{I}} \in \Delta_P$ *and* $P[\![p^{\mathcal{I}}]\!](s^{\mathcal{I}}, o^{\mathcal{I}}) \succeq \lambda$*;*

Subproperty:

1. $P[\![\mathsf{sp}^{\mathcal{I}}]\!](p, q) \otimes P[\![\mathsf{sp}^{\mathcal{I}}]\!](q, r) \preceq P[\![\mathsf{sp}^{\mathcal{I}}]\!](p, r)$*;*
2. $P[\![p^{\mathcal{I}}]\!](x, y) \otimes P[\![\mathsf{sp}^{\mathcal{I}}]\!](p, q) \preceq P[\![q^{\mathcal{I}}]\!](x, y)$*;*

Subclass:

1. $P[\![\mathsf{sc}^{\mathcal{I}}]\!](c, d) \otimes P[\![\mathsf{sc}^{\mathcal{I}}]\!](d, e) \preceq P[\![\mathsf{sc}^{\mathcal{I}}]\!](c, e)$*;*
2. $C[\![c^{\mathcal{I}}]\!](x) \otimes P[\![\mathsf{sc}^{\mathcal{I}}]\!](c, d) \preceq P[\![d^{\mathcal{I}}]\!](x)$*;*

Typing I:

1. $C[\![c]\!](x) = P[\![\mathsf{type}^{\mathcal{I}}]\!](x, c)$;
2. $P[\![\mathsf{dom}^{\mathcal{I}}]\!](p, c) \otimes P[\![p]\!](x, y) \preceq C[\![c]\!](x)$;
3. $P[\![\mathsf{range}^{\mathcal{I}}]\!](p, c) \otimes P[\![p]\!](x, y) \preceq C[\![c]\!](y)$;

Typing II:

1. *for each* $\mathsf{e} \in \{\mathsf{type}, \mathsf{sc}, \mathsf{sp}, \mathsf{dom}, \mathsf{range}\}$, $\mathsf{e}^{\mathcal{I}} \in \Delta_P$;
2. $P[\![\mathsf{sp}^{\mathcal{I}}]\!](p, q)$ *is defined only for* $p, q \in \Delta_P$;
3. $C[\![\mathsf{sc}^{\mathcal{I}}]\!](c, d)$ *is defined only for* $c, d \in \Delta_C$;
4. $P[\![\mathsf{dom}^{\mathcal{I}}]\!](p, c)$ *is defined only for* $p \in \Delta_P$ *and* $c \in \Delta_C$;
5. $P[\![\mathsf{range}^{\mathcal{I}}]\!](p, c)$ *is defined only for* $p \in \Delta_P$ *and* $c \in \Delta_C$;
6. $P[\![\mathsf{type}^{\mathcal{I}}]\!](s, c)$ *is defined only for* $c \in \Delta_C$.

Intuitively, a triple $(s, p, o)\colon \lambda$ is satisfied by $\mathcal{I}$ if (s, o) belongs to the extension of p to a "wider" extent than λ. For the temporal context, this would mean that the statement denoted by the triple is true during *at least* the timeframe denoted by the temporal annotation. For a provenance context, this would mean that the statement is true *at least* according to the source denoted by the annotation.

Entailment among annotated *ground* graphs G and H is as usual, that is G entails H iff all models of G are models of H (which we denote $G \models H$). Now, $G \models H$, where G and H may contain blank nodes, iff for any grounding G' of G there is a grounding H' of H such that $G' \models H'$, where a grounding of an annotated graph G is a ground annotated graph obtained from G by replacing blank nodes with elements of $\mathbf{I} \cup \mathbf{L}$.

5.1.3 Deductive system

The deductive system that we proposed is based on the one proposed for ρdf [Muñoz et al., 2007], but the technique for devising a deductive system for an annotated version of an entailment regime is general enough to be extended to any system that relies on deduction rules. Moreover, and most importantly, the rules are the same for any annotation domain, as long as they provide the two operators $\otimes$ and $\oplus$.

The rules are arranged in groups that capture the semantic conditions of models, A, B, C, X and Y are meta-variables representing elements in $\mathbf{I} \cup \mathbf{B} \cup \mathbf{L}$ and D, E represent elements in $\mathbf{I} \cup \mathbf{L}$. The rule set contains two rules, $(1a)$ and $(1b)$, that are the same as for the crisp case, while rules $(2a)$ to $(5b)$ are the annotated rules homologous to the crisp ones. Finally, rule (6) is specific to the annotated case.

For practical implementations, rule (6) should be destructive, that is, when applied, the premises should be removed as the conclusion is inferred. We also must assume that a rule is not applied if the consequence is of the form $\tau\colon \bot$ because all annotated triples $\tau\colon \bot$ are tautologies in our framework.

1. Simple:

   (a) $\frac{\mu(G)}{G}$ for any map μ

(b) $\dfrac{(X, D, Y)\colon \lambda_1}{(X, D, Y)\colon \lambda_2}$ for $\lambda_1 \succeq \lambda_2$

2. Subproperty:

(a) $\dfrac{(A, \mathsf{sp}, B)\colon \lambda_1, (B, \mathsf{sp}, C)\colon \lambda_2}{(A, \mathsf{sp}, C)\colon \lambda_1 \otimes \lambda_2}$

(b) $\dfrac{(D, \mathsf{sp}, E)\colon \lambda_1, (X, D, Y)\colon \lambda_2}{(X, E, Y)\colon \lambda_1 \otimes \lambda_2}$

3. Subclass:

(a) $\dfrac{(A, \mathsf{sc}, B)\colon \lambda_1, (B, \mathsf{sc}, C)\colon \lambda_2}{(A, \mathsf{sc}, C)\colon \lambda_1 \otimes \lambda_2}$

(b) $\dfrac{(A, \mathsf{sc}, B)\colon \lambda_1, (X, \mathsf{type}, A)\colon \lambda_2}{(X, \mathsf{type}, B)\colon \lambda_1 \otimes \lambda_2}$

4. Typing:

(a) $\dfrac{(D, \mathsf{dom}, B)\colon \lambda_1, (X, D, Y)\colon \lambda_2}{(X, \mathsf{type}, B)\colon \lambda_1 \otimes \lambda_2}$

(b) $\dfrac{(D, \mathsf{range}, B)\colon \lambda_1, (X, D, Y)\colon \lambda_2}{(Y, \mathsf{type}, B)\colon \lambda_1 \otimes \lambda_2}$

5. Implicit Typing:

(a) $\dfrac{(A, \mathsf{dom}, B)\colon \lambda_1, (Dspp, A)\colon \lambda_2, (X, D, Y)\colon \lambda_3}{(X, \mathsf{type}, B)\colon \lambda_1 \otimes \lambda_2 \otimes \lambda_3}$

(b) $\dfrac{(A, \mathsf{range}, B)\colon \lambda_1, (D, \mathsf{sp}, A)\colon \lambda_2, (X, D, Y)\colon \lambda_3}{(Y, \mathsf{type}, B)\colon \lambda_1 \otimes \lambda_2 \otimes \lambda_3}$

6. Generalisation:

(a) $\dfrac{(X, A, Y)\colon \lambda_1, (X, A, Y)\colon \lambda_2}{(X, A, Y)\colon \lambda_1 \oplus \lambda_2}$

This deductive system leads to a transitive relation $\vdash$ between annotated graphs such that $G \vdash G'$ for all $G' \subseteq G$ and whenever $G \vdash G'$ then $G \vdash G' \cup c$ for all conclusions c of the rules that can be applied to G'. From this, we can assert the following proposition:

Proposition 1 (Soundness and completeness) *For any annotated graphs G and G':*

- *if $G \vdash G'$ then $G \models G'$;*
- *if $G \models G'$ and G' does not contain the annotation $\bot$, then $G \vdash G'$.*

As can be seen, a condition on the entailed graph is added, which means that the deductive system is not strictly complete. Indeed, an annotated triple $(s, p, o)\colon \bot$ where s or o do not appear in G would not be derived from applying the rules to G. Yet, any triple annotated with $\bot$ is logically entailed by any annotated graph. However, such annotation is in practice uninteresting, as it says nothing about the extent to which the triple is true.

Interestingly, rules $2-5$ above can be represented concisely using the following meta-inference rule:

$$(AG) \quad \frac{\tau_1 \colon \lambda_1, ..., \tau_n \colon \lambda_n, \{\tau_1, \ldots \tau_n\} \vdash_{\mathsf{RDFS}} \tau}{\tau \colon \bigotimes_i \lambda_i}$$

Essentially, this rule says that if a classical RDFS triple τ can be inferred by applying a classical RDFS inference rule to triples $\tau_1, \ldots \tau_n$ (denoted $\{\tau_1, \ldots, \tau_n\} \vdash_{\mathsf{RDFS}} \tau$), then the annotation term of τ will be $\bigotimes_i \lambda_i$, where λ_i is the annotation of triple τ_i. It follows immediately that, using rule (AG), in addition to rules (1) and (6) from the deductive system above, it is easy to extend these rules to cover complete RDFS or even a subset of OWL, such as pD* [ter Horst, 2005].

5.1.4 Concrete annotation domains

Here are some examples of concrete annotation domains. For each presented domain, I provide additional remarks that are specific to it.

The temporal domain

Temporal annotations can be thought of as the time when a statement is true. Abstractly, we can model time as a set of time instants. The fact that an statement is true at some instants can be represented by annotating the statement with a set of instants. This makes the definition of the domain very simple:

Definition 9 *The* temporal domain *is the tuple* $\langle 2^T, \cup, \cap, \emptyset, T \rangle$ *where* T *is a set of time instants and the other components are the usual set-theoretic constructs.*

This definition is theoretically convenient: it concisely conforms to the notion of annotation domain, it only uses well known mathematical constructs, and it is independent of a chosen theory of time, as long as time is modelled as a set of instants. Time could be continuous, discrete, bounded (with a beginning of time and an end of the world) or even be a multi-dimensional topology. However, for practical reasons, it is usually assumed that a statement that holds true at some point in time is either always true, or starts and/or ends to be true at some point while staying true all the time in between. In other words, we can usually think of temporal validity as an interval of time points.

Consequently, we consider time points that can be represented as decimal values (so that it can be encoded as literals of type `xsd:dateTimeStamp`) in a set D. A temporal interval is a pair $\langle t_1, t_2 \rangle$ such that $t_1 \in D \cup \{-\infty\}$ and $t_2 \in D \cup \{+\infty\}$ and whenever $t_1, t_2 \in D$, $t_1 \leq t_2$. Temporal intervals are not sufficient to represent temporal annotations, because if, for instance, the annotated triples $\texttt{GroverCleveland}, \mathsf{type}, \texttt{POTUS} \colon [1885, 1989]$ and $\texttt{GroverCleveland}, \mathsf{type}, \texttt{POTUS} \colon [1893, 1897]$ hold, then we could conclude that $\texttt{GroverCleveland}, \mathsf{type}, \texttt{POTUS} \colon [1885, 1989] \oplus [1893, 1897]$ where the annotation cannot be a time interval.

Thus, temporal annotations are instead sets of disjoint time intervals. We denote the set of such sets T and $\bot = \{\emptyset\}$, and $\top = \{[-\infty, +\infty]\}$. Further the order relation $\preceq$ as follows:

$$t_1 \preceq t_2 \text{ if and only if for all } I_1 \in t_1 \text{ there exists } I_2 \in t_2 \text{ such that } I_1 \subseteq I_2$$

where t_1 and t_2 are sets of disjoint intervals, I_1 and I_2 are intervals. Then $\langle T, \preceq, \bot, \top \rangle$ is a bounded lattice, which in turn induces two operators:

$$t_1 \oplus t_2 = \inf\{t \mid t_1 \preceq t \wedge t_2 \preceq t\}$$

and

$$t_1 \otimes t_2 = \sup\{t \mid t \preceq t_1 \wedge t \preceq t_2\}$$

The resulting structure $\langle T, \oplus, \otimes, \bot, \top \rangle$ is an annotation domain.

Note that RDF with temporal annotations is implemented as part of the spatio-temporal triplestore Strabon [Bereta et al., 2013].

While time intervals suffices for most use cases of temporal annotations, some applications like calendars require the notion of recurring events (such as Monday meetings) that can, in theory, be easily expressed as an annotation domain, but require tricky representations and algorithms to implement their operators.

The fuzzy domain

In the fuzzy domain, annotation values are numbers between 0 and 1, with $\bot = 0$, $\top = 1$, and $\oplus = \max$. The operator $\otimes$ is not unique though. Multiple ways of combining fuzzy measures have been studied (see [Straccia, 2009]) that are called t-norms. Any continuous t-norm is appropriate for a fuzzy annotation domain. Classical examples are min and the product t-norm.

In the case of annotation domains, the term "fuzzy" may be ambiguous. In fuzzy logic, the value indicates a degree of truth of a statement, such as "Saint-Étienne is a big city", which is true to a certain extent, but less true than "Paris is a big city", and more true than "Monaco is a big city". However, the same annotation domain can be used with a different interpretation, where the value indicates the confidence in a statement, or a level of trust in a statement, or a probability (although fuzzy reasoning should not be confused with probabilistic reasoning). In particular, the way fuzzy values are attributed to triples is based on an evaluation function that implements diverse heuristics that may not be explicit. Different systems that rely on fuzzy RDF may produce annotated triples that cannot be combined naïvely.

The provenance domain

Provenance is in itself a complicated topic that goes well beyond knowledge representation and reasoning, let alone annotations. However, a simple version of provenance representation can be encoded as annotation values. Let us assume that all sources of information are independent from each others and let us assign a unique identifier to each of them. We use these identifiers as propositional logic atoms, and use the binary operators $\vee$ and $\wedge$ to indicate the combination of provenance annotations. So, an annotation $p_1 \vee p_2$ indicates that the statement is true according to p_1 as much as according to p_2, while an annotation $p_1 \wedge p_2$ indicates that the statement holds when considering both p_1 and p_2 as truthful sources.

We thus have annotations that are propositional logic formulas involving only $\vee$ and $\wedge$. However, this is not sufficient to make an annotation domain. First, we need $\bot$ and $\top$. For these, we can use the propositional formulas that correspond to "false" and "true". Still, this would not form a valid annotation domain. The annotation values must be the equivalent classes of the logic formulas formed from provenance atoms, $\bot$, $\top$ and the operators $\vee$ and $\wedge$. Then, the $\preceq$ relation coincides with $\models$.

This formalisation of provenance is rather simplistic. In particular, assuming the independence of sources of information leads to much loss of context. For instance, if a journal talks about a political scandal, the assertions made about it can be attributed to the journal, or to the article that makes the statements. The article being in the journal means that one provenance implies the other. A hierarchical provenance domain could provide a partial solution to the problem.

The access right domain

Access right can be seen as a form of context of information. Indeed, the logical conclusions that someone can derive are limited to what one can have access to. In [Lopes et al., 2012], we proposed to use annotated RDFS to model, reason with and query access-restricted triples in a triplestore. The access right domain is a bit more involved than the ones previously described.

We start with a set of *credential elements* $\mathbf{C}$. This may represent a username, a role, a group, an attribute. For a credential element $e \in \mathbf{C}$, e and $\neg e$ are both *access control elements*, where e is called a *positive* element, while $\neg e$ is a *negative* element. Positive elements are used to indicate what credentials are required to access a resource, while negative elements indicates credentials that forbid access. An *access control statement* (ACS) is a finite set of access control elements, and an *access control list* (ACL) is a finite set of ACSs.

The existence of ACSs where both a positive and negative element appear for the same credential element (that is, e and $\neg e$ belong to the same ACS) leads to a situation of conflict that must be resolved. We assume that the access control system implements an operation, called resolve that maps ACSs without conflicts to themselves, and maps ACSs with conflicts to conflict-free ACSs. The two main resolution strategies consist in either prioritising positive elements (brave resolution) or negative elements (safe resolution) by removing one or the other access control element.

Due to the resolve operation, several ACSs may result in the same access decision. Moreover, in an ACL, multiple ACSs may provide redundant information, such as $\{\{a, \neg b\}, \{a\}\}$ grants access for someone with credential a that does not have b, but also grant access to anyone with simply a. So this would be equivalent to ACL $\{\{a\}\}$. To avoid dealing with multiple representations of the same access rights, we introduce a normalisation function, defined as follows:

$$\mathsf{normalise}(A) = \{\mathsf{resolve}(S) \mid S \in A \wedge \forall S' \in A, S \neq S' \Rightarrow S \not\subseteq S'\}$$

Now we can define the annotation domain:

- The set of annotation values L is the set of all normalised ACLs;
- $\bot = \emptyset$;
- $\top = \{\emptyset\}$;
- for all $A_1, A_2 \in L$, $A_1 \oplus A_2 = \mathsf{normalise}(A_1 \cup A_2)$;
- for all $A_1, A_2 \in L$, $A_1 \otimes A_2 = \mathsf{normalise}(\{S_1 \cup S_2 \mid S_1 \in A_1 \wedge S_2 \in A_2\})$.

The access control mechanism then uses this annotation domain in combination with the Annotated Query Language AnQL (described in the next section) to determine what answers can be delivered to the user or not. The principle is that the user issues a normal SPARQL query, but based on its user credentials, the query is translated into an AnQL query that ensures only authorised data are retrieved. The formalisation of AnQL is given next.

5.2 AnQL: A query language for annotated RDFS

The Annotated Query Language (AnQL) [Lopes et al., 2010a] can be described as "SPARQL with annotations". Syntactically, it is like SPARQL, except that triple patterns are replaced by *annotated triple patterns* where any of the subject, the predicate, the object or the annotation of an annotated triple can be replaced with a variable. However, variables used in annotation position are distinguished from other variables because they, as well as annotation *values*, can be used in special constructs that only affect annotations.

5.2.1 Brief summary of the language

The details of the syntax and semantics of AnQL are not fundamental for the arguments of this thesis and the discussion that I make, so I summarise the main characteristics of AnQL briefly and refer the reader to [Zimmermann et al., 2012] for details. The basic construct is called Basic Annotated Pattern (BAP), that plays the same role as Basic Graph Patterns in SPARQL, except that it is composed of annotated triple patterns instead of triple patterns.

AnQL allows the following features: conjunction of BAPs (AND), UNION, OPTIONAL patterns, FILTER included special filters on annotations that could depend on the annotation domain (for instance, the relation "before" could be used between temporal annotations), variable assignments (ASSIGN . . . AS), ORDER BY (possibly using the natural order of an annotation domain $\preceq$), GROUP BY with or without aggregates (COUNT, AVG, SUM, MAX, MIN and the special annotation aggregates $\oplus$ and $\otimes$).

The semantics of these constructs is somewhat similar to the analogous constructs of SPARQL, but in almost all cases, the formal definition has to be adapted to account for the specificities of annotation values.

5.2.2 AnQL and mediation

The way AnQL has been designed as well as its current implementation, do not make it a loosely coupled mediation system as advocated in this thesis. In our approach to implement a proof of concept, we envisaged a single, centralised, annotated triplestore that contain all the data and context description. Moreover, we had to hardcode all the annotation domains that we wanted to make use of.

However, what I envisage is more flexible: first, it should be possible to query data distributed over the Web, following what has been done for SPARQL by Olaf Hartig in his thesis [Hartig, 2014]; second, the processing of annotation values should not be hardcoded *per annotation domain* (which completely diminishes the interest of a generic framework) but only few primitives (notably $\oplus$, $\otimes$) should be provided *dynamically* to the system when meeting data that has certain types of annotations.

The first challenge is still very much open, but the second challenge was partly addressed by the work done with Maxime Lefrançois on dynamically interpreting custom datatypes [Lefrançois and Zimmermann, 2016]. A custom datatype is a datatype conforming to [Cyganiak et al., 2014, Section 5] but that does not correspond to one of the standard datatypes of the semantic web specifications (such as XSDs). Common implementations of RDF and SPARQL cannot deal with the particularities of custom datatypes unless a dedicated module is added to their code. However, in [Lefrançois and Zimmermann, 2016], we proposed a method to dynamically fetch the required primitives to interpret custom datatypes, by exposing them at the location where their datatype IRI dereferences to.

Because annotation domains are very much like datatypes, and annotation values are like literals, it would be possible to adopt the same approach for dealing with arbitrary annotation domains. Still, there are remaining issues relating to how multiple domains of annotations can be combined, especially how to annotate the same triple with multiple annotation values. This is the topic of the next section.

5.3 Combining annotation domains

A general approach to context representation should allow the combination of different types of contextual information. For instance, it should be possible to talk about temporal validity according to a certain source, which would combine the temporal domain with the provenance domain.

There are two cases to consider: how to deal with statements that are described with two types of context? and how to represent a multi-dimensional context for a statement? In the first case, we may consider a temporally annotated statement like "Antoine works for école des mines in 2015" and another, provenance-annotated statement "Antoine works for école des mines, according to his homepage". Would this mean that, according to his homepage, Antoine works for école des mines in 2015? Would it mean that, in 2015, the homepage was saying that Antoine works for école des mines? The problem is that there is no way to know, so one has either to completely segregate different annotation domains, or rely on out of band heuristics.

The second case is about representing and reasoning with statements like "Antoine works for école des mines in 2015, according to his homepage". Syntactically, it seems obvious that this could be represented as a 5-tuple or possibly an annotated triple where the annotation is a pair: $(\texttt{az}, \texttt{worksFor}, \texttt{emse})\colon \langle[2015], \texttt{homepage}\rangle$.

It turns out that any pair of annotation domain can be combined into a single annotation domain by applying domain-wise operation like so:

Definition 10 (Domain product) *Given two annotation domains $D_1 = \langle L_1, \oplus_1, \otimes_1, \bot_1, \top_1\rangle$ and $D_2 = \langle L_2, \oplus_2, \otimes_2, \bot_2, \top_2\rangle$, the product domain of $D_1 \times D_2 = \langle L_{12}, \oplus_{12}, \otimes_{12}, \bot_{12}, \top_{12}\rangle$ of D_1 and D_2 is defined as:*

- $L_{12} = L_1 \times L_2$;
- *for all* $\lambda_1, \lambda'_1 \in L_1$ *and* $\lambda_2, \lambda'_2 \in L_2$, $\langle\lambda_1, \lambda_2\rangle \oplus_{12} \langle\lambda'_1, \lambda'_2\rangle = \langle\lambda_1 \oplus_1 \lambda'_1, \lambda_2 \oplus_2 \lambda'_2\rangle$;
- *for all* $\lambda_1, \lambda'_1 \in L_1$ *and* $\lambda_2, \lambda'_2 \in L_2$, $\langle\lambda_1, \lambda_2\rangle \otimes_{12} \langle\lambda'_1, \lambda'_2\rangle = \langle\lambda_1 \otimes_1 \lambda'_1, \lambda_2 \otimes_2 \lambda'_2\rangle$;
- $\bot_{12} = \langle\bot_1, \bot_2\rangle$;

- $\top_{12} = \langle \top_1, \top_2 \rangle$.

The domain product is clearly a domain of annotation. However, this does not correspond to the intuition that we want to convey. Take the following example:

$$(\texttt{az}, \texttt{worksFor}, \texttt{emse})\colon \langle [2015], \texttt{homepage} \rangle$$

$$(\texttt{az}, \texttt{worksFor}, \texttt{emse})\colon \langle [1924, 2089], \texttt{fakenews} \rangle$$

From Definition 10 and the semantics of Annotated RDFS, this would entail:

$$(\texttt{az}, \texttt{worksFor}, \texttt{emse})\colon \langle [1924, 2089], \texttt{homepage} \rangle$$

which is obviously not what was intended. Thus, in [Zimmermann et al., 2012], we defined a novel way to build a domain of annotation that would combine two preexisting domains into one that follow the intuition.

The main idea of the construction of compound annotation domains is that the combination of domains D_1 and D_2 is a function from the underlying set of D_1 to the one of D_2 that must follow certain properties. [Zimmermann et al., 2012] provides rationale for such construction before introducing the following definitions:

Definition 11 (Quasihomomorphism) *Let f be a function from $D_1 = \langle L_1, \oplus_1, \otimes_1, \bot_1, \top_1 \rangle$ to $D_2 = \langle L_2, \oplus_2, \otimes_2, \bot_2, \top_2 \rangle$. f is a* quasihomomorphism *of domains iff for all $x, y \in L_1$: (i) $f(x \oplus_1 y) \succeq_2 f(x) \otimes_2 f(y)$ and (ii) $f(x \otimes_1 y) \succeq_2 f(x) \oplus_2 f(y)$.*

Definition 12 (Compound annotation domain) *Given two primitive annotation domains D_1 and D_2, the* compound annotation domain *of D_1 and D_2 is the tuple $\langle L_{12}, \oplus_{12}, \otimes_{12}, \bot_{12}, \top_{12} \rangle$ defined as follows:*

- *L_{12} is the set of quasihomomorphisms from D_1 to D_2;*
- *$\bot_{12}$ is the function defined such that for all $x \in L_1$, $\bot_{12}(x) = \bot_2$;*
- *$\top_{12}$ is the function defined such that for all $x \in L_1$, $\top_{12}(x) = \top_2$;*
- *for all $\lambda, \mu \in L_{12}$, for all $x \in L_1$, $(\lambda \oplus_{12} \mu)(x) = \lambda(x) \oplus_2 \mu(x)$;*
- *for all $\lambda, \mu \in L_{12}$, for all $x \in L_1$, $(\lambda \otimes_{12} \mu)(x) = \lambda(x) \otimes_2 \mu(x)$;*

This definition, however, is not really suitable for a concrete representation of compound annotations, because quasihomomorphisms in general cannot be represented finitely. This issue is similar to the temporal domain, where a temporal annotation could be any set of time points. A lot of such sets are not representable finitely, so we restrict the temporal domain to sets of intervals, that can conveniently be represented by sets of pairs of time points. Similarly, a quasihomomorphism can be built from a finite set of pairs of annotations.

Consider the domains D_1 and D_2 and let $A \subseteq L_1 \times L_2$ be a finite set of pairs of primitive annotations. We define the function $\overline{A} : D_1 \to D_2$ as follows:

$$\forall z \in L_1, \overline{A}(z) = \mathbf{lub}\{ \bigotimes_{2\, \langle x, y \rangle \in J} y \mid J \subseteq A \text{ and } z \preceq_1 \bigoplus_{1\, \langle x, y \rangle \in J} x \} .$$

Theorem 1 *If $A \subseteq L_1 \times L_2$ is a finite set of pairs of primitive annotations, then $\overline{A}$ is a quasihomomorphism.*

However, the construction is not injective, so several sets of pairs of annotations yield the same quasihomomorphism. This problem is comparable to the temporal domain where multiple sets of intervals may represent the same set of time points. With time intervals, annotations can be normalised to a set of *disjoint* intervals. Then, one of the contributions of [Zimmermann et al., 2012] was to define a normalisation of compound annotations.

The normalisation of an arbitrary set of pairs of annotations is done by applying two algorithms consecutively: Saturate (Algorithm 1) and Reduce (Algorithm 2).

Algorithm 1 Saturate(A)

Require: $A \subseteq L_1 \times L_2$ finite
Ensure: Saturate(A)
  $R := \emptyset$;
  for all $X \subseteq 2^A$ **do**
    $R := R \cup \{\langle \bigoplus_{1\, J \in X} \bigotimes_{1\, \langle x,y \rangle \in J} x, \bigotimes_{2\, J \in X} \bigoplus_{2\, \langle x,y \rangle \in J} y \rangle\}$;
    $R := R \cup \{\langle \bigotimes_{1\, J \in X} \bigoplus_{1\, \langle x,y \rangle \in J} x, \bigoplus_{2\, J \in X} \bigotimes_{2\, \langle x,y \rangle \in J} y \rangle\}$;
  return R;

Algorithm 2 Reduce(A)

Require: $A \subseteq L_1 \times L_2$ finite and saturated
Ensure: Reduce(A)
  while $\exists \langle x,y \rangle \in A, \exists \langle x',y' \rangle \in A \setminus \{\langle x,y \rangle\}$ such that $x \preceq_1 x'$ **and** $y \preceq_2 y'$ **do**
    $R := R \setminus \{\langle x,y \rangle\}$;
  while $\exists \langle x,y \rangle \in A$ such that $x = \bot_1$ **or** $y = \bot_2$ **do**
    $R := R \setminus \{\langle x,y \rangle\}$;
  return R;

Algorithm 3 Normalise(A)

Require: $A \subseteq L_1 \times L_2$ finite
Ensure: Normalise(A)
  return Reduce(Saturate(A));

Normalised compound annotations can be used to compute the $\oplus_{12}$ and $\otimes_{12}$ operations like this:

- $A \oplus_{12} B = \mathsf{Normalise}(A \cup B)$;
- $A \otimes_{12} B = \mathsf{Normalise}(\{\langle x \otimes_1 x', y \otimes_2 y' \rangle \mid \langle x,y \rangle, \langle x',y' \rangle \in A \times B\})$.

This gives an effective way of computing compound annotations, and ensures that if two annotations are different, then they have a different normalised form. All the proofs that these operations are correctly computing the operations of compound annotation domains are given in the appendix of [Zimmermann et al., 2012].

5.4 Summary

This work was made in collaboration with Axel Polleres (team leader at DERI, Galway at this time, now Professor at University of Economics in Vienna, Austria), Nuno Lopes (PhD student at DERI, Galway at this time, now Engineer at TopQuadrant) and Umberto Straccia (researcher at CNR, Italy). Sabrina Kirrane (from Insight Galway at the time, now assistant professor in University of Economics in Vienna, Austria) contributed to the work on access control. My publications that relates to this work are: [Lopes et al., 2010a, Lopes et al., 2010b, Zimmermann et al., 2012, Lopes et al., 2012]. Figure 5.1 describes how it instantiates the general idea of a loosely coupled mediation approach that I am defending in this thesis.

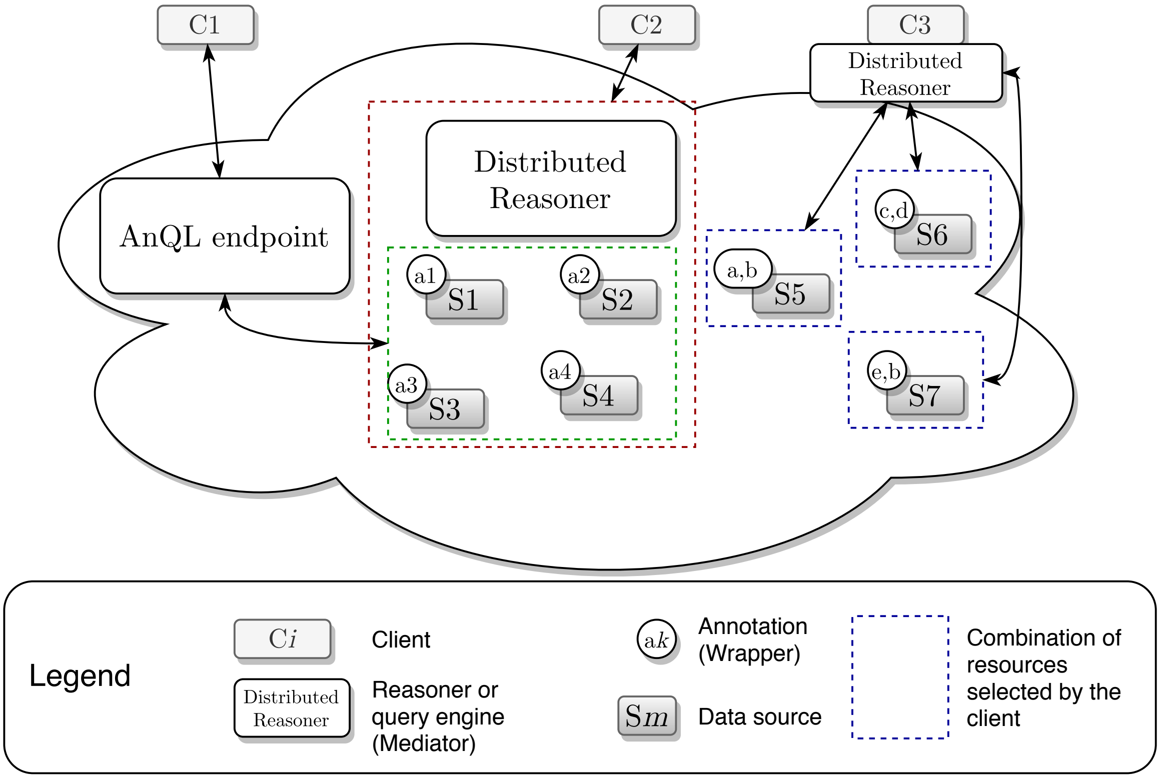


Figure 5.1: Annotated RDFS as an instance of loosely coupled mediation

Information sources are RDF data having validities in different contexts (S1 to S7, with annotations a1, ..., a4 and a, b, c, d). These sources are exploited uniformly by an annotated RDFS reasoner or an AnQL query engine that mediates across contexts. The annotations, which describe the context, are what enables the mediator to exploit the sources in a common way, so they act as the wrapper specification. Assuming that data sources are RDF graphs each available at a specific location (URI), and there exist repositories that catalogue data sources together with their contextual annotations, then the proposed approach can decouple data sources, wrappers and mediators, thus instantiating the vision.

Chapter 6

Representation of Context

Context sometimes cannot be quantified precisely. So the annotation framework of Chapter 5 cannot be applied for all kinds of contexts. In particular, it is sometimes the case that we want to relate context to other things in a qualitative way. For instance, a scientific article asserts some claim, which can be assumed to be true in the context of the article. But we may also want to add that the article, and thus the context of the claim, is endorsed by a certain institutional authority. Or that the claim is supported by evidences, or other claims, etc. As a result, context becomes an element in the domain of discourse.

Moreover, the formalisation of context as a special meta knowledge, whether done with an annotated logic as in Chapter 5 or with any other form of so called *contextual logic* (such as McCarthy logic of context [McCarthy, 1987], Giunchiglia's contextual reasoning [Giunchiglia, 1993] and variations of them) leads to the definition of non-standard logics that have never been widely adopted as part of a well used standard or application. Instead of divising special formalisms for dealing with context, one may try to inject context as a first class entity in the standard logic that describes the domain of discourse, that is, to *reify* context.

In terms of Semantic Web technologies, this could mean that we would integrate context as a node of the RDF graph that describes the domain of application, or as an individual in an OWL knowledge base. There are different ways in which the context of information can be reified, and several studies compared them to some extent. In my work with PhD student José Giménez-García, we not only proposed a new form of context reification, but also compared context reification models in a novel way, focusing on how reifying context impacts reasoning.

6.1 Overview of context reification models

Note that we talk about context reification models but it can also be seen as a statement reification model. The entity that is reified may be a statement (e.g., an entity that denotes an RDF triple or an ABox axiom) or the context itself (e.g., an entity that denotes the situation in which some claim holds). Distinguishing the two does not lead to much insight, but we will highlight when it is relevant to do so.

Consider the triple (`az`, `worksFor`, `emse`) and assume that we would like to say that this holds in 2019. If one wants to avoid introducing non standard formalisms where the temporal element can be attached directly to the statement, it is possible to express this (at least syntactically) in an RDF graph, by transforming the single triple into a pattern that can be related to the year 2019. Let us explore the different ways to do it.

6.1.1 Standard reification

Standard RDF Reification [Brickley and Guha, 2014, Section 5.3] is the approach that is recommended by the RDF standard itself in order to assert statements about statements. The principle consists in introducing a resource that will serve to identify the statement (in our example, the fact that Antoine works for école des mines). This resource is then related to the three components of the triple that denote the statement: its subject, predicate and object, as seen in Figure 6.1. A property can be used to relate the resource that reifies the statement and the description of the context (here a simple year).

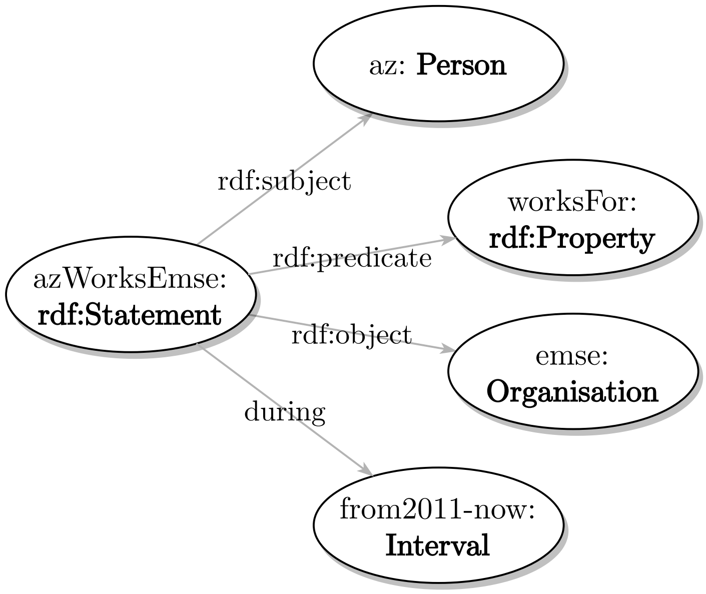


Figure 6.1: Example of a reified triple with standard reification. I simplify the graph by including the type inside an instance node

Unfortunately, RDF reification has a number of problems: first, its formal semantics is very weakly defined, such that hardly anything interesting can be inferred from it by standard reasoners; second, querying such representation is rather awkward and inefficient; third, the model is incompatible with the constraints that the Web Ontology Language puts on RDF graphs. The last point is precisely where the next model helps.

6.1.2 N-ary relations

While context sometimes can be thought of as metaknowledge on the core domain of discourse, it may also become part of the domain of discourse itself. For instance, while a web site may be concerned about presenting the current state of a company, with its current number of employees, and the relationship between employer and employee (which leaves to the mediation mechanism the task to attach (meta-)temporal information wrapping around the knowledge source), another source may be interested in describing the historical record of employment, where time is part of the model.

In the last case, employment may be seen as a ternary relation between an employer, and employee, and a time of employment. However, in Semantic Web standards, only binary relations can be directly represented. Consequently, design patterns for the representation of n-ary relations have been published by the W3C to help knowledge engineer adequately model such information [Fridman Noy and Rector, 2006]. There are essentially two proposed solutions:

As a class One may introduce a class that denotes the n-ary relation, such that an instance of the class uniquely relates to each of the components of the n-ary instantiation. For instance, Antoine being employed by école des mines in 2019 corresponds to an instance of `Employment` that relates to an `employer` (EMSE), an `employee` (AZ), and a time frame (2019). The approach is flexible enough to introduce any kind of contextual description, by interpreting `Employment` as an "employment claim" which could relate the employer/employee pair to dates, provenance, confidence measure, access control, etc.

As a split property One could consider that the relation `worksFor`, that originates from the employee, would relate to an abstract entity that combines the employer with the date. In this approach, there is still a sense of directionality that is preserved and may be found useful. It happens to be the chosen design for qualifying statements on Wikidata [Erxleben et al., 2014].

Figure 6.2 shows how the two variants compare, graphically.

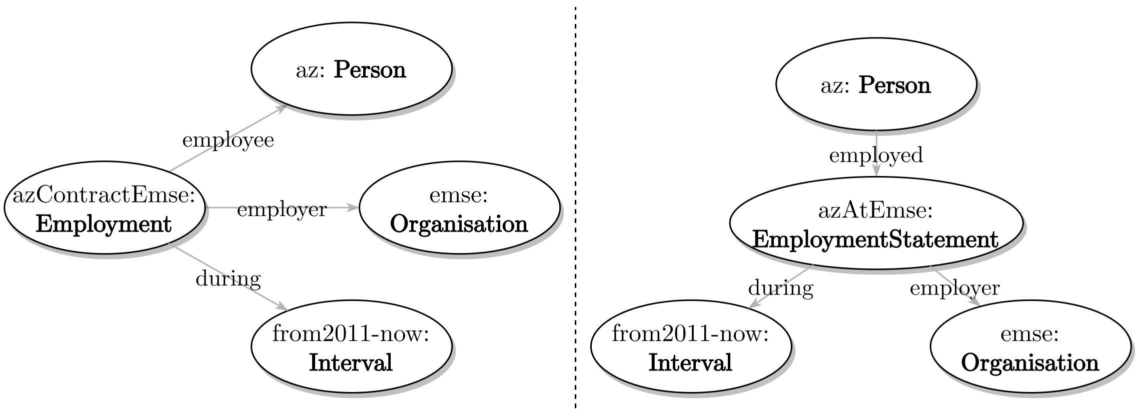


Figure 6.2: Example of nary relations: on the left, as a class; on the right, as a split property

6.1.3 Singleton properties

Another approach to modelling context is Singleton properties [Nguyen et al., 2014]. The principle of this design pattern is to introduce a property such that its domain and range are reduced to a singleton. The newly introduced property is then defined as a "singleton property of" the more general property that we want to contextualise. The approach also comes with a formal semantics that enforces the "singletonness" of the property, and makes the singleton property a subproperty of the original. Figure 6.3 exemplifies the model.

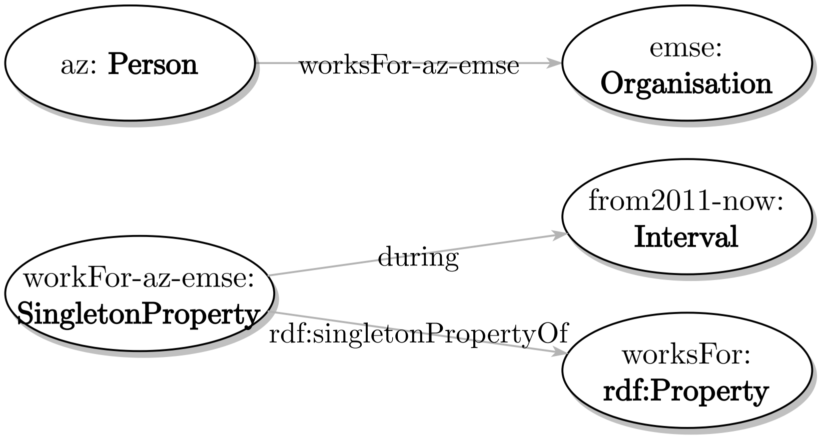


Figure 6.3: Example of a singleton property

6.1.4 Other approaches to representing context

Beyond the methods to reify RDF statements within the limits of RDF graphs, there has been different extensions of RDF that can express meta statements natively. We briefly describe them here.

Named graphs. Named graphs were introduced in [Carroll et al., 2007] as a structure that can help deal with the notion of provenance, and ultimately trust, on RDF data. They are simply defined as pairs (n, g) where n is a URI and g is an RDF graph. The concept was then integated into SPARQL [Seaborne and Prud'hommeaux, 2008]. Later, named graphs were added as a standard concept of RDF 1.1 [Cyganiak et al., 2014, Section 4], but extended to pairs (n, g) where n can be a blank node. Using multiple named graphs, one can build an RDF Dataset, a notion first defined in SPARQL: a pair (D, N) where D is an RDF graph, called the default graph, and N is a finite set of named graphs such that the names do not repeat.

No standard semantics of RDF Datasets is given, and multiple option have been identified in a note of the W3C RDF 1.1 Working Group that I wrote [Zimmermann, 2014].

Notation 3. Notation3 [Berners-Lee and Connolly, 2011] (or simply N3) is a syntax that extends RDF in syntax and in semantics: in an N3 document, the subject or object of a triple can be a graph, or even an N3 expression. Constructs writing implications (deduction rules), as well as for declaring existential and universal variables, makes this language a powerful logic. However, the formal semantics is somewhat underspecified in the current specification.

RDF*. RDF*, and its accompanying query language SPARQL* [Hartig and Thompson, 2014], is similar to Notation3 in the sense that it extends RDF with the possibility to include triples in the subject or object position, but does not add any other constructions. From a semantic point of view, the use of a triple in place of a subject or object can be thought of as syntactic sugar for RDF.

RDF with context. In [Guha et al., 2004], the authors redefine RDF semantics such that the notion of context becomes a first class citizen among the concepts of RDF and as part of RDF interpretations. Then the article argues about the advantages that such alternative model theory bring.

All these approaches that extend RDF are based on non standard formalisms. In the work presented in this chapter, the challenge that we set to ourselves was to study how we could represent contextual information without diverging from the Semantic Web standards.

6.2 Comparison

Reification models have been compared in different ways regarding performance of query answering, loading, size [Hernández et al., 2015, Frey et al., 2019, Haidar, 2019]. However, as far as I know, we are the only ones who compared the models with respect to their semantics, more specifically what inferences can be done with them.

The comparison that we made is based on the inference capabilities of reification models, studying them from the perspective of a common formal framework. Assuming that we can make deduction on non-contextual statements, do we maintain similar inference after reifying the statements. For instance, we may assume that "Antoine is a person" is not a contextual statement: it holds true all the time and from every point of view. If additionally, we know that "a person is also a human being", then we can conclude that "Antoine is a human being". If, on the contrary, we know that "Antoine is an employee of Mines Saint-Étienne" but only since 2011, and "an employee of Mines Saint-Étienne is a person", then can we conclude that "Antoine is a person", at least since 2011? It seems reasonable to do so.

Formally, if we assume a set $\mathbf{C}$ of context, and the set of RDF graphs $\mathbf{G}$, we can define a reification model as follows:

Definition 13 (Reification model) *A* reification model *is defined as a function* $\rho : \mathbf{G} \times \mathbf{C} \to \mathbf{G}$*, that maps a graph with a context to another graph.*

Arguably, more constraints on the function could be added, but this would not be relevant here. [Giménez-García et al., 2017] has more details about this. More importantly, the property that we would like to guarantee for a reification model is the following:

Definition 14 (Entailment preservation) *Let E be an entailment regime. We say that a reification model ρ* preserves *the entailment regime E iff* $\forall G_1, G_2 \in \mathbf{G}, \forall c \in \mathbf{C}, G_1 \models_E G_2 \Rightarrow \rho(G1, c) \models_E \rho(G_2, c)$.

Our study in [Giménez-García et al., 2017] showed that all reification models presented before preserve simple entailment [Hayes and Patel-Schneider, 2014, Section 5], but standard reification and n-ary properties are far from preserving the RDFS entailment regime, let alone more advance OWL features. Singleton properties, however, preserve a significant part of RDFS entailment, and some OWL entailments, but also introduce additional entailments that are not desirable and counter-intuitive from the perspective of contextual reasoning. Thus, we devised a new reification model that is described in the next section.

6.3 NdFluents

The NdFluents model is strongly influenced by the 4dFluents ontology of [Welty and Fikes, 2006], which was designed in order to represent fluents, that is, entities that evolve with time, and whose relations to other entities change. In [Welty and Fikes, 2006], a fluent, represented as an individual in OWL, is related to multiple *time slices*, other individuals representing the entity in a certain time frame. The time slice can then be used to represent a temporary relation about the fluent individual. For instance, Antoine is a fluent having a time slice Antoine-in-2012, that relates to the time slice EMSE-in-2012 of école des mines. We can assert that "Antoine-in-2012 worksFor EMSE-in-2012", but not that "Antoine worksFor EMSE", as the later is not valid in a large part of the past, and will not continue to be valid indefinitely.

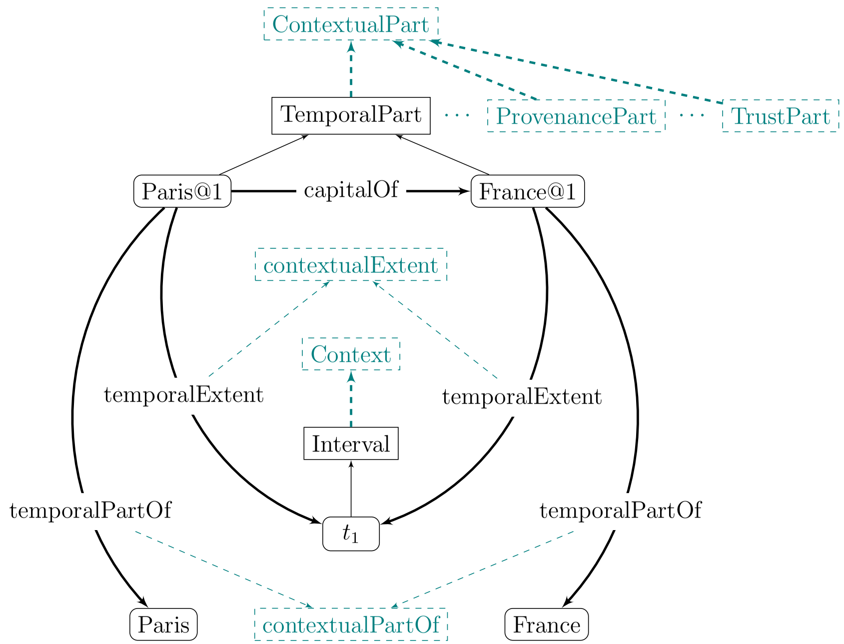
Similarly, in [Giménez-García et al., 2016b, Giménez-García et al., 2017], we adopted a similar but more general approach: an individual can have multiple *contextual parts* that each represent the same individual as seen from a specific context. For instance, there can be a contextual part of Antoine that is "Antoine-according-to-his-homepage". This technique makes use of the ontology defined in Ontology 1 and its use is illustrated on different contextual dimension in Figure 6.4.

Ontology 1 The NdFluents ontology, from [Giménez-García et al., 2017]

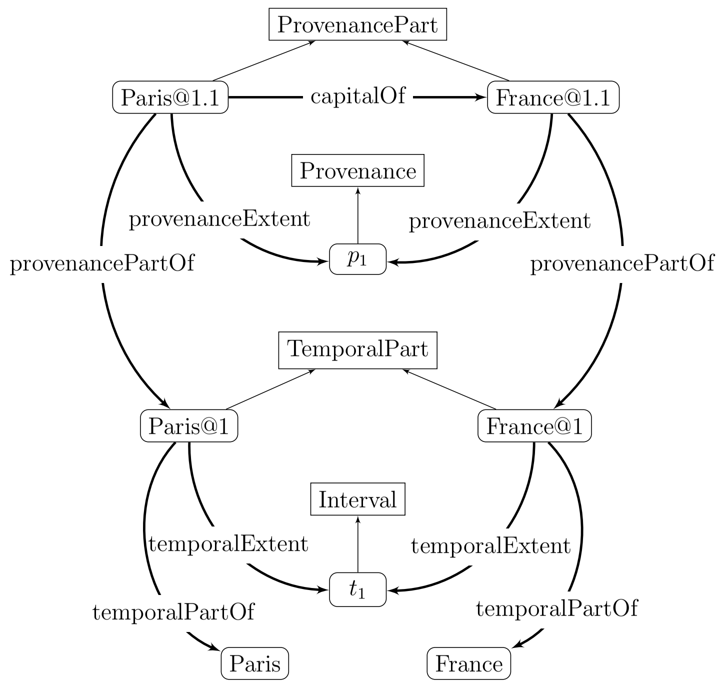
```
Prefix( nd:=<http://purl.org/NET/NdFluents#> )
Ontology( <http://purl.org/NET/NdFluents>
       Declaration( Class( nd:Context ) )
       Declaration( Class( nd:ContextualPart ) )
       DisjointClasses( nd:Context nd:ContextualPart )

       Declaration( ObjectProperty( nd:contextualProperty ) )
       ObjectPropertyDomain( nd:contextualProperty nd:ContextualPart )
       ObjectPropertyRange( nd:contextualProperty nd:ContextualPart )

       Declaration( ObjectProperty( nd:contextualExtent ) )
       ObjectPropertyDomain( nd:contextualExtent nd:ContextualPart )
       ObjectPropertyRange( nd:contextualExtent nd:Context )

       Declaration( ObjectProperty( nd:contextualPartOf ) )
       FunctionalObjectProperty( nd:contextualPartOf )
       ObjectPropertyDomain( nd:contextualPartOf nd:ContextualPart )
       ObjectPropertyRange( nd:contextualPartOf ObjectComplementOf( nd:Context ))
)
```

In [Giménez-García et al., 2017], we compared the reification models with respect to inference preservation, by listing inference rules from RDFS and pD*.[1] In this paper, we showed that NdFluents preserves more entailments than the other models, but our definition of entailment preservation was slightly different. The details are given in the paper and will not be reproduced here. In the next section, I show how we extended the approach to reach complete inference preservation for description logics.

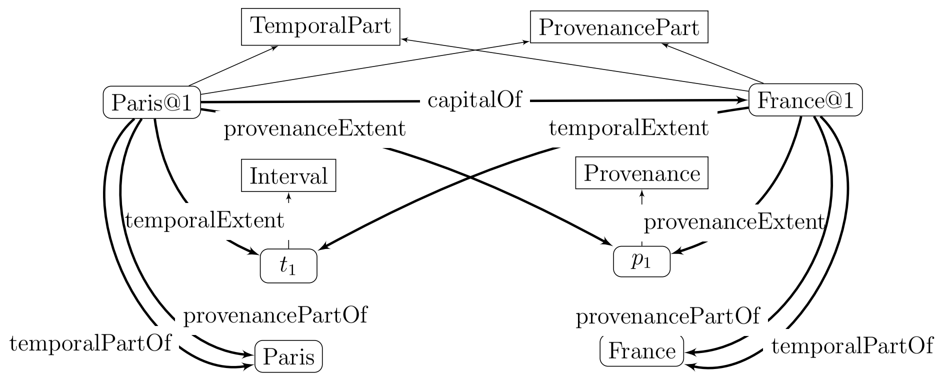
[1] pD* is a formalism that extends RDFS with a subset of the constructs from OWL [ter Horst, 2005].



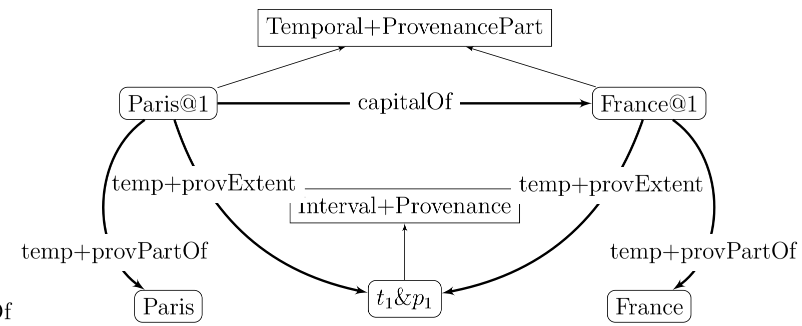
(a) Example of 4dFluents as implementation of NdFluents



(b) Contexts in context



(c) Multiple Contexts on each ContextualPart



(d) Combination of different dimensions on one contextual extent

Figure 6.4: NdFluents ontology and design patterns

6.4 Generalisation of the NdFluents approach

NdFluents is in fact an instantiation of a more general approach towards contextualising knowledge. In a nutshell, NdFuents consist in changing the names of instances when they have to be described within a context. But there is no reason one could not also contextualise classes or properties (concepts or roles in description logics) or even n-ary predicates in first order logic. However, in order to guarantee entailment preservation in general, introducing context-specific terms replacing the generic terms is not sufficient.

In [Zimmermann and Giménez-García, 2017a, Zimmermann and Giménez-García, 2017b], we introduced a reification model for description logics ontologies and knowledge bases that completely preserves entailments. In addition to introducing context-specific terms, for each context and original terms, it relies on the notion of relativisation. Simply put, relativisation consists in transforming a logical theory in such a way that everything described by the theory must be included inside a subset of the universe of discourse.

In description logics, given an ontology O, relativisation can be done by transforming every occurrence of the $\top$ concept in the axioms of O into a context-specific top concept that must be a superclass of every named concept of O, and every property of O must have a domain and range that is this context-specific top concept. Additionally, every occurrence of negation ($\neg$) and universal quantification ($\forall$) in O must be transformed into a construct that is the intersection of the context-specific top concept and the negated concept or universal quantification.

The details of the contextualisation model are rather lengthy, so I simply illustrate it on an example and refer the reader to [Zimmermann and Giménez-García, 2017a, Zimmermann and Giménez-García, 2017b] for the formal details.

Example 1 *The axiom* $\exists\texttt{capitalOf}.\top \sqsubseteq \forall\texttt{capitalOf}^{-}.\bot$ *is relativised into* $\exists\texttt{capitalOf}.\top_c \sqsubseteq \forall\texttt{capitalOf}^{-}.\bot \sqcap \top_c$*, where c denotes a context.*

Finally, the reification model consists in applying a systematic renaming on terms of the ontology, in function of the context, after relativisation.

We formally proved that this model is preserving all entailments, regardless of the description logic used, provided that *punning* is allowed (see [Golbreich and Wallace, 2012, Section 2.4.1]). That is, the logic must allow the same term to be used as an individual name, a concept name, or a role name at the same time.

Additionally, this model is inconsistency preserving, in the sense that, if an ontology is inconsistent, and all of its axioms are assumed to be part of the same context, then applying the NdFluent model leads to an inconsistency. However, if the contradiction happens from the combination of axioms that hold in different contexts, then the inconsistency is not preserved.

In [Giménez-García and Zimmermann, 2019], we proposed a variation of the approach where only the properties (binary relations, or roles) are contextualised (that is, renamed after relativisation). This leads to a model that is similar, but not equivalent, to singleton properties, but that guarantees entailment preservation and inconsistency preservation, but does not prevent inconsistencies to happen across different contexts.

6.5 Summary

This work was made in collaboration with José Giménez-García (PhD student at Université Jean-Monnet, Saint-Étienne) and Pierre Maret (Professor at Université Jean Monnet) in the context of H2020 Innovative Training Network WDAqua. My publications that relates to this work are: [Giménez-García et al., 2016b, Giménez-García et al., 2017, Zimmermann and Giménez-García, 2017a, Zimmermann and Giménez-García, 2017b, Giménez-García and Zimmermann, 2019, Giménez-García et al., 2019]. Figure 6.5 describes how it instantiates the general idea of a loosely coupled mediation approach that I am defending in this thesis.

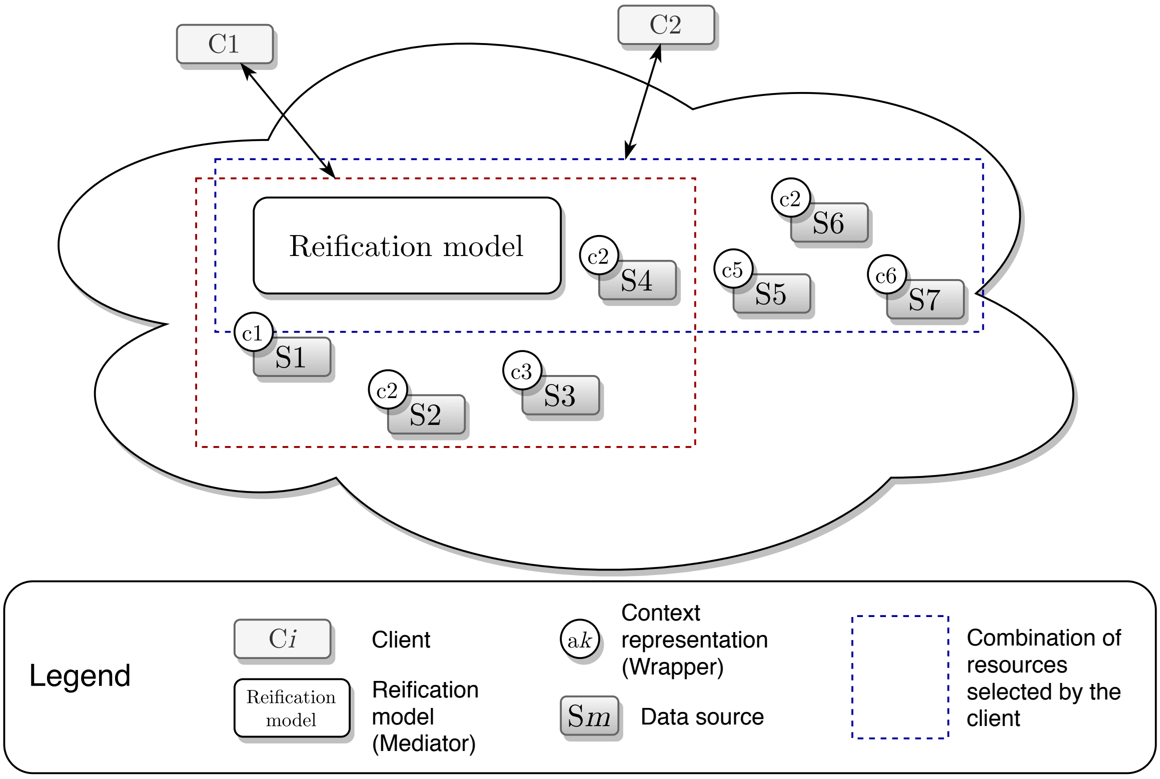


Figure 6.5: Reification models as an instance of loosely coupled mediation

Information sources are knowledge representations expressing statements in different contexts (S1 to S7 in the figure). A reification model enables the data consumer to deal with all sources together according to a single standard knowledge representation formalism. But to do so, one must have some kind of representation of the context that will be part of the reified model (c1 to c6). It is possible that several sources belong to the same context (e.g., S2 and S4 are both stated in the same context c2). The representation of the context itself can be serialised as an RDF graph separated from the data sources themselves, thereby allowing a loose coupling of this kind of mediation infrastructure.

Summary and future work

The research presented in this part, developed in order to reconcile knowledge of different origins, deals with the unification of different terminologies via ontology alignments, enabling a merged knowledge model encompassing all the domains of the ontologies; and with the delimitation and interplay of contexts, understood as *the scope of truth* of the statements composing the data.

Figure 6.6 gives an overview of how the three chapters relate to each other and conform to the vision of a loosely coupled mediation.

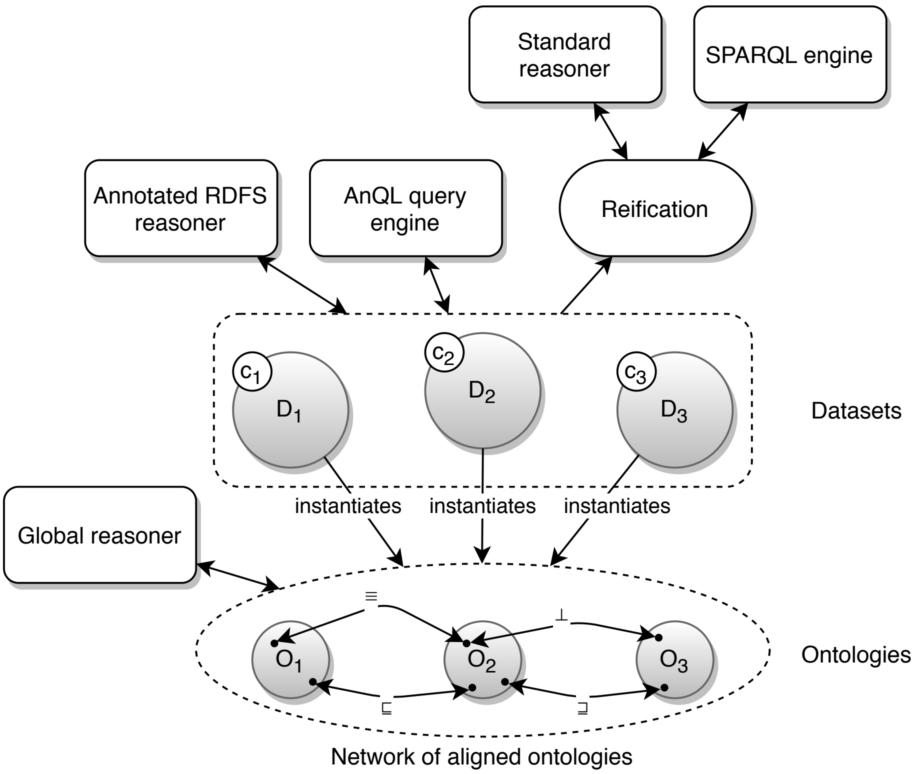


Figure 6.6: Overview of how the contributions relate to each others

First, assuming that ontologies can be aligned correctly (which is a strong assumption since achieving complete alignment in general is "virtually impossible" [Euzenat and Shvaiko, 2013]), then ontologies presenting different designs and vocabularies form a cohesive, yet modular, knowledge model in the form of a network of aligned ontologies (the bottom of Figure 6.6). With a proper semantics for such structures, complying with the problem of diverging viewpoints, a user of multiple ontologies can then view, use, and reason with an ontology network as if it were a single monolithic ontology. The ontology network is therefore the unique interface that mediates over the multiplicities of knowledge models, which make the contributions of Chapter 4 an instance of loosely coupled mediation.

Second, based on a global model, different datasets may instantiate the concepts in different ways (for instance, D1 to D3 in Figure 6.6), depending on the context they describe (c1 to c3). An annotated reasoner or the annotated query language AnQL provide a unified interface to the multiplicity of datasets. This is made possible with the contributions of Chapter 5.

Third, an alternative to the non-standard annotated logic consists in reifying the contextual datasets, such that both the data and their contexts are represented according to standard RDF or OWL. With our contribution in Chapter 6, we provide a way that RDF data or OWL ontologies can be integrated in such a way that allows inferences to be made per context, using the standard semantics of these formalisms. As far as I know, our work on NdFluent is the first one that provides objective and formal properties that a reification method should have, from a semantic point of view.

In each case, these contributions make use of declarative knowledge about the correspondences or context of each individual information source.

Among the perspectives opened by these contributions, we can identify the following:

- Networks of aligned ontologies may contain ontologies that provide contradicting views on some concepts that can be dealt with in different ways: a non standard semantics based on the principle of local model semantics can mitigate the problem, but how can the right semantics be chosen depending on the use case? Can one of such semantics be standardised? Should a different approach be used such as adopting a modal logic, paraconsistent logic, an argumentative framework, or reconciling diverging views by repairing the source ontologies? Besides, there is a need for concrete algorithms and tools to support reasoning over networks of aligned ontologies, with the added difficulty of distributedness of data sources.

- In terms of contextual reasoning, and more specifically annotated logic reasoning, the approach that I described still needs to be generalised to more expressive logics such as OWL 2 DL and its profiles. While in Annotated RDFS it is not possible to reach a contradiction (because Annotated RDFS does not contain datatype semantics and cannot express negation), in more expressive logics, contradiction is possible. However, contradicting statements in disjoint contexts should not be problematic. Another research direction relates to practical implementations of Annotated RDFS reasoners which should be able to deal with any annotation domain without the need to update the reasoner for each newly supported domain. Since annotation domains are very similar to datatypes, as defined in RDF, dealing with arbitrary domains could be done in a way similar to dealing with arbitrary custom datatypes. With colleague Maxime Lefrançois, we made a step towards this in [Lefrançois and Zimmermann, 2016].

- The problem of reifying statements, and more generally representing the context of statements, in general or in one dimension (temporal, provenance, etc.), within the RDF standard, is recurring in semantic web research. Yet, even with an attentive eye on the subject for many years, I have not seen a consensus, either by practitioners or theoreticians, on what model suits best specific use cases. With our approach NdFluents, we guarantee preservation of inferences, a property that allows one to reason in a context independently from the others, but we have not yet investigated *cross-context* reasoning. This would certainly require meta-information about contexts, preferably in the form of RDF itself, that would allow an inference engine to transfer, lift, or bridge entailments from one context to another.

More generally, I believe that the problem of exploiting knowledge in reasoning from multiple sources is one of the main challenges for the Semantic Web for years to come, though one that I would love to tackle in the future.

Part III

Decision Interoperability: Agents Interacting on the Web

Outline of the contributions

With the contributions of Part II, we can assume that knowledge is available in a unified form, whether as a monolithic ontology, a network of aligned ontologies, or combined contextual knowledge bases. Yet, this does not guarantee complete interoperability because different agents (artificial or human) can take different decisions when exposed to the same knowledge. For instance, if it is known that an electronic component is overheating, an agent may decide to switch off the power supply to avoid hazardous consequences, another may attempt to decrease the temperature by activating ventilation, and a third one may simply wait until the temperature reaches a critical maximum when the situation must be fixed urgently. It is unreasonable to pretend that all the three agents can be replaced by one another without risks: there is a lack of interoperability between them due to different processes, leading to diverging decisions.

This part addresses this level of interoperability to some extent. However, there are considerable challenges that have still to be tackled, and I have been addressing these aspects more recently, so the solutions presented here only deal with a fraction of the problem at large.

If multiple autonomous agents are available as resources to do a task and cooperate with each others, then there is the question of how to find them, how to keep track of what they are doing and how they can interact. In a truly open, decentralised, and large scale environment such as the Web, traditional multi-agent techniques are not sufficient to address these questions. Instead, I show how the combination of Web, Semantic Web and Multi-agent technologies can enable a flexible, decentralised, mashable Web of services and autonomous agents, taking advantage of social Web platforms as mediators (Chapter 7).

While the first contribution provides a generic approach towards the interoperability of processes, the second one tackles the specific problem of path finding in ubiquitous environment. I show how the process can make use of a semantic description of the environment to allow planning multiple goals along an itinerary, with robustness to latency (Chapter 8).

The work presented in this part results from: the supervision of Andrei Ciortea, jointly with Prof. Olivier Boissier and Prof. Adina Magda Florea (Chapter 7) and the supervision of Oudom Kem, jointly with Prof. Flavien Balbo (Chapter 8), both benefiting from a scholarship independent from any project.

Chapter 7

Socio Technical Networks for Hypermedia Agents

In a world where more and more objects get connected to the Internet (the Internet of Things, or IoT), it will become common that people or organisations have to manage hundreds or thousands of connected items, and coordinate them towards a common goal. This implies a heavy overhead for whoever has to monitor and control these things.

A possible solution for mitigating this problem consists in endowing objects with autonomy. However, it is still necessary to be able to easily monitor, control, and make these things communicate with each others in a flexible way.

In the work conducted by Andrei Ciortea in his PhD thesis, we envisioned that a network of things and people could better function by making use of the social network metaphor. Indeed, if sustaining social connections with hundreds of people is extremely difficult in real life, social networking platforms have allowed us to maintain enormous social networks where interactions with followers, friends, and colleagues are simplified by mediating through platforms that can filter updates, relevance, etc.

Similarly, managing thousands of objects and people together could become sustainable by interacting with them through *socio-technical networking platforms*, that is, a mediating platform of people and things. If, additionally, things are endowed with autonomy, they could proactively communicate with their owners, or between each others. This chapter provides some of the technical details of this vision.

In this chapter I start by showing an abstract architecture that supports this vision (Section 7.1). Then I show how the notion of socio-technical networks can be formalised as an ontology (Section 7.3) and as a more abstract model (Section 7.2). Then I will show how these ideas benefit from the architecture of the Web, especially hypermedia, to implement the vision (Section 7.4).

7.1 A layered architecture

The idea of a social network of people and things can be approached by separating concerns in different layers as shown in Figure 7.1, as first discussed in [Ciortea et al., 2013]. First, we assume that networking protocols are uniformised by relying on Web standards as a common ground, here described as the Web of Things, or WoT for short. Second, we assume that every participant in a system of interacting things

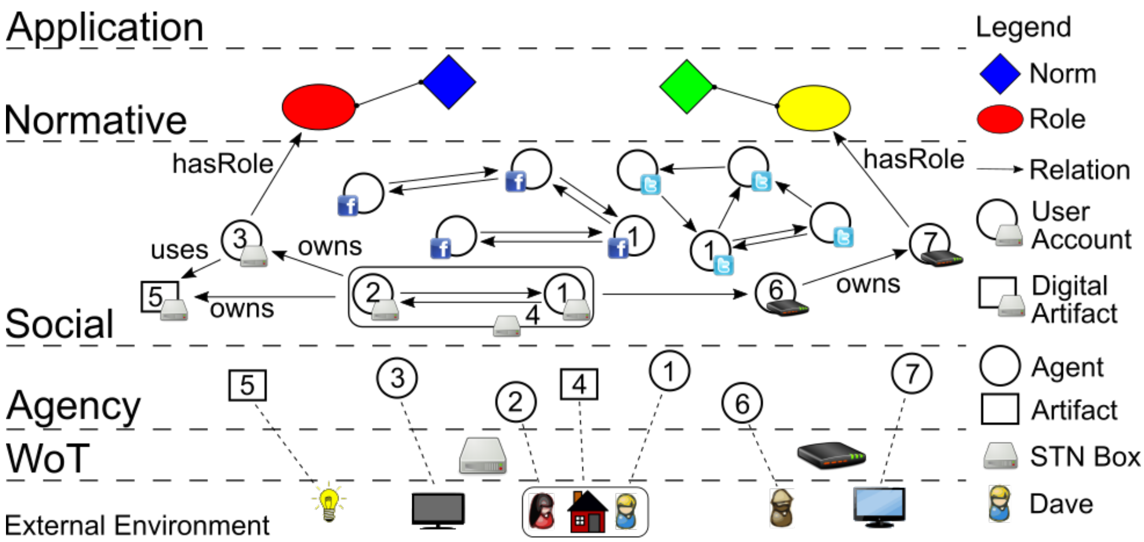


Figure 7.1: A layered model for socio-technical networks (from Andrei Ciortea's doctoral thesis [Ciortea, 2016])

and people are agentified. That is, they are abstracted as *agents* having a certain level of autonomy that allows them to take decisions on their own. However, agents evolve in an environment where they can use certain things, services or tools that do not have autonomy. Such entities lacking autonomy would be abstracted as *artifacts* instead, following the Agents & Artifacts meta-model from [Omicini et al., 2008]. Third, they should be able to create social relationships, in a broad sense of the term, to start communicating with each others. To be precise, they need to form networks of collaborating entities that can exchange information, orders, reports, and so forth. Fourth, they need to act according to agreed norms (social rules, laws, obligations) that enforce behaviours that are, on the one hand, beneficial to a common goal, and on the other hand not detrimental to what others try to achieve. Finally, there must be applications that take advantage of the other layers in order to leverage the capabilities of agents, allow them to act together as necessary, and control their joint actions according to the norms.

This architecture was implemented by Andrei Ciortea on a simple scenario where an application help a user waking up at the right time according to his schedule and the available means of interacting with him. I do not provide more details on the application layer and refer the reader to Andrei's thesis [Ciortea, 2016]. The following sections describe the lower layers, starting from the social and normative layers (Section 7.2 and Section 7.3) then going down to the agent and WoT layers (Section 7.4).

7.2 A model for socio-technical networks

A socio-technical network (STN for short) is an abstract structure that defines all the parts of an evolving, organised, network of social entities interacting together. An STN is not merely a graph that present members of the network as nodes and social relationships as edges. STNs are governed by rules that enable their evolution and enforce their regulation.

In order to understand our model of STNs, we can consider human societies as an analogy. At any given point in time, people are related according to social ties, such as being friends, siblings, spouses, etc. This forms a *social graph*. The social graph can only evolve by way of certain actions or events. Some are deliberate actions by the member of the network, such as getting married or being employed. Additionally, these actions and events cannot happen arbitrarily: they are constrained by social norms and regulations. An "employedBy" relation cannot be created by an employer and an employee if certain obligations are not respected. While many of the events and actions in a social network of people are not formalised, when it comes to endowing digital things with autonomy and socialness, one has to provide explicit information about them in order to allow digital and physical things, powered by software applications, to understand the state of the social graph, the available actions, and the regulations to take adequate decisions.

Therefore, in [Ciortea et al., 2015], we modelled an STN as an abstract structure $(G_t, \mathrm{Ops}, \mathrm{Norms}, O)$ that consists of:

- G_t is a social graph that can evolve in time;
- Ops is a set of operations that either affect the social graph, or allow autonomous agents to partially access the state of the social graph;
- Norms is the set of norms that regulate the use of operations in Ops;
- O is an ontology that defines the shared terms of the STN, that allows all participants to understand the relationships in the social graph, the operations and the norms.

There are many ways of implementing this abstraction, but as our layered architecture is ultimately grounded on the Web, we take advantage of Web standards as much as possible: the social graph is assumed to be encoded in RDF; operations are Web services documented with API description frameworks (such as Swagger,[1] Things Description [Kaebisch et al., 2020]); norms are taking advantage of multi-agent system organisation meta-models (in particular MOISE [Hübner et al., 2002b]); finally, the ontology is assumed to be an OWL ontology.

7.3 An ontology for STNs

As explained in the previous section, an ontology is an integral part of our STN model. We defined the STN ontology for online socio-technical networking platforms as a modular OWL ontology that deals with all aspects of the STN model [Ciortea et al., 2014]. Figure 7.2 shows how the STN ontology is modularised. Then Figure 7.3, Figure 7.4, and Figure 7.5 show three samples of ontology modules that give a good overview of what the ontology covers. Note that these figures are from Andrei's thesis and are more complete than the ones found in the papers that I coauthored.

If a socio-technical network is described according to this ontology, and an agent understands the vocabulary of the ontology, then it can autonomously interact with the STN. A typical description of an STN platform will inform the agent that there are operations such as register, create connections with other registered agents,

[1] Swagger's Open API specification `https://swagger.io/specification/`

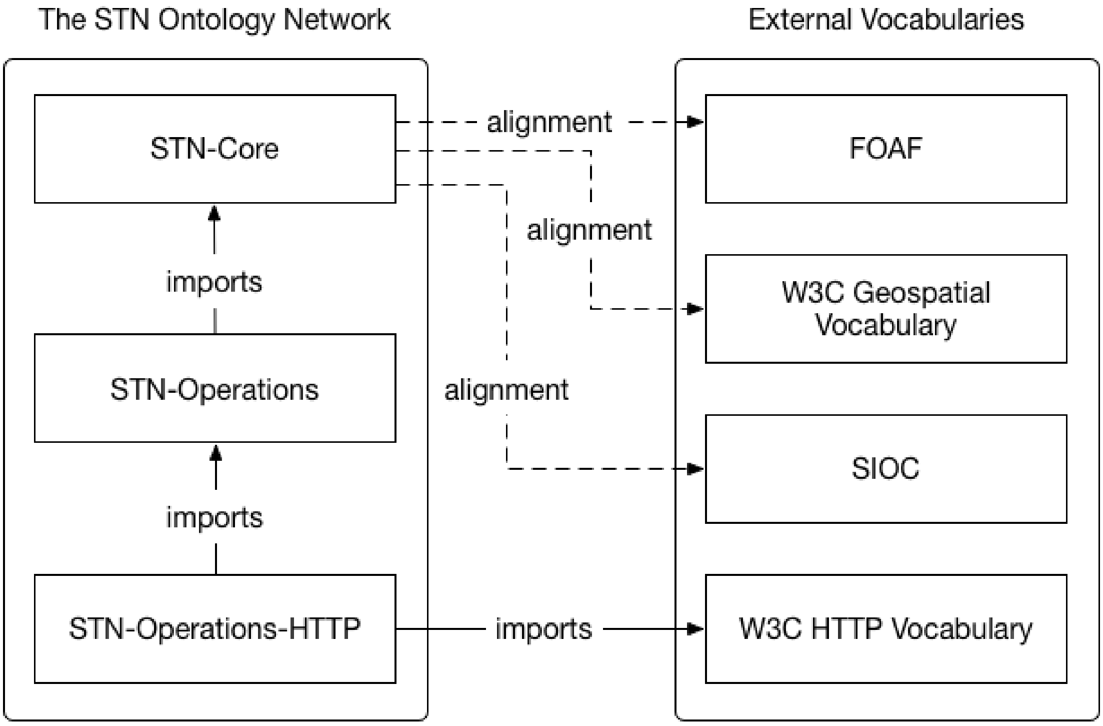


Figure 7.2: The STN network of ontologies (from Andrei Ciortea's doctoral thesis [Ciortea, 2016]). The 3 ontologies on the left can be found at `https://w3id.org/stn`

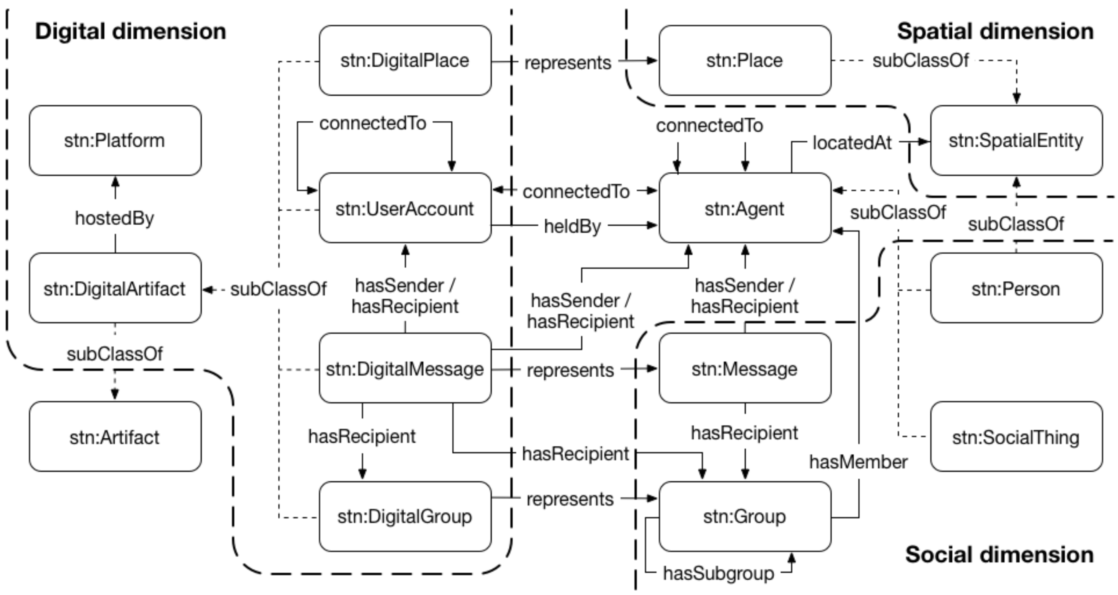


Figure 7.3: Part of the STN-core ontology (from Andrei Ciortea's doctoral thesis [Ciortea, 2016])

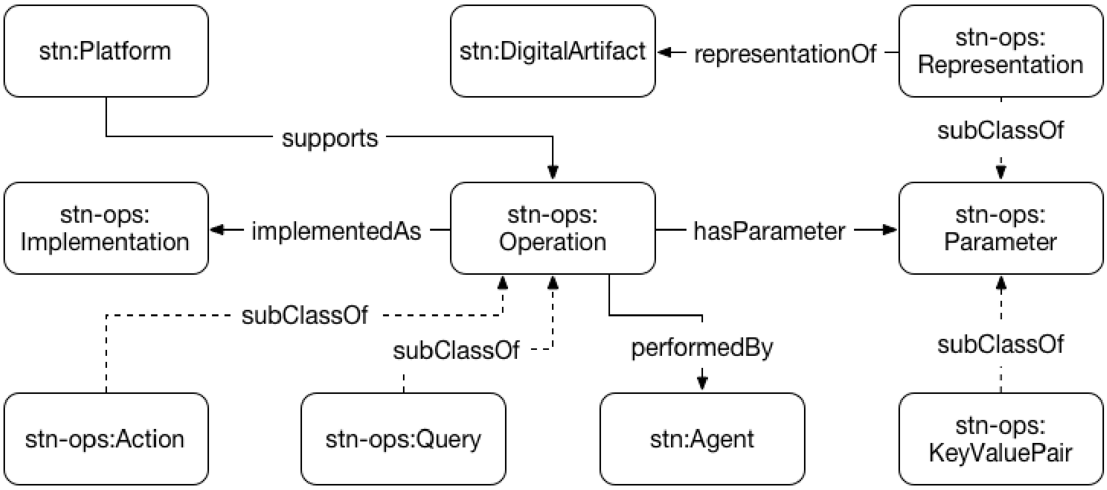


Figure 7.4: Overview of the STN-operations ontology (from Andrei Ciortea's doctoral thesis [Ciortea, 2016])

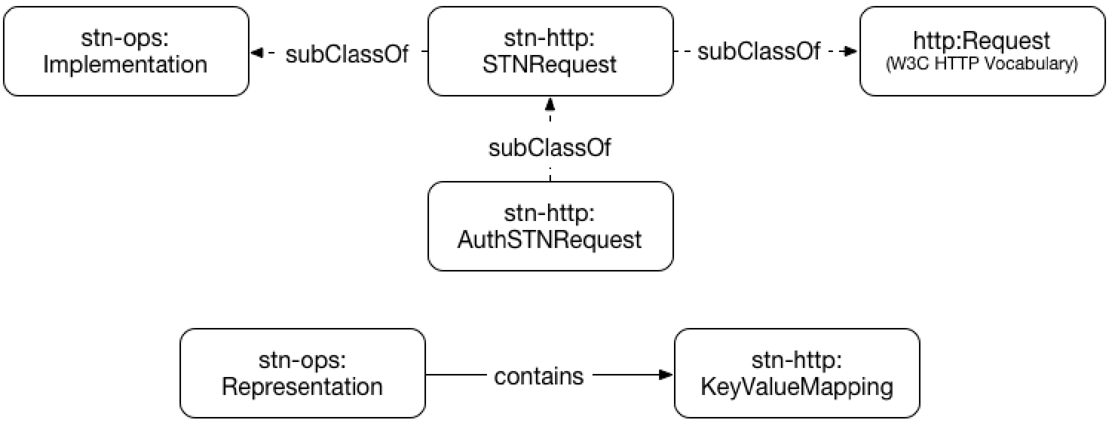


Figure 7.5: Overview of the STN-operations-HTTP ontology (from Andrei Ciortea's doctoral thesis [Ciortea, 2016])

post public messages on their profile, or send private messages to specific fellow agents. Interestingly, existing social-networking platforms such as Twitter, Facebook, LinkedIn, can be described according to the STN ontology. The description may be provided by a third party, such that a software agent with understanding of the ontology terms can automatically create an account and participate on those platforms. Therefore, even social networks that are normally targeted towards people can become socio-technical networks, because all interactions with these platforms are mediated via a Web client, which could be autonomous and embedded in things.

As a result, the STN ontology, if used consistently across multiple STN platforms (or even existing social networking platforms) can provide a uniform mediation layer between the agents and all platforms. This was partly demonstrated in [Ciortea et al., 2015] where an "STN browser" was presented where a single implementation, not depending on the human readable documentation of APIs, could interact with 3 platforms: Twitter,[2] Facebook,[3] and ThingsNet.[4]

Note that the STN ontology can be seen as a way of semantically specifying a Web service description, similarly to the goals of Semantic Web service description language WSML [de Bruijn et al., 2005] or the OWL-S ontology [Martin et al., 2004]. However, the main difference is that the STN ontology is focused on a relatively small subset of operations and interactions that characterise social networking platforms. Therefore, it is much easier to implement a generic application that can deal with many STNs, than it is to deal with any type of Web service.

7.4 Hypermedia agents

With STNs, we take advantage of the social network metaphore to improve interactions and management of autonomous things via the Web. However, we wanted to push the idea further by transitioning the multi-agent system principles towards true Web-based MAS. While in the past, multiple attempts to make agents interact on the Web were made in the multi-agent system community, they have essentially used the Web protocol HTTP as a transport layer, ignoring the fact that the Web instantiates the REST architectural style.

A particularly important aspect of REST is the HATEOAS constraint (*Hypermedia As The Engine of Application State*) which stipulates that a client obtaining a representation of a resource should be given options on where to navigate next, by providing hyperlinks in the content of the representation. Additionally, the hyperlinks should be given in a syntactic context that makes it as clear as possible what the relation between the current resource and the linked one is.

These ideas developed over the course of multiple papers to which I contributed, but that were wonderfully driven by Andrei Ciortea [Ciortea et al., 2016b, Ciortea et al., 2016a, Ciortea et al., 2018, Ciortea et al., 2017, Ciortea et al., 2019]. My personal input to this body of work is modest, so I will not try to oversell it by keeping this section brief. In addition to showing proofs of concepts of the idea of hypermedia systems of multiple agents (hypermedia MAS), we were sending a

[2]The most popular microblogging platform as of August 2020. `https://twitter.com/`

[3]The largest social networking platform as of August 2020. `https://www.facebook.com/`

[4]A minimalistic STN platform implemented by Andrei Ciortea for this experiment. `http://github.com/andreiciortea/thingsnet`

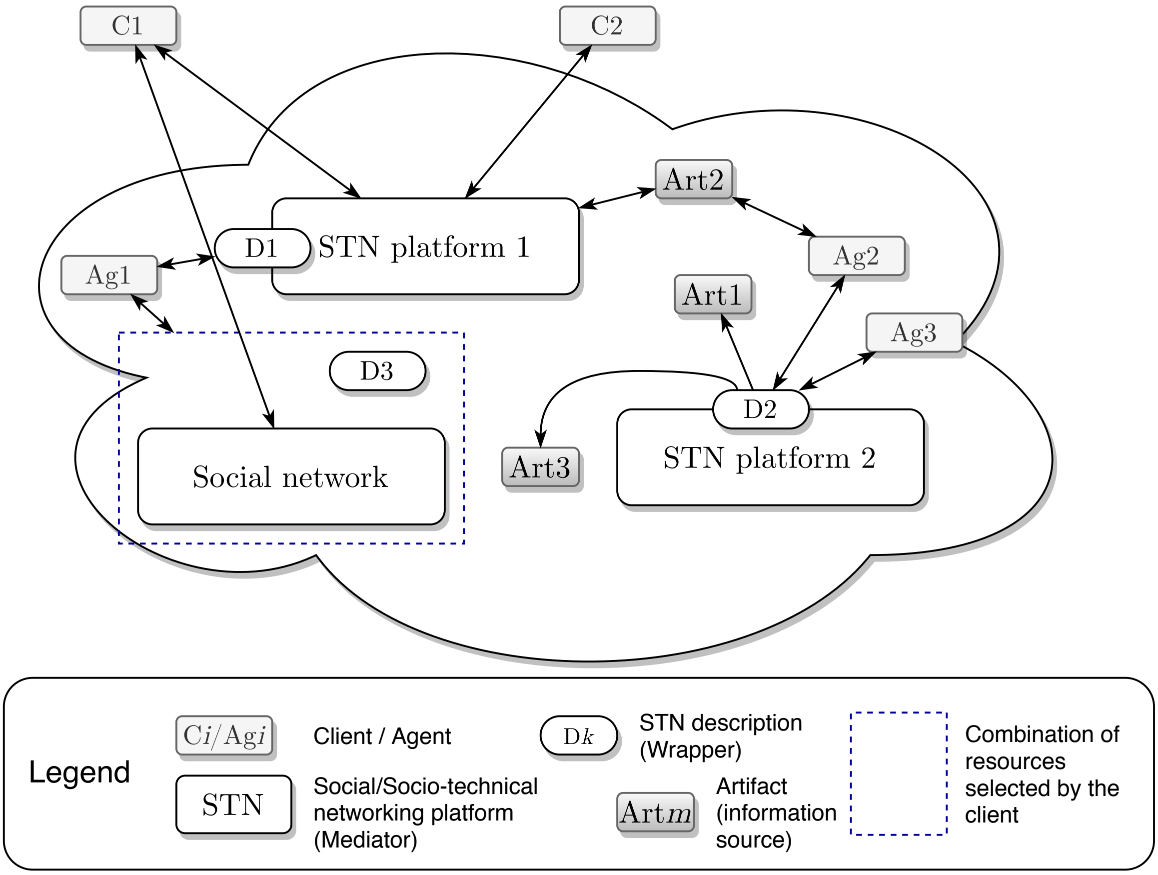


Figure 7.6: Socio-technical networks as an instantiation of loosely coupled mediation

message to the Web community as well as the multi-agent system community to work in unison. I will mention this aspect again in the perspectives in Chapter 10.

7.5 Summary

This work was made in collaboration with Andrei Ciortea (PhD student at Mines Saint-Étienne and Universitatea Politehnica din Bucureşti at the time, now post-doc at University St. Gallen, Switzerland), Olivier Boissier (Professor at Mines Saint-Étienne) and Adina Florea (Professor at Universitatea Politehnica din Bucureşti). My publications related to this work are [Ciortea et al., 2013, Ciortea et al., 2014, Ciortea et al., 2015, Ciortea et al., 2016a, Ciortea et al., 2018, Ciortea et al., 2017]. Figure 7.6 describes how it instantiates the general idea of a loosely coupled mediation approach that I am defending in this thesis.

Information sources are other agents (Ag1, Ag2, Ag3), or artifacts (Art1, Art2, Art3), with whom an agent wants to communicate or use. The STN platform acts as the mediator between them, supported by the STN description (D1, D2, D3) conforming to the STN ontology. Agents that interact with the same platform (e.g., Ag2 and Ag3) can also communicate with each other and make use of artifacts (Art1, Art3). Platforms that are not self-described according to the STN ontology can be externally described (D3), allowing agents to participate in them nonetheless. In this vision, agents can either play the role of clients that make use of an STN platform to discover and use other agents or artifacts; or they can be information sources that other agents, or clients may want to interact with. Clients C1 and C2 in the figure can be simple Web browsers that are used to connect to the STN platforms in the way they are used with traditional social networking platforms.

Chapter 8

Multi Goal Path Finding

One common activity that has greatly benefited from the Web is trip planning. Today, someone can prepare their travelling by almost exclusively interacting with online resources. With enough experience, one does not need to interact with employees of travel agencies, and hardly needs to consult offline information, for all trips to and from places that are sufficiently urbanised.

Moreover, the interactions with traveller information systems are, for the most part, very systematic and largely uniform: enter a starting location (possibly using the GPS to even avoid entering this manually), a destination, a date and time (with the default option of today and now) and this brings the itinerary. Some very common parameters are often available: which mode, with or without connection, comfort class, etc.

Yet, as soon as the trip becomes complex enough to require using multiple transportation companies, or involves additional activities along the way (eating, sleeping, visiting), many manual interventions are required. For instance, a typical scenario for a long distance trip goes as follows: one has to go from departure A (say, in Europe) to destination B (say, in America) and stay there for a few days. In such case, one would **not** provide parameters A and B to a system. Instead, one would try to find a flight by entering airport X that is closest to A as the departure, and airport Y, closest to B, as a destination. Then, *going to a different information system*, one would manually enter X as a destination for an airport shuttle, then proceed similarly at the destination. Finally, one would *change information system* again for hotel reservation, manually entering location B and arrival date that they already entered before on the destination shuttle service.

What prevents a generic trip planning application to automatise these tasks is the lack of interoperability of the various systems involved. At the time of writing this dissertation, there is no artificial intelligence system that is available to guess how to map a destination from one traveller information system to a departure input on another system. Adopting a common standard for transportation information would dramatically improve the situation, but with Oudom Kem's thesis, we proposed a loosely coupled mediation solution that works mostly with existing systems [Kem et al., 2016, Kem, 2018].

The approach relies on a semantic layer that describes the environment in terms of locations, connections between them (for path finding), resources available that can possibly change over time, and activities that can be conducted at certain locations (to allow goal satisfaction along the way). The model for describing the environment is presented in Section 8.1.

In this work, we assume that the information sources are distributed and are designed independently from the route guidance application. Therefore, the path finding plan is built by interacting with multiple autonomous systems. Therefore, we adopted a multi-agent architecture for this. Moreover, the reliance on Web-based information sources may cause latency that can reduce the efficiency, even more so that we want the system to adapt to dynamic changes in the environment, especially in the context of indoor navigation, where for instance an elevator can become out of service, or other resources for accomplishing certain goals may disappear or move. The resulting solution that addresses these issues is briefly described in Section 8.2.

The use of a combination of Web technologies, ontologies, and multi-agent systems make this domain an especially suitable application for socio-technical networks, and this is why this is presented here.

8.1 Semantic representation of the environment

In [Kem et al., 2017b, Kem et al., 2017d], we introduced a knowledge model for representing ubiquitous environment, its topology, its resources, activities, etc. Figure 8.1 shows it.

Before giving a concrete representation of the model for computation, we defined an abstract definition of an *environment*. Given that we focused on ubiquitous environments in relation to route planning, we are especially interested in locations, cyberphysical entities (in particular connected devices), in which location they are situated at any time, the activities that can be performed at which places or with which cyberphysical entities, how to navigate between locations. Also, to ensure better performance in finding the right path from a departure point to a destination, we provide a hierarchical organisation of the topology. That is, we represent mereological relations between smaller locations (a room) and larger ones (a zone, building, wing, area). If pushed to the extreme, the model could contain districts, cities, regions, countries, continents, and the world. This helps finding a path efficiently by first looking at the locations that are in the same hierarchical coverage as both the departure point and destination. I now provide the formalised definition.[1]

Definition 15 (Knowledge model of the environment) *We assume there is a fixed set of goals G and activities A, with a relation $K \subseteq G \times A$ between goals and activities. The model of the environment is an evolving tuple $E_t = (L_t, C_t, OS_t, CPSE_t, Sit_t, Act_t)$ where t is a time point and:*

- *L_t is a set of locations;*
- *$C_t \subseteq L_t \times L_t$ are the connections between locations;*
- *OS_t is the hierarchical organisation of the places, comprising locations of L_t as leaves in the hierarchy, and with only one top level place;*
- *$CPSE_t$ are cyber, physical, social entities;*
- *$Sit_t : CPSE_t \to 2^{L_t \cup C_t}$ relates CPS entities to either a location or a connection;*

[1] Here, I slightly diverge from the definition given in Oudom's thesis [Kem, 2018].

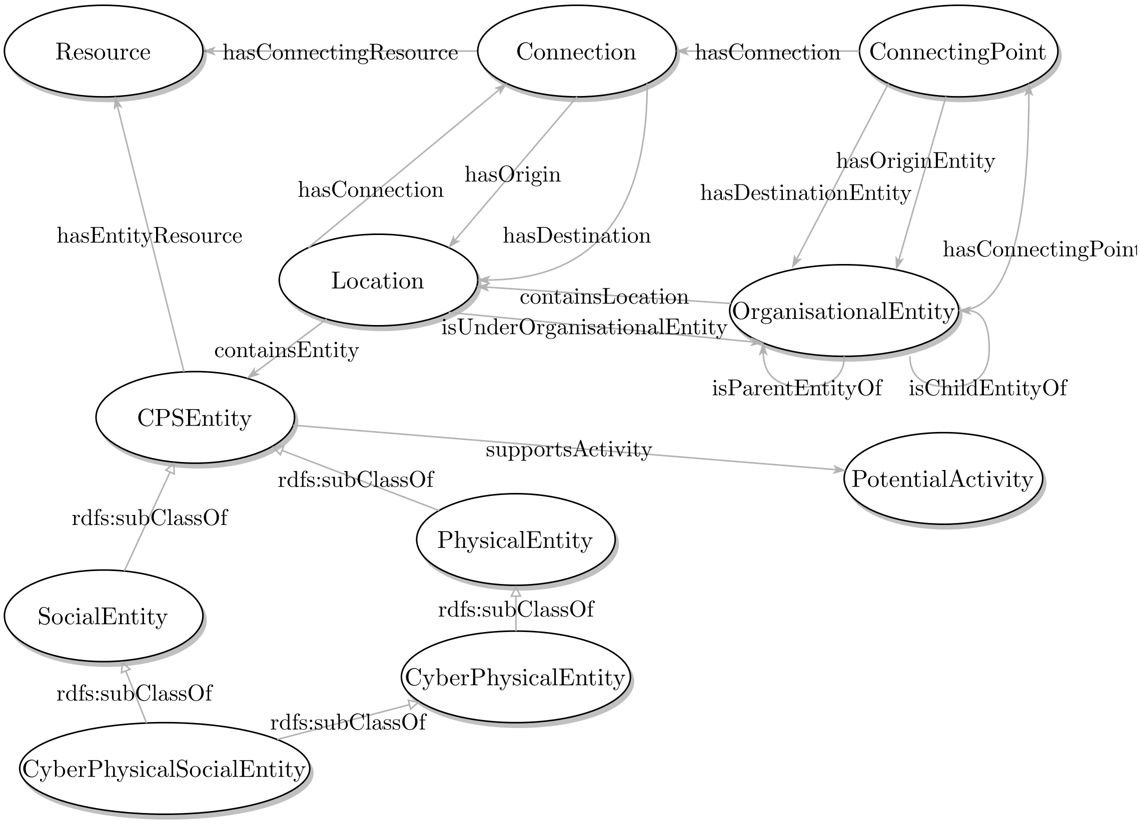


Figure 8.1: Ubiquitous Environment Abstraction Ontology, or UbiEnv for short (reproduced from Oudom Kem's thesis [Kem, 2018])

- $Act_t : CPSE_t \rightarrow 2^A$ *defines the activities that can be performed using a CPS entity.*

Note that cyber, physical, social entities may be located at several places, or connection at the same time, either because they are large, or because they are completely digital, such that their location is virtual. CPSEs can provide multiple activities at once.

In order to make possible the use of the algorithm described in Section 8.2, we make an additional assumption: that CPSEs are partitioned between location-based CPSEs (that is, $Sit_L(e) \subseteq L_t$) and connection entities (that is, $Sit_L(e) \subseteq C_t$) and only location-based CPSEs can provide activities.

The model was then made more concrete by designing an OWL ontology; the Ubiquitous Environment Abstraction Ontology (UbiEnv ontology) that is partially shown in Figure 8.1. The ontology is generic and abstract, but can be extended to provide more precise classes and relationships for certain types of environment. As an example, Oudom Kem developed a smart airport ontology module that supports the scenario discussed in next section.

8.2 Multi-goal path finding

The semantic model described previously does not do anything by itself. It must be brought forth to applications that require knowledge about the environment, especially related to travel or route guidance. To demonstrate the usefulness of the model, we devised a scenario that takes place in an airport equipped with ubiquitous software and devices.

We imagine the case of people who are not only travelling through a large airport, but in passing, have to fulfil various goals, such as having a meal, taking a trolley, finding power for their phone or laptop, going to the bathroom, etc. Even if these tasks are each easy to do for a person, it can easily become inefficient to achieve the goals in a fixed pre-defined order (as it may lead to going back from where one started), or to go directly towards the main destination with the hope that one will be able to achieve the goals along the way. A person that does not have previous knowledge of the airport is likely to choose a suboptimal plan. Furthermore, the airport environment can evolve dynamically: trolleys can be taken and move, elevators may stop to work, airport pathways are often reconfigured by displacing movable panels, etc.

In principle, a dedicated mobile application could be built and installed on the phone, with hardwired connections to a central server in the airport that could provide very efficient path finding supports, even with the additional constraints of goals. However, in our work, we wanted to add certain desirable caracteristics:

- first, the user should be able to preserve their privacy, so they should not transmit their goals to a central server, not even their destination if they do not wish to disclose it;

- second, the application that provides multi-goal path finding support should work in any environment that is described according to the model of previous section.

The first constraint means that the algorithm should be running mainly on the mobile device, and communicate as locally as possible, for instance, only with the restaurant information system that they would like to visit. The second constraint means that there should be only one application to install for every possible place that offers a semantic description of their locations and resources.

These constraints directed us towards a combination of multi-agent technologies, such that both the user agent and the ubiquitous resources can react to the evolution of the environment or situation without the need for a global, omniscient entity; and on Web of things technologies to model the cyber, plysical and social entities.

For the path finding algorithm itself, I did not take much part in its definition, so I will simply point to the relevant publications: [Kem et al., 2017c, Kem et al., 2017a]. One important aspect is that there exist search agents that are associated with locations; resource agents that can provide the status of resources or connections between locations in a uniform way, regardless of how the resource works; and network agents that are in charge of relaying information between different hierarchical places and splitting the workload as evenly as possible. An additional constraint emerged from the fact that we used Web resources dynamically during path construction, that can create latency. The algorithm is designed to avoid locking situations when a resource is occupied computing part of the solution, or retrieving information about places, goals, and so on.

Note that all this part of Oudom Kem's doctoral studies is very much focused on algorithms, not on a particular architecture, so their development is mostly orthogonal to the notion of mediation. However, a crucial aspect of the algorithms

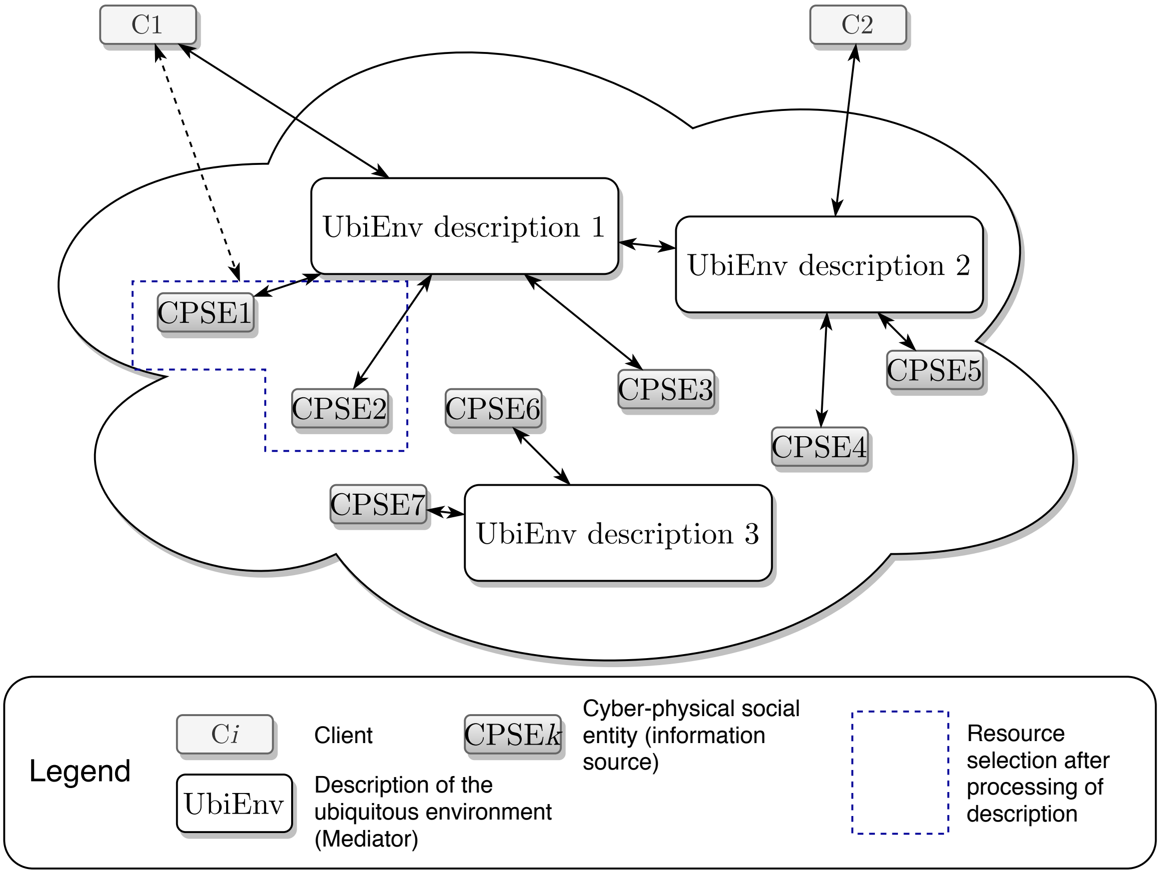


Figure 8.2: A model for uniqitous environment instantiating loosely coupled mediation

with respect to interoperability is that the initialisation requires a ubiquitous environment description in the form of an RDF graph that conforms to the UbiEnv ontology. This means that it requires a single application, and a mere description, to launch the multi-goal path finding algorithm in a completely unknown environment where resources are discovered on the fly.

8.3 Summary

This work was made in collaboration with Oudom Kem (PhD student at Mines Saint-Étienne at the time, now research engineer at CEA, France) and Flavien Balbo (Professor at Mines Saint-Étienne). My publications related to this work are [Kem et al., 2016, Kem et al., 2017a, Kem et al., 2017b, Kem et al., 2017c, Kem et al., 2017d]. Figure 8.2 describes how it instantiates the general idea of a loosely coupled mediation approach that I am defending in this thesis. However in this case, the instantiation is less evident because at runtime, after retrieving all the relevant information from the semantic model of the environment, the path finding process only requires agents to communicate directly between them, while resource agents (possibly running on the Web) become the intermediary to the relevant information sources. So I only show the situation before running the path finding process (actually, a situation that could be common to other use cases of the semantic model).

The information sources are ubiquitous entities that can tell where given goals can be achieved, or whether places are accessible from another. These sources can be exploited uniformly thanks to the semantic representation of the environment, that can be seen as the mediator between the client application (possibly a path-finding/route planning system) and digital entities in the environment.

At runtime, the path finding algorithm create search agents, network agents and resource agents, and does not need to mediate information through the environment description. Instead, the information is mediated through resource agents, that can access Web resources to tell the search agent all that is relevant to the search. In a sense, resource agents provide a uniform access to the information sources, so they play the role of mediators.

Admittedly, this view of the path-finding approach may be seen as a bit of a stretch, but I think it mostly comes from the choice of algorithm and implementation decisions. The experimental proof of concept did not require a real Web-based environment, and there was not enough time to investigate a more flexible design. However, I believe that the model for ubiquitous environment, the choice of an agent-based approach, combined with some Web resources, is very much in line with the idea of socio-technical networks and could have been implemented along the same paradigm. This is why I put this chapter in this part of the thesis.

Summary and future work

The two contributions of this part enable further interoperability at the process level, allowing better decisions. The first one, about socio-technical networks and a social Web of Things, sets the infrastructure for easier interactions, management, and monitoring of things, agents, and services. The second one can be seen as an instantiation of STNs for the use case of trip planning, especially in ubiquitous environment, where cyber-physical entities may be part of the social Web of Things.

The Social Web of Things, as envisioned in Andrei Ciortea's thesis, relies on a mediation approach by way of socio-technical networking platforms. This mediation is loosely coupled since there can coexist may such platforms, and their formal description using the STN ontology can be provided externally by a third party. The platform can provide interactions with an open and varying set of resources, some of which are themselves clients to the platform. While our work on STNs is generic, it is assumed that more specialised STN platforms would exist, that would have a more detailed description that would support the specific tasks that they provide interoperability for.

As an example, traveller information systems are good use cases for STNs, as they require interacting with various transportation systems and other resources along the way, with recurring tasks such as path finding. In Oudom Kem's doctoral study summarised in Chapter 8, we noticed the strong connections with Andrei's work, as cyber-physical social entities may be endowed with autonomy, or simply act as artifacts, available from an STN platform. However, we were not able to, in the restricted amount of time, to make the connection explicit neither in the model nor the implementation. We have nonetheless devised the specification and the algorithm with a loose coupling between its components, with a description of the topology of the environment disconnected with the descriptions of cyber-physical and social entities that can offer activities that satisfy goals.

In both cases, these contributions make use of declarative knowledge informing the clients about possible actions, thus supporting better process interoperability for decision making.

Among the perspectives opened by these contributions, we can identify the following:

- In terms of socio-technical networks, having a uniform way of discovering and interacting with social platforms is just a first step towards interoperable processes. A crucial part that have been little addressed in Andrei's thesis concerns the normative part: with a proper description of what agents must do and must not do, certain tasks could be achieved in a uniform way. As a possible example of how the normative part can be systematically described, there is the MOISE meta-model that serves to describe organisations in which agents have roles, missions, obligations. However, following the general vision of this thesis, we would need to provide a standardised organisation specification, in the form of an ontology. This was the topic of Masters student Mădălina Zarafin that I co-supervised with Olivier Boissier [Zarafin et al., 2012], but we did not investigate it further. Orthogonal to interoperability is the problem of security and regulation. STNs should enable automated processes to act on platforms, whereas most social networking platforms provide measures to prevent robots to join and interact with people. Note that the Wikipedia community is an example of an STN, with readers, editors, moderators and *bots*, where regulation is entirely defined by people and social norms. Allowing software agents to participate in the regulation would require formal languages ideally conforming to Semantic Web standards.

- In terms of path finding, and more generally, traveller information systems, it would be worthwhile adopting the social Web of Things approach, providing trip planning services as entities on STNs, along with resources such as restaurants, petrol stations, shops, etc. This would allow much more flexibility. A challenge resulting from this approach is how the system would scale. If there was a platform where all trip planning services would be registered, from local buses to international airlines, how could a journey could be composed from one end to the other, selecting and interrogating the right combination of services? The challenge becomes even harder when this is combined with multiple goals.

Part IV

Synthesis and Research Plan

Chapter 9

Synthesis

Now I summarise and relate all the contributions that I presented. As I said in the introduction, my overall goal is to empower systems with more interoperability. In this chapter, I will start by showing what each parts have achieved in this regard. This can be visualised in Figure 9.1, a diagram reproducing the layers of Figure 3 with additional details on what kind of heterogeneity is mitigated by each part. Then I will focus once more on the means towards these achievement, namely the loosely coupled mediation approach.

9.1 Interoperability achievements

In this thesis, I made the assumption that the World Wide Web provides a uniform communication medium solving the network layer of interoperability, in spite of the wide variety of information systems, software, and hardware running at different places. Then I addressed the upper layers as follows:

- In Part I, I addressed the problem of heterogeneity of data models.
  - Chapter 1 describes **SPARQL-Generate**, a declarative transformation language that can take any (combination of) data formats as input and generate RDF from it. The strengths of the language are its flexibility, relative simplicity of usage for a semantic web literate because of its extending SPARQL, and its relative simplicity of implementation on top of an existing SPARQL engine. With appropriate SPARQL-Generate transformations, all data can be seen as RDF data.
  - Chapter 2 describes **PolyWeb** a technique and system for querying in SPARQL multiple and distributed data bases that use different data models. SPARQL-Generate and PolyWeb can be viewed as two sides of the same coin: one is transforming the data to a common data model, the other is transforming the queries from a common query language. The principle of PolyWeb is to ensure that all databases can be queried as if they were all one triplestore.
  - Chapter 3 addresses a different aspect of data interoperability that can take advantage of both SPARQL-Generate and PolyWeb: that of providing a uniform API to data access in RDF using Linked Data Platforms (LDP). An LDP is providing an organised view of data pieces, and the

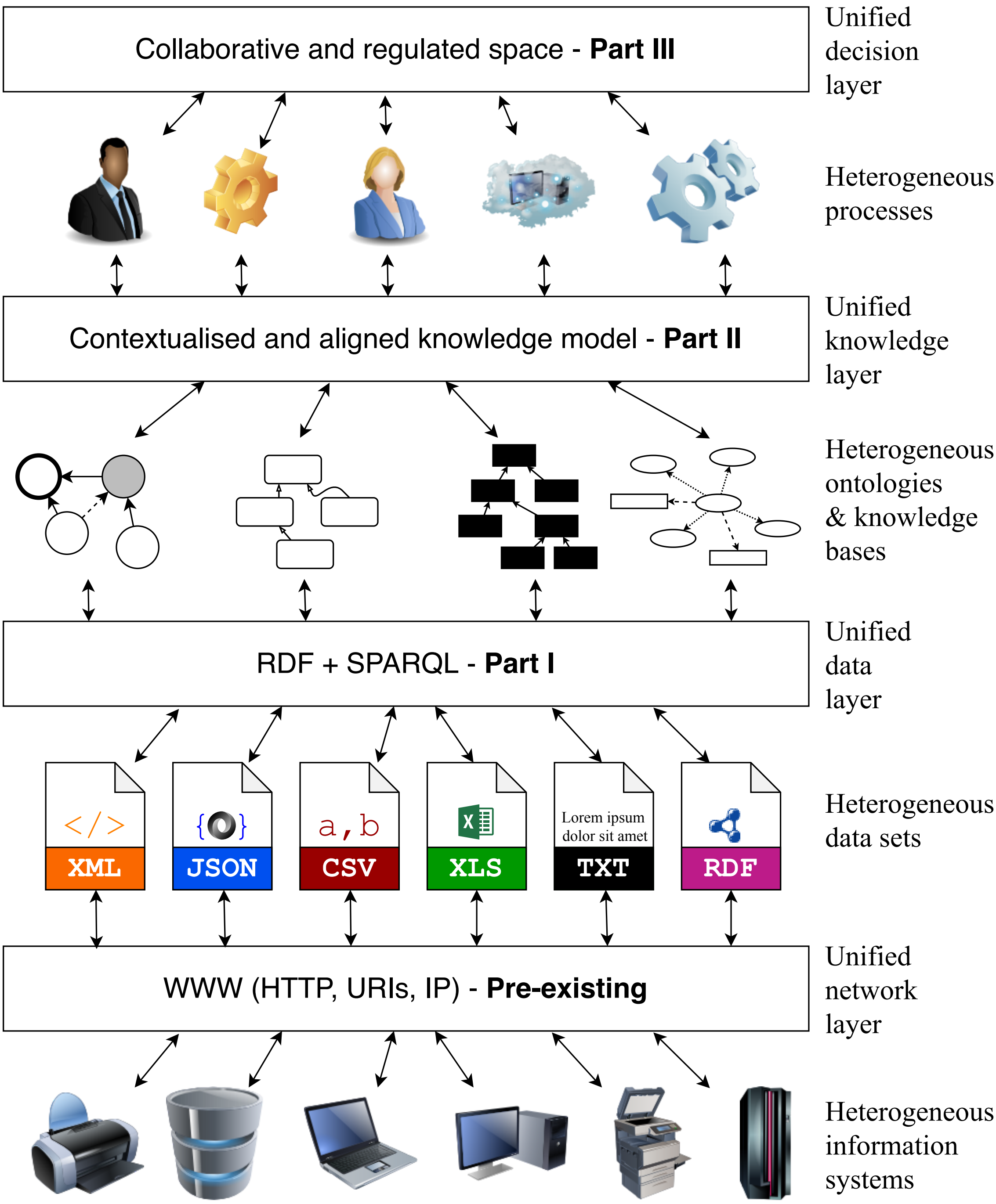


Figure 9.1: Visual presentation of the problems addressed in the dissertation

language **LDP-DL** describd in the chapter facilitates its deployment. It also enables read-only dynamic LDP, where data is always in sync with (possibly remote) data sources.

- In Part II, I addressed the problem of heterogeneity of ontologies and knowledge bases.
  - Chapter 4 presents how different ontologies can be combined by way of ontology alignments into a unified structure called **networks of aligned ontologies**. The chapter discusses the semantics of, and reasoning with,

such structures, as well as extensions towards multi-level networked knowledge. These notions allow one to view a multiple of ontologies as a single knowledge model.

  – Chapter 5 describes **Annotated RDFS** and its derived query language **AnQL**. The formalism consists in assigning a quantified element of context to every RDF statements. This enables explicit cross context reasoning and querying.
  – Chapter 6 is also related to the representation of context, but it employs the notion of reification. The chapter formalises reification and exposes desirable properties of reification methods with respect to reasoning. I present one such method, **NdFluent**, that fulfils the entailment preservation property. Both Chapter 5 and Chapter 6 deal with the representation of context, the former with the advantage of explicit cross context reasoning, the latter with the advantage of being fully compatible with the standards. Each uniformise the exploitation of multi-contextual information sources.

- In Part III, I addressed the problem of heterogeneity of processes.
  – Chapter 7 introduces the notion of **socio-technical networks** (STN) in a **social Web of Things**, where things, whether autonomous or simple artifacts, are first class citizens. Using this notion and the associated STN platforms, that generalise online social networking platforms and that are semantically described, a varity of agents can collaborate in a uniform space that, in principle, can be regulated and organised according to certain goals and norms.
  – Chapter 8 provides an example of a specific task that requires a collaborative space, namely **multi-goal path finding**, the problem of finding a path from one point to another while realising certain goals on the way, with the added difficulty that the environment is open and dynamic. Multiple agents dedicated to solving parts of the path finding or goal finding problem collaborates with a shared understanding of the description of the environment.

Now let us report on how well these contributions satisfy the vision of loosely coupled mediation.

9.2 Loosely coupled mediation

Loosely coupled mediation means that the components of the mediation architecture can be developed independently from each others, can be distributed, retrieved and composed dynamically, and usually are based on declarative specifications rather than specific implementations.

With SPARQL-Generate, transformations can be saved to files and dynamically referenced by their URI. In the current version (improved from the one described in [Lefrançois et al., 2017c]), transformations can be modularised and composed, and can select distributed data sources dynamically in function of other data. I consider the design an implementation of SPARQL-Generate an exemplary instance of loosely coupled mediation.

With PolyWeb, the proof of concept is much more tightly coupled, but the principles of the approach allows for the disconnection of data summaries and transformations. Different forms of data summaries and different transformations could co-exist to allow different optimisations on source selection time, or local query answering time, or on global join time, or data transfer sizes. At the moment, I would reconsider PolyWeb a good example of a traditional mediation architecture.

With LDP-DL and its workflow, a data platform may rely on modular and distributed design documents, that themselves tap into diverse data sources. Static deployment can work with any conforming LDP implementations. With this regard, the architecture is very well decoupled. For dynamic LDPs, we currently only have one implementation that must make use of the output of our LDPizer tool. However, the result of LDPizing is a file with an open specification that is portable to various instances of dynamic LDPs. Moreover, since LDP-DL makes use of RDF data sources or SPARQL endpoints, we can envision mediation in cascade, where SPARQL-Generate would stand between heterogeneous data sources and the LDPizer, and PolyWeb would stand between heterogeneous database systems and the LDPizer.

Networks of aligned ontologies (NAOs), are abstract structures that are not tied to a specific architecture. However, the distinction between ontologies and alignments favours a mediation approach. Distributed ontologies can be selected and, based on ontology pairs, appropriate alignments provide the global view of the network. Instead, alignments could be selected a priori, based on desired term correspondences, from which the ontologies would be referenced. Alignments can be generated with a matching process or directly from a repository, for instance thanks to an alignment server. The alignment server is then the mediator that allow to construct the ontology network from given sets of ontologies. Morever, if the semantic of NAOs is defined as in IDDL, reasoning over an NAO can be made by interacting with local reasoners as black boxes. This way, the global reasoner only knows about ontology alignments and may ignore the ontology content. Thus, we can say that the level of decoupling is dependent on the application, and more importantly, on how NAOs are interpreted.

In Annotated RDFS, decoupling can occur if the annotation can be externalised. While this is in principle possible, we did not explore this idea and the proof of concept implementation assumed that the data and their annotations are all contained together. For the query language AnQL, decoupling could be even harder, as the queries contain the annotations, and this has to be matched against the data at the query endpoint. However, with some dimensions of context, it is possible to assign annotations automatically from non-annotated sources: for instance, provenance can be attributed based on the location of the source file; access control can be affected based on the credential of the user (with SPARQL queries transformed into AnQL with access rights annotations); fuzzy or confidence values can stem from PageRank-style algorithms or machine learning induction; etc. So, although the coupling is currently tight, there is a potential for looser coupling with more advanced tooling.

Reification in general and NdFluents in particular require a description of context separate from the data. A reification method takes these two pieces and aggregate them into a single graph or knowledge base. However, since there is no standard way of describing the context of a statement, a reification technique must ensure an agreement between the structure of its input and the structure of the context description when provided by a third party. So the approach is relatively loosely coupled.

Socio-technical networks can be implemented such that the agents and the platforms can be implemented independently from one another. STN platforms simply require that a description exists conforming to the STN ontology. Agents simply have to implement the ontology specification. Even existing platforms that are not designed according to the STN concept can be described using the STN ontology, so that agent that implements the STN specification generically can connect to them. However, at the moment, there is no guidelines or constraints on how the agents can cooperate and what formal language they need to rely on for their communications through the platform.

The multi-goal path finding approach relies on resource descriptions that can be provided independently of any tasks and that are retrieved dynamically. The proof of concepts implementation for experiments used a tightly coupled design, but the task would fit well with the idea of STN. The algorithm currently does not easily allow separate developments of search agents, that would be readily available to find parts of the path for a subpart of the environment, but overall, there is a certain degree of decoupling.

In summary, I hope I have convinced that my contributions have enabled more interoperability by way of a loosely coupled mediation approach. Next, I will discuss some of the research directions I wish to pursue beyond this.

Chapter 10

Research Plan

In this final chapter, I describe a few research directions that I want to pursue in the future, that are related to the work presented before. I envisage to further still the research at the three layers of introperability: data, knowledge, and processes. But while I pinpoint specific topics, my insights gained over the years lead me to label my long term goal with the concept of *socio-technical systems*. Indeed, interoperability depends more on people accepting to adhere to a common interface than on a smart technical solution.

I wish to develop solutions that have an understandable, explicit, precise, and formal specification that allows people to agree on the interface while leaving freedom on the concrete implementation. I also want that the specification be precise and formal enough that they can be exploited by software agents, even to the point, sometimes, that they automatically and autonomously agree with one another.

To this long term aim, I propose to address first the following topics in the next 5 years:

- Personal knowledge graphs: a graph of distributed, interlinked, secured data about one person, for the consumption and usage of the person only (Section 10.1);
- Semantic content negotiation: a mechanism that ensures that a client in need for information can specify to the server constraints on what kind of content it wants, and the server can negotiate the best offering accordingly (Section 10.2);
- Agents on the Semantic Web: in spite of the contributions discussed in Chapter 7, there is still much research to be done on real hypermedia multi-agent systems (Section 10.3).

10.1 Personal Knowledge Graphs

Knowledge graphs (KGs), as commonly understood now, have been popularised by Google's 2012 announcement of the Google Knowledge Graph [Singhal, 2012]. Nowadays, people often understand the concept of KG as a large graph-like knowledge base that covers a broad range of information on a certain topic, or a broad range of topics, such as the entire knowledge of a company across its services; or comprehensive knowledge of all scholarly publications of the past centuries; or the sum of all knowledge from Wikipedia, translated to structured data as a graph.

In [Hogan et al., 2020], a broad definition of KGs was given like so: "a graph of data intended to accumulate and convey knowledge of the real world, whose nodes represent entities of interest and whose edges represent relations between these entities".

What I call *personal knowledge graphs* are KGs of personal data that encompasses all aspects of a person that leave digital traces. My hypothesis is that people would benefit much from having access to their own personal knowledge graphs. Today, people have pieces of knowledge about themselves, about their activities, distributed across countless databases, that are extremely difficult to combine: health data in patient records at different hospitals or doctors; energy consumption (water, electricity bills); commercial transactions with online stores; payrolls; taxes; bank accounts; personal address book; and much more. In many cases today, these things tend to become accessible online, but only one source can be accessed at a time. Correlations and coreferences are almost impossible to detect.

As an example, a line on a bank statement cannot be linked automatically to a receipt on an online payment site, even though the account's holder have all necessary information, in digital form, to create this link.

Personal knowledge graphs have very challenging requirements:

- They must be highly secured: access controlled, encrypted;
- They must be decentralised: parts of the personal KG must be detained by external organisations or companies (receipts, health records, payrolls);
- They must be processable, and partly sharable: its owner should be able to grant access to pieces of it for added value applications.

The first two items are satisfied by the current situation: personal and sensitive information such as health records etc. are well secured and very decentralised. However, their processing is almost completely limited to allowing the organisation in charge of the subset (hospital for health data, employer for payrol, etc.) to do what their business is supposed to do. The owner of the personal knowledge graph has almost no way to query, analyse, interlink those pieces together.

In more detail, I would envisage working on the following aspects:

- models for personal data representation (ontologies);
- ways of sharing parts of the graphs with identified third parties in a secure way;
- tracing provenance, signing subgraphs to avoid faking information;
- aligning representations between users, who may not use the same knowledge model (ontologies);
- tools to manage such data structures.

Clearly, these research directions can benefit from my past research on data and knowledge interoperability of Part I and Part II. Some of these aspects will be tackle in a coordination action called *Distributed Knowledge Graphs* funded by the COST office, started in September 2020, in which I am leading a working group on data prosumers.[1] Besides, the development of Solid,[2] a project towards the decentralisation and interlinking of personal data led by Tim Berners-Lee, should provide the software infrastructure enabling these research developments.

10.2 Semantic content negotiation

Content negotiation on the Web allows a data consumer (client) to tell a data provider (server) what it expects in terms of format, language, encoding, security [Fielding and Reschke, 2014]. In return, the server provides data that meets these expectations when it can, or indicates an alternative, for example, that the same information resource is available in another format. However, even when the client's request is satisfied, this does not mean that the client is able to interpret the data correctly. For example, for the same data format, several forms, structures or schemas may exist. The client may wish to obtain data that conforms to conventional terminology and has certain logical and structural properties.

In particular, in applications that rely on Semantic Web technologies, such as RDF and OWL, an application could expect graph-based data that conform to a specific ontology, or that fit a certain data shape, or that is compatible with a given entailment regime. In environments such as the Web of Things, strong constraints may impose requirements on the server or client side due to processing power, bandwidth, or memory limitations.

The main challenge is to find out how clients and servers can agree on the expected (client-side) or provided (server-side) content automatically, so without the developer of the client application having to contact the server manager, or read natural language documentation. The objective of this research would be to

- determine what properties the client and the server could agree on to negotiate content beyond its simple syntax;
- define the mechanism (in terms of protocol and algorithm) allowing the client to announce its expectations and how the server reacts to these requirements;
- consider making negotiation more flexible by introducing an external service in charge of mediation between client and server (data transformation, inference or validation system);
- introduce a declarative formalism allowing the server to describe the logical and structural properties of its data (possibly relying on SPARQL 1.1 Service Description [Williams, 2013], Thing Description [Kaebisch et al., 2020], or various forms of content descriptions, for instance, [Thuluva et al., 2018].

Previous and ongoing work has been and are being conducted in extending content negotiation towards more flexibility such as [Holtman and Mutz, 1998, Lefrançois, 2018, Svensson et al., 2019]. This research direction is more connected to interoperability at the knowledge layer (Part II), but may also bring it close to the decision layer (Part III), depending on what properties the negotiation focuses on.

[1] COST Action CA19134 - Distributed Knowledge Graphs `https://www.cost.eu/actions/CA19134`

[2] `https://inrupt.com/solid`

On this particular topic, I obtained funding from Mines Saint-Étienne for a PhD student that should start investigating this field before the end of 2020. As soon as I will be habilitated, I will become the main advisor of this student. Maxime Lefrançois will be co-supervising this work.

10.3 Agents on the Semantic Web

The third research direction consists in integrating autonomous agent technologies with the Web and the Semantic Web, following the goals presented in Section 7.4. In the past eight years, I worked in a team that comprises multi-agent system experts, and we tried to find synergies between our team's work on Semantic Web technologies and autonomous agents systems, as has been shown in Part III. Agents can use the web as a mediation overlay to facilitate their interactions, coordinations, goal-achievements, following the idea of hypermedia agents as in [Ciortea et al., 2019]. The HyperAgents project,[3] has just started at the time of writing this thesis and investigates these ideas.

One of the main objectives of the project is to define declarative languages and mechanisms for specifying, enacting, and regulating interactions among people and autonomous agents in Hypermedia MAS. As part of this, we are collectively trying to reach a shared understanding of different dimensions of multi-agent systems that we want to formalise as Web ontologies: core MAS concepts, interaction concepts, regulation concepts, organisational and social concepts.

In particular, I would like to extend socio-technical networks such that an STN platform can present a formal specification of its norms, so as to let autonomous agents behave accordingly. Furthermore, the platform should be able to expose an organisational specification in which agents can enact certain roles in order to achieve objectives and perform missions, with minimal intervention of humans, and as little ad hoc implementation as possible.

10.4 Side goals

In addition to the previous main goals that continue to support the thesis I am defending in this dissertation, I have multiple short-to-medium term (ongoing, planned, or potential) objectives that are not conforming exactly to the vision presented here. Among them are: defining and extending datatypes for scientific data, such as physical quantities, arrays or matrices of values with support for their processing from inside the query language SPARQL, and support for reasoning over them. This would continue previous work in [Lefrançois and Zimmermann, 2016, Lefrançois and Zimmermann, 2018]; conducting a thorough review of the concept of literals in RDF, in Web Data, in standards and specifications, and in research work, along the same lines as the article [Hogan et al., 2014] that was studying the concept of blank nodes in its most minute details; more generally, continue to seize opportunities to collaborate on research that contribute to advancing the vision of the Semantic Web, with a particular attention on defining specifications that are implementation-agnostic and strongly based on Web standards.

[3]HyperAgents: `http://hyperagents.gitlab.emse.fr/`

Conclusion

While I tried to give a cohesive view on my research over the past twelve years, I feel like I am presenting very disparate pieces of work. The techniques and formalisms involved in mediating heterogeneous data, for instance, are very different from those required for contextual reasoning. Even in this dissertation that compiles my main research areas, I left aside many publications that are not as easy to embed in this unifying thesis (such as [Hogan et al., 2010, Hogan et al., 2012] on entity reconciliation, [Sahay et al., 2011] on ontology and alignment methodology, [Sorici et al., 2015] on ambient intelligence, [Lefrançois and Zimmermann, 2016] on custom datatypes, [Audeh et al., 2017] on Web scraping, [Zimmermann, 2010] on ontology recommendation, [Lefrançois and Zimmermann, 2018] on representing units of measures, to name a few of the most cited ones). Perhaps, more focused research activities would have led me to more outstanding results. I often regret that I do not have the time and energy to dig deeper in all of the topics discussed before. I hope the last chapter of this dissertation opens enough directions of research to prove this.

However, I believe that I have always been more of a great *enabler* instead of an accomplished achiever. By this I mean that my best contributions did not come from myself alone, but from the external impulse of someone else's ideas, to whom I offered my insights. I hope that as a result I managed to improve the ideas to their best potential. With this habilitation, I wish to have even more opportunities to bring my expertise into collaborative work that will eventually impact science in the narrow space of my competencies. And perhaps, with hope and optimism, in spite of my own humility, enlarge the importance of my contributions to enable broader repercussions on society.

Bibliography

[AAM, 2002] (2002). *The First International Joint Conference on Autonomous Agents & Multiagent Systems, AAMAS 2002, July 15-19, 2002, Bologna, Italy, Proceedings*. ACM Press.

[IRI, 2006] (2006). *Proceedings of the 2006 IEEE International Conference on Information Reuse and Integration, IRI - 2006: Heuristic Systems Engineering, September 16-18, 2006, Waikoloa, Hawaii, USA*. IEEE Systems, Man, and Cybernetics Society.

[WI-, 2015] (2015). *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, WI-IAT 2015, Singapore, December 6-9, 2015 - Volume I*. IEEE Computer Society.

[WI:, 2016] (2016). *2016 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2016, Omaha, NE, USA, October 13-16, 2016*. IEEE Computer Society.

[WoT, 2016] (2016). *Proceedings of the Seventh International Workshop on the Web of Things, WoT 2016, Stuttgart, Germany, November 7, 2016*. ACM Press.

[Antoniou et al., 2011] Antoniou, G., Grobelnik, M., Simperl, E. P. B., Parsia, B., Plexousakis, D., De Leenheer, P., and Pan, J. Z., editors (2011). *The Semanic Web: Research and Applications - 8th Extended Semantic Web Conference, ESWC 2011, Heraklion, Crete, Greece, May 29 - June 2, 2011, Proceedings, Part II*, volume 6644 of *Lecture Notes in Computer Science*. Springer.

[Aroyo and Gandon, 2017] Aroyo, L. and Gandon, F., editors (2017). *Proceedings of the Doctoral Consortium at the 16th International Semantic Web Conference (ISWC 2017) co-located with 16th International Semantic Web Conference (ISWC 2017), Vienna, Austria, October 22nd, 2017*, volume 1962 of *CEUR Workshop Proceedings*. Sun SITE Central Europe (CEUR).

[Aroyo et al., 2009] Aroyo, L., Traverso, P., Ciravegna, F., Cimiano, P., Heath, T., Hyvönen, E., Mizoguchi, R., Oren, E., Sabou, M., and Paslaru Bontas Simperl, E., editors (2009). *The Semantic Web: Research and Applications, 6th European Semantic Web Conference, ESWC 2009, Heraklion, Crete, Greece, May 31-June 4, 2009, Proceedings*, volume 5554. Springer.

[Audeh et al., 2017] Audeh, B., Neigbeder, M., Zimmermann, A., Jaillon, P., and Bousquet, C. (2017). Vigi4Med Scraper: A Framework for Web Forum Structured Data Extraction and Semantic Representation. *PLOS ONE*, 12(1).

[Bail et al., 2013] Bail, S., Glimm, B., Gonçalves, R., Jiménez-Ruiz, E., Kazakov, Y., Matentzoglu, N., and Parsia, B., editors (2013). *Informal Proceedings of the 2nd International Workshop on OWL Reasoner Evaluation (ORE-2013)*, volume 1015 of *CEUR Workshop Proceedings*. Sun SITE Central Europe (CEUR).

[Bakerally, 2017] Bakerally, M. N. (2017). Towards automatic deployment of data platforms from existing data sources. In [Aroyo and Gandon, 2017].

[Bakerally, 2018] Bakerally, N. (2018). *Generation of Linked Data Platforms from existing data sources in highly decentralized information ecosystems*. PhD thesis, École des mines de Saint-Étienne, Saint-Étienne, France.

[Bakerally et al., 2016] Bakerally, N., Boissier, O., and Zimmermann, A. (2016). Smart City Artifact Web Portal. In [Sack et al., 2016b], pages 172–177.

[Bakerally and Zimmermann, 2017] Bakerally, N. and Zimmermann, A. (2017). A system to automatize the deployment of data in Linked Data Platforms. In [Nikitina et al., 2017].

[Bakerally et al., 2018a] Bakerally, N., Zimmermann, A., and Boissier, O. (2018a). A Workflow for Generation of LDP. In [Gangemi et al., 2018a], pages 142–147.

[Bakerally et al., 2018b] Bakerally, N., Zimmermann, A., and Boissier, O. (2018b). LDP-DL: A Language to Define the Design of Linked Data Platforms. In [Gangemi et al., 2018b], pages 33–49.

[Bellatreche et al., 2016] Bellatreche, L., Pastor, O., Almendros-Jiménez, J. M., and Aït Ameur, Y., editors (2016). *Model and Data Engineering - 6th International Conference, MEDI 2016, Almería, Spain, September 21-23, 2016, Proceedings*, volume 9893 of *Lecture Notes in Computer Science*. Springer.

[Bennett and Fellbaum, 2006] Bennett, B. and Fellbaum, C. D., editors (2006). *Formal Ontology in Information Systems, Proceedings of the Fourth International Conference, FOIS 2006, Baltimore, Maryland, USA, November 9-11, 2006*, volume 150 of *Frontiers in Artificial Intelligence and Applications*. IOS Press.

[Bereta et al., 2013] Bereta, K., Smeros, P., and Koubarakis, M. (2013). Representation and Querying of Valid Time of Triples in Linked Geospatial Data. In [Cimiano et al., 2013], pages 259–274.

[Berndt et al., 2017] Berndt, J. O., Petta, P., and Unland, R., editors (2017). *Multiagent System Technologies - 15th German Conference, MATES 2017, Leipzig, Germany, August 23-26, 2017, Proceedings*, volume 10413 of *Lecture Notes in Computer Science*. Springer.

[Berners-Lee et al., 2018] Berners-Lee, T., Capadisli, S., Dietze, S., Hogan, A., Janowicz, K., and Lehmann, J., editors (2018). *Workshop on Linked Data on the Web co-located with The Web Conference 2018, LDOW@WWW 2018, Lyon, France April 23rd, 2018*, volume 2073 of *CEUR Workshop Proceedings*. Sun SITE Central Europe (CEUR).

[Berners-Lee and Connolly, 2011] Berners-Lee, T. and Connolly, D. (2011). Notation3 (N3): A readable RDF syntax, W3C Team Submission 28 March 2011. W3C team submission, World Wide Web Consortium.

[Berners-Lee et al., 2001] Berners-Lee, T., Hendler, J., and Lassila, O. (2001). The Semantic Web. *Scientific American*, 284(5):34–43.

[Bernstein et al., 2009] Bernstein, A., Karger, D. R., Heath, T., Feigenbaum, L., Maynard, D., Motta, E., and Thirunarayan, K., editors (2009). *The Semantic Web - ISWC 2009, 8th International Semantic Web Conference, ISWC 2009, Chantilly, VA, USA, October 25-29, 2009. Proceedings*, volume 5823 of *Lecture Notes in Computer Science*. Springer.

[Bizer et al., 2014] Bizer, C., Heath, T., Auer, S., and Berners-Lee, T., editors (2014). *Proceedings of the Workshop on Linked Data on the Web, co-located with the 23rd International World Wide Web Conference (WWW 2014), Seoul, Korea, April 8, 2014*, volume 1184 of *CEUR Workshop Proceedings*. Sun SITE Central Europe (CEUR).

[Blomqvist et al., 2017] Blomqvist, E., Maynard, D., Gangemi, A., Hoekstra, R., Hitzler, P., and Hartig, O., editors (2017). *The Semantic Web - 14th International Conference, ESWC 2017, Portorož, Slovenia, May 28 - June 1, 2017, Proceedings, Part I*, volume 10249 of *Lecture Notes in Computer Science*. Springer.

[Borgida and Serafini, 2003] Borgida, A. and Serafini, L. (2003). Distributed Description Logics: Assimilating Information from Peer Sources. *Journal on Data Semantics*, 1:153–184.

[Borst, 1997] Borst, W. N. (1997). *Construction of Engineering Ontologies for Knowledge Sharing and Reuse.* PhD thesis, University of Twente, Enschede, Netherlands.

[Bouquet et al., 2003a] Bouquet, P., Ghidini, C., Giunchiglia, F., and Blanzieri, E. (2003a). Theories and Uses of Context in Knowledge Representation and Reasoning. *Journal of Pragmatics*, 35(3):455–484.

[Brickley and Guha, 2014] Brickley, D. and Guha, R. V. (2014). RDF Schema 1.1, W3C Recommendation 25 February 2014. W3C Recommendation, World Wide Web Consortium.

[Calvanese et al., 2007] Calvanese, D., Franconi, E., Haarslev, V., Lembo, D., Motik, B., Tessaris, S., and Turhan, A.-Y., editors (2007). *Proceedings of the 20th International Workshop on Description Logics DL'07, June 8 - 10, 2007, Brixen/Bressanone, Italy.* Bolzano University Press.

[Calvanese and Lausen, 2008] Calvanese, D. and Lausen, H., editors (2008). *Web Reasoning and Rule Systems, Second International Conference, RR 2008, Karlsruhe, Germany, October/November 2008, Proceedings*, volume 5341 of *Lecture Notes in Computer Science*. Springer.

[Capadisli et al., 2019] Capadisli, S., Cotton, F., Giménez-García, J. M., Haller, A., Kalampokis, E., Nguyen, V., Sheth, A., and Troncy, R., editors (2019). *Joint Proceedings of the International Workshops on Contextualized Knowledge Graphs, and Semantic Statistics co-located with 17th International Semantic Web Conference (ISWC 2018) Monterey, USA, October 8th, 2018*, volume 2317 of *CEUR Workshop Proceedings*. Sun SITE Central Europe (CEUR).

[Carroll et al., 2007] Carroll, J. J., Bizer, C., Hayes, P., and Stickler, P. (2007). Named graphs, provenance and trust. In [Ellis and Hagino, 2007], pages 613–622.

[Ceri et al., 2010] Ceri, S., Della Valle, E., Hendler, J., and Huang, Z., editors (2010). *Proceedings of the Workshop on New Forms of Reasoning for the Semantic Web: Scalable & Dynamic.*

[Chung et al., 2014] Chung, C.-W., Broder, A. Z., Shim, K., and Suel, T., editors (2014). *23rd International World Wide Web Conference, WWW '14, Seoul, Republic of Korea, April 7-11, 2014*. ACM Press.

[Ciancarini et al., 2016] Ciancarini, P., Poggi, F., Horridge, M., Zhao, J., Groza, T., Suárez-Figueroa, M. C., d'Aquin, M., and Presutti, V., editors (2016). *Knowledge Engineering and Knowledge Management - EKAW 2016 Satellite Events, EKM and Drift-an-LOD, Bologna, Italy, November 19-23, 2016, Revised Selected Papers*, volume 10180 of *Lecture Notes in Computer Science*. Springer.

[Cimiano et al., 2013] Cimiano, P., Corcho, Ó., Presutti, V., Hollink, L., and Rudolph, S., editors (2013). *The Semantic Web: Semantics and Big Data, 10th International Conference, ESWC 2013, Montpellier, France, May 26-30, 2013. Proceedings*, volume 7882 of *Lecture Notes in Computer Science*. Springer.

[Ciortea et al., 2013] Ciortea, A., Boissier, O., Zimmermann, A., and Magda-Florea, A. (2013). Reconsidering the social web of things: position paper. In [Mattern et al., 2013], pages 1535–1544.

[Ciortea et al., 2014] Ciortea, A., Boissier, O., Zimmermann, A., and Magda-Florea, A. (2014). Open and Interoperable Socio-technical Networks. In [Winkler et al., 2014], pages 251–257.

[Ciortea et al., 2016a] Ciortea, A., Boissier, O., Zimmermann, A., and Magda-Florea, A. (2016a). Responsive Decentralized Composition of Service Mashups for the Internet of Things. In [Schneegass et al., 2016], pages 53–61.

[Ciortea et al., 2017] Ciortea, A., Boissier, O., Zimmermann, A., and Magda-Florea, A. (2017). Give Agents Some REST: A Resource-oriented Abstraction Layer for Internet-scale Agent Environments. In [Larson et al., 2017], pages 1502–1504.

[Ciortea et al., 2018] Ciortea, A., Boissier, O., Zimmermann, A., and Magda-Florea, A. (2018). Give Agents Some REST: Hypermedia-driven Agent Environments. In [El Fallah-Seghrouchni et al., 2018], pages 125–141.

[Ciortea et al., 2019] Ciortea, A., Mayer, S., Gandon, F. L., Boissier, O., Ricci, A., and Zimmermann, A. (2019). A Decade in Hindsight: The Missing Bridge Between Multi-Agent Systems and the World Wide Web. In [Elkind et al., 2019], pages 1659–1663.

[Ciortea et al., 2015] Ciortea, A., Zimmermann, A., Boissier, O., and Magda-Florea, A. (2015). Towards a Social and Ubiquitous Web: A Model for Socio-technical Networks. In [WI-, 2015], pages 461–468.

[Ciortea et al., 2016b] Ciortea, A., Zimmermann, A., Boissier, O., and Magda-Florea, A. (2016b). Hypermedia-driven Socio-technical Networks for Goal-driven Discovery in the Web of Things. In [WoT, 2016], pages 25–30.

[Ciortea, 2016] Ciortea, A.-N. (2016). *Weaving a Social Web of Things: Enabling Autonomous and Flexible Interaction in the Internet of Things*. PhD thesis, École des mines de Saint-Étienne and Universitatea Politehnica din Bucureşti.

[Corby and Faron-Zucker, 2015] Corby, O. and Faron-Zucker, C. (2015). STTL - A SPARQL-based Transformation Language for RDF. In [Monfort et al., 2015], pages 466–476.

[Corcho et al., 2017] Corcho, Ó., Janowicz, K., Rizzo, G., Tiddi, I., and Garijo, D., editors (2017). *Proceedings of the Knowledge Capture Conference, K-CAP 2017, Austin, TX, USA, December 4-6, 2017*. ACM Press.

[Cruz et al., 2006] Cruz, I. F., Decker, S., Allemang, D., Preist, C., Schwabe, D., Mika, P., Uschold, M., and Aroyo, L., editors (2006). *The Semantic Web - ISWC 2006, 5th International Semantic Web Conference, ISWC 2006, Athens, GA, USA, November 5-9, 2006, Proceedings*, volume 4273 of *Lecture Notes in Computer Science*. Springer.

[Cyganiak et al., 2014] Cyganiak, R., Wood, D., and Lanthaler, M. (2014). RDF 1.1 Concepts and Abstract Syntax, W3C Recommendation 25 February 2014. W3C Recommendation, World Wide Web Consortium.

[da Costa et al., 2008] da Costa, P. C. G., d'Amato, C., Fanizzi, N., Laskey, K. B., Laskey, K. J., Lukasiewicz, T., Nickles, M., and Pool, M., editors (2008). *Uncertainty Reasoning for the Semantic Web I, ISWC International Workshops, URSW 2005-2007, Revised Selected and Invited Papers*, volume 5327 of *Lecture Notes in Computer Science*. Springer.

[d'Aquin et al., 2010] d'Aquin, M., García Castro, A., Lange, C., and Viljanen, K., editors (2010). *Proceedings of the 1st Workshop on Ontology Repositories and Editors for the Semantic Web, Hersonissos, Crete, Greece, May 31st, 2010*, volume 596 of *CEUR Workshop Proceedings*. Sun SITE Central Europe (CEUR).

[Das et al., 2012] Das, S., Sundara, S., and Cyganiak, R. (2012). R2RML: RDB to RDF Mapping Language. W3C Recommendation, World Wide Web Consortium.

[de Bruijn et al., 2005] de Bruijn, J., Lausen, H., Polleres, A., and Fensel, D. (2005). The Web Service Modeling Language WSML: An Overview. Technical report, Digital Enterprise Research Institute.

[Dell'Aglio et al., 2017] Dell'Aglio, D., Anicic, D., Barnaghi, P. M., Della Valle, E., McGuinness, D. L., Bozzato, L., Eiter, T., Homola, M., and Porello, D., editors (2017). *Joint Proceedings of the Web Stream Processing Workshop (WSP 2017) and the 2nd International Workshop on Ontology Modularity, Contextuality, and*

Evolution (WOMoCoE 2017) Co-Located with 16th International Semantic Web Conference (ISWC 2017), Vienna, Austria, October 22nd, 2017, volume 1936 of *CEUR Workshop Proceedings*. Sun SITE Central Europe (CEUR).

[Dietz et al., 2019] Dietz, J., Aveiro, D., and Felipe, J., editors (2019). *Proceedings of the 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management - (Volume 2), September 17-19, 2019, in Vienna, Austria*. SciTePress.

[Dimou et al., 2018] Dimou, A., Heyvaert, P., De Meester, B., and Verborgh, R. (2018). What Factors Influence the Design of a Linked Data Generation Algorithm? In [Berners-Lee et al., 2018].

[Dimou et al., 2014] Dimou, A., Vander Sande, M., Colpaert, P., Verborgh, R., Mannens, E., and Van de Walle, R. (2014). RML: A Generic Language for Integrated RDF Mappings of Heterogeneous Data. In [Bizer et al., 2014].

[Dovier and Santos Costa, 2012] Dovier, A. and Santos Costa, V., editors (2012). *Technical Communications of the 28th International Conference on Logic Programming, ICLP 2012, September 4-8, 2012, Budapest, Hungary*. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik.

[Duggan et al., 2015] Duggan, J., Elmore, A. J., Stonebraker, M., Balzinska, M., Howe, B., Kepner, J., Madden, S., Maier, D., Mattson, T. G., and Zdonik, S. B. (2015). The BigDAWG Polystore System. *SIGMOD Record*, 44(2):11–16.

[Dymytrova, 2018] Dymytrova, V. (2018). La normalisation et la durabilité de l'information au prisme des discours des acteurs de l'open data. In *Colloque COSSI 2018 Processus de normalisation et durabilitde l'information, Bordeaux, 24-25 mai 2018, ESPE Bordeaux*.

[Dymytrova and Paquienséguy, 2019] Dymytrova, V. and Paquienséguy, F. (2019). Des données aux services dans les Métropoles : enquête auprès des réutilisateurs professionnels de données ouvertes. In *CODATA France DATA VALUE CHAIN*.

[El Fallah-Seghrouchni et al., 2018] El Fallah-Seghrouchni, A., Ricci, A., and Son, T. C., editors (2018). *Engineering Multi-Agent Systems - 5th International Workshop, EMAS 2017, Sao Paulo, Brazil, May 8-9, 2017, Revised Selected Papers*, volume 10738 of *Lecture Notes in Computer Science*. Springer.

[Elkind et al., 2019] Elkind, E., Veloso, M., Agmon, N., and Taylor, M. E., editors (2019). *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS '19, Montreal, QC, Canada, May 13-17, 2019*. International Foundation for Autonomous Agents and Multiagent Systems.

[Ellis and Hagino, 2007] Ellis, A. and Hagino, T., editors (2007). *Proceedings of the 14th International Conference on World Wide Web, WWW 2005, Chiba, Japan, May 10-14, 2005*. ACM Press.

[Endris et al., 2017] Endris, K. M., Giménez-García, J. M., Thakkar, H., Demidova, E., Zimmermann, A., Lange, C., and Simperl, E. (2017). Dataset Reuse: An Analysis of References in Community Discussions, Publications and Data. In [Corcho et al., 2017], pages 5:1–5:4.

[Erxleben et al., 2014] Erxleben, F., Günther, M., Krötzsch, M., Mendez, J., and Vrandecic, D. (2014). Introducing Wikidata to the Linked Data Web. In [Mika et al., 2014], pages 50–65.

[Euzenat and Shvaiko, 2013] Euzenat, J. and Shvaiko, P. (2013). *Ontology Matching, Second Edition*. Springer.

[Fielding and Reschke, 2014] Fielding, R. T. and Reschke, J. (2014). Hypertext Transfer Protocol (HTTP/1.1): Semantics and Content. Technical report, Internet Engineering Task Force.

[Filipe and Dietz, 2010] Filipe, J. and Dietz, J. L. G., editors (2010). *KEOD 2010 - Proceedings of the International Conference on Knowledge Engineering and Ontology Development, Valencia, Spain, October 25-28, 2010*. SciTePress.

[Flouris et al., 2009] Flouris, G., Fundulaki, I., Pediaditis, P., Theoharis, Y., and Christophides, V. (2009). Coloring RDF Triples to Capture Provenance. In [Bernstein et al., 2009], pages 196–212.

[Franconi et al., 2007] Franconi, E., Kifer, M., and May, W., editors (2007). *The Semantic Web: Research and Applications, 4th European Semantic Web Conference, ESWC 2007, Innsbruck, Austria, June 3-7, 2007, Proceedings*, volume 4519 of *Lecture Notes in Computer Science*. Springer.

[Frey et al., 2019] Frey, J., Müller, K., Hellmann, S., Rahm, E., and Vidal, M. (2019). Evaluation of metadata representations in RDF stores. *Semantic Web Journal*, 10(2):205–229.

[Fridman Noy and Rector, 2006] Fridman Noy, N. and Rector, A. L. (2006). Defining N-ary Relations on the Semantic Web, W3C Working Group Note 12 April 2006. W3C Note, World Wide Web Consortium.

[Gandon and Bisson, 2017] Gandon, F. and Bisson, G., editors (2017). *17ème Journées Francophones Extraction et Gestion des Connaissances, EGC 2017, 24-27 Janvier 2017, Grenoble, France*, volume E-33 of *Revue des nouvelles technologies de l'information*. Édition RNTI.

[Gangemi et al., 2018a] Gangemi, A., Gentile, A. L., Nuzzolese, A. G., Rudolph, S., Maleshkova, M., Paulheim, H., Pan, J. Z., and Alam, M., editors (2018a). *The Semantic Web: ESWC 2018 Satellite Events - ESWC 2018 Satellite Events, Heraklion, Crete, Greece, June 3-7, 2018, Revised Selected Papers*, volume 11155 of *Lecture Notes in Computer Science*. Springer.

[Gangemi et al., 2018b] Gangemi, A., Navigli, R., Vidal, M., Hitzler, P., Raphaël, T., Hollink, L., Tordai, A., and Alam, M., editors (2018b). *The Semantic Web - 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3-7, 2018, Proceedings*, volume 10843 of *Lecture Notes in Computer Science*. Springer.

[Ghidini and Serafini, 1998] Ghidini, C. and Serafini, L. (1998). Distributed First Order Logics. In Gabbay, D. M. and de Rijke, M., editors, *Frontiers of Combining Systems 2*, volume 7 of *Studies in Logic and Computation*, pages 121–139. Research Studies Press.

[Giménez-García et al., 2019] Giménez-García, J. M., Duarte, M., Zimmermann, A., Gravier, C., Hruschka Jr., E. R., and Maret, P. (2019). NELL2RDF: Reading the Web, Tracking the Provenance, and Publishing it as Linked Data. In [Capadisli et al., 2019].

[Giménez-García et al., 2016a] Giménez-García, J. M., Thakkar, H., and Zimmermann, A. (2016a). Assessing Trust with PageRank in the Web of Data. In [Sack et al., 2016b], pages 293–307.

[Giménez-García and Zimmermann, 2019] Giménez-García, J. M. and Zimmermann, A. (2019). NdProperties: Encoding Contexts in RDF predicates with Inference Preservation. In [Capadisli et al., 2019].

[Giménez-García et al., 2016b] Giménez-García, J. M., Zimmermann, A., and Maret, P. (2016b). Representing Contextual Information as Fluents. In [Ciancarini et al., 2016], pages 119–122.

[Giménez-García et al., 2017] Giménez-García, J. M., Zimmermann, A., and Maret, P. (2017). NdFluents: An Ontology for Annotated Statements with Inference Preservation. In [Blomqvist et al., 2017], pages 638–654.

[Giunchiglia, 1993] Giunchiglia, F. (1993). Contextual Reasoning. *Epistemologica*, 16:345–364.

[Golbreich and Wallace, 2012] Golbreich, C. and Wallace, E. K. (2012). OWL 2 Web Ontology Language, New Features and Rationale (Second Edition), W3C Recommendation 11 December 2012. W3C Recommendation, World Wide Web Consortium.

[Groppe and Gruenwald, 2019] Groppe, S. and Gruenwald, L., editors (2019). *Proceedings of the International Workshop on Semantic Big Data, SBD@SIGMOD 2019, Amsterdam, The Netherlands, July 5, 2019*. ACM Press.

[Guarino et al., 2009] Guarino, N., Oberle, D., and Staab, S. (2009). What is an *Ontology*? In [Staab and Studer, 2009], pages 1–17. Second edition.

[Guha et al., 2004] Guha, R. V., McCool, R., and Fikes, R. (2004). Contexts for the Semantic Web. In [van Harmelen et al., 2004], pages 32–46.

[Gutiérrez et al., 2007] Gutiérrez, C., Hurtado, C. A., and Vaisman, A. A. (2007). Introducing Time into RDF. *IEEE Transactions on Knowledge and Data Engineering*, 19(2):207–218.

[Haesendonck et al., 2019] Haesendonck, G., Maroy, W., Heyvaert, P., Verborgh, R., and Dimou, A. (2019). Parallel RDF generation from heterogeneous big data. In [Groppe and Gruenwald, 2019], pages 1:1–1:6.

[Haidar, 2019] Haidar, A. (2019). Comparing different models to represent metadata in the Semantic Web. Intership report, Université Jean Monnet Saint-Étienne.

[Harris and Seaborne, 2013] Harris, S. and Seaborne, A. (2013). SPARQL 1.1 Query Language, W3C Recommendation 21 March 2013. W3C Recommendation, World Wide Web Consortium.

[Hartig, 2009] Hartig, O. (2009). Querying Trust in RDF Data with tSPARQL. In [Aroyo et al., 2009], pages 5–20.

[Hartig, 2014] Hartig, O. (2014). *Querying a Web of Linked Data*. PhD thesis, Humboldt University of Berlin.

[Hartig and Thompson, 2014] Hartig, O. and Thompson, B. (2014). Foundations of an Alternative Approach to Reification in RDF. Technical report, CoRR.

[Hayes and Patel-Schneider, 2014] Hayes, P. and Patel-Schneider, P. F. (2014). RDF 1.1 Semantics, W3C Recommendation 25 February 2014. W3C Recommendation, World Wide Web Consortium.

[Hernández et al., 2015] Hernández, D., Hogan, A., and Krötzsch, M. (2015). Reifying RDF: What Works Well With Wikidata? In [Liebig and Fokoue, 2015], pages 32–47.

[Hogan et al., 2014] Hogan, A., Arenas, M., Mallea, A., and Polleres, A. (2014). Everything you always wanted to know about blank nodes. *Journal of Web Semantics*, 27:42–69.

[Hogan et al., 2020] Hogan, A., Blomqvist, E., Cochez, M., d'Amato, C., de Melo, G., Gutierrez, C., Labra Gayo, J. E., Kirrane, S., Neumaier, S., Polleres, A., Navigli, R., Ngonga Ngomo, A., Rashid, S. M., Rula, A., Schmelzeisen, L., Sequeda, J. F., Staab, S., and Zimmermann, A. (2020). Knowledge Graphs. Technical report, CoRR.

[Hogan et al., 2010] Hogan, A., Polleres, A., Umbrich, J., and Zimmermann, A. (2010). Some entities are more equal than others: statistical methods to consolidate Linked Data. In [Ceri et al., 2010].

[Hogan et al., 2012] Hogan, A., Zimmermann, A., Umbrich, J., Polleres, A., and Decker, S. (2012). Scalable and distributed methods for entity matching, consolidation and disambiguation over linked data corpora. *Journal of Web Semantics*, 10:76–110.

[Holtman and Mutz, 1998] Holtman, K. and Mutz, A. (1998). Transparent Content Negotiation in HTTP. Technical report, Internet Engineering Task Force.

[Homola, 2007] Homola, M. (2007). Distributed Description Logics Revisited. In [Calvanese et al., 2007].

[Huai et al., 2008] Huai, J., Chen, R., Hon, H., Liu, Y., Ma, W., Tomkins, A., and Zhang, X., editors (2008). *Proceedings of the 17th International Conference on World Wide Web, WWW 2008, Beijing, China, April 21-25, 2008*. ACM Press.

[Hübner et al., 2002b] Hübner, J. F., Sichman, J. S., and Boissier, O. (2002b). MOISE+: towards a structural, functional, and deontic model for MAS organization. In [AAM, 2002], pages 501–502.

[Janowicz et al., 2018] Janowicz, K., Kuhn, W., Cena, F., Haller, A., and Vamvoudakis, K. G., editors (2018). *Proceedings of the 8th International Conference on the Internet of Things, IOT 2018, Santa Barbara, CA, USA, October 15-18, 2018*. ACM Press.

[Kaebisch et al., 2020] Kaebisch, S., Kamiya, T., McCool, M., Charpenay, V., and Kovatsch, M. (2020). Web of Things (WoT) Thing Description, W3C Proposed Recommendation 30 January 2020. W3C Proposed Recommendation, World Wide Web Consortium.

[Kay, 2017] Kay, M. (2017). XSL Transformations (XSLT) Version 3.0 W3C Recommendation 8 June 2017. W3C Recommendation, World Wide Web Consortium.

[Kem, 2018] Kem, O. (2018). *Modélisation et Exploitation des Connaissances de l'Environnement : Une Approche Multi-Agents pour la Recherche d'ItinÃ©raires Multi-Objectifs dans des Environnements Ubiquitaires*. PhD thesis, École des mines de Saint-Étienne, Saint-Étienne, France.

[Kem et al., 2016] Kem, O., Balbo, F., and Zimmermann, A. (2016). A Distributed Approach to Constructing Travel Solutions by Exploiting Web Resources. In [WI:, 2016], pages 713–716.

[Kem et al., 2017a] Kem, O., Balbo, F., and Zimmermann, A. (2017a). Collaborative Search for Multi-goal Pathfinding in Ubiquitous Environments. In [Berndt et al., 2017], pages 72–88.

[Kem et al., 2017b] Kem, O., Balbo, F., and Zimmermann, A. (2017b). Multi-goal pathfinding in ubiquitous environments: modeling and exploiting knowledge to satisfy goals. In [Sheth et al., 2017], pages 1147–1150.

[Kem et al., 2017c] Kem, O., Balbo, F., and Zimmermann, A. (2017c). Traveler-Oriented Advanced Traveler Information System Based on Discovery and Exploitation of Resources: Potentials and Challenges. *Transportation Research Procedia*, 22:635–644. Special issue on 19th EURO Working Group on Transportation Meeting, EWGT2016, 5-7 September 2016, Istanbul, Turkey.

[Kem et al., 2017d] Kem, O., Balbo, F., Zimmermann, A., and Nagellen, P. (2017d). Multi-goal Pathfinding in Cyber-Physical-Social Environments: Multi-layer Search over a Semantic Knowledge Graph. In [Zanni-Merk et al., 2017], pages 741–750.

[Khan et al., 2019] Khan, Y., Zimmermann, A., Jha, A., Gadepally, V., d'Aquin, M., and Sahay, R. (2019). One Size Does Not Fit All: Querying Web Polystores. *IEEE Access*, 7:9598–9617.

[Khan et al., 2017b] Khan, Y., Zimmermann, A., Jha, A., Rebholz-Schuhmann, D., and Sahay, R. (2017b). Querying web polystores. In [Nie et al., 2017], pages 3190–3195.

[Kifer and Subrahmanian, 1992] Kifer, M. and Subrahmanian, V. S. (1992). Theory of generalized annotated logic programming and its applications. *Journal of Logic Programming*, 12(3&4):335–367.

[Klai, 2016] Klai, S. (2016). *Approches flexibles pour l'intégration des ontologies hétérogènes et évolutives*. PhD thesis, UniversitÃ© Badji Mokhtar, Annaba, Annaba (Algeria).

[Klai et al., 2016a] Klai, S., Zimmermann, A., and Khadir, M. T. (2016a). Multi-level Networked Knowledge Base: DDL-Reasoning. In [Bellatreche et al., 2016], pages 118–131.

[Klai et al., 2016b] Klai, S., Zimmermann, A., and Khadir, M. T. (2016b). Multi-level networked knowledge: representation and DL-reasoning. *International Journal of Metadata, Semantics and Ontologies*, 11(1):1–15.

[Klai et al., 2019] Klai, S., Zimmermann, A., and Khadir, M. T. (2019). Networked Ontologies with Contextual Alignments. *Computing and Informatics*, 38(1):115–150.

[Kutz et al., 2004] Kutz, O., Lutz, C., Wolter, F., and Zakharyaschev, M. (2004). $\mathcal{E}$-connections of abstract description systems. *Artificial Intelligence*, 156(1):1–73.

[Larroche and Dymytrova, 2017] Larroche, V. and Dymytrova, V. (2017). Le web sémantique, un moyen de visibilité et d'exploitation des open data auprès des communautés de réutilisateurs professionnels ? In *Colloque international pluridisciplinaire Big data et visibilité en ligne : un enjeu pluridisciplinaire de l'économie numérique, Université des Antilles, Fort-de-France, Martinique, 6-8 novembre 2017*.

[Larson et al., 2017] Larson, K., Winikoff, M., Das, S., and Durfee, E. H., editors (2017). *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems, AAMAS 2017, São Paulo, Brazil, May 8-12, 2017*. ACM Press.

[Le Duc et al., 2010] Le Duc, C., Lamolle, M., and Zimmermann, A. (2010). An API for Distributed Reasoning on Networked Ontologies with Alignments. In [Filipe and Dietz, 2010], pages 295–304.

[Le Duc et al., 2013] Le Duc, C., Lamolle, M., Zimmermann, A., and Curé, O. (2013). DRAOn: A Distributed Reasoner for Aligned Ontologies. In [Bail et al., 2013], pages 81–86.

[Lefrançois et al., 2017a] Lefrançois, M., Kalaoja, J., Ghariani, T., and Zimmermann, A. (2017a). SEAS Knowledge Model. Deliverable D2.2, Smart Energy Aware Systems.

[Lefrançois and Zimmermann, 2016] Lefrançois, M. and Zimmermann, A. (2016). Supporting Arbitrary Custom Datatypes in RDF and SPARQL. In [Sack et al., 2016a], pages 371–386.

[Lefrançois and Zimmermann, 2018] Lefrançois, M. and Zimmermann, A. (2018). The Unified Code for Units of Measure in RDF: cdt:ucum and other UCUM Datatypes. In [Gangemi et al., 2018a], pages 196–201.

[Lefrançois et al., 2016b] Lefrançois, M., Zimmermann, A., and Bakerally, N. (2016b). Flexible RDF generation from RDF and heterogeneous data sources with SPARQL-Generate. In [Ciancarini et al., 2016], pages 131–135.

[Lefrançois et al., 2017b] Lefrançois, M., Zimmermann, A., and Bakerally, N. (2017b). A SPARQL Extension for Generating RDF from Heterogeneous Formats. In [Blomqvist et al., 2017], pages 35–50.

[Lefrançois et al., 2017c] Lefrançois, M., Zimmermann, A., and Bakerally, N. (2017c). Génération de RDF à partir de sources de données aux formats hétérogènes. In [Gandon and Bisson, 2017], pages 93–104.

[Lefrançois, 2018] Lefrançois, M. (2018). RDF presentation and correct content conveyance for legacy services and the web of things. In [Janowicz et al., 2018], pages 43:1–43:8.

[Liebig and Fokoue, 2015] Liebig, T. and Fokoue, A., editors (2015). *Proceedings of the 11th International Workshop on Scalable Semantic Web Knowledge Base Systems co-located with 14th International Semantic Web Conference (ISWC 2015), Bethlehem, PA, USA, October 11, 2015*, volume 1457 of *CEUR Workshop Proceedings*. Sun SITE Central Europe (CEUR).

[Lopes et al., 2012] Lopes, N., Kirrane, S., Zimmermann, A., Polleres, A., and Mileo, A. (2012). A Logic Programming approach for Access Control over RDF. In [Dovier and Santos Costa, 2012], pages 381–392.

[Lopes et al., 2010a] Lopes, N., Polleres, A., Straccia, U., and Zimmermann, A. (2010a). AnQL: SPARQLing Up Annotated RDFS. In [Patel-Schneider et al., 2010a], pages 518–533.

[Lopes et al., 2010b] Lopes, N., Zimmermann, A., Hogan, A., Lukácsy, G., Polleres, A., Straccia, U., and Stefan, D. (2010b). RDF Needs Annotations. In *RDF Next Steps, W3C Workshop*. World Wide Web Consortium.

[Maali and Erickson, 2014] Maali, F. and Erickson, J. (2014). Data Catalog Vocabulary (DCAT), W3C Recommendation 16 January 2014. W3C Recommendation, World Wide Web Consortium.

[Martin et al., 2004] Martin, D., Burstein, M., Hobbs, J., Lassila, O., McDermott, D., McIlraith, S., Narayanan, S., Paolucci, M., Parsia, B., Payne, T., Sirin, E., Srinivasan, N., and Sycara, K. (2004). OWL-S: Semantic Markup for Web Services. W3C member submission, World Wide Web Consortium.

[Mattern et al., 2013] Mattern, F., Santini, S., Canny, J. F., Langheinrich, M., and Rekimoto, J., editors (2013). *The 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp '13, Zurich, Switzerland, September 8-12, 2013 - Adjunct Publication*. ACM Press.

[Mazzieri and Dragoni, 2008] Mazzieri, M. and Dragoni, A. F. (2008). A Fuzzy Semantics for the Resource Description Framework. In [da Costa et al., 2008], pages 244–261.

[McCarthy, 1987] McCarthy, J. L. (1987). Generality in Artificial Intelligence. *Communications of the ACM*, 30(12):1029–1035.

[Meersman et al., 2009] Meersman, R., Dillon, T. S., and Herrero, P., editors (2009). *On the Move to Meaningful Internet Systems: OTM 2009 - Confederated International Conferences, CoopIS, DOA, IS, and ODBASE 2009, Vilamoura, Portugal, November 1-6, 2009, Proceedings, Part II*, volume 5871 of *Lecture Notes in Computer Science*. Springer.

[Mika et al., 2014] Mika, P., Tudorache, T., Bernstein, A., Welty, C. A., Knoblock, C. A., Vrandecic, D., Groth, P. T., Fridman Noy, N., Janowicz, K., and Goble, C. A., editors (2014). *The Semantic Web - ISWC 2014 - 13th International Semantic Web Conference, Riva del Garda, Italy, October 19-23, 2014. Proceedings, Part I*, volume 8796 of *Lecture Notes in Computer Science*. Springer.

[Monfort et al., 2015] Monfort, V., Krempels, K.-H., Majchrzak, T. A., and Turk, Z., editors (2015). *WEBIST 2015 - Proceedings of the 11th International Conference on Web Information Systems and Technologies, Lisbon, Portugal, 20-22 May, 2015*. SciTePress.

[Muñoz et al., 2007] Muñoz, S., Pérez, J., and Gutiérrez, C. (2007). Minimal Deductive Systems for RDF. In [Franconi et al., 2007], pages 53–67.

[Nguyen et al., 2014] Nguyen, V., Bodenreider, O., and Sheth, A. (2014). Don't like RDF reification?: making statements about statements using singleton property. In [Chung et al., 2014], pages 759–770.

[Nie et al., 2017] Nie, J., Obradovic, Z., Suzumura, T., Ghosh, R., Nambiar, R., Wang, C., Zang, H., Baeza-Yates, R. A., Hu, X., Kepner, J., Cuzzocrea, A., Tang, J., and Toyoda, M., editors (2017). *2017 IEEE International Conference on Big Data, BigData 2017, Boston, MA, USA, December 11-14, 2017*. IEEE Computer Society.

[Nikitina et al., 2017] Nikitina, N., Song, D., Fokoue, A., and Haase, P., editors (2017). *Proceedings of the ISWC 2017 Posters & Demonstrations and Industry Tracks co-located with 16th International Semantic Web Conference (ISWC 2017), Vienna, Austria, October 23rd - to - 25th, 2017*, volume 1963 of *CEUR Workshop Proceedings*. Sun SITE Central Europe (CEUR).

[Omicini et al., 2008] Omicini, A., Ricci, A., and Viroli, M. (2008). Artifacts in the A&A meta-model for multi-agent systems. *Autonomous Agents and Multi-Agent Systems*, 17(3):432–456.

[Ossowski et al., 2012] Ossowski, S., Toni, F., and Vouros, G., editors (2012). *Proceedings of the First International Conference on Agreement Technologies, AT 2012, Dubrovnik, Croatia, October 15-16, 2012*, volume 918 of *CEUR Workshop Proceedings*. Sun SITE Central Europe (CEUR).

[Patel-Schneider et al., 2010a] Patel-Schneider, P. F., Pan, Y., Hitzler, P., Mika, P., Zhang, Y., Pan, J. Z., Horrocks, I., and Glimm, B., editors (2010a). *The Semantic Web - ISWC 2010 - 9th International Semantic Web Conference, ISWC 2010, Shanghai, China, November 7-11, 2010, Revised Selected Papers, Part I*, volume 6496 of *Lecture Notes in Computer Science*. Springer.

[Polleres et al., 2009] Polleres, A., Krennwallner, T., Lopes, N., Kopecký, J., and Decker, S. (2009). XSPARQL Language Specification. W3C Member Submission, World Wide Web Consortium.

[Polleres and Swift, 2009] Polleres, A. and Swift, T., editors (2009). *Web Reasoning and Rule Systems, Third International Conference, RR 2009, Chantilly, VA,* *USA, October 25-26, 2009, Proceedings*, volume 5837 of *Lecture Notes in Computer Science*. Springer.

[Presutti et al., 2014] Presutti, V., d'Amato, C., Gandon, F., d'Aquin, M., Staab, S., and Tordia, A., editors (2014). *The Semantic Web: Trends and Challenges - 11th International Conference, ESWC 2014, Anissaras, Crete, Greece, May 25-29, 2014. Proceedings*, volume 8465 of *Lecture Notes in Computer Science*. Springer.

[Pugliese et al., 2008] Pugliese, A., Udrea, O., and Subrahmanian, V. S. (2008). Scaling RDF with time. In [Huai et al., 2008], pages 605–614.

[Qawasmeh et al., 2018] Qawasmeh, O., Lefrançois, M., Zimmermann, A., and Maret, P. (2018). Computer-Assisted Ontology Construction System: Focus on Bootstrapping Capabilities. In [Gangemi et al., 2018a], pages 60–65.

[Qawasmeh et al., 2019] Qawasmeh, O., Lefrançois, M., Zimmermann, A., and Maret, P. (2019). Observing the Impact and Adaptation to the Evolution of an Imported Ontology. In [Dietz et al., 2019], pages 76–86.

[Qawasmeh et al., 2020] Qawasmeh, O., Lefrançois, M., Zimmermann, A., and Maret, P. (2020). Pitfalls in Networked and Versioned Ontologies. In *Knowledge Discovery, Knowledge Engineering and Knowledge Management - 11th International Joint Conference, IC3K 2019, Vienna, Austria, November 17-19, 2019, Revised Selected Papers*, Communications in Computer and Information Science. Springer. Accepted August 2020. To appear.

[Reynolds, 2014] Reynolds, D. (2014). The Organization Ontology, W3C Recommendation 16 January 2014. W3C Recommendation, World Wide Web Consortium.

[Sack et al., 2016a] Sack, H., Blomqvist, E., d'Aquin, M., Ghidini, C., Ponzetto, S. P., and Lange, C., editors (2016a). *The Semantic Web. Latest Advances and New Domains - 13th International Conference, ESWC 2016, Heraklion, Crete, Greece, May 29 - June 2, 2016, Proceedings*, volume 9678 of *Lecture Notes in Computer Science*. Springer.

[Sack et al., 2016b] Sack, H., Rizzo, G., Steinmetz, N., Mladenic, D., Auer, S., and Lange, C., editors (2016b). *The Semantic Web - ESWC 2016 Satellite Events, Heraklion, Crete, Greece, May 29 - June 2, 2016, Revised Selected Papers*, volume 9989 of *Lecture Notes in Computer Science*. Springer.

[Sahay et al., 2011] Sahay, R., Fox, R., Zimmermann, A., Polleres, A., and Hauswirth, M. (2011). A Methodological Approach for Ontologising and Aligning Health Level Seven (HL7) Applications. In [Tjoa et al., 2011], pages 102–117.

[Sahay et al., 2013] Sahay, R., Zimmermann, A., Fox, R., Polleres, A., and Hauswirth, M. (2013). A Formal Investigation of Semantic Interoperability of HCLS Systems. In Sicilia, M.-A. and Serrano-Balazote, P., editors, *Interoperability in Healthcare Information Systems: Standards, Management, and Technology*. IGI Global.

[Saleem and Ngonga Ngomo, 2014] Saleem, M. and Ngonga Ngomo, A.-C. (2014). HiBISCuS: Hypergraph-Based Source Selection for SPARQL Endpoint Federation. In [Presutti et al., 2014], pages 176–191.

[Schenk, 2008] Schenk, S. (2008). On the Semantics of Trust and Caching in the Semantic Web. In [Sheth et al., 2008], pages 533–549.

[Schneegass et al., 2016] Schneegass, S., Schmidt, A., Michahelles, F., Kritzler, M., Ilic, A., and Kunze, K., editors (2016). *Proceedings of the 6th International Conference on the Internet of Things, IOT 2016, Stuttgart, Germany, November 7-9, 2016*. ACM Press.

[Schneider and Walther, 2012] Schneider, T. and Walther, D., editors (2012). *Proceedings of the 6th International Workshop on Modular Ontologies, Graz, Austria, July 24, 2012*, volume 875 of *CEUR Workshop Proceedings*. Sun SITE Central Europe (CEUR).

[Schwarte et al., 2011] Schwarte, A., Haase, P., Hose, K., Schenkel, R., and Schmidt, R. A. (2011). FedX: A Federation Layer for Distributed Query Processing on Linked Open Data. In [Antoniou et al., 2011], pages 481–486.

[Seaborne and Prud'hommeaux, 2008] Seaborne, A. and Prud'hommeaux, E. (2008). SPARQL Query Language for RDF. W3C Recommendation, World Wide Web Consortium.

[Sheth et al., 2017] Sheth, A. P., Ngonga, A., Wang, Y., Chang, E., Slezak, D., Franczyk, B., Alt, R., Tao, X., and Unland, R., editors (2017). *Proceedings of the International Conference on Web Intelligence, Leipzig, Germany, August 23-26, 2017*. ACM Press.

[Sheth et al., 2008] Sheth, A. P., Staab, S., Dean, M., Paolucci, M., Maynard, D., Finin, T. W., and Thirunarayan, K., editors (2008). *The Semantic Web - ISWC 2008, 7th International Semantic Web Conference, ISWC 2008, Karlsruhe, Germany, October 26-30, 2008. Proceedings*, volume 5318 of *Lecture Notes in Computer Science*. Springer.

[Singhal, 2012] Singhal, A. (2012). Introducing the Knowledge Graph: things, not strings. Google Blog.

[Sorici et al., 2015] Sorici, A., Boissier, O., Picard, G., Zimmermann, A., and Magda-Florea, A. (2015). CONSERT: Applying Semantic Web Technologies to Context Modeling in Ambient Intelligence. *Computers and Electrical Engineering*, 44:280–306.

[Speicher et al., 2015] Speicher, S., Arwe, J., and Malhotra, A. (2015). Linked Data Platform 1.0, W3C Recommendation 26 February 2015. W3C Recommendation, World Wide Web Consortium.

[Sporny et al., 2014] Sporny, M., Kellogg, G., and Lanthaler, M. (2014). JSON-LD 1.0, A JSON-based Serialization for Linked Data, W3C Recommendation 16 January 2014. Technical report, World Wide Web Consortium.

[Staab and Studer, 2009] Staab, S. and Studer, R., editors (2009). *Handbook on Ontologies*. Springer. Second edition.

[Stonebraker, 2015] Stonebraker, M. R. (2015). The case for Polystores. web published.

[Straccia, 2009] Straccia, U. (2009). A Minimal Deductive System for General Fuzzy RDF. In [Polleres and Swift, 2009], pages 166–181.

[Studer et al., 1998] Studer, R., Benjamins, V. R., and Fensel, D. (1998). Knowledge Engineering: Principles and Methods. *Data and Knowledge Engineering*, 25(1-2):161–197.

[Sure and Domingue, 2006] Sure, Y. and Domingue, J., editors (2006). *The Semantic Web: Research and Applications, 3rd European Semantic Web Conference, ESWC 2006, Budva, Montenegro, June 11-14, 2006, Proceedings*, volume 4011 of *Lecture Notes in Computer Science*. Springer.

[Svensson et al., 2019] Svensson, L. G., Atkinson, R., and Car, N. J. (2019). Content Negotiation by Profile - W3C Working Draft 26 November 2019. W3C Working Draft, World Wide Web Consortium.

[Tappolet and Bernstein, 2009] Tappolet, J. and Bernstein, A. (2009). Applied Temporal RDF: Efficient Temporal Querying of RDF Data with SPARQL. In [Aroyo et al., 2009], pages 308–322.

[Tennison and Kellogg, 2015a] Tennison, J. and Kellogg, G. (2015a). Metadata Vocabulary for Tabular Data, W3C Recommendation 17 December 2015. W3C Recommendation, World Wide Web Consortium.

[ter Horst, 2005] ter Horst, H. J. (2005). Completeness, decidability and complexity of entailment for RDF Schema and a semantic extension involving the OWL vocabulary. *Journal of Web Semantics*, 3(2–3):79–115.

[Thuluva et al., 2018] Thuluva, A. S., Anicic, D., and Rudolph, S. (2018). IoT Semantic Interoperability with Device Description Shapes. In [Gangemi et al., 2018a], pages 409–422.

[Tjoa et al., 2011] Tjoa, A. M., Quirchmayr, G., You, I., and Xu, L., editors (2011). *International Conference on Multidisciplinary Research and Practice for Business, Enterprise and Health Information Systems - MURPBES 2011 / Availability, Reliability and Security for Business, Enterprise and Health Information Systems - IFIP WG 8.4/8.9 International Cross Domain Conference and Workshop, ARES 2011, Vienna, Austria, August 22-26, 2011. Proceedings*, volume 6908 of *Lecture Notes in Computer Science*. Springer.

[Udrea et al., 2006a] Udrea, O., Reforgiato Recupero, D., and Subrahmanian, V. S. (2006a). Annotated RDF. In [Sure and Domingue, 2006], pages 487–501.

[Udrea et al., 2010] Udrea, O., Reforgiato Recupero, D., and Subrahmanian, V. S. (2010). Annotated RDF. *ACM Transactions on Computational Logics*, 11(2).

[Udrea et al., 2006b] Udrea, O., Subrahmanian, V. S., and Majkic, Z. (2006b). Probabilistic RDF. In [IRI, 2006], pages 172–177.

[van Harmelen et al., 2004] van Harmelen, F., McIlraith, S., and Plexousakis, D., editors (2004). *The Semantic Web - ISWC 2004: Third International Semantic Web Conference, Hiroshima, Japan, November 7-11, 2004. Proceedings*, volume 3298 of *Lecture Notes in Computer Science*. Springer.

[Völker et al., 2013] Völker, J., Paulheim, H., Lehmann, J., Niepert, M., and Sack, H., editors (2013). *Proceedings of the Second International Workshop on Knowledge Discovery and Data Mining Meets Linked Open Data, Montpellier, France, May 26, 2013*, volume 992 of *CEUR Workshop Proceedings*. Sun SITE Central Europe (CEUR).

[Vouros and Santipantakis, 2012] Vouros, G. and Santipantakis, G. M. (2012). Distributed Reasoning with $E^{DDL}_{HQ^+}$ SHIQ. In [Schneider and Walther, 2012].

[Welty and Fikes, 2006] Welty, C. A. and Fikes, R. (2006). A Reusable Ontology for Fluents in OWL. In [Bennett and Fellbaum, 2006], pages 226–236.

[Williams, 2013] Williams, G. T. (2013). SPARQL 1.1 Service Description, W3C Recommendation 21 March 2013. W3C Recommendation, World Wide Web Consortium.

[Winkler et al., 2014] Winkler, F., Negru, V., Ida, T., Jebelean, T., Petcu, D., Watt, S. M., and Zaharie, D., editors (2014). *16th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, SYNASC 2014, Timisoara, Romania, September 22-25, 2014*. IEEE Computer Society.

[Zanni-Merk et al., 2017] Zanni-Merk, C., Frydman, C. S., Toro, C., Hicks, Y., Howlett, R. J., and Jain, L. C., editors (2017). *Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 21st International Conference KES-2017, Marseille, France, 6-8 September 2017*, volume 112 of *Procedia Computer Science*. Elsevier Science Publishers Ltd.

[Zarafin et al., 2012] Zarafin, A.-M., Zimmermann, A., and Boissier, O. (2012). Integrating Semantic Web Technologies and Multi-Agent Systems: A Semantic Description of Multi-Agent Organizations. In [Ossowski et al., 2012], pages 296–297.

[Zimmermann, 2007] Zimmermann, A. (2007). Integrated Distributed Description Logics. In [Calvanese et al., 2007], pages 507–514.

[Zimmermann, 2008] Zimmermann, A. (2008). *Sémantiques des réseaux de connaissances : gestion de l'hétérogénéité fondée sur le principe de médiation*. PhD thesis, Université Joseph Fourier, Grenoble, Grenoble (France).

[Zimmermann, 2010] Zimmermann, A. (2010). Ontology Recommendation for the Data Publishers. In [d'Aquin et al., 2010].

[Zimmermann, 2014] Zimmermann, A. (2014). RDF 1.1: On Semantics of RDF Datasets, W3C Working Group Note 25 February 2014. W3C Working Group Note, World Wide Web Consortium.

[Zimmermann and Euzenat, 2006] Zimmermann, A. and Euzenat, J. (2006). Three Semantics for Distributed Systems and their Relations with Alignment Composition. In [Cruz et al., 2006], pages 16–29.

[Zimmermann and Giménez-García, 2017a] Zimmermann, A. and Giménez-García, J. M. (2017a). Contextualizing DL Axioms: Formalization, a New Approach, and Its Properties. In [Dell'Aglio et al., 2017], pages 74–85.

[Zimmermann and Giménez-García, 2017b] Zimmermann, A. and Giménez-García, J. M. (2017b). Integrating Context of Statements within Description Logics. Technical report, Laboratoire Hubert Curien, Saint-Étienne, France.

[Zimmermann et al., 2013] Zimmermann, A., Gravier, C., Subercaze, J., and Cruzille, Q. (2013). Nell2RDF: Read the Web, and Turn it into RDF. In [Völker et al., 2013], pages 2–8.

[Zimmermann et al., 2006] Zimmermann, A., Krötzsch, M., Euzenat, J., and Hitzler, P. (2006). Formalizing Ontology Alignment and its Operations with Category Theory. In [Bennett and Fellbaum, 2006], pages 277–288.

[Zimmermann and Le Duc, 2008b] Zimmermann, A. and Le Duc, C. (2008b). Reasoning with a Network of Aligned Ontologies. In [Calvanese and Lausen, 2008], pages 43–57.

[Zimmermann et al., 2012] Zimmermann, A., Lopes, N., Polleres, A., and Straccia, U. (2012). A general framework for representing, reasoning and querying with annotated Semantic Web data. *Journal of Web Semantics*, 11:72–95.

[Zimmermann et al., 2009] Zimmermann, A., Sahay, R., Fox, R., and Polleres, A. (2009). Heterogeneity and Context in Semantic-Web-Enabled HCLS Systems. In [Meersman et al., 2009], pages 1165–1182.

Appendix A

Teaching activities

This part first shows what responsibilities I had in teaching, then summarises how much courses I taught, as of 31 August 2020. While I regularly teach basic courses in computer science, I am also in charge of several courses in relation to my area of expertise: Semantic Web, taught to different audiences over the years, ranging from a group of 10 students, to multiple promotions of master students and engineering school students with over 70 attendees; Knowledge Representation and Reasoning, as part of a large module on Artificial Intelligence. I am also in charge of a Master programme on *Data and Connected Systems* in which I teach Semantic Web technologies.

A.1 Teaching and pedagogical responsibilities

Sep. 2019–now: Master *Informatique, parcours DSC*

Oct. 2018–now: Course Group *Informatique, Mastère MTI*

Apr. 2015–Apr. 2019: Course Group *Modélisation de systèmes*, toolbox ICM

Apr. 2015–Apr. 2018: Module *Modélisation de connaissances*, toolbox ICM

Nov. 2012–now: Module Semantic Web, in various Master/Engineer programmes

Nov. 2011–Feb. 2014: Module *Conception de systèmes d'information*, common core ICM

Feb. 2009–Sep. 2010: Reading groups, in doctoral programme NUIG (Galway)

A.2 Courses

Table A.1 shows the time spent teaching various modules. Time is split between lectures (*Heures Cours Magistraux* or HCM), tutorial classes (*Heures Travaux Dirigés* or HTD), practical classes (*Heures Travaux Pratiques* or HTP) and total amount in *Heures Équivalent Travaux Dirigés* or HETD. The title of the course is given in its original language, in French or in English, depending on the language it is taught in.

Table A.1: Summary of course, by employment status, level and subject

Status	**Level**	**Subject**	**HCM**	**HTD**	**HTP**	**HETD**
Vacataire	L1 Info. maths appli.	*Algèbre et géométrie élémentaires*	0	48	0	48
ATER	1A IUT Tech. de co.	*Bureautique : Powerpoint, Word, Excel*	18	3	27	55.5
	2A IUT Tech. de co.	*Bureautique : Excel avancé, Access*	16	8	24	54.67
Post doc.	Master, PhD, post-doc	Description logics and OWL	0	0	7.5	11.25
ATER	1A Premier cycle INSA	Databases	6	0	9	17.5
	2A Premier cycle INSA	*Programmation orientée objet*	12	10	18	45
	2A Premier cycle INSA	*Projet informatique*	0	11	48	83
Maitre assistant associé	1A Tronc commun EMSE	*Langages et concepts de programmation*	0	0	36	54
	1A Tronc commun EMSE	*Conception de systèmes d'information*	3	9	18	38
	M1 Web Intelligence	*Représentation de connaissances et Web sémantique*	4	2	6	13.67
Maitre assistant	1A Tronc commun EMSE	*TP Linux*	0	0	3	4.5
	1A Tronc commun EMSE	*Langages et concepts de programmation*	0	0	69	103.5
	1A Tronc commun EMSE	*Conception de systèmes d'information*	6	18	36	76
	1A Tronc commun EMSE	*Programmation orientée objet*	0	24	66	123
	1A Tronc commun EMSE	*Mise à niveau informatique*	0	0	10.5	15.75
	1A Tronc commun EMSE	*Logique*	0	39	0	39
	1A Tronc commun EMSE	*Systèmes d'information*	0	22.5	36	67.75
	2A Axe ISI EMSE	*Analyse et conception de systèmes d'information*	0	21	0	21
	2A Axe ISI EMSE	*Théorie des langages et compilation*	0	0	13.5	20.25
	1A/2A Toolbox EMSE Modélisation de systèmes	*Modélisation de connaissances*	0	47	0	47
	2A Toolbox EMSE Intelligence Artificielle	*Représentation des connaissances*	0	30	0	30
	2A Toolbox EMSE I2SI	Semantic and data interoperability	0	9	0	9
	2A Défi EMSE Intelligence Artificielle	Knowledge representation and reasoning	22	10	6	17.25
	M1 Web Intelligence	*Représentation de connaissances et Web sémantique*	8	4	12	27.33
	M1, M2, Ing. Civ. Mines 2A/3A, Télécom St-Étienne 3A	Semantic Web	70	20	178	317
	Mastère spécialisé MTI	*Interopérabilité des systèmes d'information*	0	30	0	30
	Mastère spécialisé MTI	*Digitalisation*	0	24	0	24
		Total:	**152**	**384**	**613**	**1392.92**

Appendix B

Research activities

Here I list activities that are inherent to the work of a researcher, excluding student supervisions that are detailed in Appendix C and project activities that are discussed in Appendix D. This also shows my involvement in the research community, nationally and internationally.

B.1 Academic responsibilities

July 2020: Organisation co-chair, Summer School on AI Technologies for Trust, Interoperability and Autonomy in Industry 4.0

Sep. 2019: Coorganiser of international workshop WOMoCoE 2019

Oct. 2018: Coorganiser of international workshop WOMoCoE 2018

6 July 2018: Session chair, national conference PFIA 2018

6 June 2018: Session chair, international conference ESWC 2018

25 Apr. 2018: Session chair, Industry track at the Web Conference 2018

2018: Guest editor, French journal ISI

3 June 2015: Session chair, international conference ESWC 2015

Feb. 2015–now: Editorial board, Journal of Web Semantics

Aug. 2014: Organisation chair, international Summer School WISS 2014

May 2014: Organisation chair, national conference JIAE 2014

2012–now: Steering committee, international conference series AT

B.2 Programme committees

Here I list the conferences and workshops for which I was part of the programme committee. This responsibility comes with a duty to review submitted papers, the number of which I indicate between parenthesis.

Year	Conferences or workshops	Reviews
2020	ISWC (3), FOIS (3), IJCAI (2), TheWebConf demo (4)	12
2019	ISWC (2), KGB (1), IJCAI (4), ESWC (3), TheWebConf demo (4), AWD (1)	15
2018	WOMoCoE (1), SKG (2), SEMANTiCS poster & demo (2), CKG (3), SSN (1), SEMANTiCS (1), FOIS (2), CNIA+RJCIA (2), ESWC (3), TheWebConf demo (4), EGC (1), AAAI (5)	27
2017	SEMANTiCS demo (1), SEMANTiCS (2), ISWC (3), IJCAI (7), ESWC (2), EGC (2)	17
2016	IC (2), SEMPER (1), LDQ (1), FOIS (3), WebSci (5), IJCAI (3), ESWC (4), WWW (3)	22
2015	AT (2), ISWC, LinkED (1), COLD (2), ICLP-DC (1), IC (1), LDQ (1), ESWC (3), WWW (3), AAAI (3)	17
2014	LDQ (1), COLD (2), ISWC (3), FOIS (3), ICLP-DC (2), ECAI (3), ESWC (3), IC (1)	18
2013	ISWC (4), ICLP-DC (3), AT (1), MASTS (1), IJCAI (2), ESWC (5)	16
2012	AOW (1), ISWC (4), ARCOE (1), ESWC (2), ICISTM (3)	11
2011	AOW (1), ISWC (5), OWLED (1), ESWC (2), ISWSA (1)	10
2010	AOW (1), ISWSA (1), WoMO (1)	3
2009	AOW (1), ISWC (4)	5
2008	WORM (2)	2
2007	WoMO (1)	1
Total:	**176 reviews**	

B.3 Other reviews

I list the number of reviews in parenthesis.

International journals: AI Review (1), Applied Ontology (3), Data and Knowledge Engineering (1), IEEE Transactions on Knowledge and Data Engineering (1), IETE Technical Review (1), Information Sciences (1), International Journal of Web Information Systems (1), Journal of Applied Artificial Intelligence (1), Journal of Autonomous Agents and Multi-Agent Systems (1), Journal of Mathematical and Computer Modelling (1), Journal of Data and Information Quality (1), Journal on Data Semantics (3), Journal of Web Semantics (3), Knowledge and Information Systems (1), Semantic Web Journal (12), Studia Logica (1), The Computer Journal (3). **Total:** 36 reviews.

Conferences and workshops: AAAI (2), AAMAS (1), AIMSA (2), CILC (1), EKAW (2), ESWC (1+1), FOIS (1+1), INCOM (1), ISWC+ASWC (1), JELIA (1), KEOD (1), KROW (1), LDOW (1), OM (1), RFIA (1), RR (1), WI (2), WWW (1), WWW demo (2+3). **Total:** 29 reviews.

Various: Workshop proposals (3), EU project presentation (1), Metareviews as track chair and senior PC member (31), as an editorial board member (15), national project proposals (3) regional project proposals (5). **Total:** 48 reviews.

Total: 123 reviews.

B.4 Other tasks

Track chair: I was Knowledge Graph track chair at ESWC 2020 for which I did 21 metareviews

Senior PC member: I did 10 metareviews at ISWC 2015

Proceedings chair: In charge of compiling the proceedings for ESWC 2015.

Metadata chair: In charge of generating and publishing the metadata of WWW 2012. I also organised a metadata challenge.

Appendix C

Supervision

In this section, I mention the students I supervised or am supervising for a medium to long period. I exclude short supervisions of course work. For each student, I add references to our joint publications during the supervision. At the end, I also add the PhD defences that I attended as a member of the jury, additionally to my own supervised students.

C.1 Defended PhD theses

Nov. 2016–Sep. 2020

Student: Omar Qawasmeh

University: Université Jean Monnet Saint-Étienne, France

Defence date: 25 Sep. 2020

Title of dissertation: *Towards A Collaborative Framework for Ontology Engineering: Impact on Ontology Evolution and Pitfalls in Ontology Networks and Versioned Ontologies*

Supervisors: Prof. Pierre Maret, Maxime Lefrançois and **Antoine Zimmermann**

Publications: [Qawasmeh et al., 2018, Qawasmeh et al., 2019, Qawasmeh et al., 2020]

Oct. 2015–Dec. 2018

Student: Mohammed Noorani Bakerally

University: École des mines de Saint-Étienne, France

Defence date: 20 Dec. 2018

Title of dissertation: *Generation of Linked Data Platforms from existing sources in highly decentralized information ecosystems*

Supervisors: Prof. Olivier Boissier and **Antoine Zimmermann**

Publications: [Bakerally et al., 2016, Lefrançois et al., 2016b, Lefrançois et al., 2017c, Lefrançois et al., 2017b, Bakerally and Zimmermann, 2017, Bakerally, 2017, Bakerally et al., 2018a, Bakerally et al., 2018b]

Current position: Post-doc at LAAS, Toulouse, France

Oct. 2014–Oct. 2018

Student: Oudom Kem

University: École des mines de Saint-Étienne, France

Defence date: 25 Oct. 2018

Title of dissertation: *Modĺisation et Exploitation des Connaissances de l'Environnement : Une Approche Multi-Agents pour la Recherche d'Itinéraires Multi-Objectifs dans des Environnements Ubiquitaires*

Supervisors: Prof. Flavien Balbo and **Antoine Zimmermann**

Publications: [Kem et al., 2016, Kem et al., 2017a, Kem et al., 2017b, Kem et al., 2017c, Kem et al., 2017d]

Current position: Post-doc at CEA Saclay, France

June 2013–Oct. 2016

Student: Sihem Klai-Soukehal

University: University Badji Mokhtar Annaba, Algeria

Defence date: 25 Oct. 2016

Title of dissertation: *Approches flexibles pour l'intégration des ontologies hétérogènes et évolutives*

Supervisors: Prof. Mohammed Tarek Khadir and **Antoine Zimmermann**

Publications: [Klai et al., 2016b, Klai et al., 2016a]

Current position: Maitre de conférences, University Badji Mokhtar Annaba, Algeria

Dec 2012–Jan. 2016

Student: Andrei-Nicolae Ciortea

University: École des mines de Saint-Étienne, France and Universitea Politehnica din Bucureşti, Romania

Defence date: 14 Jan. 2016

Title of dissertation: *Weaving the Social Web of Things: Enabling Autonomous and Flexible Interaction in the Internet of Things Encadrement*

Supervisors: Prof. Olivier Boissier, Prof. Adina Magda-Florea and **Antoine Zimmermann**

Publications: [Ciortea et al., 2013, Ciortea et al., 2014, Ciortea et al., 2015, Ciortea et al., 2016a, Ciortea et al., 2016b]

Current position: Post-doc, University Saint Gallen, Switzerland

Mar. 2009–May 2012

Student: Ratnesh Nandan Sahay

University: National University of Ireland, Galway

Defence date: 21 May 2012

Title of dissertation: *An Ontological Framework for Interoperability of Health Level Seven (HL7) Applications: the PPEPR Methodology and System*

Supervisors: Dr. Axel Polleres, Prof. Dr. Manfred Hauswirth, informally supervised by Dr. Ronan Fox and **Antoine Zimmermann**

Publications: [Zimmermann et al., 2009, Sahay et al., 2011, Sahay et al., 2013]

Current position: Associate Director, Clinical Data Science, AstraZeneca, Cambridge, United Kingdom

C.2 Current PhD supervision

Nov. 2020–now

Student: Yousouf Taghzouti

University: École des mines de Saint-Étienne, France

Supervisors: Prof. Mireille Batton-Hubert, Maxime Lefrançois and **Antoine Zimmermann**

Nov. 2020–now

Student: Gabriel Martin Lopes Cavalcante

University: École des mines de Saint-Étienne, France

Supervisors: Prof. Flavien Balbo, Maxime Lefrançois and **Antoine Zimmermann**

Oct. 2015–now

Student: José Miguel Giménez-García

University: Université Jean Monnet Saint-Étienne, France

Supervisors: Prof. Pierre Maret and **Antoine Zimmermann**

Publications: [Giménez-García et al., 2016a, Giménez-García et al., 2016b, Giménez-García et al., 2017, Zimmermann and Giménez-García, 2017a, Zimmermann and Giménez-García, 2017b, Endris et al., 2017, Giménez-García et al., 2019, Giménez-García and Zimmermann, 2019]

C.3 Supervision of Master students

Apr. 2019–June 2019

Student: Ali Haidar

University: Université Jean Monnet Saint-Étienne, France

Title of dissertation: Comparing different models to represent metadata in the Semantic Web

Supervisors: Prof. Pierre Maret, José M. Giménez-García and **Antoine Zimmermann**

Fev. 2018–Aug. 2018

Student: Alaa Daoud

University: École des mines de Saint-Étienne, France

Title of dissertation: Semantic Web Environments for MAS – Enabling agents to use Web of Things environments via semantic web

Supervisors: Prof. Olivier Boissier, Andrei Ciortea, Maxime Lefrançois and **Antoine Zimmermann**

Feb. 2014–June 2014

Student: Moncef Ben Rajeb

University: Université Jean Monnet Saint-Étienne, France

Title of dissertation: *Annotation temporelle des données DBpedia*

Supervisors: Mihaela Jugunaru-Mathieu and **Antoine Zimmermann**

Feb. 2013–June 2013

Student: Quentin Cruzille

University: Université Jean Monnet Saint-Étienne, France

Title of dissertation: Concrete Specification and Implementation of Annotated RDF(S)

Supervisors: **Antoine Zimmermann**

Publication: [Zimmermann et al., 2013].

Feb. 2012–June 2012

Student: Alexandra-Mădălina Zarafin

University: Universitea Politehnica din Bucureşti, Romania

Title of dissertation: Design of agent organizations ontology and integration with the MOISE+ organizational model

Supervisors: Prof. Olivier Boissier and **Antoine Zimmermann**

Publication: [Zarafin et al., 2012].

C.4 Member of the jury of PhD defences

3 Mar. 2020

Student: Niklas Petersen

University: University of Bonn, Germany

Title of dissertation: *Towards Semantic Integration of Supply Chain and Production Data*

Supervisors: Prof. Dr. Sören Auer

Role: related field member of the commission

6 Nov. 2017

Student: Abdullah Abbas

University: Université Grenoble Alpes, France

Title of dissertation: *Static Analysis of Semantic Web Queries with ShEx Schema Constraints*

Supervisors: Dr. Pierre Genevès and Prof. Cécile Roisin

Role: Examinateur

Appendix D

Projects

Here I list the research projects to which I participated. I split this into international projects, national projects, and I also give standardisation efforts to which I participated. My roles in these projects have spanned from simple contributor to a deliverable, to work package leader (Agreement Technologies, Work Package on Semantics; DKG, Work Package on Prosumers), to coordinator (ANR project OpenSensingCity). All of these projects were either focused on semantic web technologies, or using them as a means to achieve the objectives.

D.1 International projects

Duration	Title	Type
Sep. 2020–Sep. 2024	DKG: Distributed Knowldege Graphs (**WP leader**)	COST Action
Jan. 2015–Dec. 2018	WDAqua: Answering Questions with Web Data	H2020 ITN
Sep. 2013–Jan. 2017	SEAS: Smart Energy Aware Systems	ITEA3
May 2009–Oct. 2012	Agreement Technologies (**WP leader**)	COST Action
Sep. 2006–Nov. 2008	NeOn: Lifecycle Support for Networked Ontologies	FP6 IP
Nov. 2004–Dec. 2007	Knowledge Web	FP6 NoE

D.2 National projects

Duration	Title	Type
Apr. 2020–Apr. 2024	HyperAgents: Hypermedia Communities of People and Autonomous Agents	ANR/SNSF
Feb. 2020–Feb. 2024	CoSWoT: Constrained Semantic Web of Things	ANR
June 2018–May 2019	Bilateral contract with Engie	Direct contract
Apr. 2016–Mar. 2017	Bilateral contract with Engie (**PI**)	Direct contract
Mar. 2015–Sep. 2018	OpenSensingCity: Fostering Uses and Usages of Open Sensor Data in Smart Cities (**coordinator**)	ANR
Jan. 2014–Jui. 2018	Ethicaa: Ethics and Autonomous Agents	ANR
Dec. 2013–May 2017	Vigi4Med: *Recherche et analyse des effets indésirables rapportés par les patients dans les réseaux sociaux* (**PI**)	ANSM
Feb. 2009–Sep.2010	Líon 2	Science Foundation Ireland

D.3 Standardisation

Duration	Title
Feb. 2017–now	W3C Web of Things Working Group
Jan. 2015–Oct. 2017	W3C Spatial Data on the Web Working Group
Jan. 2011–June 2014	W3C RDF Working Group
Mar.2009–Dec.2009	W3C Web Ontology Language Working Group

Appendix E

Publications

E.1 Edited conference proceedings

[Gandon et al., 2015a] Gandon, F., Cabrio, E., Stankovic, M., and Zimmermann, A., editors (2015a). *Semantic Web Evaluation Challenges - Second SemWebEval Challenge at ESWC 2015, Portorož, Slovenia, May 31 - June 4, 2015, Revised Selected Papers*, volume 548 of *Communications in Computer and Information Science*. Springer.

[Gandon et al., 2015b] Gandon, F., Guéret, C., Villata, S., Breslin, J. G., Faron-Zucker, C., and Zimmermann, A., editors (2015b). *The Semantic Web: ESWC 2015 Satellite Events - ESWC 2015 Satellite Events Portorož, Slovenia, May 31 - June 4, 2015, Revised Selected Papers*, volume 9341 of *Lecture Notes in Computer Science*. Springer.

[Gandon et al., 2015c] Gandon, F., Sabou, M., Sack, H., d'Amato, C., Cudré-Mauroux, P., and Zimmermann, A., editors (2015c). *The Semantic Web. Latest Advances and New Domains - 12th European Semantic Web Conference, ESWC 2015, Portoroz, Slovenia, May 31 - June 4, 2015. Proceedings*, volume 9088 of *Lecture Notes in Computer Science*. Springer.

E.2 Edited workshop proceedings

[Barton et al., 2019] Barton, A., Seppälä, S., Porello, D., Hahmann, T., Peñaloza, R., Schulz, S., Guizzardi, G., Kutz, O., Troquard, N., Hedblom, Maria M.and Righetti, G., Masolo, C., Ravelli, A. A., Basile, V., Caselli, T., Radicioni, D. P., Mosca, A., Confalonieri, R., Calvanese, D., Grüninger, M., Terkaj, W., Galton, A., Borgo, S., Loebe, F., Neuhaus, F., Boeker, M., Jansen, L., Melone, R. S., Hinterwaldner, I., Brochhausen, M., Guarino, N., Adrian, K., Euzenat, J., Gromann, D., Jiménez-Ruiz, E., Schorlemmer, M., Tamma, V., Asmundo, M. N., Ferrario, R., Sanfilippo, E. M., Bozzato, L., Mossakowski, T., and Zimmermann, A., editors (2019). *Proceedings of the Joint Ontology Workshops 2019, Episode V: The Styrian Autumn of Ontology, Graz, Austria, September 23-25, 2019*, volume 2518 of *CEUR Workshop Proceedings*. Sun SITE Central Europe (CEUR).

[Čyras et al., 2018] Čyras, K., Oliveira, T., Williams, M., Bozzato, L., Homola, M., Mossakowski, T., and Zimmermann, A., editors (2018). *Proceedings of the Joint*

Proceedings of Reasoning with Ambiguous and Conflicting Evidence and Recommendations in Medicine (MedRACER 2018) and the 3rd International Workshop on Ontology Modularity, Contextuality, and Evolution (WOMoCoE 2018) co-located with the 16th International Conference on Principles of Knowledge Representation and Reasoning (KR 2018), Tempe, Arizona, USA, October 29th, 2018, volume 2237 of *CEUR Workshop Proceedings*. Sun SITE Central Europe (CEUR).

E.3 International journal articles

[Audeh et al., 2017] Audeh, B., Neigbeder, M., Zimmermann, A., Jaillon, P., and Bousquet, C. (2017). Vigi4Med Scraper: A Framework for Web Forum Structured Data Extraction and Semantic Representation. *Public Library of Science*, 12(1).

[Gillani et al., 2019] Gillani, S., Zimmermann, A., Picard, G., and Laforest, F. (2019). A query language for semantic complex event processing: Syntax, semantics and implementation. *Semantic Web Journal*, 10(1):53–93.

[Gravier et al., 2015] Gravier, C., Subercaze, J., and Zimmermann, A. (2015). Conflict resolution when axioms are materialized in semantic-based smart environments. *Journal of Ambient Intelligence and Smart Environments*, 7(2):187–199.

[Gyrard et al., 2018] Gyrard, A., Zimmermann, A., and Sheth, A. P. (2018). Building IoT-Based Applications for Smart Cities: How Can Ontology Catalogs Help? *IEEE Internet of Things Journal*, 5(5):3978–3990.

[Hogan et al., 2012] Hogan, A., Zimmermann, A., Umbrich, J., Polleres, A., and Decker, S. (2012). Scalable and distributed methods for entity matching, consolidation and disambiguation over linked data corpora. *Journal of Web Semantics*, 10:76–110.

[Kem et al., 2017] Kem, O., Balbo, F., and Zimmermann, A. (2017). Traveler-Oriented Advanced Traveler Information System Based on Discovery and Exploitation of Resources: Potentials and Challenges. *Transportation Research Procedia*, 22:635–644. Special issue on 19th EURO Working Group on Transportation Meeting, EWGT2016, 5-7 September 2016, Istanbul, Turkey.

[Khan et al., 2019] Khan, Y., Zimmermann, A., Jha, A., Gadepally, V., d'Aquin, M., and Sahay, R. (2019). One Size Does Not Fit All: Querying Web Polystores. *IEEE Access*, 7:9598–9617.

[Klai et al., 2016] Klai, S., Zimmermann, A., and Khadir, M. T. (2016). Multi-level networked knowledge: representation and DL-reasoning. *International Journal of Metadata, Semantics and Ontologies*, 11(1):1–15.

[Klai et al., 2019] Klai, S., Zimmermann, A., and Khadir, M. T. (2019). Networked Ontologies with Contextual Alignments. *Computing and Informatics*, 38(1):115–150.

[Sorici et al., 2015] Sorici, A., Boissier, O., Picard, G., Zimmermann, A., and Magda-Florea, A. (2015). CONSERT: Applying Semantic Web Technologies to

Context Modeling in Ambient Intelligence. *Computers and Electrical Engineering*, 44:280–306.

[Zimmermann et al., 2012] Zimmermann, A., Lopes, N., Polleres, A., and Straccia, U. (2012). A general framework for representing, reasoning and querying with annotated Semantic Web data. *Journal of Web Semantics*, 11:72–95.

E.4 Book chapters

[Laborie and Zimmermann, 2009] Laborie, S. and Zimmermann, A. (2009). On Using Information Retrieval Techniques for Semantic Media Adaptation. In Angelides, M. C., editor, *Advances in Semantic Media Adaptation and Personalization, Volume 2*, pages 137–157. Auerbach Publications.

[Qawasmeh et al., 2019] Qawasmeh, O., Lefrançois, M., Zimmermann, A., and Maret, P. (2019). Pitfalls in Networked and Versioned Ontologies. In Fred, A. L. N., Salgado, A., Aveiro, D., Dietz, J., Bernardino, J., and Felipe, J., editors, *Knowledge Discovery, Knowledge Engineering and Knowledge Management - 11th International Joint Conference, IC3K 2019, Vienna, Austria, September 17-19, 2019, Revised Selected Papers*, Communications in Computer and Information Science, pages 185–212. Springer.

[Sahay et al., 2013] Sahay, R., Zimmermann, A., Fox, R., Polleres, A., and Hauswirth, M. (2013). A Formal Investigation of Semantic Interoperability of HCLS Systems. In Sicilia, M.-A. and Serrano-Balazote, P., editors, *Interoperability in Healthcare Information Systems: Standards, Management, and Technology*. IGI Global.

[Zimmermann, 2013] Zimmermann, A. (2013). Logical formalisms for Agreement Technologies. In Ossowski, S., editor, *Agreement Technologies*, Law, Governance and Technology Series. Springer.

E.5 International conference articles

[Alkhateeb and Zimmermann, 2007] Alkhateeb, F. and Zimmermann, A. (2007). Query Answering in Distributed Description Logics. In Labiod, H. and Badra, M., editors, *New Technologies, Mobility and Security - Proceedings of NTMS'2007 Conference*, pages 523–534. Springer.

[Bakerally et al., 2016] Bakerally, N., Boissier, O., and Zimmermann, A. (2016). Smart City Artifact Web Portal. In Sack, H., Rizzo, G., Steinmetz, N., Mladenic, D., Auer, S., and Lange, C., editors, *The Semantic Web - ESWC 2016 Satellite Events, Heraklion, Crete, Greece, May 29 - June 2, 2016, Revised Selected Papers*, volume 9989 of *Lecture Notes in Computer Science*, pages 172–177. Springer.

[Bakerally and Zimmermann, 2017] Bakerally, N. and Zimmermann, A. (2017). A system to automatize the deployment of data in Linked Data Platforms. In Nikitina, N., Song, D., Fokoue, A., and Haase, P., editors, *Proceedings of the ISWC 2017 Posters & Demonstrations and Industry Tracks co-located with 16th*

International Semantic Web Conference (ISWC 2017), Vienna, Austria, October 23rd - to - 25th, 2017, volume 1963 of *CEUR Workshop Proceedings*. Sun SITE Central Europe (CEUR).

[Bakerally et al., 2018a] Bakerally, N., Zimmermann, A., and Boissier, O. (2018a). A Workflow for Generation of LDP. In Gangemi, A., Gentile, A. L., Nuzzolese, A. G., Rudolph, S., Maleshkova, M., Paulheim, H., Pan, J. Z., and Alam, M., editors, *The Semantic Web: ESWC 2018 Satellite Events - ESWC 2018 Satellite Events, Heraklion, Crete, Greece, June 3-7, 2018, Revised Selected Papers*, volume 11155 of *Lecture Notes in Computer Science*, pages 142–147. Springer.

[Bakerally et al., 2018b] Bakerally, N., Zimmermann, A., and Boissier, O. (2018b). LDP-DL: A Language to Define the Design of Linked Data Platforms. In Gangemi, A., Navigli, R., Vidal, M., Hitzler, P., Raphaël, T., Hollink, L., Tordai, A., and Alam, M., editors, *The Semantic Web - 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3-7, 2018, Proceedings*, volume 10843 of *Lecture Notes in Computer Science*, pages 33–49. Springer.

[Belloni et al., 2014] Belloni, A., Berger, A., Besson, V., Boissier, O., Bonnet, G., Bourgne, G., Chardel, P.-A., Cotton, J.-P., Evreux, N., Ganascia, J.-G., Jaillon, P., Mermet, B., Picard, G., Reber, B., Simon, G., De Swarte, T., Tessier, C., Vexler, F., Voyer, R., and Zimmermann, A. (2014). Towards A Framework To Deal With Ethical Conflicts In Autonomous Agents And Multi-Agent Systems. In *Proceedings of CEPE 2014: Well-being, Flourishing, and ICTs*.

[Bennara et al., 2020] Bennara, M., Lefrançois, M., Zimmermann, A., and Messalti, N. (2020). Interoperability of Semantically-Enabled Web Services on the WoT: Challenges and Prospects. In Indrawan-Santiago, M., Pardede, E., Salvadori, I. L., Steinbauer, M., Khalil, I., and Kotsis, G., editors, *Proceedings of the 22nd International Conference on Information Integration and Web-based Applications & Services, iiWAS 2020, Chiang Mai, Thailand, November 30-December 2, 2020*, page 149–153. ACM Press.

[Calvier et al., 2016] Calvier, F., Kammoun, A., Zimmermann, A., Singh, K., and Fayolle, J. (2016). Ontology Driven Complex Event Pattern Definition (Short Paper). In Debruyne, C., Panetto, H., Meersman, R., Dillon, T. S., Kühn, e., O'Sullivan, D., and Ardagna, C. A., editors, *On the Move to Meaningful Internet Systems: OTM 2016 Conferences - Confederated International Conferences: CoopIS, C&TC, and ODBASE 2016, Rhodes, Greece, October 24-28, 2016, Proceedings*, volume 10033 of *Lecture Notes in Computer Science*, pages 522–530. Springer.

[Ciortea et al., 2014] Ciortea, A., Boissier, O., Zimmermann, A., and Magda-Florea, A. (2014). Open and Interoperable Socio-technical Networks. In Winkler, F., Negru, V., Ida, T., Jebelean, T., Petcu, D., Watt, S. M., and Zaharie, D., editors, *16th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, SYNASC 2014, Timisoara, Romania, September 22-25, 2014*, pages 251–257. IEEE Computer Society.

[Ciortea et al., 2016] Ciortea, A., Boissier, O., Zimmermann, A., and Magda-Florea, A. (2016). Responsive Decentralized Composition of Service Mashups for the

Internet of Things. In Schneegass, S., Schmidt, A., Michahelles, F., Kritzler, M., Ilic, A., and Kunze, K., editors, *Proceedings of the 6th International Conference on the Internet of Things, IOT 2016, Stuttgart, Germany, November 7-9, 2016*, pages 53–61. ACM Press.

[Ciortea et al., 2017] Ciortea, A., Boissier, O., Zimmermann, A., and Magda-Florea, A. (2017). Give Agents Some REST: A Resource-oriented Abstraction Layer for Internet-scale Agent Environments. In Larson, K., Winikoff, M., Das, S., and Durfee, E. H., editors, *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems, AAMAS 2017, São Paulo, Brazil, May 8-12, 2017*, pages 1502–1504. ACM Press.

[Ciortea et al., 2019] Ciortea, A., Mayer, S., Gandon, F. L., Boissier, O., Ricci, A., and Zimmermann, A. (2019). A Decade in Hindsight: The Missing Bridge Between Multi-Agent Systems and the World Wide Web. In Elkind, E., Veloso, M., Agmon, N., and Taylor, M. E., editors, *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS '19, Montreal, QC, Canada, May 13-17, 2019*. International Foundation for Autonomous Agents and Multiagent Systems.

[Ciortea et al., 2015] Ciortea, A., Zimmermann, A., Boissier, O., and Magda-Florea, A. (2015). Towards a Social and Ubiquitous Web: A Model for Socio-technical Networks. In *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, WI-IAT 2015, Singapore, December 6-9, 2015 - Volume I*, pages 461–468. IEEE Computer Society.

[Dumitrache et al., 2011] Dumitrache, A., Mileo, A., Zimmermann, A., Polleres, A., Obermeier, P., and Friel, O. (2011). Enabling Privacy-Preserving Semantic Presence in Instant Messaging Systems. In Beigl, M., Christiansen, H., Roth-Berghofer, T., Kofod-Petersen, A., Coventry, K. R., and Schmidtke, H. R., editors, *Modeling and Using Context - 7th International and Interdisciplinary Conference, CONTEXT 2011, Karlsruhe, Germany, September 26-30, 2011. Proceedings*, volume 6967 of *Lecture Notes in Computer Science*, pages 82–96. Springer.

[Endris et al., 2017] Endris, K. M., Giménez-García, J. M., Thakkar, H., Demidova, E., Zimmermann, A., Lange, C., and Simperl, E. (2017). Dataset Reuse: An Analysis of References in Community Discussions, Publications and Data. In Corcho, Ó., Janowicz, K., Rizzo, G., Tiddi, I., and Garijo, D., editors, *Proceedings of the Knowledge Capture Conference, K-CAP 2017, Austin, TX, USA, December 4-6, 2017*, pages 5:1–5:4. ACM Press.

[Giménez-García et al., 2016] Giménez-García, J. M., Zimmermann, A., and Maret, P. (2016). Representing Contextual Information as Fluents. In Ciancarini, P., Poggi, F., Horridge, M., Zhao, J., Groza, T., Suárez-Figueroa, M. C., d'Aquin, M., and Presutti, V., editors, *Knowledge Engineering and Knowledge Management - EKAW 2016 Satellite Events, EKM and Drift-an-LOD, Bologna, Italy, November 19-23, 2016, Revised Selected Papers*, volume 10180 of *Lecture Notes in Computer Science*, pages 119–122. Springer.

[Giménez-García et al., 2017] Giménez-García, J. M., Zimmermann, A., and Maret, P. (2017). NdFluents: An Ontology for Annotated Statements with Inference

Preservation. In Blomqvist, E., Maynard, D., Gangemi, A., Hoekstra, R., Hitzler, P., and Hartig, O., editors, *The Semantic Web - 14th International Conference, ESWC 2017, Portorož, Slovenia, May 28 - June 1, 2017, Proceedings, Part I*, volume 10249 of *Lecture Notes in Computer Science*, pages 638–654. Springer.

[Hauswirth et al., 2010] Hauswirth, M., Euzenat, J., Friel, O., Griffin, K., Hession, P., Jennings, B., Groza, T., Handschuh, S., Podnar Zarko, I., Polleres, A., and Zimmermann, A. (2010). Towards Consolidated Presence. In Aberer, K. and Joshi, J., editors, *The 6th International Conference on Collaborative Computing: Networking, Applications and Worksharing, CollaborateCom 2010, Chicago, IL, USA, 9-12 October 2010*, pages 1–10. IEEE Computer Society.

[Jung et al., 2007] Jung, J. J., Zimmermann, A., and Euzenat, J. (2007). Concept-Based Query Transformation Based on Semantic Centrality in Semantic Peer-to-Peer Environment. In Dong, G., Lin, X., Wang, W., Yang, Y., and Yu, J. X., editors, *Advances in Data and Web Management, Joint 9th Asia-Pacific Web Conference, APWeb 2007, and 8th International Conference, on Web-Age Information Management, WAIM 2007, Huang Shan, China, June 16-18, 2007, Proceedings*, volume 4505 of *Lecture Notes in Computer Science*, pages 622–629. Springer.

[Kem et al., 2016] Kem, O., Balbo, F., and Zimmermann, A. (2016). A Distributed Approach to Constructing Travel Solutions by Exploiting Web Resources. In *2016 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2016, Omaha, NE, USA, October 13-16, 2016*, pages 713–716. IEEE Computer Society.

[Kem et al., 2017a] Kem, O., Balbo, F., and Zimmermann, A. (2017a). Collaborative Search for Multi-goal Pathfinding in Ubiquitous Environments. In Berndt, J. O., Petta, P., and Unland, R., editors, *Multiagent System Technologies - 15th German Conference, MATES 2017, Leipzig, Germany, August 23-26, 2017, Proceedings*, volume 10413 of *Lecture Notes in Computer Science*, pages 72–88. Springer.

[Kem et al., 2017b] Kem, O., Balbo, F., Zimmermann, A., and Nagellen, P. (2017b). Multi-goal Pathfinding in Cyber-Physical-Social Environments: Multi-layer Search over a Semantic Knowledge Graph. In Zanni-Merk, C., Frydman, C. S., Toro, C., Hicks, Y., Howlett, R. J., and Jain, L. C., editors, *Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 21st International Conference KES-2017, Marseille, France, 6-8 September 2017*, volume 112 of *Procedia Computer Science*, pages 741–750. Elsevier Science Publishers Ltd.

[Klai et al., 2016] Klai, S., Zimmermann, A., and Khadir, M. T. (2016). Multi-level Networked Knowledge Base: DDL-Reasoning. In Bellatreche, L., Pastor, O., Almendros-Jiménez, J. M., and Aït Ameur, Y., editors, *Model and Data Engineering - 6th International Conference, MEDI 2016, Almería, Spain, September 21-23, 2016, Proceedings*, volume 9893 of *Lecture Notes in Computer Science*, pages 118–131. Springer.

[Le Duc et al., 2010] Le Duc, C., Lamolle, M., and Zimmermann, A. (2010). An API for Distributed Reasoning on Networked Ontologies with Alignments. In Filipe, J.

and Dietz, J. L. G., editors, *KEOD 2010 - Proceedings of the International Conference on Knowledge Engineering and Ontology Development, Valencia, Spain, October 25-28, 2010*, pages 295–304. SciTePress.

[Lefrançois and Zimmermann, 2016] Lefrançois, M. and Zimmermann, A. (2016). Supporting Arbitrary Custom Datatypes in RDF and SPARQL. In Sack, H., Blomqvist, E., d'Aquin, M., Ghidini, C., Ponzetto, S. P., and Lange, C., editors, *The Semantic Web. Latest Advances and New Domains - 13th International Conference, ESWC 2016, Heraklion, Crete, Greece, May 29 - June 2, 2016, Proceedings*, volume 9678 of *Lecture Notes in Computer Science*, pages 371–386. Springer.

[Lefrançois and Zimmermann, 2018] Lefrançois, M. and Zimmermann, A. (2018). The Unified Code for Units of Measure in RDF: cdt:ucum and other UCUM Datatypes. In Gangemi, A., Gentile, A. L., Nuzzolese, A. G., Rudolph, S., Maleshkova, M., Paulheim, H., Pan, J. Z., and Alam, M., editors, *The Semantic Web: ESWC 2018 Satellite Events - ESWC 2018 Satellite Events, Heraklion, Crete, Greece, June 3-7, 2018, Revised Selected Papers*, volume 11155 of *Lecture Notes in Computer Science*, pages 196–201. Springer.

[Lefrançois et al., 2016] Lefrançois, M., Zimmermann, A., and Bakerally, N. (2016). Flexible RDF generation from RDF and heterogeneous data sources with SPARQL-Generate. In Ciancarini, P., Poggi, F., Horridge, M., Zhao, J., Groza, T., Suárez-Figueroa, M. C., d'Aquin, M., and Presutti, V., editors, *Knowledge Engineering and Knowledge Management - EKAW 2016 Satellite Events, EKM and Drift-an-LOD, Bologna, Italy, November 19-23, 2016, Revised Selected Papers*, volume 10180 of *Lecture Notes in Computer Science*, pages 131–135. Springer.

[Lefrançois et al., 2017] Lefrançois, M., Zimmermann, A., and Bakerally, N. (2017). A SPARQL Extension for Generating RDF from Heterogeneous Formats. In Blomqvist, E., Maynard, D., Gangemi, A., Hoekstra, R., Hitzler, P., and Hartig, O., editors, *The Semantic Web - 14th International Conference, ESWC 2017, Portorož, Slovenia, May 28 - June 1, 2017, Proceedings, Part I*, volume 10249 of *Lecture Notes in Computer Science*, pages 35–50. Springer.

[Lopes et al., 2012] Lopes, N., Kirrane, S., Zimmermann, A., Polleres, A., and Mileo, A. (2012). A Logic Programming approach for Access Control over RDF. In Dovier, A. and Santos Costa, V., editors, *Technical Communications of the 28th International Conference on Logic Programming, ICLP 2012, September 4-8, 2012, Budapest, Hungary*, pages 381–392. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik.

[Lopes et al., 2010] Lopes, N., Polleres, A., Straccia, U., and Zimmermann, A. (2010). AnQL: SPARQLing Up Annotated RDFS. In Patel-Schneider, P. F., Pan, Y., Hitzler, P., Mika, P., Zhang, Y., Pan, J. Z., Horrocks, I., and Glimm, B., editors, *The Semantic Web - ISWC 2010 - 9th International Semantic Web Conference, ISWC 2010, Shanghai, China, November 7-11, 2010, Revised Selected Papers, Part I*, volume 6496 of *Lecture Notes in Computer Science*, pages 518–533. Springer.

[Passant et al., 2010] Passant, A., Zimmermann, A., Schneider, J., and Breslin, J. G. (2010). A semantic framework for modelling quotes in email conversations.

In Alnsour, A. and Aljawarneh, S., editors, *Proceedings of the 1st International Conference on Intelligent Semantic Web-Services and Applications, ISWSA 2010, Amman, Jordan, June 14-16, 2010*. ACM Press.

[Petersen et al., 2016] Petersen, N., Grangel-González, I., Coskun, G., Auer, S., Frommhold, M., Tramp, S., Lefrançois, M., and Zimmermann, A. (2016). SCOR-Voc: Vocabulary-Based Information Integration and Exchange in Supply Networks. In *Tenth IEEE International Conference on Semantic Computing, ICSC 2016, Laguna Hills, CA, USA, February 4-6, 2016*, pages 132–139. IEEE Computer Society.

[Qawasmeh et al., 2018] Qawasmeh, O., Lefrançois, M., Zimmermann, A., and Maret, P. (2018). Computer-Assisted Ontology Construction System: Focus on Bootstrapping Capabilities. In Gangemi, A., Gentile, A. L., Nuzzolese, A. G., Rudolph, S., Maleshkova, M., Paulheim, H., Pan, J. Z., and Alam, M., editors, *The Semantic Web: ESWC 2018 Satellite Events - ESWC 2018 Satellite Events, Heraklion, Crete, Greece, June 3-7, 2018, Revised Selected Papers*, volume 11155 of *Lecture Notes in Computer Science*, pages 60–65. Springer.

[Qawasmeh et al., 2019] Qawasmeh, O., Lefrançois, M., Zimmermann, A., and Maret, P. (2019). Observing the Impact and Adaptation to the Evolution of an Imported Ontology. In Dietz, J., Aveiro, D., and Felipe, J., editors, *Proceedings of the 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management - (Volume 2), September 17-19, 2019, in Vienna, Austria*, pages 76–86. SciTePress.

[Sahay et al., 2011] Sahay, R., Fox, R., Zimmermann, A., Polleres, A., and Hauswirth, M. (2011). A Methodological Approach for Ontologising and Aligning Health Level Seven (HL7) Applications. In Tjoa, A. M., Quirchmayr, G., You, I., and Xu, L., editors, *International Conference on Multidisciplinary Research and Practice for Business, Enterprise and Health Information Systems - MURPBES 2011 / Availability, Reliability and Security for Business, Enterprise and Health Information Systems - IFIP WG 8.4/8.9 International Cross Domain Conference and Workshop, ARES 2011, Vienna, Austria, August 22-26, 2011. Proceedings*, volume 6908 of *Lecture Notes in Computer Science*, pages 102–117. Springer.

[Szmeja and Zimmermann, 2012] Szmeja, P. and Zimmermann, A. (2012). Reengineering the Core Grid Ontology. In Ossowski, S., Toni, F., and Vouros, G., editors, *Proceedings of the First International Conference on Agreement Technologies, AT 2012, Dubrovnik, Croatia, October 15-16, 2012*, volume 918 of *CEUR Workshop Proceedings*, pages 298–299. Sun SITE Central Europe (CEUR).

[Zarafin et al., 2012] Zarafin, A.-M., Zimmermann, A., and Boissier, O. (2012). Integrating Semantic Web Technologies and Multi-Agent Systems: A Semantic Description of Multi-Agent Organizations. In Ossowski, S., Toni, F., and Vouros, G., editors, *Proceedings of the First International Conference on Agreement Technologies, AT 2012, Dubrovnik, Croatia, October 15-16, 2012*, volume 918 of *CEUR Workshop Proceedings*, pages 296–297. Sun SITE Central Europe (CEUR).

[Zimmermann and Euzenat, 2006] Zimmermann, A. and Euzenat, J. (2006). Three Semantics for Distributed Systems and their Relations with Alignment Composition. In Cruz, I. F., Decker, S., Allemang, D., Preist, C., Schwabe, D., Mika, P., Uschold, M., and Aroyo, L., editors, *The Semantic Web - ISWC 2006, 5th International Semantic Web Conference, ISWC 2006, Athens, GA, USA, November 5-9, 2006, Proceedings*, volume 4273 of *Lecture Notes in Computer Science*, pages 16–29. Springer.

[Zimmermann et al., 2006] Zimmermann, A., Krötzsch, M., Euzenat, J., and Hitzler, P. (2006). Formalizing Ontology Alignment and its Operations with Category Theory. In Bennett, B. and Fellbaum, C. D., editors, *Formal Ontology in Information Systems, Proceedings of the Fourth International Conference, FOIS 2006, Baltimore, Maryland, USA, November 9-11, 2006*, volume 150 of *Frontiers in Artificial Intelligence and Applications*, pages 277–288. IOS Press.

[Zimmermann and Le Duc, 2008] Zimmermann, A. and Le Duc, C. (2008). Reasoning with a Network of Aligned Ontologies. In Calvanese, D. and Lausen, H., editors, *Web Reasoning and Rule Systems, Second International Conference, RR 2008, Karlsruhe, Germany, October/November 2008, Proceedings*, volume 5341 of *Lecture Notes in Computer Science*, pages 43–57. Springer.

[Zimmermann et al., 2009] Zimmermann, A., Sahay, R., Fox, R., and Polleres, A. (2009). Heterogeneity and Context in Semantic-Web-Enabled HCLS Systems. In Meersman, R., Dillon, T. S., and Herrero, P., editors, *On the Move to Meaningful Internet Systems: OTM 2009 - Confederated International Conferences, CoopIS, DOA, IS, and ODBASE 2009, Vilamoura, Portugal, November 1-6, 2009, Proceedings, Part II*, volume 5871 of *Lecture Notes in Computer Science*, pages 1165–1182. Springer.

E.6 International workshop articles

[Belloni et al., 2015] Belloni, A., Berger, A., Boissier, O., Bonnet, G., Bourgne, G., Chardel, P.-A., Cotton, J.-P., Evreux, N., Ganascia, J.-G., Jaillon, P., Mermet, B., Picard, G., Reber, B., Simon, G., De Swarte, T., Tessier, C., Vexler, F., Voyer, R., and Zimmermann, A. (2015). Dealing With Ethical Conflicts In Autonomous Agents And Multi-Agent Systems. In Walsh, T., editor, *Artificial Intelligence and Ethics, Papers from the 2015 AAAI Workshop, Austin, Texas, USA, January 25, 2015*, volume WS-15-02 of *AAAI Workshops*. AAAI Press.

[Bezerra et al., 2008] Bezerra, C., Freitas, F., Euzenat, J., and Zimmermann, A. (2008). ModOnto: A tool for modularizing ontologies. In Freitas, F., Stuckenschmidt, H., Pinto, H. S., Malucelli, A., and Corcho, Ó., editors, *WONTO 2008, 3rd Workshop on Ontologies and their Applications, Proceedings of the 3rd Workshop on Ontologies and their Applications Salvador, Bahia, Brazil, October 26, 2008*, volume 427 of *CEUR Workshop Proceedings*. Sun SITE Central Europe (CEUR).

[Bezerra et al., 2009] Bezerra, C., Freitas, F., Euzenat, J., and Zimmermann, A. (2009). An Approach for Ontology Modularization. In Navaux, P. O. A., editor, *Proceedings of Colibri - Colloquium of Computation: Brazil / INRIA, Cooperations, Advances and Challenges*.

[Ciortea et al., 2013] Ciortea, A., Boissier, O., Zimmermann, A., and Magda-Florea, A. (2013). Reconsidering the social web of things: position paper. In Mattern, F., Santini, S., Canny, J. F., Langheinrich, M., and Rekimoto, J., editors, *The 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp '13, Zurich, Switzerland, September 8-12, 2013 - Adjunct Publication*, pages 1535–1544. ACM Press. Presented at the workshop on the Web of Things.

[Ciortea et al., 2018] Ciortea, A., Boissier, O., Zimmermann, A., and Magda-Florea, A. (2018). Give Agents Some REST: Hypermedia-driven Agent Environments. In El Fallah-Seghrouchni, A., Ricci, A., and Son, T. C., editors, *Engineering Multi-Agent Systems - 5th International Workshop, EMAS 2017, Sao Paulo, Brazil, May 8-9, 2017, Revised Selected Papers*, volume 10738 of *Lecture Notes in Computer Science*, pages 125–141. Springer.

[Ciortea et al., 2016] Ciortea, A., Zimmermann, A., Boissier, O., and Magda-Florea, A. (2016). Hypermedia-driven Socio-technical Networks for Goal-driven Discovery in the Web of Things. In *Proceedings of the Seventh International Workshop on the Web of Things, WoT 2016, Stuttgart, Germany, November 7, 2016*, pages 25–30. ACM Press.

[Euzenat et al., 2007] Euzenat, J., Zimmermann, A., and Freitas, F. (2007). Alignment-based modules for encapsulating ontologies. In Cuenca-Grau, B., Honavar, V. G., Schlicht, A., and Wolter, F., editors, *Proceedings of the 2nd International Workshop on Modular Ontologies (WoMO-2007), Whistler, Canada, October 28, 2007*, volume 315 of *CEUR Workshop Proceedings*, pages 32–45. Sun SITE Central Europe (CEUR).

[Gillani et al., 2014] Gillani, S., Picard, G., Laforest, F., and Zimmermann, A. (2014). Towards Efficient Semantically Enriched Complex Event Processing and Pattern Matching (Position Paper). In Celino, I., Corcho, O., Dell'Aglio, D., Della Valle, E., Krötzsch, M., and Schlobach, S., editors, *Proceedings of the 3rd International Workshop on Ordering and Reasoning, Co-located with the 13th International Semantic Web Conference (ISWC 2014), Riva del Garda, Italy, October 20th, 2014*, volume 1303 of *CEUR Workshop Proceedings*, pages 47–54. Sun SITE Central Europe (CEUR).

[Giménez-García et al., 2019] Giménez-García, J. M., Duarte, M., Zimmermann, A., Gravier, C., Hruschka Jr., E. R., and Maret, P. (2019). NELL2RDF: Reading the Web, Tracking the Provenance, and Publishing it as Linked Data. In Capadisli, S., Cotton, F., Giménez-García, J. M., Haller, A., Kalampokis, E., Nguyen, V., Sheth, A., and Troncy, R., editors, *Joint Proceedings of the International Workshops on Contextualized Knowledge Graphs, and Semantic Statistics co-located with 17th International Semantic Web Conference (ISWC 2018) Monterey, USA, October 8th, 20180*, volume 2317 of *CEUR Workshop Proceedings*. Sun SITE Central Europe (CEUR).

[Giménez-García et al., 2016] Giménez-García, J. M., Thakkar, H., and Zimmermann, A. (2016). Assessing Trust with PageRank in the Web of Data. In Sack, H., Rizzo, G., Steinmetz, N., Mladenic, D., Auer, S., and Lange, C., editors, *The Semantic Web - ESWC 2016 Satellite Events, Heraklion, Crete, Greece, May 29*

- *June 2, 2016, Revised Selected Papers*, volume 9989 of *Lecture Notes in Computer Science*, pages 293–307. Springer. Extended version of a workshop article presented at PROFILES 2016.

[Giménez-García and Zimmermann, 2019] Giménez-García, J. M. and Zimmermann, A. (2019). NdProperties: Encoding Contexts in RDF predicates with Inference Preservation. In Capadisli, S., Cotton, F., Giménez-García, J. M., Haller, A., Kalampokis, E., Nguyen, V., Sheth, A., and Troncy, R., editors, *Joint Proceedings of the International Workshops on Contextualized Knowledge Graphs, and Semantic Statistics co-located with 17th International Semantic Web Conference (ISWC 2018) Monterey, USA, October 8th, 20180*, volume 2317 of *CEUR Workshop Proceedings*. Sun SITE Central Europe (CEUR).

[Hogan et al., 2010] Hogan, A., Polleres, A., Umbrich, J., and Zimmermann, A. (2010). Some entities are more equal than others: statistical methods to consolidate Linked Data. In Ceri, S., Della Valle, E., Hendler, J., and Huang, Z., editors, *Proceedings of the Workshop on New Forms of Reasoning for the Semantic Web: Scalable & Dynamic*.

[Kem et al., 2017] Kem, O., Balbo, F., and Zimmermann, A. (2017). Multi-goal pathfinding in ubiquitous environments: modeling and exploiting knowledge to satisfy goals. In Sheth, A. P., Ngonga, A., Wang, Y., Chang, E., Slezak, D., Franczyk, B., Alt, R., Tao, X., and Unland, R., editors, *Proceedings of the International Conference on Web Intelligence, Leipzig, Germany, August 23-26, 2017*, pages 1147–1150. ACM Press. Presented at 10th Workshop on Natural Language Processing and Ontology Engineering (NLPOE 2017).

[Khan et al., 2017] Khan, Y., Zimmermann, A., Jha, A., Rebholz-Schuhmann, D., and Sahay, R. (2017). Querying web polystores. In Nie, J., Obradovic, Z., Suzumura, T., Ghosh, R., , Wang, C., Zang, H., Baeza-Yates, R. A., Hu, X., Kepner, J., Cuzzocrea, A., Tang, J., and Toyoda, M., editors, *2017 IEEE International Conference on Big Data, BigData 2017, Boston, MA, USA, December 11-14, 2017*, pages 3190–3195. IEEE Computer Society. Presented at Workshop on Methods to Manage Heterogenous Data and Polystore Databases.

[Laborie and Zimmermann, 2007] Laborie, S. and Zimmermann, A. (2007). A Framework for Media Adaptation Using the Web and the Semantic Web. In Mylonas, P., Wallace, M., and Angelides, M., editors, *Second International Workshop on Semantic Media Adaption and Personalisation, SMAP 2007, London, United Kingdom, December 17-18, 2007*, pages 32–37. IEEE Computer Society.

[Le Duc et al., 2013] Le Duc, C., Lamolle, M., Zimmermann, A., and Curé, O. (2013). DRAOn: A Distributed Reasoner for Aligned Ontologies. volume 1015 of *CEUR Workshop Proceedings*, pages 81–86. Sun SITE Central Europe (CEUR).

[Lopes et al., 2010] Lopes, N., Zimmermann, A., Hogan, A., Lukácsy, G., Polleres, A., Straccia, U., and Stefan, D. (2010). RDF Needs Annotations. In *RDF Next Steps, W3C Workshop*. World Wide Web Consortium.

[Sorici et al., 2013] Sorici, A., Boissier, O., Picard, G., and Zimmermann, A. (2013). Applying semantic web technologies to context modeling in ambient intelligence.

In O'Grady, M. J., Vahdat-Nejad, H., Wolf, K.-H., Dragone, M., Ye, J., Röcker, C., and O'Hare, G. M., editors, *Evolving Ambient Intelligence - AmI 2013 Workshops, Dublin, Ireland, December 3-5, 2013. Revised Selected Papers*, volume 413 of *Communications in Computer and Information Science*, pages 217–229. Springer. Revised version of article presented at 3rd International Workshop on Pervasive and Context-Aware Middleware (PerCAM13).

[Zimmermann, 2007] Zimmermann, A. (2007). Integrated Distributed Description Logics. In Calvanese, D., Franconi, E., Haarslev, V., Lembo, D., Motik, B., Tessaris, S., and Turhan, A.-Y., editors, *Proceedings of the 20th International Workshop on Description Logics DL'07, June 8 - 10, 2007, Brixen/Bressanone, Italy*, pages 507–514. Bolzano University Press.

[Zimmermann, 2010] Zimmermann, A. (2010). Ontology Recommendation for the Data Publishers. In d'Aquin, M., García Castro, A., Lange, C., and Viljanen, K., editors, *Proceedings of the 1st Workshop on Ontology Repositories and Editors for the Semantic Web, Hersonissos, Crete, Greece, May 31st, 2010*, volume 596 of *CEUR Workshop Proceedings*. Sun SITE Central Europe (CEUR).

[Zimmermann, 2011] Zimmermann, A. (2011). Leveraging the Linked Data Principles for Electronic Communications. In Hübner, J. F., Petit, J., and Suzuki, E., editors, *Proceedings of the 2011 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology - Workshops, WI-IAT 2011, Campus Scientifique de la Doua, Lyon, France, August 22-27, 2011*, pages 385–388. IEEE Computer Society. Presented at First International Workshop on Next Trends in Email (NextMail'11).

[Zimmermann and Giménez-García, 2017] Zimmermann, A. and Giménez-García, J. M. (2017). Contextualizing DL Axioms: Formalization, a New Approach, and Its Properties. In Dell'Aglio, D., Anicic, D., Barnaghi, P. M., Della Valle, E., McGuinness, D. L., Bozzato, L., Eiter, T., Homola, M., and Porello, D., editors, *Joint Proceedings of the Web Stream Processing Workshop (WSP 2017) and the 2nd International Workshop on Ontology Modularity, Contextuality, and Evolution (WOMoCoE 2017) Co-Located with 16th International Semantic Web Conference (ISWC 2017), Vienna, Austria, October 22nd, 2017*, volume 1936 of *CEUR Workshop Proceedings*, pages 74–85. Sun SITE Central Europe (CEUR).

[Zimmermann et al., 2013] Zimmermann, A., Gravier, C., Subercaze, J., and Cruzille, Q. (2013). Nell2RDF: Read the Web, and Turn it into RDF. In Völker, J., Paulheim, H., Lehmann, J., Niepert, M., and Sack, H., editors, *Proceedings of the Second International Workshop on Knowledge Discovery and Data Mining Meets Linked Open Data, Montpellier, France, May 26, 2013*, volume 992 of *CEUR Workshop Proceedings*, pages 2–8. Sun SITE Central Europe (CEUR).

E.7 French conference and workshop articles

[Alkhateeb and Zimmermann, 2007] Alkhateeb, F. and Zimmermann, A. (2007). Répondre à des requêtes dans un système distribué à base de connaissances. In Demazeau, Y., Euzenat, J., Jacquenet, F., and Vercouter, L., editors, *Atelier Intelligence Artificielle et Web Intelligence*.

[Antoniazzi et al., 2020] Antoniazzi, F., Atemezing, G., Badeig, F., Bennara, M., Bernard, S., Champin, P., Chanet, J., Gravier, C., Gripay, Y., Laforest, F., Lefrançois, M., Médini, L., Moiroux, L., Roussey, C., Servigne, S., Singh, K., Subercaze, J., and Zimmermann, A. (2020). Interopérabilité et raisonnement dans le Web Sémantique des objets: le projet CoSWoT. In Ferré, S., editor, *IC 2020 : 31es Journées francophones d'Ingénierie des Connaissances (Proceedings of the 31st French Knowledge Engineering Conference), Angers, France, June 29 - July 3, 2020*, pages 155–158.

[Bakerally et al., 2019] Bakerally, M. N., Zimmermann, A., and Boissier, O. (2019). LDP-DL: un langage pour définir la conception des plateformes de données liées. In Hernandez, N., editor, *IC 2019: 30es Journées francophones d'Ingénierie des Connaissances (Proceedings of the 30th French Knowledge Engineering Conference), Toulouse, France, July 2-4, 2019*, page 142.

[Lefrançois and Zimmermann, 2015] Lefrançois, M. and Zimmermann, A. (2015). LinkedVocabularyEditor: une extension MediaWiki pour l'édition collaborative et la publication de vocabulaires liés. In Abel, M., editor, *IC 2015 : 26es Journées francophones d'Ingénierie des Connaissances (Proceedings of the 26th French Knowledge Engineering Conference), Rennes, France, July 1-3, 2015.*

[Lefrançois and Zimmermann, 2017] Lefrançois, M. and Zimmermann, A. (2017). Support uniforme de types de données personnalisées dans RDF et SPARQL. In Gandon, F. and Bisson, G., editors, *17ème Journées Francophones Extraction et Gestion des Connaissances, EGC 2017, 24-27 Janvier 2017, Grenoble, France*, volume E-33 of *Revue des nouvelles technologies de l'information*, pages 321–326. Édition RNTI.

[Lefrançois et al., 2017] Lefrançois, M., Zimmermann, A., and Bakerally, N. (2017). Génération de RDF à partir de sources de données aux formats hétérogènes. In *17ème Journées Francophones Extraction et Gestion des Connaissances, EGC 2017, 24-27 Janvier 2017, Grenoble, France*, volume E-33 of *Revue des nouvelles technologies de l'information*, pages 93–104. Édition RNTI.

E.8 Technical reports

[d'Aquin et al., 2008a] d'Aquin, M., Haase, P., Le Duc, C., and Zimmermann, A. (2008a). Reasoning over Distributed Networked Ontologies and Data Sources. Deliverable D1.1.4, IST NeOn IP.

[d'Aquin et al., 2008b] d'Aquin, M., Haase, P., Rudolph, S., Euzenat, J., Zimmermann, A., Dzbor, M., Iglesias, M., Jacques, Y., Caracciolo, C., Buil Aranda, C., and Gomez, J. M. (2008b). NeOn Formalisms for Modularization: Syntax, Semantics, Algebra. Deliverable D1.1.3, IST NeOn IP.

[Euzenat et al., 2007a] Euzenat, J., Scharffe, F., and Zimmermann, A. (2007a). Expressive alignment language and implementation. Deliverable D2.2.10, Knowledge Web NoE.

[Euzenat et al., 2007b] Euzenat, J., Zimmermann, A., Sabou, M., and d'Aquin, M. (2007b). Matching ontologies for context. Deliverable D3.3.1, IST NeOn IP.

[Giménez-García et al., 2018] Giménez-García, J. M., Duarte, M., Zimmermann, A., Gravier, C., Hruschka Jr., E. R., and Maret, P. (2018). NELL2RDF: Reading the Web, and Publishing it as Linked Data. Technical report, CoRR.

[Giménez-García et al., 2016] Giménez-García, J. M., Zimmermann, A., and Maret, P. (2016). NdFluents: A Multi-dimensional Contexts Ontology. Technical report, CoRR.

[Hitzler et al., 2005] Hitzler, P., Euzenat, J., Krötzsch, M., Serafini, L., Stuckenschmidt, H., Wache, H., and Zimmermann, A. (2005). Integrated view and comparison of alignment semantics. Deliverable D2.2.5, Knowledge Web NoE.

[Hogan et al., 2020] Hogan, A., Blomqvist, E., Cochez, M., d'Amato, C., de Melo, G., Gutierrez, C., Labra Gayo, J. E., Kirrane, S., Neumaier, S., Polleres, A., Navigli, R., Ngonga Ngomo, A., Rashid, S. M., Rula, A., Schmelzeisen, L., Sequeda, J. F., Staab, S., and Zimmermann, A. (2020). Knowledge Graphs. Technical report, CoRR.

[Lefrançois et al., 2017] Lefrançois, M., Kalaoja, J., Ghariani, T., and Zimmermann, A. (2017). SEAS Knowledge Model. Deliverable D2.2, ITEA2 12004 Smart Energy Aware Systems.

[Lopes et al., 2010] Lopes, N., Luckczy, G., Polleres, A., Straccia, U., and Zimmermann, A. (2010). A General Framework for Representing, Reasoning and Querying with Annotated Semantic Web Data. Technical report, Digital Enterprise Research Institute.

[Lopes et al., 2011] Lopes, N., Polleres, A., Straccia, U., and Zimmermann, A. (2011). A General Framework for Representing, Reasoning and Querying with Annotated Semantic Web Data. Technical report, CoRR.

[Trimponias et al., 2008] Trimponias, G., , Le Duc, C., Zimmermann, A., and Schenk, S. (2008). Reasoning over Distributed Networked Ontologies and Data Sources. Deliverable D1.1.4, IST NeOn IP.

[Zimmermann, 2014] Zimmermann, A. (2014). RDF 1.1: On Semantics of RDF Datasets, W3C Working Group Note 25 February 2014. W3C Working Group Note, World Wide Web Consortium.

[Zimmermann, 2017] Zimmermann, A. (2017). Modèle d'enrichissement et d'intégration des données. Deliverable, ANR 14-CE24-0029 OpenSensingCity.

[Zimmermann and Giménez-García, 2017] Zimmermann, A. and Giménez-García, J. M. (2017). Integrating Context of Statements within Description Logics. Technical report, CoRR.

[Zimmermann and Le Duc, 2008] Zimmermann, A. and Le Duc, C. (2008). Reasoning with a Network of Aligned Ontologies. Research report D1.1.3, Institut Nationale de Recherche en Informatique et Automatique.

E.9 Theses

[Zimmermann, 2004] Zimmermann, A. (2004). Fusion de descriptions hétérogènes en RDF. Master's thesis, Université Joseph Fourier, Grenoble (France).

[Zimmermann, 2008] Zimmermann, A. (2008). *Sémantiques des réseaux de connaissances : gestion de l'hétérogénéité fondée sur le principe de médiation*. PhD thesis, Université Joseph Fourier, Grenoble (France).

E.10 Other publications

[Capadisli et al., 2018] Capadisli, S., Cochez, M., Gutierrez, C., Harth, A., and Zimmermann, A. (2018). Knowledge Graphs and the Web. In Bonatti, P. A., Cochez, M., Decker, S., Polleres, A., and Presutti, V., editors, *Knowledge Graphs: New Directions for Knowledge Representation on the Semantic Web*, pages 106–110.

[d'Amato et al., 2018] d'Amato, C., Kirrane, S., Bonatti, P., Rudolph, S., Krötzsch, M., van Erp, M., and Zimmermann, A. (2018). Foundations. In Bonatti, P. A., Cochez, M., Decker, S., Polleres, A., and Presutti, V., editors, *Knowledge Graphs: New Directions for Knowledge Representation on the Semantic Web*, pages 79–89.

[Hogan et al., 2018] Hogan, A., Brickley, D., Gutierrez, C., Polleres, A., and Zimmermann, A. (2018). (Re)Defining Knowledge Graphs. In Bonatti, P. A., Cochez, M., Decker, S., Polleres, A., and Presutti, V., editors, *Knowledge Graphs: New Directions for Knowledge Representation on the Semantic Web*, pages 74–79.

Résumé étendu en français

De nos jours, les systèmes d'information modernes sont généralement accessibles à distance au moyen de matériels et de logiciels courants. Dans la plupart des cas, un simple navigateur web accédant à une page d'authentification suffit désormais pour accder à l'intérieur du système d'une organisation. Toutes les interactions sont effectuées facilement par des appels HTTP via le navigateur Web. De nombreuses personnes peuvent se connecter en m ême temps, en utilisant des logiciels et du matériel provenant de nombreux fournisseurs différents qui n'ont pas eu à se mettre d'accord au préalable sur les fonctionnalités sous-jacentes. L'interopérabilité entre les clients individuels et les serveurs de l'organisation est assurée par conception par les normes et l'architecture du Web.

Parfois, le système d'information lui-même doit se connecter à des ressources distantes fournies par des tiers afin d'automatiser certaines tâches (par exemple, se mettre à jour avec les dernières corrections de bogues). Cependant, alors que l'interaction humaine avec les ressources du Web est devenue extrêmement portable, où les gens peuvent utiliser presque tous les types de dispositifs informatisés – ordinateur portable, téléphone, tablette, voiture, réfrigérateur, etc. – comme interface entre eux et des ressources distantes de toute sorte – texte, images, sons, films, jeux, applications –, le niveau d'interopérabilité requis pour automatiser les interactions – c'est-à-dire laisser un dispositif autonome agir sur les ressources distantes sans intervention humaine – n'est pas encore atteint. Mon travail au cours des quinze dernières années a consisté à doter les systèmes d'une plus grande interopérabilité pour plus de capacités d'automatisation. Ce mémoire est un compte-rendu de mes contributions à cette fin, ainsi qu'une thèse défendant **la médiation faiblement couplée** comme approche pour atteindre cet objectif.

Le type de problèmes que nous voulons résoudre suppose que nous avons un besoin d'information qui nécessite la combinaison de plusieurs ressources d'information. Cela est grandement facilité par l'uniformité du protocole de réseau (Internet) et la tendance à utiliser le Web comme couche d'application de prédilection. Cela permet un accès interopérable aux ressources d'information, puisqu'il suffit d'une URL pour obtenir une ressource. En outre, les ressources peuvent fournir davantage d'URL pour naviguer de ressources en ressources via les liens hypermédia.

Mais l'interopérabilité ne s'arrête pas au bas de la couche applicative. Pour assurer une forte interopérabilité (sémantique), le projet initial pour le web sémantique était d'avoir une couche de données uniforme avec XML comme syntaxe de surface pour tous les échanges de données, RDF comme modèle de données pour la description des ressources et l'interconnexion, un langage d'ontologie standard (qui sera finalement nommé OWL) pour permettre la description des connaissances de base afin de comprendre plus facilement les graphes RDF et de faire des inférences, etc. Cette vision a été représentée sous la forme du "gâteau en couches du web sémantique" de la figure E.1, qui a été révisé à plusieurs reprises.

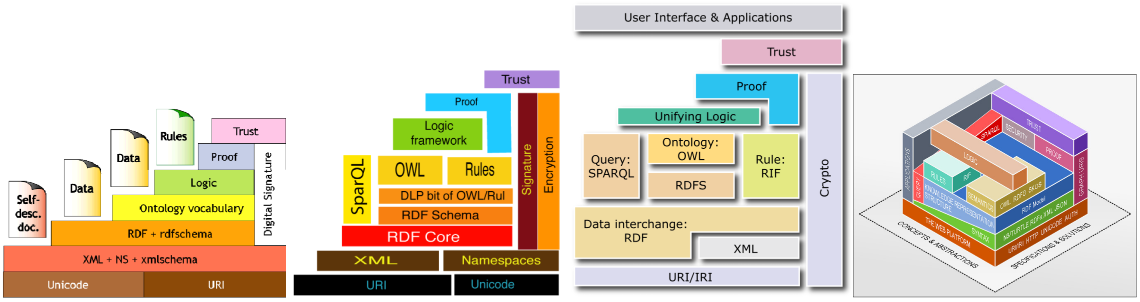


Figure E.1: Le gâteau en couches du web sémantique et son évolution. De gauche à droite : de [Berners-Lee et al., 2001] en 2001 ; version 2004 du W3C ; version 2007 de Wikipédia ; version 2009 publiée sur Flickr

Malheureusement, si des éléments de la pile du Web sémantique se sont peu à peu normalisés, ils ne sont pas tous devenus des normes *de facto*. Toutes les couches du Web ont une diversité de technologies qui coexistent. Il peut être surprenant que, alors que le Web a été rapidement adopté comme une couche d'application uniforme, le reste de la pile technologique qui aurait dû permettre l'interopérabilité sémantique ne s'est pas aussi bien répandu. À mon avis, une raison importante est que le Web a comblé un vide qui n'était pas bien rempli par les technologies existantes. Il a rendu l'accès aux ressources en ligne si facile, contrairement, par exemple, au FTP où les adresses IP doivent être connues, la structure des répertoires doit être explorée afin de trouver une ressource pertinente. Ensuite, rien dans l'écosystème FTP ne permet de naviguer facilement d'un serveur à l'autre. Au contraire, lorsque XML, RDF, etc. ont été conçus comme des couches pour le web sémantique, en plus des normes de base du web, les systèmes d'information disposaient déjà de moyens pour consommer et exporter des données. Le modèle de données dominant utilisé était principalement relationnel, qui n'est pas facile à transformer ou à migrer vers RDF. RDF nécessite une modélisation spéciale que les spécialistes des bases de données ne connaissent pas. L'ingénierie des ontologies est une compétence complexe et peu répandue.

En conséquence, nous avons actuellement une grande hétérogénéité dans les couches supérieures des technologies de base du Web qui entrave une forte interopérabilité. Cependant, je prétends qu'au lieu de lutter pour l'uniformité à tous les niveaux des opérations, nous pouvons **embrasser la diversité et permettre l'interopérabilité par le biais d'une approche de médiation peu couplée**. Le fait qu'elle soit faiblement couplée est important car c'est ce qui distingue cette approche de la façon dont le web lui-même sert de couche de médiation entre des systèmes d'information très hétérogènes et des clients web qui fonctionnent eux-mêmes sur des architectures logicielles et matérielles très hétérogènes. Les serveurs web servant de médiateur entre les infrastructures locales et les clients web distants sont mis en œuvre *à la source d'information*, mais une couche de médiation pour les données, les connaissances et les décisions pourrait être fournie par des tiers à des endroits très différents. Cette vision est illustrée dans la figure E.2, où la médiation classique est comparée à la médiation faiblement couplée.

Dans la médiation classique, les adaptateurs sont attachées aux sources, et un médiateur assure l'interface entre l'application utilisateur et les sources par le biais des adaptateurs. Dans le cadre d'une médiation à faible couplage, les médiateurs, les adaptateurs et les sources d'information sont des ressources sur le Web. Les médiateurs peuvent ou non être attachés à des adaptateurs. Les clients peuvent décider de connecter dynamiquement un médiateur (tel que M2) à différents adaptateurs, et de même décider de connecter un adaptateurs (tel que A2) à différentes sources d'information. Cette sélection par le client est visualisée sous forme d'ellipses pointillées dans la figure E.2.

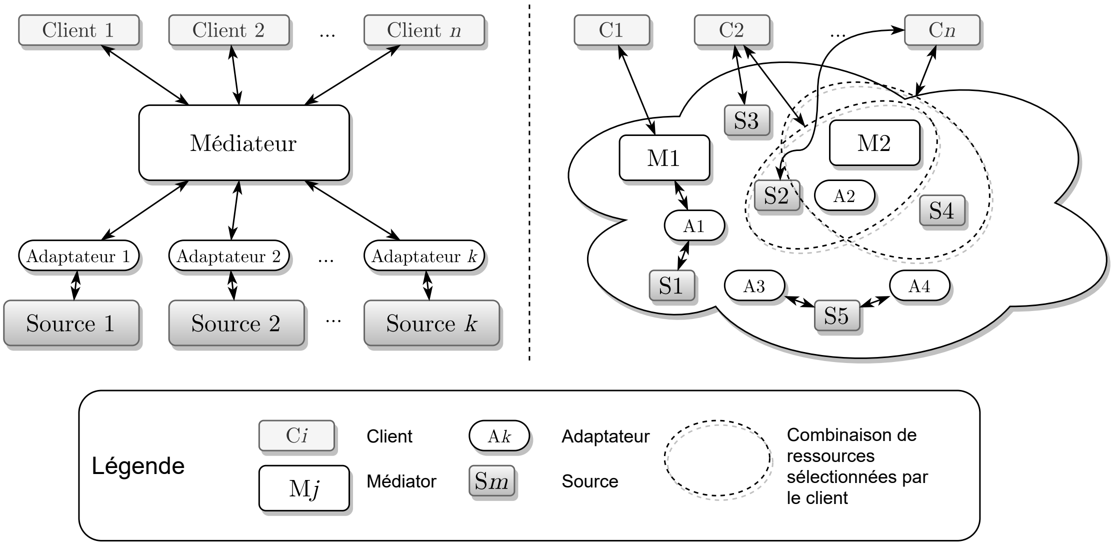


Figure E.2: À gauche, la médiation classique. À droite, la médiation faiblement couplée au dessus du Web.

La majorité de mon travail instancie cette vision de différentes manières. Dans de nombreux cas, le médiateur et les adaptateurs dans mes contributions ne sont pas décrits comme des composants logiciels : le médiateur est un modèle ou un langage que les adaptateurs instancient, tandis qu'un moteur générique se charge de la médiation opérationnelle. Afin de fournir une vision cohérente de mes contributions selon cette vision, je définis des couches d'interopérabilité comme dans la figure E.3 qui commencent en bas avec la couche réseau qui permet la communication entre les participants des systèmes distribués ou décentralisés. Je pars du principe que toutes les communications passent par le Web et je ne m'intéresse donc pas à cette couche. En outre, la couche de données vise à fournir un modèle de données uniforme et un accès aux données uniforme. Actuellement, aucun modèle ni méthode d'accès de ce type ne prévaut, malgré l'existence d'un modèle de données standard pour le Web, à savoir RDF. Mon travail à cette couche consiste en deux approches : fournir un médiateur flexible qui traduit tous les types de données en RDF, ou fournir un mécanisme d'accès uniforme à des données hétérogènes, via une interface RESTful uniforme ou un langage d'interrogation unique. Grâce à ces travaux, nous pouvons supposer que la couche supérieure peut être entièrement définie par-dessus le modèle de données RDF. Même avec un modèle de données commun, il est possible d'avoir des modèles de connaissances hétérogènes, en fonction des points de vue, des objectifs, des contextes. Dans mon travail, la médiation à cette couche est rendue possible par l'ajout de méta-informations sur les modèles de connaissance, soit sous la forme de correspondances explicites entre les terminologies, soit comme représentation du contexte de la connaissance. L'exploitation de ces méta-

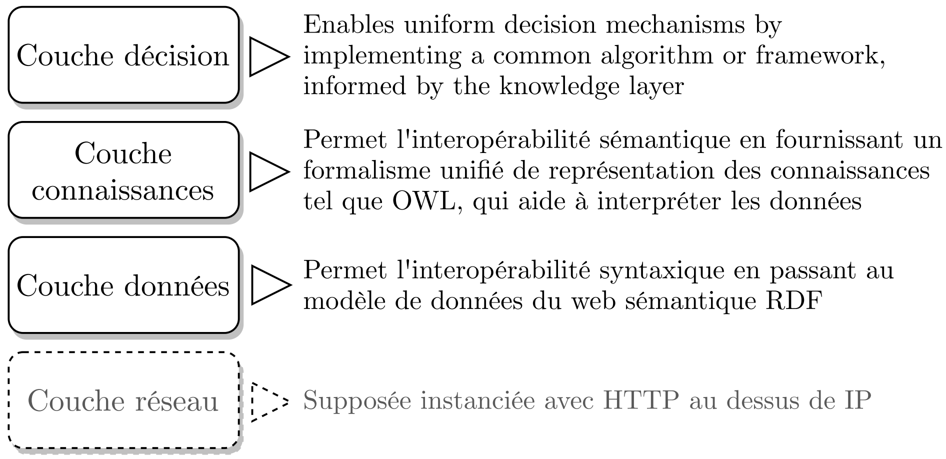


Figure E.3: Un modèle général en couches pour l'interopérabilité

informations permet de raisonner à travers les contextes et les points de vue. Enfin, à la couche supérieure, les processus qui utilisent un modèle de connaissance commun peuvent encore diverger. Plus précisément, deux agents qui reçoivent exactement les mêmes connaissances peuvent entreprendre des actions différentes, prendre des décisions différentes, et donc ne pas se coordonner. Mon travail à cette couche consiste à faciliter le processus de coopération par le biais d'une plateforme Web, rendue possible par des connaissances explicites.

Dans cette thèse, je montre comment mes contributions scientifiques s'inscrivent dans cette vision générale. Mes contributions sont structurées selon les trois couches supérieures de la figure E.3. Chaque couche est associée à une partie de cette thèse, ordonnée de la couche inférieure à la couche supérieure, divisée en chapitres qui correspondent chacun à une contribution faite dans le cadre d'un projet collaboratif ou d'un encadrement de doctorant. Je présente la structure de ce document en mettant en évidence les collaborations et la supervision, ainsi que le principal défi abordé.

- La première partie se concentre sur la couche de données de la figure E.3, avec trois approches pour concilier les formats de données et l'interopérabilité des données :
  - Dans le chapitre 1, je présente un langage de transformation des données – SPARQL-Generate – qui peut être utilisé pour donner une vue uniforme sur des données hétérogènes sous forme de graphes RDF. Le défi consiste à fournir un mécanisme flexible pour la transformation de données de n'importe quel format de données vers le modèle de données RDF du web sémantique. La transformation est le mécanisme de médiation entre les sources de données hétérogènes et l'utilisateur des données RDF. Ceci a été rendu possible par ma coordination du projet ANR OpenSensingCity et mon implication dans le projet ITEA SEAS. Les publications pertinentes pour ce travail sont : [Lefrançois et al., 2016b, Lefrançois et al., 2017c, Lefrançois et al., 2017b].
  - Dans le chapitre 2, je présente un moteur d'interrogation fédéré – PolyWeb – qui s'appuie sur des transformations de données et des métadonnées pour interroger de multiples sources de données en utilisant leur langage

d'interrogation de modèle de données natif. Le défi consiste à définir comment une requête originale peut être divisée et traduite vers différents moteurs de base de données qui utilisent des langages différents. Le moteur de fédération sert d'intermédiaire entre l'utilisateur avec son langage d'interrogation unique et les systèmes de bases de données hétérogènes et distribués. Ceci est le résultat d'une collaboration avec des collègues du centre de recherche Insight Galway. Les publications pertinentes pour ce travail sont : [Khan et al., 2017b, Khan et al., 2019].

  – Dans le chapitre 3, je présente un flux de travaux (*workflow*), un langage et un moteur pour les données RDF existantes d'une manière plus accessible et plus navigable, en suivant la norme Linked Data Platform 1.0. Le défi consiste à faciliter le déploiement des données sur des plateformes faciles à exploiter par des développeurs de données ouvertes. Les outils du workflow servent d'intermédiaire entre les données brutes et leur exposition sous forme de données liées structurées sur une plateforme Web. Ceci est le résultat de ma supervision du doctorant Noorany Bakerally, rendue possible par le financement du projet OpenSensingCity sous ma coordination. Les publications pertinentes pour ce travail sont : [Bakerally et al., 2016, Bakerally and Zimmermann, 2017, Bakerally et al., 2018a, Bakerally et al., 2018b] et la thèse de Noorani [Bakerally, 2018].

- La deuxième partie se concentre sur la couche de connaissance de la figure E.3, avec une attention particulière sur la représentation du contexte et le raisonnement :

  – Dans le chapitre 4, je présente un résumé des contributions de ma thèse de doctorat, suivi d'une prolongation plus récente de celle-ci qui consiste à formaliser des réseaux multi-niveaux d'ontologies alignées. Le défi est de pouvoir traiter les réseaux d'ontologies alignées comme une seule ontologie qui peut être alignée à son tour avec d'autres réseaux d'ontologies alignées, créant ainsi des niveaux d'alignements. De cette façon, un ensemble d'ontologies hétérogènes peut être utilisé comme un modèle de connaissance unifié. Les alignements aident à la médiation entre l'utilisateur des connaissances en réseau et les ontologies individuelles qui les composent. Ceci a été réalisé dans le cadre de la supervision de Sihem Klai de l'Université d'Annaba. Les publications pertinentes pour ce travail sont : [Klai et al., 2016a, Klai et al., 2016b] ainsi que la thèse de Sihem [Klai, 2016].

  – Dans le chapitre 5, je présente un formalisme – RDFS annoté – pour rendre le contexte d'un énoncé logique explicite comme une valeur d'une structure algébrique, avec sa sémantique, une méthode de raisonnement et le langage de requête associés. Le défi est d'avoir un traitement uniforme de nombreux types d'annotation contextuelle (temporelle, de provenance, floue, etc.). Cela a été rendu possible grâce à une collaboration lancée dans le cadre l'action européenne COST *Agreement Technologies*, dans laquelle j'étais l'un des responsable de groupe de travail. Les publications pertinentes pour ce travail sont : [Lopes et al., 2010a, Lopes et al., 2010b, Zimmermann et al., 2012, Lopes et al., 2012].

- Dans le chapitre 6, je présente un modèle – NdFluents – pour réifier les informations contextuelles dans des formalismes standards du web sémantique de telle sorte que le raisonnement puisse être effectué sur le modèle réifié de la même manière qu'il est fait dans un cadre non contextuel. Le défi est de fournir une représentation du contexte qui soit entièrement dans le formalisme standard de représentation de la connaissance, tout en permettant une certaine forme de raisonnement contextuel. Ceci contraste avec le chapitre précédent où un formalisme non standard est introduit. La connaissance multi-contextuelle peut servir de médiateur grâce à un mécanisme de transformation qui rassemble les sources en un modèle unifié. Ce travail a été rendu possible par ma participation au projet H2020 WDAqua et par ma supervision du doctorant José Giménez-García. Les publications pertinentes pour ce travail sont : [Giménez-García et al., 2016b, Giménez-García et al., 2017, Zimmermann and Giménez-García, 2017a, Zimmermann and Giménez-García, 2017b].

- La troisième partie fait l'hypothèse que nous avons une couche de données et de connaissances plus uniforme, et en tire profit pour deux cas spécifiques :

  - Dans le chapitre 7, l'interaction avec les agents sociaux et les objets sur le Web, en tirant parti des plateformes de réseaux sociaux qui sont décrites sémantiquement pour permettre une prise de décision autonome. Le défi consiste à rendre les interactions entre les choses et les agents plus faciles et plus systématiques dans le contexte de l'internet des objets. Les plateformes sociales sont décrites sémantiquement, fournissant un point de médiation à travers lequel les agents peuvent communiquer et se coordonner. Ce travail est le résultat de ma supervision du doctorant Andrei Ciortea, qui était dans une *cotutelle* entre les Mines Saint-Étienne et l'Université polytechnique de Bucarest. Les publications pertinentes pour ce travail sont : [Ciortea et al., 2013, Ciortea et al., 2014, Ciortea et al., 2015, Ciortea et al., 2016b, Ciortea et al., 2016a, Ciortea et al., 2018, Ciortea et al., 2017, Ciortea et al., 2019] et la thèse d'Andrei [Ciortea, 2016].
  - Dans le chapitre 8, la recherche d'itinéraires à buts multiples dans des environnements ubiquitaires, où les ontologies et les graphes de connaissances sont utilisés pour déterminer des itinéraires qui satisfont aussi des objectifs au passage. Le défi consiste à disposer d'un algorithme de recherche de chemin qui puisse exploiter la description formelle d'un environnement dynamique, avec des problèmes éventuels de latence et de recalcul au moment du voyage. Les communications entre les acteurs de ce type de système peuvent tirer parti du réseau social des objets décrit dans le chapitre précédent. Ce travail est le résultat de ma supervision du doctorant Oudom Kem. Les publications pertinentes pour ce travail sont : [Kem et al., 2016, Kem et al., 2017d, Kem et al., 2017a, Kem et al., 2017b, Kem et al., 2017c] et la thèse d'Oudom [Kem, 2018].

- La quatrième partie résume les contributions et présente les orientations de recherche que je souhaite poursuivre à l'avenir :

  - Dans le chapitre 9, je fournis une synthèse qui met en évidence les relations et les dépendances entre les différentes parties de cette thèse, en me concentrant sur les deux aspects principaux, à savoir *l'interopérabilité* et *la médiation faiblement couplée*.
  - Dans le chapitre 10, je donne plusieurs pistes de recherche que je souhaite approfondir afin d'aller au-delà de mes contributions actuelles. Si je me concentre sur quelques composants de base, le plan global à long terme consiste à développer des systèmes socio-techniques interopérables sur le Web.

- Enfin, les annexes présentent un résumé de ma carrière passée, en termes d'enseignement (appendice A), de recherche (appendice B), de supervision (appendice C), de projets (appendice D) et de publications (appendice E). Cette synthèse finale, essentiellement quantitative, de mes activités professionnelles devrait témoigner de mes capacités à mener des recherches indépendantes, à les communiquer de manière appropriée, à collaborer à leur développement, à former d'autres personnes pour les faire progresser et à participer aux communautés en partageant les réflexions et les charges.