RESEARCH CENTRE
Sophia Antipolis - Méditerranée

IN PARTNERSHIP WITH:
CNRS, Université Côte d'Azur

2021
ACTIVITY REPORT

Project-Team
WIMMICS

Web-Instrumented Man-Machine Interactions, Communities and Semantics

IN COLLABORATION WITH: Laboratoire informatique, signaux systèmes de Sophia Antipolis (I3S)

DOMAIN
Perception, Cognition and Interaction

THEME
Data and Knowledge Representation and Processing

Contents

Project-Team WIMMICS

Creation of the Project-Team: 2013 July 01

Keywords

Computer sciences and digital sciences

A1.2.9. – Social Networks

A1.3.1. – Web

A1.3.4. – Peer to peer

A2.1. – Programming Languages

A2.1.1. – Semantics of programming languages

A3.1.1. – Modeling, representation

A3.1.2. – Data management, quering and storage

A3.1.3. – Distributed data

A3.1.4. – Uncertain data

A3.1.5. – Control access, privacy

A3.1.6. – Query optimization

A3.1.7. – Open data

A3.1.9. – Database

A3.1.10. – Heterogeneous data

A3.1.11. – Structured data

A3.2. – Knowledge

A3.2.1. – Knowledge bases

A3.2.2. – Knowledge extraction, cleaning

A3.2.3. – Inference

A3.2.4. – Semantic Web

A3.2.5. – Ontologies

A3.2.6. – Linked data

A3.3.1. – On-line analytical processing

A3.3.2. – Data mining

A3.4. – Machine learning and statistics

A3.4.1. – Supervised learning

A3.4.6. – Neural networks

A3.4.8. – Deep learning

A3.5. – Social networks

A3.5.1. – Analysis of large graphs

A3.5.2. – Recommendation systems

A5.1. – Human-Computer Interaction

A5.1.1. – Engineering of interactive systems

A5.1.2. – Evaluation of interactive systems

A5.1.9. – User and perceptual studies

A5.2. – Data visualization

A5.7.2. – Music

A5.8. – Natural language processing

A7.1.3. – Graph algorithms

A7.2.2. – Automated Theorem Proving

A8.2.2. – Evolutionary algorithms

A9.1. – Knowledge

A9.2. – Machine learning

A9.4. – Natural language processing

A9.6. – Decision support

A9.7. – AI algorithmics

A9.8. – Reasoning

A9.9. – Distributed AI, Multi-agent

A9.10. – Hybrid approaches for AI

Other research topics and application domains

B1.2.2. – Cognitive science

B2. – Health

B5.1. – Factory of the future

B5.6. – Robotic systems

B5.8. – Learning and training

B6.3.1. – Web

B6.3.2. – Network protocols

B6.3.4. – Social Networks

B6.4. – Internet of things

B6.5. – Information systems

B8.5. – Smart society

B8.5.1. – Participative democracy

B9. – Society and Knowledge

B9.1. – Education

B9.1.1. – E-learning, MOOC

B9.1.2. – Serious games

B9.2. – Art

B9.2.1. – Music, sound

B9.3. – Medias

B9.5.1. – Computer science

B9.5.6. – Data science

B9.6. – Humanities

B9.6.1. – Psychology

B9.6.2. – Juridical science

B9.6.5. – Sociology

B9.6.7. – Geography

B9.6.8. – Linguistics

B9.6.9. – Political sciences

B9.6.10. – Digital humanities

B9.7. – Knowledge dissemination

B9.7.1. – Open access

B9.7.2. – Open data

B9.9. – Ethics

B9.10. – Privacy

1 Team members, visitors, external collaborators

Research Scientists

- Fabien Gandon [Team leader, Inria, Senior Researcher, from 2004, http://fabien.info, HDR]

- Olivier Corby [Inria, Researcher]

- Damien Graux [Inria, Starting Faculty Position]

- Franck Michel [CNRS, Researcher]

- Serena Villata [CNRS, Researcher, HDR]

Faculty Members

- Michel Buffa [Univ Côte d'Azur, Professor]

- Elena Cabrio [Univ Côte d'Azur, Associate Professor, HDR]

- Catherine Faron [Univ Côte d'Azur, Associate Professor, HDR]

- Nhan Le Thanh [Univ Côte d'Azur, Professor]

- Amaya Nogales Gómez [Univ Côte d'Azur, Associate Professor, from Oct 2021]

- Peter Sander [Univ Côte d'Azur, Professor]

- Andrea Tettamanzi [Univ Côte d'Azur, Professor]

- Marco Winckler [Univ Côte d'Azur, Professor, HDR]

Post-Doctoral Fellows

- Molka Dhouib [Inria, from May 2021]

- Raphaël Gazzotti [Inria, until Oct 2021]

- Pierre Maillot [Inria]

- Aline Menin [Univ. Côte d'Azur, until November, 30th]

- Anaïs Ollagnier [Univ Côte d'Azur, from Jan 2021]

- Iliana Petrova [Inria]

- Stefan Sarkadi [Inria]

- Nadia Yacoubi Ayadi [CNRS, from Sep 2021]

PhD Students

- Ali Ballout [Univ Côte d'Azur]

- Lucie Cadorel [Kinaxia SA, CIFRE]

- Dupuy Rony Charles [Doriane SA, CIFRE]

- Molka Dhouib [Silex, CIFRE, until Apr 2021]

- Ahmed Elamine Djebri [Algeria (Ministère de la défense)]

- Antonia Ettorre [Univ Côte d'Azur]

- Remi Felin [Univ Côte d'Azur, from Sep 2021]

- Pierpaolo Goffredo [CNRS, from Sep 2021]

- Nicholas Halliwell [Inria]

- Mina Ayse Ilhan [Univ Côte d'Azur]

- Adnane Mansour [Ecole Nationale Supérieure des Mines de Saint Etienne]

- Santiago Marro [Univ Côte d'Azur]

- Benjamin Molinet [Univ Côte d'Azur, from Sep 2021]

- Thu Huong Nguyen [Aix-Marseille Université, until Aug 2021]

- Shihong Ren [Aix-Marseille Université]

- Maroua Tikat [Univ Côte d'Azur]

- Mahamadou Toure [Université Gaston Berger, Sénégal]

- Vorakit Vorakitphan [Inria]

Technical Staff

- Anna Bobasheva [Inria, Engineer]

- Remi Ceres [Inria, Engineer, from Mar 2021]

- Celian Ringwald [Inria, Engineer, from Oct 2021]

Interns and Apprentices

- Saad El Din Ahmed [CNRS, from May 2021 until Aug 2021]

- Manon Audren [Univ Côte d'Azur, from Feb 2021 until Jun 2021]

- Yessine Ben El Bey [Inria, from May 2021 until Aug 2021]

- Minh Nhat Do [Univ Côte d'Azur, from Apr 2021 until Sep 2021]

- Remi Felin [Univ Côte d'Azur, from Mar 2021 until Aug 2021]

- Julien Kaplan [Inria, from May 2021 until Aug 2021]

- Mathis Le Quiniou [Inria, Jan 2021]

- Hadi Mahmoudi [Univ Côte d'Azur, from May 2021 until Jun 2021]

- Youssef Mekouar [Inria, from Mar 2021 until Aug 2021]

- Frederic Metereau [Inria, from Jun 2021 until Aug 2021]

- Benjamin Molinet [Inria, Apprentice, until Sep 2021]

- Aymeric Rebuffel [Inria, from Jun 2021 until Aug 2021]

- Pierre Saunders [Inria, from Apr 2021 until Sep 2021]

- Ekaterina Sviridova [CNRS, Jun 2021]

- Qingwen Ye [Univ Côte d'Azur, from Jun 2021 until Jul 2021]

Administrative Assistant

- Christine Foggia [Inria]

Visiting Scientist

- Dario Malchiodi [Université de Milan-Bicocca - Italie, from Oct 2021]

External Collaborators

- Andrei Ciortea [University of St.Gallen]

- Alain Giboin [Retired]

- Freddy Lecue [CortAIx, Thalès]

- Oscar Rodríguez Rocha [TeachOnMars]

2   Overall objectives

2.1   Context and Objectives

The Web became a virtual place where persons and software interact in mixed communities. The Web has the potential of becoming the collaborative space for natural and artificial intelligence, raising the problem of supporting these worldwide interactions. These large scale mixed interactions create many problems that must be addressed with multidisciplinary approaches [85].

One particular problem is to reconcile formal semantics of computer science (e.g. logics, ontologies, typing systems, protocols, etc.) on which the Web architecture is built, with soft semantics of people (e.g. posts, tags, status, relationships, etc.) on which the Web content is built.

Wimmics proposes models and methods to bridge formal semantics and social semantics on the Web [86] in order to address some of the challenges in building a Web as a universal space linking many different kinds of intelligence.

From a formal modeling point of view, one of the consequences of the evolutions of the Web is that the initial graph of linked pages has been joined by a growing number of other graphs. This initial graph is now mixed with sociograms capturing the social network structure, workflows specifying the decision paths to be followed, browsing logs capturing the trails of our navigation, service compositions specifying distributed processing, open data linking distant datasets, etc. Moreover, these graphs are not available in a single central repository but distributed over many different sources. Some sub-graphs are small and local (e.g. a user's profile on a device), some are huge and hosted on clusters (e.g. Wikipedia), some are largely stable (e.g. thesaurus of Latin), some change several times per second (e.g. social network statuses), etc. Moreover, each type of network of the Web is not an isolated island. Networks interact with each other: the networks of communities influence the message flows, their subjects and types, the semantic links between terms interact with the links between sites and vice-versa, etc.

Not only do we need means to represent and analyze each kind of graphs, we also do need means to combine them and to perform multi-criteria analysis on their combination. Wimmics contributes to these challenges by: (1) proposing multidisciplinary approaches to analyze and model the many aspects of these intertwined information systems, their communities of users and their interactions; (2) formalizing and reasoning on these models using graphs-based knowledge representation from the semantic Web to propose new analysis tools and indicators, and to support new functionalities and better management. In a nutshell, the first research direction looks at models of systems, users, communities and interactions while the second research direction considers formalisms and algorithms to represent them and reason on their representations.

2.2   Research Topics

The research objectives of Wimmics can be grouped according to four topics that we identify in reconciling social and formal semantics on the Web:

Topic 1 - users modeling and designing interaction on the Web and with knowledge graphs: The general research question addressed by this objective is "*How do we improve our interactions with a semantic and social Web more and more complex and dense ?*". Wimmics focuses on specific sub-questions: "How can we capture and model the users' characteristics?" "How can we represent and reason with the users' profiles?" "How can we adapt the system behaviors as a result?" "How can we design new interaction means?" "How can we evaluate the quality of the interaction designed?". This topic includes a long-term research direction in Wimmics on information visualization of semantic graphs on the Web. The general research question addressed in this last objective is "*How to represent the inner and complex relationships between data obtained from large and multivariate knowledge graph?*". Wimmics focuses on several sub-questions: "Which visualization techniques are suitable (from a user point of view) to support the exploration and the analysis of large graphs?" How to identify the new knowledge created by users during the exploration of knowledge graph ?" "How to formally describe the dynamic transformations allowing to convert raw data extracted from the Web into meaningul visual representations?" "How to guide the analysis of graphs that might contain data with diverse levels of accuracy, precision and interestingness to the users?"

Topic 2 - communities and social interactions and content analysis on the Web: The general question addressed in this second objective is "*How can we manage the collective activity on social media?*". Wimmics focuses on the following sub-questions: "How do we analyze the social interaction practices and the structures in which these practices take place?" "How do we capture the social interactions and structures?" "How can we formalize the models of these social constructs?" "How can we analyze and reason on these models of the social activity ?"

Topic 3 - vocabularies, semantic Web and linked data based knowledge extraction and representation with knowledge graphs on the Web: The general question addressed in this third objective is "*What are the needed schemas and extensions of the semantic Web formalisms for our models?*". Wimmics focuses on several sub-questions: "What kinds of formalism are the best suited for the models of the previous section?" "What are the limitations and possible extensions of existing formalisms?" "What are the missing schemas, ontologies, vocabularies?" "What are the links and possible combinations between existing formalisms?" We also address the question of knowledge extraction and especially AI and NLP methods to extract knowledge from text.In a nutshell, an important part of this objective is to formalize as typed graphs the models identified in the previous objectives and to populate thems in order for software to exploit these knowledge graphs in their processing (in the next objective).

Topic 4 - artificial intelligence processing: learning, analyzing and reasoning on heterogeneous semantic graphs on the Web: The general research question addressed in this objective is "*What are the algorithms required to analyze and reason on the heterogeneous graphs we obtained?*". Wimmics focuses on several sub-questions: "How do we analyze graphs of different types and their interactions?" "How do we support different graph life-cycles, calculations and characteristics in a coherent and understandable way?" "What kind of algorithms can support the different tasks of our users?".

3   Research program

3.1   Users Modeling and Designing Interaction on the Web and with AI systems

Wimmics focuses on interactions of ordinary users with ontology-based knowledge systems, with a preference for semantic Web formalisms and Web 2.0 applications. We specialize interaction design and evaluation methods to Web application tasks such as searching, browsing, contributing or protecting data. The team is especially interested in using semantics in assisting the interactions. We propose knowledge graph representations and algorithms to support interaction adaptation, for instance for context-awareness or intelligent interactions with machine. We propose and evaluate Web-based visualization techniques for linked data, querying, reasoning, explaining and justifying. Wimmics also integrates natural language processing approaches to support natural language based interactions. We rely on cognitive studies to build models of the system, the user and the interactions between users

through the system, in order to support and improve these interactions. We extend the user modeling technique known as *Personas* where user models are represented as specific, individual humans. *Personas* are derived from significant behavior patterns (i.e., sets of behavioral variables) elicited from interviews with and observations of users (and sometimes customers) of the future product. Our user models specialize *Personas* approaches to include aspects appropriate to Web applications. Wimmics also extends user models to capture very different aspects (e.g. emotional states).

3.2   Communities and Social Media Interactions and Content Analysis on the Web and Linked Data

The domain of social network analysis is a whole research domain in itself and Wimmics targets what can be done with typed graphs, knowledge representations and social models. We also focus on the specificity of social Web and semantic Web applications and in bridging and combining the different social Web data structures and semantic Web formalisms. Beyond the individual user models, we rely on social studies to build models of the communities, their vocabularies, activities and protocols in order to identify where and when formal semantics is useful. We propose models of collectives of users and of their collaborative functioning extending the collaboration personas and methods to assess the quality of coordination interactions and the quality of coordination artifacts. We extend and compare community detection algorithms to identify and label communities of interest with the topics they share. We propose mixed representations containing social semantic representations (e.g. folksonomies) and formal semantic representations (e.g. ontologies) and propose operations that allow us to couple them and exchange knowledge between them. Moving to social interaction we develop models and algorithms to mine and integrate different yet linked aspects of social media contributions (opinions, arguments and emotions) relying in particular on natural language processing and argumentation theory. To complement the study of communities we rely on multi-agent systems to simulate and study social behaviors. Finally we also rely on Web 2.0 principles to provide and evaluate social Web applications.

3.3   Vocabularies, Semantic Web and Linked Data Based Knowledge Representation and Extraction of Knowledge Graphs on the Web

For all the models we identified in the previous sections, we rely on and evaluate knowledge representation methodologies and theories, in particular ontology-based modeling. We also propose models and formalisms to capture and merge representations of different levels of semantics (e.g. formal ontologies and social folksonomies). The important point is to allow us to capture those structures precisely and flexibly and yet create as many links as possible between these different objects. We propose vocabularies and semantic Web formalizations for all the aspects that we model and we consider and study extensions of these formalisms when needed. The results have all in common to pursue the representation and publication of our models as linked data. We also contribute to the extraction, transformation and linking of existing resources (informal models, databases, texts, etc.) to publish knowledge graphs on the Semantic Web and as Linked Data. Examples of aspects we formalize include: user profiles, social relations, linguistic knowledge, bio-medical data, business processes, derivation rules, temporal descriptions, explanations, presentation conditions, access rights, uncertainty, emotional states, licenses, learning resources, etc. At a more conceptual level we also work on modeling the Web architecture with philosophical tools so as to give a realistic account of identity and reference and to better understand the whole context of our research and its conceptual cornerstones.

3.4   Artificial Intelligence Processing: Learning, Analyzing and Reasoning on Heterogeneous Knowledge Graphs

One of the characteristics of Wimmics is to rely on graph formalisms unified in an abstract graph model and operators unified in an abstract graph machine to formalize and process semantic Web data, Web resources, services metadata and social Web data. In particular Corese, the core software of Wimmics, maintains and implements that abstraction. We propose algorithms to process the mixed representations of the previous section. In particular we are interested in allowing cross-enrichment between them and in exploiting the life cycle and specificity of each one to foster the life-cycles of the others. Our results all

have in common to pursue analyzing and reasoning on heterogeneous knowledge graphs issued from social and semantic Web applications. Many approaches emphasize the logical aspect of the problem especially because logics are close to computer languages. We defend that the graph nature of Linked Data on the Web and the large variety of types of links that compose them call for typed graphs models. We believe the relational dimension is of paramount importance in these representations and we propose to consider all these representations as fragments of a typed graph formalism directly built above the Semantic Web formalisms. Our choice of a graph based programming approach for the semantic and social Web and of a focus on one graph based formalism is also an efficient way to support interoperability, genericity, uniformity and reuse.

4 Application domains

4.1 Social Semantic Web

A number of evolutions have changed the face of information systems in the past decade but the advent of the Web is unquestionably a major one and it is here to stay. From an initial wide-spread perception of a public documentary system, the Web as an object turned into a social virtual space and, as a technology, grew as an application design paradigm (services, data formats, query languages, scripting, interfaces, reasoning, etc.). The universal deployment and support of its standards led the Web to take over nearly all of our information systems. As the Web continues to evolve, our information systems are evolving with it.

Today in organizations, not only almost every internal information system is a Web application, but these applications more and more often interact with external Web applications. The complexity and coupling of these Web-based information systems call for specification methods and engineering tools. From capturing the needs of users to deploying a usable solution, there are many steps involving computer science specialists and non-specialists.

We defend the idea of relying on Semantic Web formalisms to capture and reason on the models of these information systems supporting the design, evolution, interoperability and reuse of the models and their data as well as the workflows and the processing.

4.2 Linked Data on the Web and on Intranets

With billions of triples online (see Linked Open Data initiative), the Semantic Web is providing and linking open data at a growing pace and publishing and interlinking the semantics of their schemas. Information systems can now tap into and contribute to this Web of data, pulling and integrating data on demand. Many organisations also started to use this approach on their intranets leading to what is called linked enterprise data.

A first application domain for us is the publication and linking of data and their schemas through Web architectures. Our results provide software platforms to publish and query data and their schemas, to enrich these data in particular by reasoning on their schemas, to control their access and licenses, to assist the workflows that exploit them, to support the use of distributed datasets, to assist the browsing and visualization of data, etc.

Examples of collaboration and applied projects include: SMILK Joint Laboratory, Corese, DBpedia.fr.

4.3 Assisting Web-based Epistemic Communities

In parallel with linked open data on the Web, social Web applications also spread virally (e.g. Facebook growing toward 1.5 billion users) first giving the Web back its status of a social read-write media and then putting it back on track to its full potential of a virtual place where to act, react and interact. In addition, many organizations are now considering deploying social Web applications internally to foster community building, expert cartography, business intelligence, technological watch and knowledge sharing in general.

By reasoning on the Linked Data and the semantics of the schemas used to represent social structures and Web resources, we provide applications supporting communities of practice and interest and fostering their interactions in many different contexts (e-learning, business intelligence, technical watch, etc.).

We use typed graphs to capture and mix: social networks with the kinds of relationships and the descriptions of the persons; compositions of Web services with types of inputs and outputs; links between documents with their genre and topics; hierarchies of classes, thesauri, ontologies and folksonomies; recorded traces and suggested navigation courses; submitted queries and detected frequent patterns; timelines and workflows; etc.

Our results assist epistemic communities in their daily activities such as biologists exchanging results, business intelligence and technological watch networks informing companies, engineers interacting on a project, conference attendees, students following the same course, tourists visiting a region, mobile experts on the field, etc. Examples of collaboration and applied projects: EduMICS, OCKTOPUS, Vigiglobe, Educlever, Gayatech.

4.4   Linked Data for a Web of Diversity

We intend to build on our results on explanations (provenance, traceability, justifications) and to continue our work on opinions and arguments mining toward the global analysis of controversies and online debates. One result would be to provide new search results encompassing the diversity of viewpoints and providing indicators supporting opinion and decision making and ultimately a Web of trust. Trust indicators may require collaborations with teams specialized in data certification, cryptography, signature, security services and protocols, etc. This will raise the specific problem of interaction design for security and privacy. In addition, from the point of view of the content, this requires to foster the publication and coexistence of heterogeneous data with different points of views and conceptualizations of the world. We intend to pursue the extension of formalisms to allow different representations of the world to co-exist and be linked and we will pay special attention to the cultural domain and the digital humanities. Examples of collaboration and applied projects: Zoomathia, Seempad, SMILK,

4.5   Artificial Web Intelligence

We intend to build on our experience in artificial intelligence (knowledge representation, reasoning) and distributed artificial intelligence (multi-agent systems - MAS) to enrich formalisms and propose alternative types of reasoning (graph-based operations, reasoning with uncertainty, inductive reasoning, non-monotonic, etc.) and alternative architectures for linked data with adequate changes and extensions required by the open nature of the Web. There is a clear renewed interest in AI for the Web in general and for Web intelligence in particular. Moreover, distributed AI and MAS provide both new architectures and new simulation platforms for the Web. At the macro level, the evolution accelerated with HTML5 toward Web pages as full applications and direct Page2Page communication between browser clearly is a new area for MAS and P2P architectures. Interesting scenarios include the support of a strong decentralization of the Web and its resilience to degraded technical conditions (downscaling the Web), allowing pages to connect in a decentralized way, forming a neutral space, and possibly going offline and online again in erratic ways. At the micro level, one can imagine the place RDF and SPARQL could take as data model and programming model in the virtual machines of these new Web pages and, of course, in the Web servers. RDF is also used to serialize and encapsulate other languages and becomes a pivot language in linking very different applications and aspects of applications. Example of collaboration and applied projects: MoreWAIS, Corese, Vigiglobe collaboration.

4.6   Human-Data Interaction (HDI) on the Web

We need more interaction design tools and methods for linked data access and contribution. We intend to extend our work on exploratory search coupling it with visual analytics to assist sense making. It could be a continuation of the Gephi extension that we built targeting more support for non experts to access and analyze data on a topic or an issue of their choice. More generally speaking SPARQL is inappropriate for common users and we need to support a larger variety of interaction means with linked data. We also believe linked data and natural language processing (NLP) have to be strongly integrated to support natural language based interactions. Linked Open Data (LOD) for NLP, NLP for LOD and Natural Dialog Processing for querying, extracting and asserting data on the Web is a priority to democratize its use. Micro accesses and micro contributions are important to ensure public participation and also call for

customized interfaces and thus for methods and tools to generate these interfaces. In addition, the user profiles are being enriched now with new data about the user such as her current mental and physical state, the emotion she just expressed or her cognitive performances. Taking into account this information to improve the interactions, change the behavior of the system and adapt the interface is a promising direction. And these human-data interaction means should also be available for "small data", helping the user to manage her personal information and to link it to public or collective one, maintaining her personal and private perspective as a personal Web of data. Finally, the continuous knowledge extractions, updates and flows add the additional problem of representing, storing, querying and interacting with dynamic data. Examples of collaboration and applied projects: QAKIS, Sychonext collaboration, ALOOF, DiscoveryHub, WASABI, MoreWAIS.

4.7   Web-augmented interactions with the world

The Web continues to augment our perception and interaction with reality. In particular, Linked Open Data enable new augmented reality applications by providing data sources on almost any topic. The current enthusiasm for the Web of Things, where every object has a corresponding Web resource, requires evolutions of our vision and use of the Web architecture. This vision requires new techniques as the ones mentioned above to support local search and contextual access to local resources but also new methods and tools to design Web-based human devices interactions, accessibility, etc. These new usages are placing new requirements on the Web Architecture in general and on the semantic Web models and algorithms in particular to handle new types of linked data. They should support implicit requests considering the user context as a permanent query. They should also simplify our interactions with devices around us jointly using our personal preferences and public common knowledge to focus the interaction on the vital minimum that cannot be derived in another way. For instance, the access to the Web of data for a robot can completely change the quality of the interactions it can offer. Again, these interactions and the data they require raise problems of security and privacy. Examples of collaboration and applied projects: ALOOF, AZKAR, MoreWAIS.

4.8   Analysis of scientific co-authorship

Over the last decades, scientific research has matured and diversified. In all areas of knowledge, we observe an increasing number of scientific publications, a rapid development of ever more specialized conferences and journals, and the creation of dynamic collaborative networks that cross borders and evolve over time. In this context, analyzing scientific publications and the resulting inner co-authorship networks is a major issue for the sustainability of scientific research. To illustrate this, let us consider what happens in the context of the COVID-19 pandemics, when the whole scientific community engaged numerous fields of research to contribute in a common effort to study, understand and fight the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). In order to support the scientific community, many datasets covering the publications about coronaviruses and related diseases have been compiled. In a short time, the number of publications available (over 200,000+ and still increasing) suggests that it is impossible for any researcher to examine every publication and extract the relevant information.

By reasoning on the Linked Data and Web semantic schemas, we investigate methods and tools to assist users on finding relevant publications to answer their research questions. Hereafter we present some example of typical domain questions and how we can contributed to the matter.

- How to find relevant publication in huge datasets? We investigate the use of association rules as a suitable solution to identify relevant scientific publications. By extracting association rules that determine the co-occurrence between terms in a text, it is possible to create clusters of scientific publications that follow a certain pattern; users can focus the search on clusters that contain the terms of interests rather than search the whole dataset.

- How to explain the contents of scientific publications? By reasoning on the Linked Data and Web semantic schemas, we investigate methods for the creation and exploration of argurment graphs that describe association and development of ideas in scientific papers.

- How to understand the impact of co-authorship (collaboration of one or more authors) in the development of scientific knowledge? For that, we proposed visualization techniques that allows

the description of co-authorship networks describing the clusters of collaborations that evolve over time. Co-authorship networks can inform both collaboration between authors and institutions.

Currently, the analysis of co-publications has been performed over two majors datasets: Hal open archive and the Covid-on-the-Web datasets.

5   Highlights of the year

5.1   General news

- Michel Buffa was promoted Full Professor.

- Elena Cabrio became the fourth researcher in Wimmics to obtain a 3IA Chair from 3IA Côte d'Azur.

- Serena Villata became the Deputy Scientific Director of the 3IA Côte d'Azur Institute.

- We welcome a new Inria permanent researcher on a starting faculty position: Damien Graux.

- We welcome a new full-time researcher in the team: Amaya Nogales Gómez.

- Starting from November 2021, Serena Villata is mandated (*chargée de mission*) for the Culture Ministry for the mission on "Conversational Agents", together with Célia Zolynski and Karine Favro.[1]

5.2   Awards

- Serena Villata received the "Prix Inria – Académie des sciences jeunes chercheurs et jeunes chercheuses 2021".

- Marco Winckler received IFIP TC Pioneer Award 2021 for his active participation in IFIP Technical Committees and outstanding contributions to the educational, theoretical, technical, commercial, or professional aspects of analysis, design, construction, evaluation and use of interactive systems.

- The team received the "IC 2021 - mention spéciale prix du meilleur article publié à l'international IC 2021" i.e. best international paper at IC 2021 [63] for its article about the CovidOnTheWeb project at ISWC 2021 [18].

- Michel Buffa and Shihong Ren received the best Paper Award at the international WebAudio Conference 2021 for their work on "WebAudio And JavaScript Web Application Using JSPatcher: A Web-Based Visual Programming Editor" [58].

- The team and its members received several awards at the UCA Prix d'excellence 2021 for the results of this year.

6   New software and platforms

The following list of software was extracted from BIL.

6.1   New software

6.1.1   CORESE

Name: COnceptual REsource Search Engine

Keywords: Semantic Web, Search Engine, RDF, SPARQL

---

[1]The details about this mission are available here.

Functional Description: Corese is a Semantic Web Factory, it implements W3C RDF, RDFS, OWL RL, SHACL, SPARQL 1 .1 Query and Update as well as RDF Inference Rules.

Furthermore, Corese query language integrates original features such as approximate search and extended Property Path. It provides STTL: SPARQL Template Transformation Language for RDF graphs. It also provides LDScript: a Script Language for Linked Data. Corese provides distributed federated query processing.

URL: http://project.inria.fr/corese

Contact: Olivier Corby

Participants: Erwan Demairy, Fabien Gandon, Fuqi Song, Olivier Corby, Olivier Savoie, Virginie Bottollier

Partners: I3S, Mnemotix

6.1.2 DBpedia

Name: DBpedia

Keywords: RDF, SPARQL

Functional Description: DBpedia is an international crowd-sourced community effort to extract structured information from Wikipedia and make this information available on the semantic Web as linked open data. The DBpedia triple stores then allow anyone to solve sophisticated queries against Wikipedia extracted data, and to link the different data sets on these data. The French chapter of DBpedia was created and deployed by Wimmics and is now an online running platform providing data to several projects such as: QAKIS, Izipedia, zone47, Sépage, HdA Lab., JocondeLab, etc.

Release Contributions: The new release is based on updated Wikipedia dumps and the inclusion of the DBpedia history extraction of the pages.

URL: http://wiki.dbpedia.org/

Contact: Fabien Gandon

Participants: Fabien Gandon, Elmahdi Korfed

6.1.3 Fuzzy labelling argumentation module

Name: Fuzzy labelling algorithm for abstract argumentation

Keywords: Artificial intelligence, Multi-agent, Knowledge representation, Algorithm

Functional Description: The goal of the algorithm is to compute the fuzzy acceptability degree of a set of arguments in an abstract argumentation framework. The acceptability degree is computed from the trustworthiness associated with the sources of the arguments.

Contact: Serena Villata

Participant: Serena Villata

6.1.4 Corese Server

Name: Corese Server

Keywords: Semantic Web, RDF, SPARQL

Scientific Description: This library provides a Web server to interact with Corese via HTTP requests. In includes a SPARQL endpoint and the STTL display engine to generate portals from linked data (RDF).

Functional Description: This library provides a Web server to interact with Corese via HTTP requests. In includes a SPARQL endpoint and the STTL display engine to generate portals from linked data (RDF).

Contact: Olivier Corby

Participants: Alban Gaignard, Fuqi Song, Olivier Corby

Partner: I3S

6.1.5 CREEP semantic technology

Keywords: Natural language processing, Machine learning, Artificial intelligence

Scientific Description: The software provides a modular architecture specifically tailored at the classification of cyberbullying and offensive content on social media platforms. The system can use a variety of features (ngrams, different word embeddings, etc) and all the netwok parameters (number of hidden layers, dropout, etc) can be altered by using a configuration file.

Functional Description: The software uses machine learning techniques to classify cyberbullying instances in social media interactions.

Release Contributions: Attention mechanism, Hyperparameters for emoji in config file, Predictions output, Streamlined labeling of arbitrary files

Publications: hal-01906096v1, hal-01920266v1

Contact: Michele Corazza

Participants: Michele Corazza, Elena Cabrio, Serena Villata

6.1.6 Licentia

Keywords: Right, License

Scientific Description: In order to ensure the high quality of the data published on the Web of Data, part of the self-description of the data should consist in the licensing terms which specify the admitted use and re-use of the data by third parties. This issue is relevant both for data publication as underlined in the "Linked Data Cookbook" where it is required to specify an appropriate license for the data, and for the open data publication as expressing the constraints on the reuse of the data would encourage the publication of more open data. The main problem is that data producers and publishers often do not have extensive knowledge about the existing licenses, and the legal terminology used to express the terms of data use and reuse. To address this open issue, we present Licentia, a suite of services to support data producers and publishers in data licensing by means of a user-friendly interface that masks to the user the complexity of the legal reasoning process. In particular, Licentia offers two services: i) the user selects among a pre-defined list those terms of use and reuse (i.e., permissions, prohibitions, and obligations) she would assign to the data and the system returns the set of licenses meeting (some of) the selected requirements together with the machine readable licenses' specifications, and ii) the user selects a license and she can verify whether a certain action is allowed on the data released under such license. Licentia relies on the dataset of machine-readable licenses (RDF, Turtle syntax, ODRL vocabulary and Creative Commons vocabulary) available at http://datahub.io/dataset/rdflicense. We rely on the deontic logic presented by Governatori et al. to address the problem of verifying the compatibility of the licensing terms in order to find the license compatible with the constraints selected by the user. The need for licensing compatibility checking is high, as shown by other similar services (e.g., Licensius5 or Creative Commons Choose service6). However, the advantage of Licentia with respect to these services is twofold: first, in these services compatibility is pre-calculated among a pre-defined and small set of licenses, while in Licentia compatibility is computed at runtime and we consider more than 50 heterogeneous licenses, second, Licentia provides a further service that is not considered

by the others, i.e., it allows to select a license from our dataset and verify whether some selected actions are compatible with such license.

Functional Description: Licentia is a web service application with the aim to support users in licensing data. Our goal is to provide a full suite of services to help in the process of choosing the most suitable license depending on the data to be licensed.

The core technology used in our services is powered by the SPINdle Reasoner and the use of Defeasible Deontic Logic to reason over the licenses and conditions.

The dataset of RDF licenses we use in Licentia is the RDF licenses dataset where the Creative Commons Vocabulary and Open Digital Rights Language (ODRL) Ontology are used to express the licenses.

URL: http://licentia.inria.fr/

Contact: Serena Villata

Participant: Cristian Cardellino

6.1.7 SPARQL micro-services

Name: SPARQL micro-services

Keywords: Web API, SPARQL, Microservices, LOD - Linked open data, Data integration

Functional Description: The approach leverages the micro-service architectural principles to define the SPARQL Micro-Service architecture, aimed at querying Web APIs using SPARQL. A SPARQL micro-service is a lightweight SPARQL endpoint that typically provides access to a small, resource-centric graph. Furthermore, this architecture can be used to dynamically assign dereferenceable URIs to Web API resources that do not have URIs beforehand, thus literally "bringing" Web APIs into the Web of Data. The implementation supports a large scope of JSON-based Web APIs, may they be RESTful or not.

URL: https://github.com/frmichel/sparql-micro-service

Publications: hal-02060966, hal-01722792, hal-01947589, hal-02168164

Author: Franck Michel

Contact: Franck Michel

6.1.8 ACTA

Name: A Tool for Argumentative Clinical Trial Analysis

Keywords: Artificial intelligence, Natural language processing, Argument mining

Functional Description: Argumentative analysis of textual documents of various nature (e.g., persuasive essays, online discussion blogs, scientific articles) allows to detect the main argumentative components (i.e., premises and claims) present in the text and to predict whether these components are connected to each other by argumentative relations (e.g., support and attack), leading to the identification of (possibly complex) argumentative structures. Given the importance of argument-based decision making in medicine, ACTA is a tool for automating the argumentative analysis of clinical trials. The tool is designed to support doctors and clinicians in identifying the document(s) of interest about a certain disease, and in analyzing the main argumentative content and PICO elements.

URL: http://ns.inria.fr/acta/

Contact: Serena Villata

6.1.9 WebAudio tube guitar amp sims CLEAN, DISTO and METAL MACHINEs

Name: Tube guitar amplifier simulators for Web Browser : CLEAN MACHINE, DISTO MACHINE and METAL MACHINE

Keyword: Tube guitar amplifier simulator for web browser

Scientific Description: This software is one of the only ones of its kind to work in a web browser. It uses "white box" simulation techniques combined with perceptual approximation methods to provide a quality of guitar playing in hand comparable to the best existing software in the native world.

Functional Description: Software programs for creating real-time simulations of tube guitar amplifiers that behave most faithfully like real hardware amplifiers, and run in a web browser. In addition, the generated simulations can run within web-based digital audio workstations as plug-ins. The "CLEAN MACHINE" version specializes in the simulation of acoustic guitars when playing electric guitars. The DISTO machine specializes in classic rock tube amp simulations, and METAL MA-CHINE targets metal amp simulations. These programs are one of the results of the ANR WASABI project.

Release Contributions: First stable version, delivered and integrated into the ampedstudio.com software. Two versions have been delivered: a limited free version and a commercial one.

News of the Year: Best paper at WebAudio Conference 2020.

Publications: hal-01721463, hal-01893681, hal-02337828, hal-03087768, hal-01721483, hal-01735478, hal-02366725, hal-02557901, hal-01589330, hal-03087763, hal-01893660, hal-01589229

Contact: Michel Buffa

Participant: Michel Buffa

Partner: Amp Track Ltd, Finland

6.1.10 Morph-xR2RML

Name: Morph-xR2RML

Keywords: RDF, Semantic Web, LOD - Linked open data, MongoDB, SPARQL

Functional Description: The xR2RML mapping language that enables the description of mappings from relational or non relational databases to RDF. It is an extension of R2RML and RML.

Morph-xR2RML is an implementation of the xR2RML mapping language, targeted to translate data from the MongoDB database, as well as relational databases (MySQL, PostgreSQL, MonetDB). Two running modes are available: (1) the graph materialization mode creates all possible RDF triples at once, (2) the query rewriting mode translates a SPARQL 1.0 query into a target database query and returns a SPARQL answer. It can run as a SPARQL endpoint or as a stand-alone application.

Morph-xR2RML was developed by the I3S laboratory as an extension of the Morph-RDB project which is an implementation of R2RML.

URL: https://github.com/frmichel/morph-xr2rml/

Publications: hal-01207828, hal-01330146, hal-01280951

Author: Franck Michel

Contact: Franck Michel

6.1.11   ARViz

Name: Association Rules Visualization

Keyword: Information visualization

Scientific Description: ARViz supports the exploration of data from named entities knowledge graphs based on the joint use of association rule mining and visualization techniques. The former is a widely used data mining method to discover interesting correlations, frequent patterns, associations or casual structures among transactions in a variety of contexts. An association rule is an implication of the form X -> Y, where X is an antecedent itemset and Y is a consequent itemset, indicating that transactions containing items in set X tend to contain items in set Y. Although the approach helps to reduce and focus the exploration of large datasets, analysts are still confronted with the inspection of hundreds of rules in order to grasp valuable knowledge. Moreover, when extracting association rules from named entities knowledge graphs, the items are NEs that form antecedent -> consequent links, which the user should be able to cross to recover information. In this context, information visualization can help analysts to visually identify interesting rules that are worthy of further investigation, while providing suitable visual representation to communicate the relationships between itemsets and association rules.

Functional Description: ARViz supports the exploration of thematic attributes describing association rules (e.g. confidence, interestingness, and symmetry) through a set of interactive, synchronized, and complementary visualisation techniques (i.e. a chord diagram, an association graph, and a scatter plot). Furthermore, the interface allows the user to recover the scientific publications related to rules of interest.

Release Contributions: Visualization of association rules within the scientific literature of COVID-19.

URL: http://covid19.i3s.unice.fr:8080/arviz/

Publication: hal-03292140

Contact: Marco Antonio Alba Winckler

Participants: Aline Menin, Lucie Cadorel, Andrea Tettamanzi, Alain Giboin, Fabien Gandon, Marco Antonio Alba Winckler

6.1.12   MGExplorer

Name: Multivariate Graph Explorer

Keyword: Information visualization

Scientific Description: MGExplorer (Multidimensional Graph Explorer) allows users to explore different perspectives to a dataset by modifying the input graph topology, choosing visualization techniques, arranging the visualization space in meaningful ways to the ongoing analysis and retracing their analytical actions. The tool combines multiple visualization techniques and visual querying while representing provenance information as segments connecting views, which each supports selection operations that help define subsets of the current dataset to be explored by a different view. The adopted exploratory process is based on the concept of chained views to support the incremental exploration of large, multidimensional datasets. Our goal is to provide visual representation of provenance information to enable users to retrace their analytical actions and to discover alternative exploratory paths without loosing information on previous analyses.

Functional Description: MGExplorer is an information visualization tool suite that integrates many information visualization techniques aimed at supporting the exploration of multivariate graphs. MGExplorer allows users to choose and combine the information visualization techniques creating a graph that describes the exploratory path of dataset.

Release Contributions: Visualization of data extracted from linked data datasets.

URL: http://covid19.i3s.unice.fr:8080/

Publications: hal-03292172, hal-03404572, hal-03404580

Contact: Marco Antonio Alba Winckler

Participants: Aline Menin, Marco Antonio Alba Winckler, Olivier Corby

Partner: Universidade Federal do Rio Grande do Sul

7 New results

7.1 Users Modeling and Designing Interaction

7.1.1 MGExplorer: A Visual Approach for Representing Analytical Provenance in Exploration Processes

Participants:	Marco Winckler, Aline Menin, Olivier Corby, Catherine Faron, Alain Giboin, Fabien Gandon, Maroua Tikat, Michel Buffa.

Visualization techniques are useful tools to explore datasets, enabling the discovery of meaningful patterns and causal relationships. Nonetheless, the discovery process is often exploratory and requires multiple views to support analyzing different or complementary perspectives to the data. In this context, analytic provenance shows great potential to understand users' reasoning process through the study of their interactions on multiple view systems. Thus, we present an approach based on the concept of chained views to support the incremental exploration of large, multidimensional datasets. Our goal is to provide visual representation of provenance information to enable users to retrace their analytical actions and to discover alternative exploratory paths without loosing information on previous analyses. We demonstrate that our implementation of the approach, MGExplorer (Multidimensional Graph Explorer), allows users to explore different perspectives to a dataset by modifying the input graph topology, choosing visualization techniques, arranging the visualization space in meaningful ways to the ongoing analysis and retracing their analytical actions. The tool combines multiple visualization techniques and visual querying while representing provenance information as segments connecting views, which each supports selection operations that help define subsets of the current dataset to be explored by a different view. The resulting publication [55] presents a flexible visualization approach based on the concept of chained views, capable of depicting analytical provenance via a sequence of views, while supporting one or more visualization techniques applied to one or more datasets. Thus, it supports visual analysis via multiple alternative exploration scenarios that can be retraced and modified. We compare our solution with other existing approaches, highlighting the need for innovative and flexible tools for exploring large datasets, and demonstrate the feasibility of the chained views approach through an interactive visual tool called MGExplorer, created to assist the exploration of multivariate networks, and available online.

7.1.2 LDViz: Linked Data Visualization

Participants:	Marco Winckler, Aline Menin, Olivier Corby, Catherine Faron, Alain Giboin, Fabien Gandon, Maroua Tikat, Michel Buffa.

Over the recent years, Linked Open Data (LOD) has been increasingly used to support decision-making processes in various application domains. For that purpose, an increasing interest in information visualization has been observed in the literature as a suitable solution to communicate the knowledge described in LOD data sources. Nonetheless, transforming raw LOD data into a graphical representation (the so-called visualization pipeline) is not a straightforward process and often requires a set of operations to transform data into meaningful visualizations that suit users' needs. Contrariwise to typical visualizations, which uses specific datasets whose structure and nature are known, enabling one to

easily define indicators and visualization techniques that are suitable to support the data exploration, visualizing linked data requires a preceding RDF graph processing to retrieve suitable data that may originate from different endpoints. Moreover, it may sometimes require combining data from different endpoints, resulting in datasets that may contain quality issues (e.g., missing data, inconsistency) and which structure and nature are unknown to the visualization. Although the design process of every visualization tool follows a well-known pipeline (i.e., import -> transform -> map -> render -> interact), we could not find any definition of these stages and the issues that arise from applying such visualization pipeline for LOD exploration. Particularly, a visualization pipeline for LOD data should take into account the linked nature of these datasets by leveraging/supporting/exploiting these links, while being capable of processing and visualizing the data appropriately. This requires a high level of flexibility in every step of the pipeline, which could be seen in the drafting of SPARQL queries in a way that appropriately addresses the links in the linked data, the possibility of tuning the parameters of the graphic display and the associated interaction, and the availability of multiple visualization techniques that can help users see data according to diverse and complementary viewpoints.

In the paper [56] we propose a LOD generic visualization pipeline and discuss the implications of the internal operations for creating meaningful visualizations of LOD datasets. To demonstrate the feasibility of this generic visualization pipeline, we implement it as the tool LDViz (Linked Data Visualizer), which integrates a SPARQL query management interface, a data transformation engine, and a visualization interface to support the automatic visualization of data extracted from any SPARQL endpoint. Our implementation allows any expert user to access the SPARQL endpoint of their choice, perform searches with SPARQL queries, and visualize the results via a visualization interface, MGExplorer, designed to assist the exploration of any multivariate network. The tool is available online.

7.1.3   FollowUp Queries: Incremental exploration of linked data

Participants:	Marco Winckler, Aline Menin, Olivier Corby, Catherine Faron, Alain Giboin, Maroua Tikat, Michel Buffa.

Information visualization techniques are useful to discover patterns and causal relationships within LOD datasets. However, since the discovery process is often exploratory (i.e., users have no predefined goal and do not expect a particular outcome), when users find something interesting, they should be able to (i) retrace their exploratory path to explain how results have been found, and (ii) branch out the exploratory path to compare data observed in different views or found in different datasets. Furthermore, as most of LOD datasets are very specialized, users often need to explore multiple datasets to obtain the knowledge required to support decision-making processes. Thus, the design of visualization tools is confronted with two main challenges: the visualization system should provide multiple views to enable the exploration of different or complementary perspectives to the data; and the system should support the combination of diverse data sources during the exploration process. To our knowledge, the existing tools before our work, are limited to visualizing a single dataset at a time and, often, use static and preprocessed data. Thus, we proposed the concept of follow-up queries to allow users to create queries on demand during the exploratory process while connecting multiple LOD datasets with chained views. Our approach relies on a exploration process supported by the use of predefined SPARQL queries that the user can select on-the-fly to retrieve data from different SPARQL endpoints. It enables users to enrich the ongoing analysis by bringing external and complementary data to the exploration process, while also supporting the visual analysis and comparison of different subsets of data (from the same or different SPARQL endpoints) and, thus, the incremental exploration of the LOD cloud. The resulting paper [31] (accepted for publication at IJHCI) presents a generic visualization approach to assist the analysis of multiple LOD datasets based on the concepts of chained views and follow-up queries. We demonstrate the feasibility of our approach via four use case scenarios and a formative evaluation where we explore scholarly data described by RDF graphs publicly available through SPARQL endpoints. These scenarios demonstrate how the tool supports (i) composing, running, and visualizing the results of a query; (ii) subsetting the data and exploring it via different visualization techniques; (iii) instantiating a follow-up query to retrieve external data; and (iv) querying a different database and compare datasets. The usability and usefulness of the proposed approach is confirmed by results obtained with a series of semi-structured

interviews. The results are encouraging while showing the relevance of the approach to explore big linked data.

7.1.4   Association Rules Visualization

> **Participants:**   Marco Winckler, Aline Menin, Lucie Cadorel, Andrea Tettamanzi,
> Alain Giboin, Fabien Gandon.

In order to extend the palette of tools for exploring the COVID-19 litterature (which part is described by the CovidOnTheWeb dataset), an association rules mining approach was proposed [81] to assist the discovery of interesting knowledge, hidden by the deluge of data, that could be helpful for assisting decision-making processes. This data mining method is widely used to discover interesting correlations, frequent patterns, associations, or casual structures among transactions in a variety of contexts. An association rule is an implication of the form X -> Y, where X is an antecedent itemset and Y is a consequent itemset, indicating that transactions containing items in set X tend to contain items in set Y.

Although the approach helps to reduce and focus the exploration of large datasets, analysts are still confronted with the inspection of hundreds of rules in order to grasp valuable knowledge. Moreover, when extracting association rules from named entities knowledge graphs, the items are NEs that form antecedent -> consequent links, which the user should be able to cross to recover information. In this context, information visualization can help analysts to visually identify interesting rules that are worthy of further investigation, while providing suitable visual representation to communicate the relationships between itemsets and association rules. Thus, we proposed a visualization interface, called ARViz (Association Rules Visualization), to assist the exploration of association rules over RDF knowledge graphs and their measures of interest, supporting tasks of comparison, identification, and overview of items and rules. The resulting publication [54] presents the proposed visualization approach, as well as a comparative review of existing visualization interfaces/techniques for association rule exploration regarding task support, and the results of a formative evaluation performed with expert users in Semantic Web and biomedical research to assess the feasibility and usefulness of our approach. The tool is available covid19.i3s.unice.fr:8080/arviz/. Although ARViz is currently applied to the CovidOntheWeb RDF dataset, its architecture is generic enough to represent association rules datasets extracted from data of different type (beyond RDF) and describing different phenomena (i.e. relationship between products frequently bought together in a supermarket).

7.1.5   Interactive WebAudio applications

> **Participants:**   Michel Buffa, Shihong Ren.

In the context of the WASABI research project, we built a 2M song database made of metadata collected from the Web of Data and from the analysis of song lyrics [84] of the audio files provided by Deezer. We designed a WebAudio plugin standard, new tools for developing high performances plugins in the browser [89], and new methods for real-time tube guitar amplifier simulations that run in the browser [79]. Some of these results are unique in the world as in 2022, and have been acclaimed by several awards in international conferences. The guitar amp simulations are now commercialized by the CNRS SATT service and are available in the online collaborative Digital Audio Workstation ampedstudio.com [80]. Some other tools we designed are linked to the WASABI knowledge base, that allow, for example, songs to be played along with sounds similar to those used by artists. An ongoing PhD proposes a visual language for music composers to create instruments and effects linked to the WASABI corpus content [90, 58], and has been awarded by a best paper at the WebAudio Conference 2022.

7.1.6   Timelining Knowledge Graphs in the browser

Participant: Damien Graux.

Knowledge graphs, available on the Web via SPARQL endpoints, provide practitioners with various kinds of information from general considerations to more specific ones such as temporal data. In this effort, we propose a light-weight solution to visually grasp, navigate, and compare, in a Web browser, temporal information available from SPARQL endpoints.

In particular, we wanted our approach to be lightweight and to do so we exploited SPARQL as much as possible. Indeed, the information necessary to build the timelines is extracted through a pair of SPARQL queries. This allows the approach to be easy deployed on alternative endpoints and even to visually compare temporal information coming from various sources (as shown on the online demonstrator).

This solution was presented at the Voila! workshop co-located with ISWC'21 [45]. Furthermore, we use the Wikidata public SPARQL endpoint to demonstrate our solution and allow users to navigate Wikidata's temporal information.

7.2   Communities and Social Interactions Analysis

7.2.1   Autonomous agents in a social and ubiquitous Web

Participants: Andrei Ciortea, Olivier Corby, Fabien Gandon, Franck Michel.

Recent W3C recommendations for the Web of Things (WoT) and the Social Web are turning hypermedia into a homogeneous information fabric that interconnects heterogeneous resources: devices, people, information resources, abstract concepts, etc. The integration of multi-agent systems with such hypermedia environments now provides a means to distribute autonomous behavior in worldwide pervasive systems. A central problem then is to enable autonomous agents to discover heterogeneous resources in world wide and dynamic hypermedia environments. This is a problem in particular in WoT environments that rely on open standards and evolve rapidly—thus requiring agents to adapt their behavior at runtime in pursuit of their design objectives. To this end, we developed a hypermedia search engine for the WoT that allows autonomous agents to perform approximate search queries in order to retrieve relevant resources in their environment in (weak) real time. The search engine crawls dynamic WoT environments to discover and index device metadata described with the W3C WoT Thing Description, and exposes a SPARQL endpoint that agents can use for approximate search. To demonstrate the feasibility of our approach, we implemented a prototype application for the maintenance of industrial robots in worldwide manufacturing systems. The prototype demonstrates that our semantic hypermedia search engine enhances the flexibility and agility of autonomous agents in a social and ubiquitous Web [82].

7.2.2   Abusive language detection

Participants: Elena Cabrio, Serena Villata, Anais Ollagnier.

Recent studies have highlighted the importance to reach a fine-grained online hate speech characterisation to provide appropriate solutions to curb online abusive behaviours. In this direction, we have proposed a full pipeline that enables to capture targeting characteristics in hatred contents (i.e., types of hate, such as race and religion) aiming at improving the understanding on how hate is conveyed on Twitter. Our contribution is threefold: (1) we leverage multiple data views of a different nature to contrast different kinds of abusive behaviours expressed towards targets; (2) we develop a full pipeline relying on a multi-view clustering technique to address the task of hate speech target characterisation; and (3) we propose a methodology to assess the quality of generated hate speech target communities. Relying on multiple data views built from multilingual pre-trained language models (i.e., multilingual BERT and multilingual Universal Sentence Encoder) and the Multi-view Spectral Clustering (MvSC) algorithm,

the experiments conducted on a freely available multilingual dataset of tweets (i.e., the MLMA hate speech dataset) show that most of the configurations of the proposed pipeline significantly outperforms state-of-the-art clustering algorithms on all the tested clustering quality metrics on both French and English (paper under submission).

In addition, we have carried out a data collection in the context of the UCA IDEX OTESIA project "Artificial Intelligence to prevent cyberviolence, cyberbullying and hate speech online." [2] to created a dataset of aggressive chats in French collected through a role-playing game in high-schools. The collected conversations have been annotated with the participant roles (victim, bully and bystanders), the presence of hate speech (content that mocks, insults, or discriminates against a person or group based on specific characteristics such as colour, ethnicity, gender, sexual orientation, nationality, religion, or other characteristics) and whether utterances use different humour figurative devices such as sarcasm or irony. Moreover, we have also introduced a new annotation layer referring to the different types of verbal abuse present in the message, defined as a type of psychological/mental abuse that involves the use of written language relying on derogatory terms, the delivery of statements intended to demean, humiliate, blame or threaten the victim with the aim of decreasing its self-confidence and making it feel powerless. The identification of the different types of verbal abuse will allow to investigate (and learn, from a computational perspective) strategies used by cyberhate perpetrators to cause emotional harm, and to reach insights about how victims respond to bullying/victimisation.

7.2.3 Propaganda detection and classification

Participants: Vorakit Vorakitphan, Elena Cabrio, Serena Villata.

One of the mechanisms through which disinformation is spreading online, in particular through social media, is by employing propaganda techniques. These include specific rhetorical and psychological strategies, ranging from leveraging on emotions to exploiting logical fallacies. The goal of our work is to push forward research on propaganda detection based on text analysis, given the crucial role these methods may play to address this main societal issue. More precisely, we propose a supervised approach to classify textual snippets both as propaganda messages and according to the precise applied propaganda technique, as well as a detailed linguistic analysis of the features characterising propaganda information in text (e.g., semantic, sentiment, and argumentation features). Extensive experiments conducted on two available propagandist resources (i.e., NLP4IF'19 and SemEval'20-Task 11 datasets) show that the proposed approach, leveraging different language models and the investigated linguistic features, achieves very promising results on propaganda classification, both at sentence- and at fragment-level [61]. The best performing approach has been implemented into a demo system called PROTECT (PROpaganda Text dEteCTion), that automatically detects propagandist messages and classifies them along with the propaganda techniques employed. PROTECT is designed as a full pipeline to firstly detect propaganda text snippets from the input text, and then classify the technique of propaganda, taking advantage of semantic and argumentation features [62].

7.2.4 Using Agent-Based Modeling to explore the role of socio-environmental interactions on Ancient Settlement Dynamics

Participant: Andrea Tettamanzi.

Within the framework of a mult-disciplinary project involving archaeologists, economists, geographers, and computer scientists, we used Agent-Based Modelling to explore the respective impacts of environmental and social factors on the settlement pattern and dynamics during the Roman period in South-Eastern France [64].

---

[2]OTESIA Project Launch

7.2.5 Web Futures: Inclusive, Intelligent, Sustainable The 2020 Manifesto for Web Sciences

Participant: Fabien Gandon.

We co-authored the manifesto produced from the Perspectives Workshop 18262 entitled "10 Years of Web Science" that took place at Schloss Dagstuhl from June 24-29, 2018. At the Workshop, we revisited the origins of Web Science, explored the challenges and opportunities of the Web, and looked ahead to potential futures for both the Web and Web Science. We outline the need to extend Web Science as the science that is devoted to the analysis and engineering of the Web, to strengthen our role in shaping the future of the Web and present five key directions for capacity building that are necessary to achieve this: (i), supporting interdisciplinarity, (ii), supporting collaboration, (iii), supporting the sustainable Web, (iv), supporting the Intelligent Web, and (v), supporting the Inclusive Web. Our writing reflects our background in several disciplines of the social and technical sciences and that these disciplines emphasize topics to various extents. [26]

7.2.6 Analysing the use of OpenStreetMap solutions in H2020 European projects

Participant: Damien Graux.

Since 1984, the European Commission has been supporting research through various successive programmes. Recently, from 2014 to 2020, the EU invested approximately 80 billion euros into its eighth programme, named Horizon 2020. Among various focuses such as the excellence of science or industrial secondments, H2020 emphasised on supporting an open access policy for all the research results. Moreover, H2020 projects were strongly encouraged to use open source software and tools.

In this study, we systematically analysed all the available H2020 deliverables, searching for cartographic service references, with a specific focus on OpenStreetMap. Our efforts show that OSM is the most used cartographic service in European H2020 projects in terms of mentions in the deliverable's texts, followed by GoogleMaps with one order of magnitude less mentions. It is worth noting that these projects involving OSM were backed by almost 4 billion euros of public money.

We presented our results at the State-of-the-Map event [44] (the international forum dedicated the OpenStreetMap community); and made our results available online.

7.3 Vocabularies, Semantic Web and Linked Data Based Knowledge Representation and Artificial Intelligence Formalisms on the Web

7.3.1 Publication and exploitation of the Covid-on-the-Web dataset

Participants: Franck Michel, Fabien Gandon, Valentin Ah-Kane, Anna Bobasheva, Elena Cabrio, Olivier Corby, Raphaël Gazzotti, Alain Giboin, Santiago Marro, Tobias Mayer, Serena Villata, Marco Winckler.

The Covid-on-the-Web project aims to allow biomedical researchers to access, query, and make sense of COVID-19 related literature. Launched in Mars 2020, it involved multiple skills of the team in knowledge representation, text, data and argument mining, as well as data visualization and exploration. A first outcome was the publication of the Covid-on-the-Web RDF dataset genetared by analyzing and enriching the "COVID-19 Open Research Dataset" (100K+ full-text scientific articles related to the coronaviruses), comprising (1) named entities mentioned in the articles and (2) arguments extracted using the ACTA platform.

In 2021, we carried on this work with a paper at the 32es Journées francophones d'Ingénierie des Connaissances that received a best paper award [63]. We also pushed forward the reflection on the adaptation of visualization and exploration methods for linked data sources [32]. This was implemented

in the LDViz tool that assists the exploration of different views of the data by combining a querying management interface, which enables the definition of meaningful subsets of data through SPARQL queries, and a visualization interface based on a set of six visualization techniques integrated in a chained visualization concept, which also supports the tracking of provenance information.

Finnaly, we also proposed a method to identify `owl:sameAs` relationships of a resource relying on online SPARQL querying of distributed datasets and to correct results using declarative curation rules. We also exploit and inspect the quality of owl:InverseFunctionalProperty and owl:FunctionalProperty relationships, using the definitions given by their schemata, endpoints and a voting approach. We evaluated our method on an existing benchmark and compare to state of the art baselines. We showed that a heuristic approach can retrieve high quality equivalence links without requiring the extraction of all the alleged existing equivalence relations. [40]

7.3.2 Publication of the WASABI dataset

Participants: Michel Buffa, Franck Michel, Fabien Gandon, Elena Cabrio, Alain Giboin, Marco Winckler, Maroua Tikat, Michael Fell.

Since 2017, a two-million song database consisting of metadata collected from multiple open data sources and automatically extracted information has been constructed in the context of the WASABI project. The goal is to build a knowledge graph linking collected metadata (artists, discography, producers, dates, etc.) with metadata generated by the analysis of both the songs' lyrics (topics, places, emotions, structure, etc.) and audio signal (chords, sound, etc.). It relies on natural language processing and machine learning methods for extraction, and semantic Web frameworks for integration. The dataset describes more than 2 millions commercial songs, 200K albums and 77K artists. It can be exploited by music search engines, music professionals or scientists willing to analyze popular music published since 1950. It is available under an open license in multiple formats and is accompanied by online applications and open source software including an interactive navigator, a REST API and a SPARQL endpoint. This work has been described in [65]

Wasabi dataset Web site.

7.3.3 Semantic Web for Biodiversity

Participants: Franck Michel, Catherine Faron.

This activity addresses the challenges of exploiting knowledge representation and semantic Web technologies to enable data sharing and integration in the biodiversity area. The collaboration with the "Muséum National d'Histoire Naturelle" of Paris (MNHN) goes on along several axes.

First, since 2019 the MNHN has been using our SPARQL Micro-Services architecture and framework to help biologists in editing taxonomic information by confronting multiple, heterogeneous data sources [88]. This collaboration is going on and has been strengthened over 2020 and 2021, the MNHN now heavily relies on those services for daily activities.

Second, the work initiated within the Bioschemas.org W3C community group seeks the definition and adoption of common biology-related markup terms. The schema.org/Taxon term is now officially published and new term TaxonName is being pushed forward.

Lastly, we co-organized the symposium "Connecting biodiversity data with knowledge graphs" at the yearly TDWG 2021 international conference, and we presented advances of our work in terms of biodiversity knowledge modeling [57].

7.3.4 Enriching the WASABI Song Corpus with Lyrics Annotations.

Participants: Elena Cabrio, Michael Fell, Michel Buffa.

The WASABI Song Corpus is a large corpus of songs enriched with metadata extracted from music databases on the Web, and resulting from the processing of song lyrics and from audio analysis. Given that lyrics encode an important part of the semantics of a song, we have focused on the design and application of methods to extract relevant information from the lyrics, such as their structure segmentation, their topics, the explicitness of the lyrics content, the salient passages of a song and the emotions conveyed. So far, the corpus contains 1.73M songs with lyrics (1.41M unique lyrics) annotated at different levels with the output of the above mentioned methods. Such corpus labels and the provided methods can be exploited by music search engines and music professionals (e.g. journalists, radio presenters) to better handle large collections of lyrics, allowing an intelligent browsing, categorization and recommendation of songs.

7.3.5 Ontology alignment in the sourcing domain

Participants: Molka Dhouib, Catherine Faron, Andrea Tettamanzi.

Ontology alignement plays a key role in the management of heterogeneous data sources and metadata. In this context, various ontology techniques have been proposed to discover correspondences between the entities of different ontologies. We proposed a new ontology alignement approach based on a set of rules exploiting the embedding space and measuring clusters of labels to discover the relationship between entites. We tested our system on several open datasets from the Ontology Alignment Evaluation Initiative (OAEI) benchmark and then applied it to aligning ontologies in a real-world case study provided by Silex cie and the "Office National d'Information sur les Enseignements et les Professions" (ONISEP) [34].

7.3.6 A feature-based comparative analysis of legal ontologies

Participants: Serena Villata.

Ontologies represent the standard way to model the knowledge about specific domains. This holds also for the legal domain where several ontologies have been put forward to model specific kinds of legal knowledge. Both for standard users and for law scholars, it is often difficult to have an overall view on the existing alternatives, their main features, and their interlinking with the other ontologies. To answer this need, we address an analysis of the state-of-the-art in legal ontologies and we characterize them along with some distinctive features. This work aims to guide generic users and law experts in selecting the legal ontology that better fits their needs and in understanding its specificity so that proper extensions to the selected model could be investigated [87].

7.3.7 Evolutionary agent-based evaluation of the sustainability of different knowledge sharing strategies in open multi-agent systems

Participants: Stefan Sarkadi, Fabien Gandon.

The advancement of agent technologies and their deployment in various fields of application has brought numerous benefits w.r.t., knowledge or data gathering and processing. However, one of the key challenges in deploying artificial intelligent agents in an open environment like the Web is their interoperability. Even tough research and development of agent technologies on the Semantic Web has advanced significantly,

artificial agents live on the Web in silos, that is in very limited domains, isolated from other systems and agents that live on the Web. In this work we setup a simulation framework and evaluation based on evolutionary agent-based modeling to empirically test how sustainable different strategies are for knowledge sharing in open multi-agent systems and to see which of these strategies could actually enable global interoperability between Web agents. The first results are showing the interest of translation-based approaches and the need for further incentives to support these.

7.4 Analyzing and Reasoning on Heterogeneous Semantic Graphs

7.4.1 Uncertainty Evaluation for Linked Data

Participants: Ahmed Elamine Djebri, Fabien Gandon, Andrea Tettamanzi.

For data sources to ensure providing reliable linked data, they need to indicate information about the (un)certainty of their data based on the views of their consumers. In addition, uncertainty information in terms of Semantic Web has also to be encoded into a readable, publishable, and exchangeable format to increase the interoperability of systems. We introduced a novel approach to evaluate the uncertainty of data in an RDF dataset based on its links with other datasets. We proposed to evaluate uncertainty for sets of statements related to user-selected resources by exploiting their similarity interlinks with external resources. Our data-driven approach translates each interlink into a set of links referring to the position of a target dataset from a reference dataset, based on both object and predicate similarities. We showed how our approach can be implemented and present an evaluation with real-world datasets. Finally, we discussed updating the publishable uncertainty values. Details are available online.

7.4.2 Extended SPARQL Service

Participants: Olivier Corby, Catherine Faron, Fabien Gandon, Franck Michel, Damien Graux.

In the context of the D2KAB ANR project, we have investigated federated SPARQL queries. We have introduced the concept of Extended SPARQL Service where the URL of the service endpoint can be provided with URL parameters in order to tune client and server behavior.

In addition, we have designed a service execution mode which enables us to provide a service execution report in such a way that the user can trace the execution of service clauses. We have also introduced the concept of ”Linked Result” where the SPARQL Query Results are provided with additional information in the form of linked documents, e.g. a geographical map generated for geolocalized data.

This work has been published at WEBIST [36].

7.4.3 Tracking RDF updates using SPARQL hashing features

Participant: Damien Graux.

The recent increase of RDF usage has witnessed a rising need of "verification" around data obtained from SPARQL endpoints. It is now possible to deploy Semantic Web pipelines and to adapt them to a wide range of needs and use-cases. Practically, these complex ETL pipelines relying on SPARQL endpoints to extract relevant information often have to be relaunched from scratch every once in a while in order to refresh their data. Such a habit adds load on the network and is heavy resource-wise, while sometimes unnecessary if data remains untouched.

In this effort, we present a useful method to help data consumers (and pipeline designers) identify when data has been updated in a way that impacts the pipeline's result set. This method is based on standard SPARQL 1.1 features and relies on digitally signing parts of query result sets to inform data

consumers about their eventual change. Additionally, we also declined this approach to the case of SPARQL queries using federation features with the SERVICE keyword. In particular, our solution divides the coming query into pieces following the SERVICE calls and generated a sort of Merkle-tree to better locate the potential RDF data updates.

This effort has been published at ISWC'21 as a poster [47] and in MEPDaW'21 [46]. In addition, an online interface, implemented in JavaScript, is available online: dgraux.github.io/SPARQL-hash/

7.4.4 Using Formal Concept Analysis to compress RDF data versions

Participant: Damien Graux.

Recent years have witnessed the increase of openly available RDF knowledge graphs online. With this availability of information comes the challenge of coping with dataset versions as information may change in time and therefore deprecates the former knowledge graph. Several solutions have been proposed to deal with data versioning, mainly based on computing data deltas and having an incremental approach to keep track of the version history. In this effort, we describe a novel method that relies on aggregating graph versions to obtain one single complete graph. Our solution semantically compresses similar and common edges together to obtain a final graph smaller than the sum of the distinct versioned ones. Technically, our method takes advantage of FCA to match graph elements together. We also describe how this compressed graph can be queried without being unzipped, using standard methods.

This effort has been published in the ninth FCA4AI workshop co-located with IJCAI 2021 [41].

7.4.5 A decentralized triplestore managed via the Ethereum blockchain

Participant: Damien Graux.

The growing Web of data warrants better data management strategies. Data silos are single points of failure and they face availability problems which lead to broken links. Furthermore the dynamic nature of some datasets increases the need for a versioning scheme. In this work, we propose a novel architecture for a linked open data infrastructure, built on open decentralized technologies. IPFS is used for storage and retrieval of data, and the public Ethereum blockchain is used for naming, versioning and storing metadata of datasets.

We furthermore exploit two mechanisms for maintaining a collection of relevant, high-quality datasets in a distributed manner in which participants are incentivised. The platform is shown to have a low barrier to entry and censorship-resistance. It benefits from the fault-tolerance of its underlying technologies. Furthermore, we validate the approach by implementing our solution.

These results were presented at SEMANTiCS [67].

7.4.6 A Semantic Model for Meteorological Knowledge Graphs

Participants: Nadia Yacoubi Ayadi, Catherine Faron, Franck Michel, Olivier Corby, Fabien Gandon.

The great interest advocated by the agronomy and biodiversity communities in the development of crop models coupled to weather and climate models has led to the need for datasets of meteorological observations in which data are semantically described and integrated. For this purpose, in the context of the D2KAB ANR project, we have proposed a semantic model to represent and publish meteorological observational data as Linked Data. Our model reuses a network of existing ontologies to capture the semantics of data, it covers multiple dimensions of meteorological data including geospatial, temporal, observational and provenance characteristics. Our proposition provides also a SKOS vocabulary of terms

to describe domain-specific observable properties and features. We paid specific attention to propose a model that adheres to LD best practices and standards, thereby allowing for its extension and re-use by several meteorological data producers, and making it capable of accommodating multiple application domains. Therefore, we implemented and made available a software pipeline that is reproducible to generate RDF-based datasets compliant with the proposed semantic model. For instance, we generate the first release of an RDF dataset built from the Météo-France archives.

7.4.7 Corese Semantic Web Factory

Participants: Rémi Ceres, Olivier Corby.

In the context of the National research program in artificial intelligence (PNRIA) [3] and in collaboration with the Mnemotix cooperative [4] we work on the "industrialization" of Corese.

We introduced the concept of external Property to tune Corese behavior.

We improved Corese GUI by adding graphical editors for Turtle and SHACL.

We have implemented the RDF4J [5] model API in Corese. This allows users to request and manipulate a Corese Graph through a standard RDF4J API.

We have designed a "Broker" architecture to plug the SPARQL interpreter to external RDF graph implementations. This will allow the use of Corese SPARQL engine with triple storage system and prepare for the implementation of persistency.

7.4.8 Linked Data Crawling

Participants: Fabien Gandon, Hai Huang.

A Linked Data crawler performs a selection to focus on collecting linked RDF (including RDFa) data on the Web. From the perspectives of throughput and coverage, given a newly discovered and targeted URI, the key issue of Linked Data crawlers is to decide whether this URI is likely to dereference into an RDF data source and therefore if it is worthy downloading the representation it points to. Current solutions adopt heuristic rules to filter irrelevant URIs. But when the heuristics are too restrictive this hampers the coverage of crawling. We proposed and compared approaches to learn strategies for crawling Linked Data on the Web by predicting whether a newly discovered URI will lead to an RDF data source or not. We detailed the features used in predicting the relevance and the methods we evaluated including a promising adaptation of FTRL-proximal online learning algorithm. We compared several options through extensive experiments including existing crawlers as baseline methods to evaluate their efficiency [53].

7.4.9 Semantic Overlay Network for Linked Data Access

Participants: Fabien Gandon, Mahamadou Toure.

We proposed and evaluated MoRAI (Mobile Read Access in Intermittent internet connectivity), a distributed peer-to-peer architecture organized in three levels dedicated to RDF data exchanges by mobile contributors. We presented the conceptual and technical aspects of this architecture as well as a theoretical analysis of the different characteristics. We then evaluated it experimentally and results show the relevance of considering geographical positions during data exchanges and of integrating RDF graph replication to ensure data availability in terms of requests completion rate and resistance to crash scenarios [24].

---

[3] National research program in artificial intelligence (PNRIA)
[4] Mnemotix cooperative
[5] RDF4J Documentation

7.4.10   SHACL Extension

Participants: Olivier Corby, Iliana Petrova, Fabien Gandon, Catherine Faron.

In the context of a collaboration with Stanford University, we worked on extensions of W3C SHACL Shape Constraint Language [6].

We conducted a study on large, active, and recognized ontology projects (ex. Gene Ontology, Human Phenotype Ontology, Mondo Disease Ontology, Ontology for Biomedical Investigations, OBO Foundry, etc.) as well as an analysis of several existing tools, methodologies and guidelines for ontological engineering.

As a result we identified several sets of ontology validation constraints that fall into six big clusters: **i)** formalization/modeling checks; **ii)** terminological/writing checks; **iii)** documentation/ editorial practices, terminology-level checks; **iv)** coherence between terminology and formalization; **v)** metamodel-based checks; **vi)** integration/interoperability/data checking, and can be further refined depending on whether they are specific to RDFS/OWL meta-model, domain/ontology specific, or Linked Data specific. This precise categorization of the ontology validation constraints allowed us to analyse the needs and impact of the extension we are targeting in terms of semantic expressiveness, computational complexity of the validation and current syntax of the SHACL language.

We then concentrated on the formalization of the semantic extensions and their validation methods and came up with a proposal of a corresponding syntactic extensions of SHACL.

The formal specification of the identified extensions enabled us to proceed with the implementation of a prototype plugin for Protégé (Stanford's widely used ontology editor) based on the Corese engine and which extends the SHACL standard with these newly proposed capabilities.

7.4.11   Injection of Knowledge in a Sourcing Recommender System

Participants: Molka Dhouib, Catherine Faron, Andrea Tettamanzi.

In the framework of a collaborative project with Silex cie aiming to propose a decision support to recommend relevant providers for a service request, we proposed a sourcing recommender system approach that exploits knowledge extracted from textual descriptions of providers and service requests to automatically suggest the best providers. In this work, we study the benefits of using ontological knowledge to improve our recommender process. We focus especially on the enrichment of the vector representation of service requests and providers with domain knowledge. The result of this work is presented in the doctoral thesis [23].

7.4.12   Identifying argumentative structures in clinical trials

Participants: Elena Cabrio, Serena Villata, Tobias Mayer, Santiago Marro.

In the latest years, the healthcare domain has seen an increasing interest in the definition of *intelligent* systems to support clinicians in their everyday tasks and activities. Among others, also the field of Evidence-Based Medicine is impacted by this twist, with the aim to combine the reasoning frameworks proposed thus far in the field with mining algorithms to extract structured information from clinical trials, clinical guidelines, and Electronic Health Records. In this work, we go beyond the state of the art by proposing a new end-to-end pipeline to address argumentative outcome analysis on clinical trials. More precisely, our pipeline is composed of *(i)* an Argument Mining module to extract and classify argumentative components (i.e., evidence and claims of the trial) and their relations (i.e., support,

---

[6]see SHACL Recommendation

attack), and *(ii)* an outcome analysis module to identify and classify the effects (i.e., improved, increased, decreased, no difference, no occurrence) of an intervention on the outcome of the trial, based on PICO elements. We annotated a dataset composed of more than 500 abstracts of Randomized Controlled Trials (RCT) from the MEDLINE database, leading to a labeled dataset with 4198 argument components, 2601 argument relations, and 3351 outcomes on five different diseases (i.e., *neoplasm, glaucoma, hepatitis, diabetes, hypertension*). We experiment with deep bidirectional transformers in combination with different neural architectures (i.e., LSTM, GRU and CRF) and obtain a macro F1-score of .87 for component detection and .68 for relation prediction, outperforming current state-of-the-art end-to-end Argument Mining systems, and a macro F1-score of .80 for outcome classification [30].

7.4.13 Qualitative evaluation of arguments in persuasive essais

Participants: Elena Cabrio, Serena Villata, Santiago Marro.

To generate good argument-based explanations we need to first assess the argument(ation) quality. In 2021 we worked in this direction with the definition of guidelines to annotate the different quality dimensions present in arguments. We decided to start from the annotation of persuasive essays, as they represent prototypical instances of argumentation. We then plan to move to other argumentation genres, in particular clinical texts, where we are aware further quality dimensions will require to be included.

Following these guidelines, we created a new quality annotated dataset that enables us to propose different neural network models with the objective to automatically assess the argument(ation) quality in persuasive essays. The results obtained with the proposed models are promising and are still being adjusted for further publication.

7.4.14 Identification of the information captured by Knowledge Graph Embeddings

Participants: Antonia Ettorre, Anna Bobasheva, Catherine Faron, Franck Michel.

The recent growth in the utilization of Knowledge Graphs has been powered by the expanding landscape of Graph Embedding techniques, which facilitates the manipulation of the vast and sparse information described by such Knowledge Graphs. Although the effectiveness of Knowledge Graph Embeddings has been proved on many occasions and for many contexts, the interpretability of such vector representations remains an open issue. To tackle it, we provided a systematic approach to decode and make sense of the knowledge captured by Graph Embeddings. We proposed a technique for verifying whether Graph Embeddings are able to encode certain properties of the graph elements they represent and we presented a categorization for such properties [38].

7.4.15 Embedding Knowledge Graphs Attentive to Positional and Centrality Qualities

Participant: Damien Graux.

Knowledge graphs embeddings (KGE) are lately at the center of many artificial intelligence studies due to their applicability for solving downstream tasks, including link prediction and node classification. However, most Knowledge Graph embedding models encode, into the vector space, only the local graph structure of an entity, i.e., information of the 1-hop neighborhood. Capturing not only local graph structure but global features of entities is crucial for prediction tasks on Knowledge Graphs.

This work proposes a novel KGE method named Graph Feature Attentive Neural Network (GFA-NN) that computes graphical features of entities. As a consequence, the resulting embeddings are attentive to two types of global network features. First, the relative centrality of nodes is based on the observation that some of the entities are more "prominent" than the others. Second, comes the relative position of

entities in the graph. GFA-NN computes several centrality values per entity, generates a random set of reference nodes' entities, and computes a given entity's shortest path to each entity in the reference set. It then learns this information through optimization of objectives specified on each of these features.

We investigate GFA-NN on several link prediction benchmarks in the inductive and transductive setting and show that GFA-NN achieves on-par or better results than state-of-the-art KGE solutions. We presented our findings at ECML-PKDD [68].

7.4.16 RDF Mining

Participants: Ali Ballout, Rémi Felin, Thu Huong Nguyen, Andrea Tettamanzi.

In the wake of Nguyen Thu Huong's thesis [19], which was successfully defended at the beginning of July, a new PhD student, Rémi Felin, was hired after an internship, during which he refactored existing software developed within this project and applied it to discover OWL class disjointness axioms involving complex class expressions [39].

On the other hand, our evolutionary approach critically relies on (candidate) axiom scoring. In practice, testing an axiom boils down to computing an acceptability score, measuring the extent to which the axiom is compatible with the recorded facts. Methods to approximate the semantics of given types of axioms have been thoroughly investigated in the last decade, but a promising alternative to their direct computation is to train a surrogate model on a sample of candidate axioms for which the score is already available, to learn to predict the score of a novel, unseen candidate axiom. This is the main objective Ali Ballout's thesis, whose first results are forthcoming.

7.4.17 Capturing Geospatial Knowledge from Real-Estate Advertisements

Participants: Lucie Cadorel, Andrea Tettamanzi.

In the framework of a CIFRE thesis with Kinaxia, we have proposed a workflow to extract geographic and spatial entities based on a BiLSTM-CRF architecture with a concatenation of several text representations and to extract spatial relations, to build a structured Geospatial knowledge base. This pipeline has been applied it to the case of French housing advertisements, which generally provide information about a property's location and neighbourhood. Our results show that the workflow tackles French language and the variability and irregularity of housing advertisements, generalizes Geoparsing to all geographic and spatial terms, and successfully retrieves most of the relationships between entities from the text [35].

7.4.18 ISSA: semantic indexing of scientific articles and advanced services

Participants: Franck Michel, Marco Winckler, Anna Bobasheva, Olivier Corby.

Project ISSA started in October 2020. In 2021 we have recruited a part-time engineer and supervised a 6-month Master 2 intership. We have set up a framework for the semantic indexing of the scientific publications from Agritrop, Cirad's scientific archive. Indexing is performed on two levels: thematic and geographic descriptors characterizing the article as a whole, and named entities extracted from the articles text. Descriptors and named entities are linked with reference knowlegde bases such as Wikidata, DBpedia, Geonames, and Agrovoc. We have deployed a pipeline to transform the outcome of this indexing phase to produce a knowledge graph that shall be made public in the following of the project in 2022. We have also prototyped basic visualization services that render the entities and descriptors extracted during the indexing process.

7.4.19 IndeGx: A Model and a Framework for Indexing Linked Datasets and theirKnowledge Graphs with SPARQL-based Test Suits

Participants:	Pierre Maillot, Olivier Corby, Catherine Faron, Fabien Gandon, Franck Michel.

The joint exploitation of RDF datasets relies on the knowledge of their content, of their endpoints, and of what they have in common. Yet, not every dataset contains a self-description, and not every endpoint can handle the complex queries used to generate such a description.

As part of the ANR DeKaloG, we proposed a standard-based approach to generate the description of a dataset. The description generated as well as the process of its computation are expressed using standard vocabularies and languages. We have implemented our approach into a framework, called IndeGx, to automatically generate the description of datasets and endpoints and collect them in an index. We have experimented IndeGx on a set of 197 active knowledge bases.

Several visualisations wer also generated from IndeGx and are available online: IndeGx Web Site.

7.4.20 Evaluation of Explanations for Relational Graph Convolutional Network Link Prediction on Knowledge Graph

Participants:	Nicholas Halliwell, Fabien Gandon, Freddy Leccue.

This collaboration is considering the fundamental problem of evaluating the quality of explanations generated by some of the latest AI systems[48].

We first proposed a rule-based and ontology based generator for a simplified benchmark focusing on providing non-ambiguous explanations for knowledge graph link prediction using Relational Graph Convolutional Networks (RGCN) [50, 49, 51].

We then proposed an extended method to support user-scored evaluation of non-unique explanations for link prediction by RGCNs integrating the constraint of having multiple possible explanations for a prediction of different value for a user [52].

7.4.21 Extending electronic medical records vector models with knowledge graphs to improve hospitalization prediction

Participants:	Raphael Gazzoti, Catherine Faron, Fabien Gandon.

We proposed to address the problem of hospitalization prediction for patients with an approach that enriches vector representation of EMRs with information extracted from different knowledge graphs before learning and predicting. In addition, we performed an automatic selection of features resulting from knowledge graphs to distinguish noisy ones from those that can benefit the decision making. We evaluted our results with experiments on the PRIMEGE PACA database that contains more than 600,000 consultations carried out by 17 general practitioners (GPs). A statistical evaluation shows that our proposed approach improves hospitalization prediction. More precisely, injecting features extracted from cross-domain knowledge graphs in the vector representation of EMRs given as input to the prediction algorithm significantly increases the F1 score of the prediction. By injecting knowledge from recognized reference sources into the representation of EMRs, it is possible to significantly improve the prediction of medical events. [29]

8 Bilateral contracts and grants with industry

8.1 Bilateral contracts with industry

PREMISSE Collaborative Project

Participants:	Molka Dhouib, Catherine Faron, Andrea Tettamanzi.

Partner: SILEX France.

This collaborative project with the SILEX France company started in March 2017, funded by the ANRT (CIFRE PhD). It ended in December 2021. SILEX France is developing a B2B platform where service providers and consumers upload their service offers or requests in free natural language; the platform is intended to recommend service providers to the applicant, which are likely to fit his/her service request. The aim of this project was to propose a decision support system by exploiting the semantic knowledge that is extracted from the textual descriptions of requests for services and providers, in order to recommend relevant providers for a service request.

HealthPredict Collaborative Project

Participants:	Raphaël Gazzotti, Catherine Faron, Fabien Gandon.

Partner: Synchronext.

This collaborative project with the Synchronext company started in April 2017, funded by the ANRT (CIFRE PhD). It ended in April 2021. Synchronext is a startup aiming at developing Semantic Web business solutions. The aim of this project was to design a digital health solution for the early management of patients through consultations with their general practitioner and health care circuit. The goal was to develop a predictive Artificial Intelligence interface that allows to cross the data of symptoms, diagnosis and medical treatments of the population in real time to predict the hospitalization of a patient. [29, 10]

Curiosity Collaborative Project

Participants:	Catherine Faron, Oscar Rodríguez Rocha, Molka Dhouib.

Partner: TeachOnMars.

This collaborative project with the TeachOnMars company started in October 2019. TeachOnMars is developping a platform for mobile learning. The aim of this project is to develop an approach for automatically indexing and semantically annotating heterogeneous pedagogical resources from different sources to build up a knowledge graph enabling to compute training paths, that correspond to the learner's needs and learning objectives.

CIFRE Contract with Doriane

Participants:	Andrea Tettamanzi, Rony Dupuy Charles.

Partner: Doriane.

This collaborative contract for the supervision of a CIFRE doctoral scholarship, relevant to the PhD of Rony Duput Charles, is part of Doriane's Fluidity Project (Generalized Experiment Management), the feasibility phase of which has been approved by the Terralia cluster and financed by the Région Sud-Provence Alpes Côte d'Azur and BPI France in March 2019. The objective of the thesis is to develop machine learning methods for the field of agro-vegetation-environment. To do so, this research work will take into account and address the specificities of the problem, i.e. data with mainly numerical characteristics, scalability of the study object, small data, availability of codified background knowledge,

need to take into account the economic stakes of decisions, etc., as explained in the section on the context of the project. To enable the exploitation of ontological resources, the combination of symbolic and connective approaches will be studied, among others. Such resources can be used, on the one hand, to enrich the available datasets and, on the other hand, to restrict the search space of predictive models and better target learning methods.

The PhD student will develop original methods for the integration of background knowledge in the process of building predictive models and for the explicit consideration of uncertainty in the field of agro-plant environment.

CIFRE Contract with Kinaxia

> **Participants:** Andrea Tettamanzi, Lucie Cadorel.

Partner: Kinaxia.

This thesis project is part of a collaboration with Kinaxia that began in 2017 with the Incertimmo project. The main theme of this project was the consideration of uncertainty for a spatial modeling of real estate values in the city. It involved the computer scientists of the Laboratory and the geographers of the ESPACE Laboratory. It allowed the development of an innovative methodological protocol to create a mapping of real estate values in the city, integrating fine-grained spatiality (the street section), a rigorous treatment of the uncertainty of knowledge, and the fusion of multi-source (with varying degrees of reliability) and multi-scale (parcel, street, neighbourhood) data.

This protocol was applied to the Nice-Côte d'Azur metropolitan area case study, serving as a test bed for application to other metropolitan areas.

The objective of this thesis, which will be carried out by Lucie Cadorel with the advice of Andrea Tettamanzi, is, on the one hand, to study and adapt the application of methods for extracting knowledge from texts (or text mining) to the specific case of real estate ads written in French, before extending them to other languages, and, on the other hand, to develop a methodological framework that makes it possible to detect, explicitly qualify, quantify and, if possible, reduce the uncertainty of the extracted information, in order to make it possible to use it in a processing chain that is finalized for recommendation or decision making, while guaranteeing the reliability of the results.

8.2 Bilateral grants with industry

> **Participants:** Freddy Lecue, Fabien Gandon, Nicholas Halliwell.

Partner: Accenture.

Accenture gifts (June 2017 - January 2022): Wimmics has received two gifts from Accenture. Together with additional funds from another project these gifts have been used to fund the Engineer position and then the PhD Grant (June 2017 - January 2022) of Nicholas Halliwell on a topic agreed with Accenture: "Interpretable and explainable predictions"

9 Partnerships and cooperations

9.1 International initiatives

9.1.1 Associate Teams in the framework of an Inria International Lab or in the framework of an Inria International Program

PROTEMICS

Title: Protégé and SHACL extension to support ontology validation

Duration: 2020 -> 2024

Coordinator: Fabien Gandon

Partners:

- School of Computing, Stanford (United States)

Inria contact: Fabien Gandon

Summary: We propose to investigate the extension of the structure-oriented SHACL validation to include more semantics, and to support ontology validation and the modularity and reusability of the associated constraints. Where classical Logical (OWL) schema validation focuses on checking the semantic coherence of the ontology, we propose to explore a language to capture ontology design patterns as extended SHACL shapes organized in modular libraries. The overall objective of our proposed work is to augment the Protégé editor with fundamental querying and reasoning capabilities provided by CORESE, in order to assist ontology developers in performing ontology quality assurance throughout the life-cycle of their ontologies PROTEMICS is an associate team, SHACL-S is an Exploratory Action (AEx) and CoP4Pro is a Development Action (ADT) and these three complementary projects are adressing the research, collaboration and development aspects of the same topic.

9.1.2 Participation in other International Programs

NOMOS

Title: A Model-Based Approach for Designing Territorial User Interfaces

duration : 2020-2021

Coordinator : Marco Winckler (France) and Jean Vanderdonckt (Belgium)

partners : Université Côte d'Azur and Universté catholique de Louvain-la-Neuve

Contact : Marco Winckler

Summary : NOMOS (the French acronym for Nouvelle Organisation de Modèles Orientés Surfaces pour la conception de systèmes de systèmes interactifs basés sur la territorialité) is an international cooperation project funded by the program Tournesol. The research questions of NOMOS are articulated around the developement of a model-based approach for designing graphical user interfaces that are delineated based on the concept of territoriality. A territorial user interface is referred to as the set of interaction and physical surfaces, considered as parts or wholes, owned by a user involved in a dynamically-changing group collaboration in a given environment. For this purpose, we investigate five models covering the domain, the collaborative tasks, the users and the roles that play in the collaboration, the interaction surfaces involved in the collaboration, and the environment in which the collaboration takes places. For each model, intra-model relationships characterize static and dynamic relations. Across models, inter-model relationships dynamically map respective concepts.

9.2 European initiatives

9.2.1 FP7 & H2020 projects

AI4EU

Title: A European AI On Demand Platform and Ecosystem

Duration: *2019 - 2021*

Coordinator: THALES

Partners:

- AGENCIA ESTATAL CONSEJO SUPERIOR DEINVESTIGACIONES CIENTIFICAS (Spain)
- ALMA MATER STUDIORUM - UNIVERSITA DI BOLOGNA (Italy)
- ARISTOTELIO PANEPISTIMIO THESSALONIKIS (Greece)
- ASSOCIACAO DO INSTITUTO SUPERIOR TECNICO PARA A INVESTIGACAO E DESENVOLVI-MENTO (Portugal)
- BARCELONA SUPERCOMPUTING CENTER - CENTRO NACIONAL DE SUPERCOMPUTA-CION (Spain)
- BLUMORPHO SAS (France)
- BUDAPESTI MUSZAKI ES GAZDASAGTUDOMANYI EGYETEM (Hungary)
- BUREAU DE RECHERCHES GEOLOGIQUES ET MINIERES (France)
- CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE CNRS (France)
- CINECA CONSORZIO INTERUNIVERSITARIO (Italy)
- COMMISSARIAT A L ENERGIE ATOMIQUE ET AUX ENERGIES ALTERNATIVES (France)
- CONSIGLIO NAZIONALE DELLE RICERCHE (Italy)
- DEUTSCHES FORSCHUNGSZENTRUM FUR KUNSTLICHE INTELLIGENZ GMBH (Germany)
- DEUTSCHES ZENTRUM FUR LUFT - UND RAUMFAHRT EV (Germany)
- EOTVOS LORAND TUDOMANYEGYETEM (Hungary)
- ETHNIKO KAI KAPODISTRIAKO PANEPISTIMIO ATHINON (Grecce)
- ETHNIKO KENTRO EREVNAS KAI TECHNOLOGIKIS ANAPTYXIS (Greece)
- EUROPEAN ORGANISATION FOR SECURITY (Belgium)
- FONDATION DE L'INSTITUT DE RECHERCHE IDIAP (Switzerland)
- FONDAZIONE BRUNO KESSLER (Italy)
- FORUM VIRIUM HELSINKI OY (Finland)
- FRANCE DIGITALE (France)
- FRAUNHOFER GESELLSCHAFT ZUR FOERDERUNG DER ANGEWANDTEN FORSCHUNG E.V. (Germany)
- FUNDACION CARTIF (Spain)
- FUNDINGBOX ACCELERATOR SP ZOO (Poland)
- FUNDINGBOX RESEARCH APS (Denmark)
- GOODAI RESEARCH SRO (Czech Republic)
- Hochschule für Technik und Wirtschaft Berlin (Germany)
- IDRYMA TECHNOLOGIAS KAI EREVNAS (Greece)
- IMT TRANSFERT (France)
- INSTITUT JOZEF STEFAN (Slovenia)
- INSTITUT POLYTECHNIQUE DE GRENOBLE (France)
- INTERNATIONAL DATA SPACES EV (Germany)
- KARLSRUHER INSTITUT FUER TECHNOLOGIE (Germany)
- KNOW-CENTER GMBH RESEARCH CENTER FOR DATA-DRIVEN BUSINESS & BIG DATA ANALYTICS (Austria)
- NATIONAL CENTER FOR SCIENTIFIC RESEARCH "DEMOKRITOS" (Greece)
- NATIONAL UNIVERSITY OF IRELAND GALWAY (Ireland)

- NORGES TEKNISK-NATURVITENSKAPELIGE UNIVERSITET NTNU (Norway)
- OFFICE NATIONAL D'ETUDES ET DE RECHERCHES AEROSPATIALES (France)
- ORANGE SA (France)
- OREBRO UNIVERSITY (Sweden)
- QWANT (France)
- TECHNICKA UNIVERZITA V KOSICIACH (Slovakia)
- TECHNISCHE UNIVERSITAET MUENCHEN (Germany)
- TECHNISCHE UNIVERSITAET WIEN (Austria)
- TECHNISCHE UNIVERSITAT BERLIN (Germany)
- THALES (France)
- THALES ALENIA SPACE FRANCE SAS (France)
- THALES SIX GTS FRANCE SAS (France)
- THOMSON LICENSING (France)
- TILDE SIA (Latvia)
- TWENTY COMMUNICATIONS SRO (Slovakia)
- UNIVERSIDAD POLITECNICA DE MADRID (Spain)
- UNIVERSIDADE DE COIMBRA (Portugal)
- UNIVERSITA CA' FOSCARI VENEZIA (Italy)
- UNIVERSITA DEGLI STUDI DI SIENA (Italy)
- UNIVERSITAT POLITECNICA DE CATALUNYA (Spain)
- UNIVERSITE DE LORRAINE (France)
- UNIVERSITE GRENOBLE ALPES (France)
- UNIVERSITY COLLEGE CORK - NATIONAL UNIVERSITY OF IRELAND, CORK (Ireland)
- UNIVERSITY OF LEEDS (UK)
- VRIJE UNIVERSITEIT BRUSSEL (Belgium)
- WAVESTONE (France)
- WAVESTONE ADVISORS (France)
- WAVESTONE LUXEMBOURG SA (Luxembourg)

Inria contact: Olivier Corby (for Wimmics)

Summary: In January 2019, the AI4EU consortium was established to build the first European Artificial Intelligence On-Demand Platform and Ecosystem with the support of the European Commission under the H2020 programme. The activities of the AI4EU project include:

- The creation and support of a large European ecosystem spanning the 28 countries to facilitate collaboration between all Europeans actors in AI (scientists, entrepreneurs, SMEs, Industries, funding organizations, citizens…);
- The design of a European AI on-Demand Platform to support this ecosystem and share AI resources produced in European projects, including high-level services, expertise in AI research and innovation, AI components and datasets, high-powered computing resources and access to seed funding for innovative projects using the platform;
- The implementation of industry-led pilots through the AI4EU platform, which demonstrates the capabilities of the platform to enable real applications and foster innovation;
- Research activities in five key interconnected AI scientific areas (Explainable AI, Physical AI, Verifiable AI, Collaborative AI, Integrative AI), which arise from the application of AI in real-world scenarios;

- The funding of SMEs and start-ups benefitting from AI resources available on the platform (cascade funding plan of €3m) to solve AI challenges and promote new solutions with AI;

- The creation of a European Ethical Observatory to ensure that European AI projects adhere to high ethical, legal, and socio-economical standards; click here to know more

- The production of a comprehensive Strategic Research Innovation Agenda for Europe

- The establishment of an AI4EU Foundation that will ensure a handover of the platform in a sustainable structure that supports the European AI community in the long run.

In the context of the AI4EU European project, we have translated the Thales Knowledge Graph into an RDF graph and have defined a set of SPARQL queries to query and navigate the graph. This has been integrated into the AI4EU endpoint prototype.

Web site: AI4EU Project

AI4Media

Title: AI4Media

Duration: 2020 - 2024

Coordinator: The Centre for Research and Technology Hellas (CERTH)

Partners:
   see consortium

Inria contact: through 3IA

Summary: AI4Media is a 4-year-long project. Funded under the European Union's Horizon 2020 research and innovation programme, the project aspires to become a Centre of Excellence engaging a wide network of researchers across Europe and beyond, focusing on delivering the next generation of core AI advances and training to serve the Media sector, while ensuring that the European values of ethical and trustworthy AI are embedded in future AI deployments. AI4Media is composed of 30 leading partners in the areas of AI and media (9 Universities, 9 Research Centres, 12 industrial organisations) and a large pool of associate members, that will establish the networking infrastructure to bring together the currently fragmented European AI landscape in the field of media, and foster deeper and long-running interactions between academia and industry.

9.2.2 Other european programs/initiatives

HyperAgents - SNSF/ANR project

Title: HyperAgents

Duration: 2020 - 2024

Coordinator: Olivier Boissier, MINES Saint-Étienne

Partners:

- MINES Saint-Étienne (FR)
- Inria (FR)
- Univ. of St. Gallen (HSG, Switzerland)

Inria contact: Fabien Gandon

Summary: The HyperAgents project, Hypermedia Communities of People and Autonomous Agents, aims to enable the deployment of world-wide hybrid communities of people and autonomous agents on the Web. For this purpose, HyperAgents defines a new class of multi-agent systems that use hypermedia as a general mechanism for uniform interaction. To undertake this investigation, the project consortium brings together internationally recognized researchers actively contributing to research on autonomous agents and MAS, the Web architecture, Semantic Web, and to the standardization of the Web. Project Web site: HyperAgents Project

ANTIDOTE - CHIST-ERA project

Title: ANTIDOTE

Duration: 2020 - 2024

Coordinator: Elena Cabrio, Serena Villata

Partners:

- University of the Cote d'Azur (Wimmics Team)
- Fondazione Bruno Kessler (IT)
- Universitu of the Basque Country (ES)
- University of Leuven (Belgium)
- University of Lisbon (PT)

Summary: Providing high quality explanations for AI predictions based on machine learning requires to combine several interrelated aspects, including, among the others: selecting a proper level of generality/specificity of the explanation, considering assumptions about the familiarity of the explanation beneficiary with the AI task under consideration, referring to specific elements that have contributed to the decision, making use of additional knowledge (e.g. metadata) which might not be part of the prediction process, selecting appropriate examples, providing evidences supporting negative hypothesis, and the capacity to formulate the explanation in a clearly interpretable, and possibly convincing, way. According to the above considerations, ANTIDOTE fosters an integrated vision of explainable AI, where low level characteristics of the deep learning process are combined with higher level schemas proper of the human argumentation capacity. ANTIDOTE will exploit cross-disciplinary competences in three areas, i.e. deep learning, argumentation and interactivity, to support a broader and innovative view of explainable AI. Although we envision a general integrated approach to explainable AI, we will focus on a number of deep learning tasks in the medical domain, where the need for high quality explanations, both to clinicians and to patients, is perhaps more critical than in other domains. Project Web site: Antidote Project

9.3   National initiatives
PIA GDN ANSWER

Participants: Fabien Gandon, Hai Huang, Vorakit Vorakitphan, Serena Villata, Elena Cabrio.

ANSWER stands for Advanced aNd Secured Web Experience and seaRch. It is a GDN project (Grands Défis du Numérique) from the PIA program (Programme d'Investissements d'Avenir) on Big Data. The project is between four Inria research teams and the Qwant company.

The aim of the ANSWER project is to develop the new version of the Qwant search engine by introducing radical innovations in terms of search criteria as well as indexed content and users' privacy.

The purpose is to strengthen everyone's confidence in the search engine and increase the effectiveness of Web search. Building trust in the search engine is based on innovations in (1) Security: computer

security, privacy; (2) Completeness: completeness and heterogeneity of (re)sources; and (3) Neutrality: analysis, extraction, indexing, and classification of data.

Increasing the effectiveness of Web-based research relies on innovations related to (1) Relevance: variety and value of content taken into account, measurement of emotions carried by query results; (2) Interaction with the user: adaptation of the interfaces to the types of research; and (3) Performance: perceived relevance of results and response time.

The proposed innovations include:

- Design and develop models and tools for the detection of emotions in query results:

  - Ontology, thesaurus, linguistic resources
  - Metrics, indicators, classification of emotions

- Design and develop new crawling algorithms:

  - Dynamic crawling strategies
  - Crawlers and indexes for linked open data

- Ensure respect for privacy:

  - Detection of Internet tracking
  - Preventive display of tracing techniques
  - Certified security of automatic adaptation of ads to keywords entered by the user

Ministry of Culture: MonaLIA 3.0

Participants: Anna Bobasheva, Fabien Gandon, Frédéric Precioso.

The objective of the MonaLIA project is to exploit the crossover of the automatic learning methods particularly applied to image analysis and knowledge-based representation and reasoning, in particular for the semantic indexing of annotated works and images in JocondeLab. The goal is to identify automated or semi-automatable tasks to improve the annotation. This project follows the preliminary project "MonaLIA 1" which established the state of the art in order to evaluate the potential and the combination of learning (notably deep learning) and the semantization of annotations on the case of JocondeLab. In the project MonaLIA we now want to go beyond the preliminary study and to design and build a prototype and the methods assisting the creation, the improvement, and the maintenance of the metadata of the image database in order to assist the actors of the cultural world in their daily tasks. The preliminary study identified several possible coupling points between deep learning from non-necessarily structured data and reasoning from structured data. This project proposes to select the most promising of them to carry out a proof of concept combining these methods by focusing on the assistance to the annotation and curation tasks of the metadata of a real base to improve the contents, the course and exploitation thereafter.

ANR WASABI

Participants: Michel Buffa, Elena Cabrio, Catherine Faron, Alain Giboin.

The ANR project WASABI started in January 2017 with IRCAM, Deezer, Radio France and the SME Parisson, consists in building a 2 million songs knowledge base of commercial popular music (rock, pop, etc.). Its originality is the joint use of audio-based music information extraction algorithms, song lyrics analysis algorithms (natural language processing), and the use of the Semantic Web. Web Audio technologies will then explore these bases of musical knowledge by providing innovative applications for composers, musicologists, music schools, and sound engineers, music broadcasters and journalists. This

project is in its mid-execution and gave birth to many publications in international conferences as well as some mainstream coverage (i.e for "la fête de la Science"). Participation in the ANR OpenMiage project aimed at offering online Bachelor and Master degrees.

The project also led to industrial transfer of some of the results (partnership with AmpedStudio.com/Amp Track company) for the integration of our software into theirs), SATT PACA.

Web site: Wasabi HomePage

ANR SIDES 3.0

Participants:	Catherine Faron, Olivier Corby, Antonia Ettore, Anna Bobasheva, Fabien Gandon, Alain Giboin, Franck Michel.

Partners: Université Grenoble Alpes, Inria, Ecole Normale Supérieure de Lyon, Viseo, Theia.

SIDES 3.0 is an ANR project which started in fall 2017. It is led by Université Grenoble Alpes (UGA) and its general objective is to introduce semantics within the existing SIDES educational platform [7] for medicine students, in order to provide them with added value educational services. Within this project we are developing an approach to predict the success of students on training quizzes based on the knowledge graph representing their interactions with the pedagogical resources within the SIDES platform.

Web site: SIDES 3.0 Project

ANR D2KAB

Participants:	Olivier Corby, Catherine Faron, Fabien Gandon, Franck Michel, Nadia Yacoubi Ayadi.

Partners: LIRMM, INRAE, IRD, ACTA

D2KAB is an ANR project which started in June 2019, led by the LIRMM laboratory (UMR 5506). Its general objective is to create a framework to turn agronomy and biodiversity data into knowledge –semantically described, interoperable, actionable, open– and investigate scientific methods and tools to exploit this knowledge for applications in science and agriculture. Within this project the Wimmics team is contributing to the lifting of heterogeneous dataset related to agronomy coming from the different partners of the project and is responsible to develop a unique entry point with semantic querying and navigation services providing a unified view on the lifted data.

Web site: D2KAB Project

ANR DeKaloG

Participants:	Olivier Corby, Catherine Faron, Fabien Gandon, Pierre Maillot, Franck Michel.

Partners: Université Nantes, INSA Lyon, INRIA Sophia Antipolis-Méditerranée

DeKaloG (Decentralized Knowledge Graphs) aims to: (1) propose a model to provide fair access policies to KGs without quota while ensuring complete answers to any query. Such property is crucial for enabling web automation, i.e. to allow agents or bots to interact with KGs. Preliminary results on web preemption open such perspective, but scalability issues remain; (2) propose models for capturing different levels of transparency, a method to query them efficiently, and especially, techniques to enable web automation of transparency. (3) propose a sustainable index for achieving the findability principle.

DeKaloG Web site

---

[7]SIDES Platform

DBpedia.fr

Participants:	Fabien Gandon, Franck Michel, Célian Ringwald.

The DBpedia.fr project ensures the creation and maintenance of a French chapter of the DBpedia knowledge base. This project was the first project of the Semanticpedia convention signed by the Ministry of Culture, the Wikimedia foundation and Inria.

A new project proposal was selected in 2021 between Inria and the Ministry of Culture to support evolutions and long-term sustaining of this project.

Convention between Inria and the Ministry of Culture

Participant:	Fabien Gandon.

We supervise the research convention with the Ministry of Culture to foster research and development at the crossroad of culture and digital sciences. This convention signed between Inria and the Ministry of Culture provides a framework to support projects at the crossroad of the cultural domain and the digital sciences.

Qwant-Inria Joint Laboratory

Participant:	Fabien Gandon.

We supervise the Qwant-Inria Joint Laboratory where joint teams are created and funded to contribute to the search engine research and development. The motto of the joint lab is Smart Search and Privacy with five research directions:

- Crawling, Indexing, Searching

- Execution platform, privacy by design, security, ethics

- Maps and navigation

- Augmented interaction, connected objects, chatbots, personnal assistants

- Education technologies (EdTech)

We released the final, but confidential, report of the Qwant-Culture short-term project. This project aimed at identifying possibilities of exploiting the Qwant search engine to improve the search for information in the digital cultural resources of the French Ministry of Culture. Some possibilities have been selected to be the subject of research actions in the context a long-term project.

CovidOnTheWeb - Covid Inria program

Participants:	Valentin Ah-Kane, Anna Bobasheva, Lucie Cadorel, Olivier Corby, Elena Cabrio, Jean-Marie Dormoy, Fabien Gandon, Raphaël Gazzotti, Alain Giboin, Santiago Marro, Tobias Mayer, Aline Menin, Franck Michel, Andrea Tettamanzi, Serena Villata, Marco Winckler.

The Covid-On-The-Web project aims to allow biomedical researchers to access, query and make sense of COVID-19 scholarly literature. To do so, we designed and implemented a pipeline levereding our skills in knowledge representation, text mining, argument mining and visualization techniques to process, analyze and enrich the COVID-19 Open Research Dataset (CORD-19) that gathers 100,000+ full-text scientific articles related to the coronaviruses.

The generated RDF dataset comprises the Linked Data description of (1) named entities (NE) mentioned in the CORD-19 corpus and linked to DBpedia, Wikidata and other BioPortal vocabularies, and (2) arguments extracted using ACTA, a tool automating the extraction and visualization of argumentative graphs, meant to help clinicians analyze clinical trials and make decisions.

On top of this dataset, we have adapted visualization and exploration tools (MGExplorer, Arviz) to provide Linked Data visualizations that meet the expectations of the biomedical community.
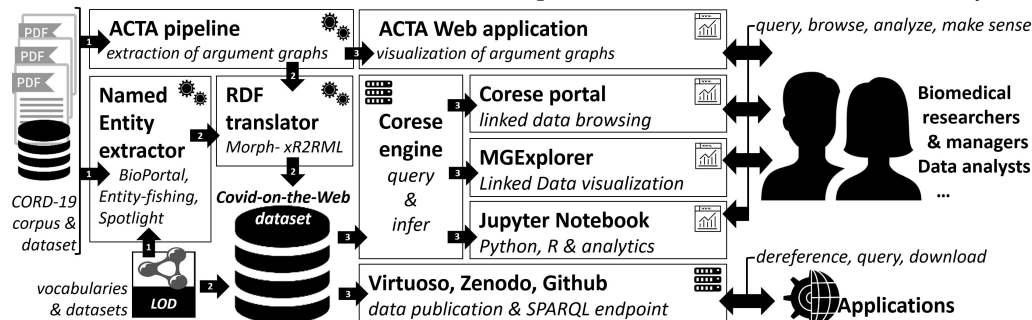
ISSA (AAP Collex-Persée)

Participants:	Franck Michel, Marco Winckler, Anna Bobasheva, Olivier Corby.

Partners: CIRAD, Mines d'Alès

The ISSA project started in October 2020 and is led by the CIRAD. It aims to set up a framework for the semantic indexing of scientific publications with thematic and geographic keywords from terminological resources. It also intends to demonstrate the interest of this approach by developing innovative search and visualization services capable of exploiting this semantic index. Agritrop, Cirad's open publications archive, serves as a use case and proof of concept throughout the project. In this context, the primarily semantic resources are the Agrovoc thesaurus, Wikidata and GeoNames.

Wimmics team is responsible for (1) the generation and publication of the knowledge graph representing the indexed entities, and (2) the development of search/visualization tools intended for researchers and/or information

9.4 Regional initiatives

3IA Côte d'Azur

Participants:	Elena Cabrio, Catherine Faron, Fabien Gandon, Freddy Limpens, Andrea Tettamanzi, Serena Villata.

3IA Côte d'Azur is one of the four "Interdisciplinary Institutes of Artificial Intelligence"[8] that were created in France in 2019. Its ambition is to create an innovative ecosystem that is influential at the local, national and international level. The 3IA Côte d'Azur institute is led by Université Côte d'Azur in partnership with major higher education and research partners in the region of Nice and Sophia Antipolis: CNRS, Inria, INSERM, EURECOM, ParisTech MINES and SKEMA Business School. The 3IA Côte d'Azur institute is also supported by ECA, Nice University Hospital Center (CHU Nice), CSTB, CNES, Data Science Tech Institute and INRAE. The project has also secured the support of more than 62 companies and start-ups.

---

[8]3IA Côte d'Azur Web site

We have four 3IA chairs for tenured researchers of Wimmics and several grants for PhD and postdocs.

We also have an industrial 3IA Affiliate Chair with the company Mnemotix focused on the industrialisation and scalability of the CORESE software.

UCA IDEX OTESIA project "Artificial Intelligence to prevent cyberviolence and hate speech online"

Participants: Elena Cabrio, Serena Villata, Anais Ollagnier.

The project will cross the approaches of mediation / remediation: it is planned to develop a software to detect hate messages based on an analysis of natural language, but also to understand their argumentative structure (not a simple detection of isolated words, insults) and to develop the critical spirit of the victims and therefore, to do this, to develop a counter-speech. Hence interventions in 6 secondary schools for role-playing games that will serve as a basis for data collection. This data, once analysed, will support the development of software to detect hate and violent speech online. As part of a restitution of the work carried out, the institutions will participate in a collaborative manner in the development of counter-speech.

10 Dissemination

10.1 Promoting scientific activities

Member of organizing committees

Participants: Michel Buffa, Elena Cabrio, Damien Graux, Benjamin Molinet, Serena Villata, Marco Winckler.

- Damien Graux: organizer of the 7th Workshop on Managing the Evolution and Preservation of the Data Web (MEPDaW) co-located with the 20th International Semantic Web Conference (ISWC 2021), organizer of the 1st Ph.D. Workshop on Big Data Analytics from the LAMBDA Network.

- Michel Buffa and Shihong Ren : organizers of the "WebAudio Plugin" workshop at the WebAudio Conference 2022 (WAC 2022), July 5-7, 2022.

- Serena Villata: co-chair of the "Sister Conference Best Papers" track of the 30th International Joint Conference on Artificial Intelligence (IJCAI-2021).[9].

- Elena Cabrio: co-chair of the "Natural Language for Artificial Intelligence workshop" (NL4AI@AI*IA 2021).

- Benjamin Molinet: Technical staff of the Inria hackAtech 24 - 26 October 2021 event, presentation and support for the argumentative mining tool ACTA based on this work.

- Marco Winckler: co-chair of the track Demonstrations/posters for the International Conference on Web Engineering (ICWE 2021), May 18-21, 2021 - Biarritz, France. Also member of the advisory board of the (INTERACT 2021) - IFIP TC 13 International Conference on Human Computer Interaction, August 30th-September 3rd 2021 - Bari, Ital.

---

[9]IJCAI-2021 Organising Committee

10.1.1   Scientific events: selection

Participants: Michel Buffa, Elena Cabrio, Olivier Corby, Catherine Faron, Fabien Gandon, Alain Giboin, Damien Graux, Pierre Maillot, Aline Menin, Amaya Nogales Gómez, Shihong Ren, Andrea Tettamanzi, Serena Villata, Marco Winckler.

Member of conference program committees

- Olivier Corby: PC member of TheWebConf 2021, ESWC 2021 (European Semantic Web Conference), ICCS 2021 (Int. Conference on Conceptual Structures), ISWC, KCAP, S4BioDiv, IC (Ingénierie des Connaissances).

- Catherine Faron: PC member of TheWebConf 2022, ESWC 2021 (European Semantic Web Conference), Semantics 2021 (Int. conf. on Semantic Systems), ICCS 2021 (Int. Conference on Conceptual Structures), S4BioDiv 2021 (Int. Workshop on Semantics for Biodiversity), IC 2021 (Ingénierie des Connaissances).

- Fabien Gandon : PC member of AAAI 2021 (Association for the Advancement of Artificial Intelligence), IJCAI 2021 (International Joint Conference on Artificial Intelligence), ISWC 2021 (International Semantic Web Conference).

- Damien Graux : PC member of Simplify 2021, SEMANTiCS 2021, ISWC 2021 (International Semantic Web Conference).

- Pierre Maillot: PC member of ISWC 2021 MEPDaW workshop (Managing the Evolution and Preservation of the Data Web), eKNOW 2021 (International Conference on Information, Process, and Knowledge Management).

- Michel Buffa : PC member of WAC 2022, ESWC 2021, W3C/SMPTE Workshop on Professional Media Production on the Web, November 2021.

- Alain Giboin: PC member of IC 2021 (Ingénierie des Connaissances), VOILA 2021 (Visualization and Interaction for Ontologies and Linked Data)

- Serena Villata: Action Editor for the ACL Rolling Review, Area Chair for "Sentiment Analysis, Stylistic Analysis and Argument Mining" at ACL-IJCNLP 2021, Area Chair at IJCAI 2021, Area chair for "Sentiment Analysis and Argument Mining" at EACL 2021, senior PC member at AAAI 2021, PC member at KR 2021.

- Elena Cabrio: Action Editor for the ACL Rolling Review, Area Chair at IJCAI 2021, senior PC member at AAAI 2021.

- Andrea Tettamanzi: ESWC 2021, Evo* 2021, UAI 2021, FUZZ-IEEE 2021, ICAART 2022, AAAI-22, SAC 2022, ESWC 2022, UAI 2022.

- Marco Winckler: PC member of ACM EICS 2021, ACM IUI 2021, Brazilian IHC'2021, HuFaMo 2021 workshop, IEEE ETFA 2021, IEEE INDIN 2021, INTERACT 2021, ICWE 2021, IFIP IOT 2021, IS-EUD 2021, ISD 2021, ManComp 2021, MIDI2021, SVR 2021.

Reviewer

- Aline Menin : reviewer for IEEE Conference on Virtual Reality and 3D User Interfaces (VR), International Conference in Human-Computer Interaction (INTERACT), and Engineering Interactive Computing Systems (EICS).

10.1.2 Journal

> **Participants:** Michel Buffa, Catherine Faron, Fabien Gandon, Alain Giboin, Aline Menin, Andrea Tettamanzi, Serena Villata, Marco Winckler.

Member of editorial boards

- Catherine Faron: board member of Revue Ouverte d'Intelligence Artificielle (ROIA)

- Serena Villata: member of the Editorial Board of the journal "Artificial Intelligence and Law".[10], member of the Editorial Board of the journal "Argument and Computation".[11], member of the Editorial Board of the journal "Journal of Web Semantics".[12]

- Marco Winckler: member of the editorial board of: Interacting with Computers (Oxford Press), (Multimodal Technologies and Interaction), Behaviour and Information Technology (Taylor and Francis), IFIP Advances in Information and Communication Technology (Springer).

Reviewer - reviewing activities

- Catherine Faron: reviewer for Journal of Web Semantics JWS, International Journal of Information Management Data Insights IJIMDI, Data Science Journal, Future Generation Computer Systems FGCS.

- Fabien Gandon : reviewer for ACM TOIT (Transactions on Internet Technology).

- Aline Menin : reviewer for Springer Virtual Reality Journal (VIRE).

- Michel Buffa: reviewer for Journal of Web Semantics JWS.

- Alain Giboin: reviewer for Le Travail Humain (a bilingual and multi-disciplinary journal in human factors)

- Andrea Tettamanzi: Knowledge-Based Systems, Retailing and Consumer Services, Transactions on Fuzzy Systems, Applied Soft Computing.

- Marco Winckler: International Journal of Human-Computer Studies - IJHCS (Elsevier), Entertainment Computing (Elsevier), FRONTIER, Journal of the Brazilian Computer Society (Springer), Knowledge-Based Systems, ACM Transcation on Human-Computer Interaction.

10.1.3 Invited talks

> **Participants:** Michel Buffa, Elena Cabrio, Olivier Corby, Fabien Gandon, Damien Graux, Franck Michel, Shihong Ren, Serena Villata.

- Olivier Corby:

  - Inrae workshop on Semantic Linked Data, Sète, 11-13 october 2021:
    * Tutorial: Semantic Web languages.
    * Talk: Extended SPARQL Service for federated queries.

- Michel Buffa and Shihong Ren:

  - "Etat de l'art sur les plugins WebAudio", Journées d'Informatique Musicale 2021 (JIM 2021)

---

[10] AI & Law journal
[11] A&C journal
[12] J. of Web Semantics

- "WebAudio Modules 2.0, a standard for WebAudio plugins", W3C TPAC conference 2022.

- Damien Graux:

  - "Hints to Save Time when Dealing with Big Data", keynote at the 1st Ph.D. Workshop on Big Data Analytics from the LAMBDA Network

- Shihong Ren:

  - Post-doctoral seminar "Intelligently generate Faust digital audio processors from block diagrams" at Shanghai Conservatory of Music (China)
  - Seminar "From Diagram to Code: a Web-based Interactive Graph Editor for Digital Sound Processing Design and Code Generation" (Inria)

- Franck Michel:

  - "Bioschemas: Marking up biodiversity websites for data discovery & integration", TDWG webinars series.

  - "Marking up biodiversity web pages with structured information to make them discoverable and reusable", Mini-symposium Open Science and Interoperability in Bioinformatics (OSIBIO), Journées Ouvertes en Biologie, Informatique et Mathématiques (JOBIM).

  - "Web Pages, Databases, Knowledge Graphs... Considering Separated Web Data Sources as a Continuum". Minisymposium Toward Semantic Integration of Biological Resources, PASC 2021 conference

- Fabien Gandon:

  - Keynote at ENDORSE 2021 on "Web open standards for linked data and knowledge graphs as enablers of EU digital sovereignty", 19/03/2021
  - short talk at online workshop ADEME-Inria "demain l'IA" 18/06/2021
  - short talk at online workshop CWI-Inria " Data and Intelligent Systems" 05/07/2021
  - short talk at online workshop La Poste-Inria "SOLID" 13/01/2021
  - short talk 3IA Côte d'Azur days 30/11/2021

- Serena Villata:

  - Keynote speaker at the 3rd International Workshop on Argument Strength on "Towards assessing natural language argument strength: results and open challenges", October 11th-13th, 2021, online event.

  - Invited speaker at the 2nd International Workshop on Deceptive AI (co-located with IJCAI-2021) on "Deceptive Argumentation: identification, reasoning and ethical challenges", August 19th, 2021, online event.

  - Keynote speaker of the EuropeaN Data conference On Reference data and SEmantics (ENDORSE-2021): "Key findings and future challenges in AI & Law", invited by the Publications Office of the European Union, March 17, 2021, online event.

  - Invited lecture (advanced course) at the Institut d'Automne en Intelligence Artificielle (IA2 2021) on "Argumentation, natural language, and explanation", September 27th, 2021, Paris.

  - Invited lecture at the 1st International Munich Legal Tech Summer School on "Argument Mining", August 4th, 2021, online event.

  - Invited lecture at the SKEMA Hakathon 2021 on "Algorithmic bias", September 13th-17th, 2021.

  - Invited talk at the APPLY TV on "Towards argument-based explanatory dialogues: from argument mining to (explanatory) argument generation", April 28th, 2021, online event.

- – Invited talk at the CLORA webinair on "New research in the fight against disinformation", June 7th, 2021.

- Elena Cabrio:

  - – Invited speaker at third edition of the Symposium MaDICS on "L'intelligence artificielle au service de la prévention de la cyberviolence, du cyberharcèlement et de la haine en ligne", July 5-8 2021, online event.

10.1.4   Leadership within the scientific community

Participants: Fabien Gandon, Marco Winckler.

Fabien Gandon is a member of Semantic Web Science Association (SWSA) a non-profit organisation for promotion and exchange of the scholarly work in Semantic Web and related fields throughout the world ans steering committee of the ISWC conference.

Marco Winckler is French representative at the IFIP TC13 and member of the Association Francophone pour l'Interaction Homme-Machine.

10.1.5   Scientific expertise

Participants: Michel Buffa, Catherine Faron, Fabien Gandon.

- Catherine Faron: member of the ANR scientific evaluation committe "Artificial Intelligence" (CE23) ; reviewer for the HAISCoDe programme of Normandie region; scientific referent of the Inria Learning Lab.

- Michel Buffa : member of the scientific council of the GRAME laboratory (Lyon).

- Fabien Gandon is member of the Choose France committee ; member of the evaluation committee for 3IA Côte d'Azur chairs ; member of the jury DR0 Inria.

10.1.6   Research administration

Participants: Elena Cabrio, Olivier Corby, Catherine Faron, Fabien Gandon, Andrea Tettamanzi, Serena Villata, Marco Winckler.

- Olivier Corby: member of the working group on environmental issue at I3S laboratory.

- Catherine Faron: board member of the French Society for Artificial Intelligence (AFIA) ; member of the steering committee of the AFIA college on Knowledge Engineering.

- Fabien Gandon : Leader of the Wimmics team ; Vice-director of Research Inria Sophia Antipolis ; co-president of scientific and pedagogical council of the Data Science Technical Institure (DSTI) ; vice-president of the Jury for CRCN-ISFP Inria Côte d'Azur 2021 contest ; Supervision of evaluation of two teams in HPC and DIG domains of Inria ; Member of the Evaluation Committee of Inria ; W3C Advisory Committee Representative (AC Rep) for Inria ; Leader of the convention Inria - Ministry of Culture ; Leader of Workgroup for the evaluation of the team proposal BOREAL.

- Serena Villata: Deputy Scientific Director of 3IA Côte d'Azur Institute.

- Elena Cabrio is member of the Conseil d'Administration (CA) of the French Association of Computational Linguistics (ATALA) ; Member of the Bureau of the Académie 1 of IDEX UCA JEDI.

- Andrea Tettamanzi is the leader of the SPARKS team at I3S laboratory.

- Marco Winckler is joint-director of the SPARKS team at the I3S Laboratory, Secretary for the IFIP TC13 on Human-Computer Interaction, member of the Steering Committee of INTERACT.

10.2   Teaching - Supervision - Juries

10.2.1   Teaching

Participants: Michel Buffa, Elena Cabrio, Olivier Corby, Catherine Faron, Fabien Gandon, Damien Graux, Aline Menin, Amaya Nogales Gómez, Andrea Tettamanzi, Serena Villata, Marco Winckler.

- Michel Buffa:
  - Licence 3, Master 1, Master 2 Méthodes Informatiques Appliquées à la Gestion des Entreprises (MIAGE) : Web Technologies, Web Components, etc. 192h.
  - DS4H Masters 3D games programming on Web, JavaScript Introduction: 40h.

- Olivier Corby:
  - Licence 3 IOTA UCA 25 hours Semantic Web
  - Licence 3 IA DS4H UCA 25 hours Semantic Web

- Catherine Faron :
  - Master 2/5A SI PNS: Web of Data, 32 h
  - Master 2/5A SI PNS: Semantic Web 32h
  - Master 2/5A SI PNS: Ingénierie des connaissances 15h
  - Master DSAI UCA: Web of Data, 30h
  - Master 1/4A SI PNS and Master2 IMAFA/5A MAM PNS: Web languages, 28h
  - Licence 3/3A SI PNS and Master 1/4A MAM PNS: Relational Databases, 60h
  - Master DSTI: Data pipeline, 50h.

- Fabien Gandon : Master: Integrating Semantic Web technologies in Data Science developments, 78 h, M2, DSTI, France.

- Aline Menin :
  - Master 1, Data Sciences & Artificial Intelligence, UCA, 15h (CM/TP), Data visualization.
  - Master 1, Mobiliquité, Big Data et intégration de systèmes, UCA, 9h (CM/TP), Data visualization.
  - Master 1, Méthodes Informatiques Appliquées à la Gestion des Entreprises (MIAGE), UCA, 12h (TP), Javascript programming.
  - Licence 3, Licence MATHÉMATIQUES ET INFORMATIQUE APPLIQUÉES AUX SCIENCES HUMAINES ET SOCIALES Parcours Méthodes Informatiques Appliquées à la Gestion des Entreprises (MIASHS/MIAGE), UCA, 48h (CM/TD/TP), Human-computer interaction.

- Amaya Nogales Gómez:
  - Master 1, Data Sciences & Artificial Intelligence, UCA, 20h (CM/TD), Security and Ethical Aspects of Data.
  - Licence 2, Licence Informatique, UCA, 36h (TP), Structures de données et programmation C.

- Serena Villata:

- Master II Droit de la Création et du Numérique - Sorbonne University: Approche de l'Elaboration et du Fonctionnement des Logiciels, 15 hours (CM), 20 students.

- Master 2 MIAGE IA - University Côte d'Azur: I.A. et Langage : Traitement automatique du langage naturel, 28 hours (CM+TP), 30 students.

- Master Communication et Langage Politique - University Côte d'Azur: Argumentation, 15 hours (CM+TD), 10 students.

- Elena Cabrio:

  - Master I Computer Science, Text Processing in AI, e. 30 hours (eq. TD).

  - Master 2 MIAGE IA - University Côte d'Azur: I.A. et Langage : Traitement automatique du langage naturel, 28 hours (CM+TP), 30 students.

  - Master 1 EUR CREATES, Parcours Linguistique, traitements informatiques du texte et processus cognitifs. Introduction to Computational Linguistics, 30 hours.

  - Master 1 EUR CREATES, Parcours Linguistique, traitements informatiques du texte et processus cognitifs. Textual Data Analysis, 30 hours.

  - Master Modeling for Neuronal and Cognitive Systems. Text analysis, deep learning and statistics, 18.5 hours.

  - License 2, Computer Science. Web Technologies, 54 hours.

- Andrea Tettamanzi

  - Licence: Andrea Tettamanzi, Introduction à l'Intelligence Artificielle, 45 h ETD, L2, UCA, France.

  - Master: Andrea Tettamanzi, Logic for AI, 30 h ETD, M1, UCA, France.

  - Master: Andrea Tettamanzi, Web, 30 h ETD, M1, UCA, France.

  - Master: Andrea Tettamanzi, Algorithmes Évolutionnaires, 24.5 h ETD, M2, UCA, France.

  - Master: Andrea Tettamanzi, Modélisation del l'Incertitude, 24.5 h ETD, M2, UCA, France.

- Marco Winckler

  - Licence 3: Introduction to Human-Computer Interaction, 45 h ETD, UCA, Polytech Nice, France.

  - Master 1: Accessibility and Universal Design, 10 h ETD, UCA, DS4H, France.

  - Master 1: Methods and tools for technical and scientific writing, Master DSAI, 15 h ETD, UCA, DS4H, France.

  - Master 1: Introduction to Information Visualization, Master DSAI, 15 h ETD, UCA, DS4H, France.

  - Master 2: Introduction to Scientific Research, 10 h ETD, UCA, DS4H, France.

  - Master 2: Introduction to Scientific Research, 15 h ETD, UCA, Polytech Nice, France.

  - Master 2: Data Mining Visualisation, 8 h ETD, UCA, Polytech Nice, France.

  - Master 2: Data Visualization, 15 h ETD, UCA, MBDS DS4H, France.

  - Master 2: Design and Evaluation of User Interfaces, 45 ETD, UCA, Polytech Nice, France.

  - Master 2: Multimodal Interaction Techniques, 15 ETD, UCA, Polytech Nice, France.

  - Master 2: coordination of the TER (Travaux de Fin d'Etude), UCA, Polytech Nice, France.

  - Master 2: coordination of the track on Human-Computer Interaction at the Informatics Department, UCA, Polytech Nice, France.

E-learning

- Mooc: Fabien Gandon, Olivier Corby & Catherine Faron, Web of Data and Semantic Web (FR), 7 weeks, FUN, Inria, France Université Numérique, self-paced course 41002, Education for Adults, 14164 learners registered at the time of this report, MOOC page.

- Mooc: Fabien Gandon, Olivier Corby & Catherine Faron, Introduction to a Web of Linked Data (EN), 4 weeks, FUN, Inria, France Université Numérique, self-paced course 41013, Education for Adults, 5614 learners registered at the time of this report, MOOC page.

- Mooc: Fabien Gandon, Olivier Corby & Catherine Faron, Web of Data (EN), 4 weeks, Coursera, self-paced course Education for Adults, 4224 learners registered at the time of this report, MOOC page.

- Mooc: Michel Buffa, HTML5 coding essentials and best practices, 6 weeks, edX MIT/Harvard, self-paced course Education for Adults, more than 500k learners at the time of this report (2015-2022), MOOC page.

- Mooc: Michel Buffa, HTML5 Apps and Games, 5 weeks, edX MIT/Harvard, self-paced course Education for Adults, more than 150k learners at the time of this report (2015-2022), MOOC page.

- Mooc: Michel Buffa, JavaScript Introduction, 5 weeks, edX MIT/Harvard, self-paced course Education for Adults, more than 250k learners at the time of this report (2015-2022), MOOC page.

  Curriculum

  - Damien Graux: Participation to the creation of a curriculum dedicated to teaching Big Data and associated technologies and use case. Initially dedicated to the West Balkan region, the curriculum was later (re-)thought as a set of lectures *globally* useful. This initiative as been presented at the 26th Annual Conference on Innovation and Technology in Computer Science Education (ITiCSE) [43]. This curriculum is also available online through an open platform, also presented at ITiCSE 2021 [42].

10.2.2 Supervision

Participants: Michel Buffa, Elena Cabrio, Catherine Faron, Fabien Gandon, Damien Graux, Aline Menin, Andrea Tettamanzi, Serena Villata, Marco Winckler.

- PhD in progress: Ali Ballout, *Active Learning for Axiom Discovery*, Andrea Tettamanzi, UCA.

- PhD in progress: Lucie Cadorel, *Localisation sur le territoire et prise en compte de l'incertitude lors de l'extraction des caractéristiques de biens immobiliers à partir d'annonces*, Andrea Tettamanzi, UCA.

- PhD in progress: Ahmed El Amine Djebri, Uncertainty in Linked Data, UCA, Andrea Tettamanzi, Fabien Gandon.

- PhD in progress: Rony Dupuy Charles, *Combinaison d'approches symboliques et connexionnistes d'apprentissage automatique pour les nouvelles méthodes de recherche et développement en agro-végétale-environnement*, Andrea Tettamanzi, UCA.

- PhD in progress: Antonia Ettore, Artificial Intelligence for Education and Training: Knowledge Representation and Reasoning for the development of intelligent services in pedagogical environments, UCA, Catherine Faron, Franck Michel.

- PhD in progress: Rémi Felin, Découverte évolutive d'axiomes à partir de graphes de connaissances, UCA, Andrea Tettamanzi, Catherine Faron.

- PhD in progress: Nicholas Halliwell, Explainable and Interpretable Prediction, UCA, Fabien Gandon.

- PhD in progress: Santiago Marro, Argument-based Explanatory Dialogues for Medicine, UCA 3IA, Elena Cabrio and Serena Villata.

- PhD in progress: Benjamin Molinet, Explanatory argument generation for healthcare applications, UCA 3IA, Elena Cabrio and Serena Villata.

- PhD in progress: Pierpaolo Goffredo, Natural Language Counter-argumentation to Fight Online Disinformation, UCA 3IA, Elena Cabrio and Serena Villata.

- PhD defended: Thu Huong Nguyen, Mining the Semantic Web for OWL axioms, Andrea Tettamanzi, July 2 [70].

- PhD in progress: Maroua Tikat, Visualisation multimédia interactive pour l'exploration d'une base de métadonnées multidimensionnelle de musiques populaires. Michel Buffa, Marco Winckler.

- PhD in progress: Florent ROBERT, Analyzing and Understanding Embodied Interactions in Extended Reality Systems. Co-supervision with Hui-Yin Wu, Lucile Sassatelli, and Marco Winckler.

- Post-Doc in progress: Aline MENIN, Interactive Visulization of Semantic Databases of Publications about Covid-19. (18 months)

- PhD defended: Molka Tounsi Dhouib, Knowledge engineering in the sourcing domain for the recommendation of providers, UCA, Catherine Faron, Andrea Tettamanzi, March [23].

- PhD defended: Mahamadou Toure, Local peer-to-peer mobile access to linked data in resource-constrained networks, UCA, Fabien Gandon, Moussa Lo (UGB, Senegal). [24]

- PhD defended: Vorakit Vorakitphan, Fine-Grained Classification of Polarized and Propagandist Text in News Articles and Political Debates, UCA, Elena Cabrio and Serena Villata.

Internship

- Master 1 Internship: **Minh Nhat Do**, A visualization approach based on chained views and follow-up queries to explore large datasets. UCA. Marco Winckler & Aline Menin.

- Master 2 Internship (MIAGE parcours IA2): **Rémi Felin**, Evolutionary Axiom Discovery from Populated Knowledge Bases. Supervised by Andrea Tettamanzi.

- Master 2 Internship (MIAGE parcours IA2): **Pierre Saunders**, Towards a Holistic Benchmarking Platform for Semantic Web Data Storage Systems. Supervised by Damien Graux.

- Master 2 Internship (MIAGE): **Manon Audren**, Annotation of hate speech messages. Supervised by Elena Cabrio.

- Licence 3 Internship (MIAGE): **Yessine Ben El Bey**, A demo system for propaganda detection. Supervised by Elena Cabrio and Serena Villata.

- Licence 3 Internship (MIAGE): **Saad El Din Ahmed**, Named Entity Recognition in Legal Texts. Supervised by Serena Villata.

- Master 2 internship: **Youssef MEKOUAR** Development of augmented visualisation for Web contents. Co-superved by Franck Michel, Anne Toulet, and Marco Winckler.

- Master 2 internship: **Florent Robert** ROBERT, Florent. "Creating a virtual reality serious game using a DSL for interactive 3D environments". Co-supervised by Hui-Yin Wu, Lucile Sassatelli, and Marco Winckler.

10.2.3 Juries

Participants: Elena Cabrio, Catherine Faron, Fabien Gandon, Aline Menin, Andrea Tettamanzi, Marco Winckler.

- Catherine Faron:

  - examiner of Federico Ulliana's HDR thesis entitled "Rule-based languages for reasoning on data: analysis, design, and applications", Université de Montpellier.

  - reviewer of Yu Du's PhD thesis entitled "Des données aux connaissances : vers des recommandations plus pertinentes, diversifiées et transparentes", Université de Montpellier, IMT Mines Alès.

  - reviewer of Stella Zevio's PhD thesis entitled "Découverte et enrichissement de connaissances à partir de textes pour la recherche d'experts", Université Sorbonne Paris Nord.

  - examiner of Huong Nguyen's PhD thesis entitled "Mining the semantic web for OWL axioms", Université Côte d'Azur.

- Fabien Gandon:

  - reviewer of Hicham Hossayni's PhD thesis entitled "Web Preemption for Querying the Linked Open Data", defended on December 17th, 2021 at Telecom SudParis, Institut Polytechnique de Paris.

  - president of Vorakit Vorakitphan's PhD thesis entitled "Fine-Grained Classification of Polarized and Propagandist Text in News Articles and Political Debates", defended on December 15th, 2021 at Université Côte d'Azur.

  - reviewer of the HDR Thesis of Antoine Zimmermann entitled "Interoperability for the Semantic Web: A Loosely Coupled Mediation Approach", University Jean Monner - Saint-Etienne, 23/02/2021.

  - examiner HDR Claudia-Lavinia Ignat entitled "Large Scale Trustworthy Distributed Collaborative Systems", Université de Lorraine, 23/04/2021.

  - president PhD jury of Thu Huong NGUYEN on "Mining the Semantic Web for OWL Axioms", at Université Côte d'Azur, 02/07/2021.

- Aline Menin : reviewer of Arthur Back's MSc thesis entitled "The Landscape of XR Evaluation: Tertiary Review and Visualizations", Federal University of Rio Grande do Sul, Brésil. Defended on December 14th, 2021 ; online.

- Andrea Tettamanzi: examiner of seven PhD theses of the University of Milan Bicocca:

  - Giulia Bernardini, "Combinatorial Methods for the Analysis of Related Genomic Sequences", February 24, 2021.

  - Anna Ferrari, "Personalization of Human Activity Recognition Methods using Inertial Data", February 24, 2021.

  - Adriano De Marino, "iSwap: a bioinformatics pipeline for index switching in Illumina sequencing platforms", April 28, 2021.

  - Intissar Khalifa, "Deep psychology recognition based on automatic analysis of non-verbal behaviors", April 28, 2021.

  - Riccardo Perego, "Automated Deep Learning through Constrained Bayesian Optimization", April 28, 2021.

  - Davide Ginelli, "Understanding and Improving Automatic Program Repair: A Study of Code-removal Patches and a New Exception-driven Fault Localization Approach", April 28, 2021.

– Vincenzo Cutrona, "Semantic Table Annotation for Large-Scale Data Enrichment", April 28, 2021.

• Elena Cabrio:

– Reviewer and member of the PhD committee of Pablo Accuosto "Mining arguments in scientific abstracts and its application to argumentative quality assessment.", Universitat Pompeu Fabra (Spain), November 26 2021.

– Reviewer and member of the PhD committee of Francois Torregrossa "Representation des mots et des connaissances", Université de Rennes (France), December 16, 2021.

– Reviewer and member of the PhD committee of Patricia Chiril "Hate speech detection on social media", Université de Toulouse (France), November 16 2021.

– Reviewer and member of the PhD committee of M. Mozafari "Hate Speech and Offensive Language Detection using Transfer Learning Approaches", Institut Polytechnique de Paris, Mines-Telecom, May 28 2021.

– Member of the PhD committee of Karim Ibrahim "Personalized Audio Auto-tagging as a Proxy for Context-aware Music Recommendation", Telecom ParisTech, December 16, 2021.

– Member of the PhD committee of Lucia Siciliani "Question Answering over Knowledge Graphs", University of Bari (Italy), February 22 2021.

• Marco Winckler: examiner of 3 PhD theses and 1 master thesis overseas :

– Jingya YUAN. "Data centered Usage based Protection in a SMACIT context". July 8th, 2021, at the INSA Lyon, France. (President of the Jury).

– Itziar OTADUY IGARTUA. "Promoting End-User Involvement in Web-based tasks: A Model-Driven Engineering Approach to Form-filling and User-Acceptance Testing". February 26th 2021, at the Universidad del Pais Vasco (UPV/EHU), Spain. (Rapporteur).

– Elodie BOUZEKRI. "Notation et processus outillé pour la description, l'analyse et la compréhension de l'automatisation dans les systèmes de commande et contrôle". January 14th 2021, Université of Toulouse, Université Paul Sabatier, Toulouse, France (President of the Jury).

– Carlos Victor QUIJANO CHAVES. An immersive Approach for Exploring Multiple Coordinated 3D Visualizations in Immersive Virtual Environments. Presented on November 30th, 2021. (Virtual) Universidade Federal do Rio Grande do Sul, UFRGS, Porto Alegre, Brazil.

10.3 Popularization

10.3.1 Articles and contents

Participants: Michel Buffa, Elena Cabrio, Fabien Gandon, Serena Villata.

• Fabien Gandon:

– article Binaire, Le Monde, "Dessine-moi un graphe de connaissances" [75].

– interview Chut! Magazine, N°7 - Lost in election "Le CERN a fait don du web au monde", 23/09/2021.

– podcast Chut! Radio "La Puce à l'Oreille, ep 5 : Faut-il sauver le world wide web ?" 29/09/2021.

– Interview in Comics / Bande Dessinée "Les défis de l'intelligence artificielle : un reporter dans les labos de recherche" [83].

– short talk at online Café-In "Mission Covid-19" about "CovidOnTheWeb", 06/05/2021.

– Inria news article IA : bâtir la confiance et garantir la souveraineté, 09/11/2021.

- contributor of the white paper/book of Inria "Internet of Things (IoT): Societal Challenges & Scientific Research Fields for IoT" [78].

- Michel Buffa:

  - Talk at online Inria Café-In about "WebAudio, WebMidi, WebComponents, WebAssembly, panorama de la MAO dans le browser", 06/05/2021.

- Serena Villata:

  - Participation to the round table on "Gender bias in AI" organised by the Institut français de Slovaquie, June 2021.

  - Participation to the online conference "Women and Artificial Intelligence", September 2021.

  - TV interview together with Elena Cabrio, FranceInfo, "L'intelligence artificielle pour lutter contre le harcèlement scolaire sur les réseaux sociaux", November 2021.

- Elena Cabrio:

  - Comment prevenir la cyberviolence, cyberharcelement et la haine en ligne ?, Festival de la Science de Nice, September 2021.

  - Interview as Argument Mining expert "Nature Podcast, March 2021 "The AI that argues back".

11 Scientific production

11.1 Major publications

[1] D. Allemang, J. Hendler and F. Gandon. *Semantic Web for the Working Ontologist*. 3. ACM, June 2020. DOI: 10.1145/3382097. URL: https://hal.inria.fr/hal-02939606.

[2] A. Ben othmane. 'CARS - A multi-agent framework to support the decision making in uncertain spatio-temporal real-world applications'. Université Côte d'Azur, 12th Oct. 2017. URL: https://hal.archives-ouvertes.fr/tel-01680984.

[3] F. Berthelon. 'Emotion modelization and detection from expressive and contextual data'. Université Nice Sophia Antipolis, 16th Dec. 2013. URL: https://tel.archives-ouvertes.fr/tel-00917416.

[4] K. R. Bouzidi. 'Semantic web models to support the creation of technical regulatory documents in building industry'. Université Nice Sophia Antipolis, 11th Sept. 2013. URL: https://tel.archives-ouvertes.fr/tel-00876366.

[5] E. Cabrio. 'Artificial Intelligence to Extract, Analyze and Generate Knowledge and Arguments from Texts to Support Informed Interaction and Decision Making'. Habilitation à diriger des recherches. Université Côte d'Azur, Oct. 2020. URL: https://hal.inria.fr/tel-03084380.

[6] L. Costabello. 'Context-aware access control and presentation of linked data'. Université Nice Sophia Antipolis, 29th Nov. 2013. URL: https://tel.archives-ouvertes.fr/tel-00934617.

[7] P. F. Diallo. 'Sociocultural and temporal aspects in ontologies dedicated to virtual communities'. COMUE Université Côte d'Azur (2015 - 2019); Université de Saint-Louis (Sénégal), 16th Sept. 2016. URL: https://tel.archives-ouvertes.fr/tel-01402394.

[8] M. Fell. 'Natural language processing for music information retrieval : deep analysis of lyrics structure and content'. Theses. Université Côte d'Azur, May 2020. URL: https://tel.archives-ouvertes.fr/tel-02587910.

[9] F. Gandon. 'Distributed Artificial Intelligence And Knowledge Management: Ontologies And Multi-Agent Systems For A Corporate Semantic Web'. Université Nice Sophia Antipolis, 7th Nov. 2002. URL: https://tel.archives-ouvertes.fr/tel-00378201.

[10] R. Gazzotti. 'Knowledge graphs based extension of patients' files to predict hospitalization'. Theses. Université Côte d'Azur, Apr. 2020. URL: https://hal.archives-ouvertes.fr/tel-02959420.

[11] R. Hasan. 'Predicting query performance and explaining results to assist Linked Data consumption'. Université Nice Sophia Antipolis, 4th Nov. 2014. URL: https://tel.archives-ouvertes.fr/tel-01127124.

[12] M. Lefrançois. 'Meaning-Text Theory lexical semantic knowledge representation : conceptualization, representation, and operationalization of lexicographic definitions'. Université Nice Sophia Antipolis, 24th June 2014. URL: https://tel.archives-ouvertes.fr/tel-01071945.

[13] A. Macina. 'SPARQL distributed query processing over linked data'. COMUE Université Côte d'Azur (2015 - 2019), 17th Dec. 2018. URL: https://tel.archives-ouvertes.fr/tel-02340700.

[14] N. Marie. 'Linked data based exploratory search'. Université Nice Sophia Antipolis, 12th Dec. 2014. URL: https://tel.archives-ouvertes.fr/tel-01130622.

[15] T. Mayer. 'Argument Mining on Clinical Trials'. Université Côte d'Azur, 16th Dec. 2020. URL: https://tel.archives-ouvertes.fr/tel-03209489.

[16] Z. Meng. 'Temporal and semantic analysis of richly typed social networks from user-generated content sites on the web'. Université Côte d'Azur, 7th Nov. 2016. URL: https://hal.inria.fr/tel-01402612.

[17] F. Michel. 'Integrating heterogeneous data sources in the Web of data'. Université Côte d'Azur, 3rd Mar. 2017. URL: https://tel.archives-ouvertes.fr/tel-01508602.

[18] F. Michel, F. Gandon, V. Ah-Kane, A. Bobasheva, E. Cabrio, O. Corby, R. Gazzotti, A. Giboin, S. Marro, T. Mayer, M. Simon, S. Villata and M. Winckler. 'Covid-on-the-Web: Knowledge Graph and Services to Advance COVID-19 Research'. In: ISWC 2020 - 19th International Semantic Web Conference. Athens / Virtual, Greece, 1st Nov. 2020. DOI: 10.1007/978-3-030-62466-8_19. URL: https://hal.archives-ouvertes.fr/hal-02939363.

[19] T. H. Nguyen. 'Mining the semantic Web for OWL axioms'. Université Côte d'Azur, 2nd July 2021. URL: https://tel.archives-ouvertes.fr/tel-03406784.

[20] C. Pasquier, C. Da Costa Pereira and A. G. B. Tettamanzi. 'Extending a Fuzzy Polarity Propagation Method for Multi-Domain Sentiment Analysis with Word Embedding and POS Tagging'. In: *Frontiers in Artificial Intelligence and Applications*. ECAI 2020 - 24th European Conference on Artificial Intelligence. Vol. 325. Santiago de Compostela, Spain: IOS Press, 29th Aug. 2020, pp. 2140–2147. DOI: 10.3233/FAIA200338. URL: https://hal.inria.fr/hal-02936130.

[21] T. A. Pham. 'OntoApp : a declarative approach for software reuse and simulation in early stage of software development life cycle'. Université Côte d'Azur, 22nd Sept. 2017. URL: https://tel.archives-ouvertes.fr/tel-01680766.

[22] O. Seye. 'Sharing and reusing rules for the Web of data'. Université Nice Sophia Antipolis; Université Gaston Berger de Saint Louis, 15th Dec. 2014. URL: https://hal.inria.fr/tel-01096306.

[23] M. Tounsi Dhouib. 'Knowledge engineering in the sourcing domain for the recommendation of providers'. Université Côte d'Azur, 26th Mar. 2021. URL: https://tel.archives-ouvertes.fr/tel-03336353.

[24] M. Toure. 'Local peer-to-peer mobile access to linked data in resource-constrained networks'. Université Côte d'Azur; Université de Saint-Louis (Sénégal), 4th Oct. 2021. URL: https://tel.archives-ouvertes.fr/tel-03514271.

[25] D. M. Tran. 'Discovering multi-relational association rules from ontological knowledge bases to enrich ontologies'. Université Côte d'Azur; Université de Danang (Vietnam), 23rd July 2018. URL: https://tel.archives-ouvertes.fr/tel-01926812.

11.2 Publications of the year

International journals

[26] B. Berendt, F. Gandon, S. Halford, W. Hall, J. Hendler, K. E. Kinder-Kurlanda, E. Ntoutsi and S. Staab. 'Web Futures: Inclusive, Intelligent, Sustainable The 2020 Manifesto for Web Science'. In: *Dagstuhl Manifestos* (2021). DOI: 10.4230/DagMan.9.1.1. URL: https://hal.inria.fr/hal-03189474.

[27]   M. Fell, Y. Nechaev, G. Meseguer-Brocal, E. Cabrio, F. Gandon and G. Peeters. 'Lyrics segmentation via bimodal text–audio representation'. In: *Natural Language Engineering* (2021), pp. 1–20. DOI: 10.1017/S1351324921000024. URL: https://hal.archives-ouvertes.fr/hal-03295581.

[28]   D. A. Furman, S. Marro, C. Cardellino, D. N. Popa and L. Alonso Alemany. 'You can simply rely on communities for a robust characterization of stances'. In: *Florida Artificial Intelligence Research Society* 34.1 (18th Apr. 2021). DOI: 10.32473/flairs.v34i1.128515. URL: https://hal.archives-ouvertes.fr/hal-03260142.

[29]   R. Gazzotti, C. Faron Zucker, F. Gandon, V. Lacroix-Hugues and D. Darmon. 'Extending electronic medical records vector models with knowledge graphs to improve hospitalization prediction'. In: *Journal of Biomedical Semantics* (2021). URL: https://hal.archives-ouvertes.fr/hal-03502001.

[30]   T. Mayer, S. Marro, S. Villata and E. Cabrio. 'Enhancing Evidence-Based Medicine with Natural Language Argumentative Analysis of Clinical Trials'. In: *Artificial Intelligence in Medicine* (7th May 2021), p. 102098. DOI: 10.1016/j.artmed.2021.102098. URL: https://hal.archives-ouvertes.fr/hal-03264761.

[31]   A. Menin, M. N. Do, C. Dal Sasso Freitas, O. Corby, C. Faron Zucker, A. Giboin and M. Winckler. 'Using Chained Views and Follow-up Queries to Assist the Visual Exploration of the Web of Big Linked Data'. In: *International Journal of Human-Computer Interaction* (2022). URL: https://hal.archives-ouvertes.fr/hal-03518845.

[32]   A. Menin, F. Michel, F. Gandon, R. Gazzotti, E. Cabrio, O. Corby, A. Giboin, S. Marro, T. Mayer, S. Villata and M. Winckler. 'Covid-on-the-Web: Exploring the COVID-19 Scientific Literature through Visualization of Linked Data from Entity and Argument Mining'. In: *Quantitative Science Studies* (5th Nov. 2021). DOI: 10.1162/qss_a_00164. URL: https://hal.archives-ouvertes.fr/hal-03404580.

[33]   A. Menin, R. Torchelsen and L. Nedel. 'The effects of VR in training simulators: Exploring perception and knowledge gain'. In: *Computers and Graphics* (Oct. 2021). DOI: 10.1016/j.cag.2021.09.015. URL: https://hal.archives-ouvertes.fr/hal-03369680.

[34]   M. Tounsi Dhouib, C. Faron and A. G. B. Tettamanzi. 'Measuring Clusters of Labels in an Embedding Space to Refine Relations in Ontology Alignment'. In: *Journal on Data Semantics* (16th Oct. 2021). DOI: 10.1007/s13740-021-00137-8. URL: https://hal.inria.fr/hal-03403125.

International peer-reviewed conferences

[35]   L. Cadorel, A. Blanchi and A. G. B. Tettamanzi. 'Geospatial Knowledge in Housing Advertisements: Capturing and Extracting Spatial Information from Text'. In: *K-CAP '21: Knowledge Capture Conference, Virtual Event, USA, December 2-3, 2021*. K-CAP '21: Knowledge Capture Conference. Virtual Event USA, United States: ACM, 2nd Dec. 2021, pp. 41–48. DOI: 10.1145/3460210.3493547. URL: https://hal.inria.fr/hal-03518717.

[36]   O. Corby, C. Faron, F. Gandon, D. Graux and F. Michel. 'Beyond Classical SERVICE Clause in Federated SPARQL Queries: Leveraging the Full Potential of URI Parameters'. In: International Conference on Web Information Systems and Technologies (WEBIST). Online, Portugal, 26th Oct. 2021. URL: https://hal.inria.fr/hal-03404125.

[37]   A. E. A. Djebri, A. Ettorre and J. Mortara. 'Towards a Linked Open Code'. In: *The Semantic Web*. ESWC 2021 - 18th Extended Semantic Web Conference. ESWC 2021. Heraklion / Virtual, Greece, 31st May 2021. DOI: 10.1007/978-3-030-77385-4_29. URL: https://hal.archives-ouvertes.fr/hal-03190617.

[38]   A. Ettorre, A. Bobasheva, C. Faron and F. Michel. 'A systematic approach to identify the information captured by Knowledge Graph Embeddings'. In: IEEE/WIC/ACM International Conference on Web Intelligence (WI-IAT '21). IEEE/WIC/ACM International Conference on Web Intelligence (WI-IAT '21). ESSENDON, VIC, Australia, 14th Dec. 2021. URL: https://hal.archives-ouvertes.fr/hal-03476795.

[39] R. Felin and A. G. B. Tettamanzi. 'Using Grammar-Based Genetic Programming for Mining Subsumption Axioms Involving Complex Class Expressions'. In: *IEEE/WIC/ACM International Conference on Web Intelligence (WI-IAT '21)*. 20th IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. Melbourne, Australia, 2022. DOI: 10.1145/3486622.3494025. URL: https://hal.archives-ouvertes.fr/hal-03519365.

[40] R. Gazzotti and F. Gandon. 'When owl:sameAs is the Same: Experimenting Online Resolution of Identity with SPARQL queries to Linked Open Data Sources'. In: WEBIST 2021 - 17th International Conference on Web Information Systems and Technologies. Virtual, France, 26th Oct. 2021. URL: https://hal.archives-ouvertes.fr/hal-03301330.

[41] D. Graux, D. Collarana and F. Orlandi. 'Formal Concept Analysis for Semantic Compression of Knowledge Graph Versions'. In: FCA4AI 2021 - 9th International Workshop "What can FCA do for Artificial Intelligence?" Montréal (Virtual Event), Canada, 21st Aug. 2021. URL: https://hal.inria.fr/hal-03516393.

[42] D. Graux, V. Janev, H. Jabeen and E. Sallinger. 'A Big Data Learning Platform for the West Balkans and Beyond'. In: ITiCSE 2021: 26th ACM Conference on Innovation and Technology in Computer Science Education. Virtual Event, Germany: ACM, 26th June 2021, pp. 617–618. DOI: 10.1145/3456565.3460026. URL: https://hal.inria.fr/hal-03516263.

[43] D. Graux, V. Janev, H. Jabeen and E. Sallinger. 'Deploying a Strategy to Unlock Big Data Research and Teaching Activities in the West Balkan Region'. In: ITiCSE 2021: 26th ACM Conference on Innovation and Technology in Computer Science Education. Virtual Event, Germany: ACM, 26th June 2021, pp. 491–497. DOI: 10.1145/3430665.3456325. URL: https://hal.inria.fr/hal-03516267.

[44] D. Graux and T. Michel. 'Involvement of OpenStreetMap in European H2020 Projects'. In: Proceedings of the Academic Track at State of the Map 2021. Virtual Event, Germany, 9th July 2021. URL: https://hal.inria.fr/hal-03516296.

[45] D. Graux, F. Orlandi, T. Kaushik, D. Kavanagh, H. Jiang, B. Bredican, M. Grouse and D. Geary. 'Timelining Knowledge Graphs in the Browser'. In: VOILA! 2021- 6th International Workshop on the Visualization and Interaction for Ontologies and Linked Data. Virtual Event, United States, 25th Oct. 2021. URL: https://hal.inria.fr/hal-03516420.

[46] D. Graux, F. Orlandi and D. O'sullivan. 'De-icing federated SPARQL pipelines: a method for assessing the "freshness" of result sets'. In: MEPDaW 2021 - 7th Workshop on Managing the Evolution and Preservation of the Data Web. Virtual Event, United States, 25th Oct. 2021. URL: https://hal.inria.fr/hal-03516459.

[47] D. Graux, F. Orlandi and D. O'sullivan. 'Hash-ssessing the freshness of SPARQL pipelines'. In: ISWC 2021 - International Semantic Web Conference : Posters, Demos and Industry Tracks. Virtual Event, United States, 24th Oct. 2021. URL: https://hal.inria.fr/hal-03516436.

[48] N. Halliwell. 'Evaluating Explanations of Relational Graph Convolutional Network Link Predictions on Knowledge Graphs'. In: AAAI. Vancouver, Canada, 22nd Feb. 2022. URL: https://hal.archives-ouvertes.fr/hal-03454121.

[49] N. Halliwell, F. Gandon and F. Lecue. 'A Simplified Benchmark for Ambiguous Explanations of Knowledge Graph Link Prediction using Relational Graph Convolutional Networks'. In: 36th AAAI Conference on Artificial Intelligence. Vancouver, Canada, 22nd Feb. 2022. URL: https://hal.archives-ouvertes.fr/hal-03434544.

[50] N. Halliwell, F. Gandon and F. Lecue. 'A Simplified Benchmark for Non-ambiguous Explanations of Knowledge Graph Link Prediction using Relational Graph Convolutional Networks'. In: International Semantic Web Conference. Proceedings of International Semantic Web Conference. Troy, United States, 24th Oct. 2021. URL: https://hal.archives-ouvertes.fr/hal-03339562.

[51] N. Halliwell, F. Gandon and F. Lecue. 'Linked Data Ground Truth for Quantitative and Qualitative Evaluation of Explanations for Relational Graph Convolutional Network Link Prediction on Knowledge Graphs'. In: International Conference on Web Intelligence and Intelligent Agent Technology. Melbourne, Australia, 14th Dec. 2021. DOI: 10.1145/3486622.3493921. URL: https://hal.archives-ouvertes.fr/hal-03430113.

[52] N. Halliwell, F. Gandon and F. Lecue. 'User Scored Evaluation of Non-Unique Explanations for Relational Graph Convolutional Network Link Prediction on Knowledge Graphs'. In: International Conference on Knowledge Capture. Virtual Event, United States, 2nd Dec. 2021. DOI: 10.1145/3460210.3493557. URL: https://hal.archives-ouvertes.fr/hal-03402766.

[53] H. Huang and F. Gandon. 'Learning URI Selection Criteria to Improve the Crawling of Linked Open Data (Extended Abstract)'. In: IJCAI 2020 - 29th International Joint Conference on Artificial Intelligence. Yokohama, Japan, 7th Jan. 2021. URL: https://hal.inria.fr/hal-03064912.

[54] A. Menin, L. Cadorel, A. G. B. Tettamanzi, A. Giboin, F. Gandon and M. Winckler. 'ARViz: Interactive Visualization of Association Rules for RDF Data Exploration'. In: IV 2021 - 25th International Conference Information Visualisation. Vol. 25. 2021 25th International Conference Information Visualisation (IV). Melbourne / Virtual, Australia, 2021, pp. 13–20. DOI: 10.1109/IV53921.2021.00013. URL: https://hal.archives-ouvertes.fr/hal-03292140.

[55] A. Menin, R. Cava, C. M. Dal Sasso Freitas, O. Corby and M. Winckler. 'Towards a Visual Approach for Representing Analytical Provenance in Exploration Processes'. In: IV 2021 - 25th International Conference Information Visualisation. Vol. 25. 2021 25th International Conference Information Visualisation (IV). Melbourne / Virtual, Australia, 2021, pp. 21–28. DOI: 10.1109/IV53921.2021.00014. URL: https://hal.archives-ouvertes.fr/hal-03292172.

[56] A. Menin, C. Faron Zucker, O. Corby, C. M. Dal Sasso Freitas, F. Gandon and M. Winckler. 'From Linked Data Querying to Visual Search: Towards a Visualization Pipeline for LOD Exploration'. In: Proceedings of the 17th International Conference on Web Information Systems and Technologies (WEBIST). Proceedings of the 17th International Conference on Web Information Systems and Technologies (WEBIST). Online Streaming, France, 2021. DOI: 10.5220/0010654600003058. URL: https://hal.archives-ouvertes.fr/hal-03404572.

[57] F. Michel, A. Ettorre, C. Faron, J. Kaplan and O. Gargominy. 'Biodiversity Knowledge Graphs: Time to move up a gear!' In: TDWG 2021 annual virtual conference. Vol. 5. Biodiversity Information Science and Standards. Virtual, France: Pensoft Publishers; Sofia : Pensoft Publishers, 2017-, 31st Aug. 2021. DOI: 10.3897/biss.5.73699. URL: https://hal.archives-ouvertes.fr/hal-03373536.

[58] S. Ren, L. Pottier and M. Buffa. 'Build WebAudio and JavaScript Web Applications using JSPatcher: A Web-based Visual Programming Editor'. In: Web Audio Conference 2021. Barcelone, Spain, 5th June 2021. URL: https://hal.inria.fr/hal-03519504.

[59] F. A. S. Robert, H.-Y. Wu, M. Winckler and L. Sassatelli. 'Creating serious game with a domain specific language for embodied experience in virtual reality'. In: Les journées Françaises de l'Informatique Graphique. Biot, France, 24th Nov. 2021. URL: https://hal.inria.fr/hal-03464373.

[60] M. Tikat, M. Winckler and M. Buffa. 'Interactive multimedia visualization for exploring and fixing a multi-dimensional metadata base of popular musics'. In: 7th MEPDaW Workshop at ISWC'21. 7th MEPDaW Workshop at ISWC'21. Virtual, France, 2021. URL: https://hal.archives-ouvertes.fr/hal-03425819.

[61] V. Vorakitphan, E. Cabrio and S. Villata. '"Don't discuss": Investigating Semantic and Argumentative Features for Supervised Propagandist Message Detection and Classification'. In: RANLP 2021 - Recent Advances in Natural Language Processing. Varna / Virtual, Bulgaria, 1st Sept. 2021. URL: https://hal.archives-ouvertes.fr/hal-03314797.

[62] V. Vorakitphan, E. Cabrio and S. Villata. 'PROTECT - A Pipeline for Propaganda Detection and Classification'. In: Eighth Italian Conference on Computational Linguistics (CLIC-it 2021). Milan, Italy, 26th Jan. 2022. URL: https://hal.archives-ouvertes.fr/hal-03417019.

National peer-reviewed Conferences

[63] F. Michel, F. Gandon, V. Ah-Kane, A. Bobasheva, E. Cabrio, O. Corby, R. Gazzotti, A. Giboin, S. Marro, T. Mayer, M. Simon, S. Villata and M. Winckler. 'Covid-on-the-Web: Graphe de Connaissances et Services pour faire Progresser la Recherche sur la COVID-19'. In: IC 2021 - 32es Journées francophones d'Ingénierie des Connaissances (32st French Knowledge Engineering Conference). Bordeaux, France, 28th June 2021, pp. 1–9. URL: https://hal.archives-ouvertes.fr/hal-03230741.

Scientific book chapters

[64]   F. Bertoncello, M.-J. Ouriachi, C. Da Costa Pereira, A. G. B. Tettamanzi, L. Purdue and R. Ajroud. 'Using ABM to explore the role of socio-environmental interactions on Ancient Settlement Dynamics.' In: *Human history and digital future, Proceeding of the 46th Computer Applications and Quantitative Methods in Archaeology international conference (CAA 2018), Tübingen, mars 2018.* 2021. URL: https://halshs.archives-ouvertes.fr/halshs-03098473.

[65]   M. Buffa, E. Cabrio, M. Fell, F. Gandon, A. Giboin, R. Hennequin, F. Michel, J. Pauwels, G. Pellerin, M. Tikat and M. Winckler. 'The WASABI Dataset: Cultural, Lyrics and Audio Analysis Metadata About 2 Million Popular Commercially Released Songs'. In: *The Semantic Web. ESWC 2021. Lecture Notes in Computer Science, vol 12731.* 31st May 2021, pp. 515–531. DOI: 10.1007/978-3-030-77385-4_31. URL: https://hal.archives-ouvertes.fr/hal-03282619.

[66]   A. GIBOIN. 'Common frame of reference'. In: *Ergonomie : 150 notions clés.* Dunod, Sept. 2021, pp. 412–415. URL: https://hal.archives-ouvertes.fr/hal-03498776.

[67]   D. Graux and S. Mahmoodi. 'A Fully Decentralized Triplestore Managed via the Ethereum Blockchain'. In: *Further with Knowledge Graphs.* Studies on the Semantic Web. IOS Press, 31st Aug. 2021. DOI: 10.3233/SSW210044. URL: https://hal.inria.fr/hal-03516322.

[68]   A. Sadeghi, D. Collarana, D. Graux and J. Lehmann. 'Embedding Knowledge Graphs Attentive to Positional and Centrality Qualities'. In: *Machine Learning and Knowledge Discovery in Databases. Research Track.* Vol. 12976. Lecture Notes in Computer Science. Springer International Publishing, 10th Sept. 2021, pp. 548–564. DOI: 10.1007/978-3-030-86520-7_34. URL: https://hal.inria.fr/hal-03516350.

[69]   H.-Y. Wu, J. Delachambre, L. Sassatelli and M. Winckler. 'Through the Eyes of Women in Engineering: An immersive VR experience'. In: *Texts of Discomfort.* Carnegie Mellon University: ETC Press, 23rd Nov. 2021, pp. 387–414. DOI: 10.1184/R1/17071172. URL: https://hal.inria.fr/hal-03515116.

Doctoral dissertations and habilitation theses

[70]   T. H. Nguyen. 'Mining the semantic Web for OWL axioms'. Université Côte d'Azur, 2nd July 2021. URL: https://tel.archives-ouvertes.fr/tel-03406784.

[71]   M. Tounsi Dhouib. 'Knowledge engineering in the sourcing domain for the recommendation of providers'. Université Côte d'Azur, 26th Mar. 2021. URL: https://tel.archives-ouvertes.fr/tel-03336353.

[72]   M. Toure. 'Local peer-to-peer mobile access to linked data in resource-constrained networks'. Université Côte d'Azur; Université de Saint-Louis (Sénégal), 4th Oct. 2021. URL: https://tel.archives-ouvertes.fr/tel-03514271.

Reports & preprints

[73]   N. Abbas, K. Alghamdi, M. Alinam, F. Alloatti, G. Amaral, M. Beno, F. Bensmann, C. d'Amato, L. Asprino, R. Biswas et al. *Knowledge Graphs Evolution and Preservation – A Technical Report from ISWS 2019.* 15th Jan. 2021. URL: https://hal.archives-ouvertes.fr/hal-03111661.

Other scientific publications

[74]   R. Felin and A. G. B. Tettamanzi. 'RDFMiner : Grammar-Based Genetic Programming for Mining OWL 2 Axioms'. In: 3IA Annual Scientific Meeting. Biot, France, 30th Nov. 2021. URL: https://hal.archives-ouvertes.fr/hal-03517059.

[75]   F. Gandon. *Dessine-moi un graphe de connaissances !* 5th Oct. 2021. URL: https://hal.inria.fr/hal-03376942.

[76]   Y. Mekouar. 'Visualisation augmentée d'articles scientifiques s'appuyant sur un index sémantique.' Inria Sophia Antipolis - Méditerranée, Université Côte d'Azur, 1st Sept. 2021. URL: https://hal.archives-ouvertes.fr/hal-03469855.

[77]   A. OLLAGNIER, E. Cabrio and S. Villata. 'Multi-view Clustering for Hate Speech and Target Community Detection on Social Media'. In: Soph.I.A Summit 2021. sophia antipolis, France, 17th Nov. 2021. URL: https://hal.archives-ouvertes.fr/hal-03466433.

11.3   Cited publications

[78]   E. Baccelli. *Internet of Things (IoT): Societal Challenges & Scientific Research Fields for IoT*. Dec. 2021. URL: https://hal.inria.fr/hal-03474888.

[79]   M. Buffa and J. Lebrun. 'A FAUST-based re-creation of the power amp stage for WebAudio-based simulations of guitar tube amplifiers'. In: *IFC 2020 - Second International Functional Audio Stream (Faust) Conference*. Saint-Denis / Virtual, France, Dec. 2020. URL: https://hal.inria.fr/hal-03087768.

[80]   M. Buffa, J. Lebrun, S. Ren, S. Letz, Y. Orlarey, R. Michon and D. Fober. 'Emerging W3C APIs opened up commercial opportunities for computer music applications'. In: *The Web Conference 2020 - DevTrack*. Taipei, Taiwan, Apr. 2020. URL: https://hal.inria.fr/hal-02557901.

[81]   L. Cadorel and A. G. B. Tettamanzi. 'Mining RDF Data of COVID-19 Scientific Literature for Interesting Association Rules'. In: *WI-IAT'20 - IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*. The 2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT'20), 14-17 December 2020, a fully virtual conference. Melbourne, Australia, Dec. 2020. URL: https://hal.inria.fr/hal-03084029.

[82]   A. Ciortea, S. Mayer, S. Bienz, F. Gandon and O. Corby. 'Autonomous search in a social and ubiquitous Web'. In: *Personal and Ubiquitous Computing* (June 2020). DOI: 10.1007/s00779-020-01415-1. URL: https://hal.inria.fr/hal-03135088.

[83]   J. Dres. *Les défis de l'intelligence artificielle : un reporter dans les labos de recherche*. Paris: First, 2021.

[84]   M. Fell, E. Cabrio, E. Korfed, M. Buffa and F. Gandon. 'Love Me, Love Me, Say (and Write!) that You Love Me: Enriching the WASABI Song Corpus with Lyrics Annotations'. In: *LREC 2020 - 12th edition of the Language Resources and Evaluation Conference*. Due to COVID 19 pandemic, the 12th edition is cancelled. Next edition, the 13th, LREC 2022 will take place in Pharo on June 16-24, 2022. Marseille, France, May 2020. URL: https://hal.archives-ouvertes.fr/hal-03133188.

[85]   F. Gandon. 'The three 'W' of the World Wide Web call for the three 'M' of a Massively Multidisciplinary Methodology'. In: *10th International Conference, WEBIST 2014*. Ed. by V. Monfort and K.-H. Krempels. Vol. 226. Web Information Systems and Technologies. Barcelona, Spain: Springer International Publishing, Apr. 2014. DOI: 10.1007/978-3-319-27030-2. URL: https://hal.inria.fr/hal-01223236.

[86]   F. Gandon, M. Buffa, E. Cabrio, O. Corby, C. Faron Zucker, A. Giboin, N. Le Thanh, I. Mirbel, P. Sander, A. G. B. Tettamanzi and S. Villata. 'Challenges in Bridging Social Semantics and Formal Semantics on the Web'. In: *5h International Conference, ICEIS 2013*. Ed. by S. Hammoudi, J. Cordeiro, L. Maciaszek and J. Filipe. Vol. 190. Angers, France: Springer, July 2013, pp. 3–15. URL: https://hal.inria.fr/hal-01059273.

[87]   V. Leone, L. Di Caro and S. Villata. 'Taking stock of legal ontologies: a feature-based comparative analysis'. In: *Artificial Intelligence and Law* 28.2 (2020), pp. 207–235. DOI: 10.1007/s10506-019-09252-1. URL: https://hal.archives-ouvertes.fr/hal-03065217.

[88]   F. Michel, S. Tercerie, A. Ettorre, O. Gargominy and C. Faron Zucker. 'Assisting Biologists in Editing Taxonomic Information by Confronting Multiple Data Sources using Linked Data Standards'. In: *Biodiversity Next*. Vol. 3. Biodiversity Information Science and Standards 37421. Leiden, Netherlands, Oct. 2019. DOI: 10.3897/biss.3.37421. URL: https://hal.archives-ouvertes.fr/hal-02168164.

[89]   S. Ren, S. Letz, Y. Orlarey, R. Michon, D. Fober, M. Buffa and J. Lebrun. 'Using Faust DSL to Develop Custom, Sample Accurate DSP Code and Audio Plugins for the Web Browser'. In: *Journal of the Audio Engineering Society* 68.10 (Nov. 2020). URL: https://hal.inria.fr/hal-03087763.

[90] S. Ren, L. Pottier and M. Buffa. 'From Diagram to Code: a Web-based Interactive Graph Editor for Faust DSP Design and Code Generation'. In: *IFC 2020 - Second International Functional Audio Stream (Faust) Conference*. Saint-Denis / Virtual, France, Dec. 2020. URL: https://hal.inria.fr/hal-03087778.