Francesca Bugiotti

Assistant professor at CentraleSupélec and member of Large-scale Heterogeneous DAta and Knowledge (LaHDAK) team of LISN (Laboratoire Interdisciplinaire des Sciences du Numérique) of Paris Saclay University.

Contacts

Current Position
Enseignant-chercheur
CentraleSupélec - LISN

Web Page
`www.bugiotti.it`

CentraleSupélec
3, rue Joliot-Curie
91192, Gif-sur-Yvette
France

LISN
1, rue Raimond Castaing
91190, Gif-sur-Yvette
France

E-mail
francesca.bugiotti@centralesupelec.fr

francesca.bugiotti@lisn.upsaclay.fr

Working Activity

CentraleSupélec
(*March 2015 - present*)
Research activity on Data Modeling and Data integration of NoSQL databases and Artificial Intelligence.

Inria - Institut National de Recherche en Informatique et en Automatique
(*November 2013 - February 2015*)
Research activity on RDF datasets indexing in the Amazon Cloud Computing Environment.

Università Roma Tre
(*April 2012 - October 2013*)
Post graduate research activity on model management in databases and No-SQL data stores integration.

Inria - Institut National de Recherche en Informatique et en Automatique
(*April 2011 - July 2011*)
Research activity on RDF datasets indexing in the Amazon Cloud Computing Environment.

Università Roma Tre
(*January 2008 - November 2008*)
Research activity on model management problems in databases.

ISA s.r.l. [1]
(*December 2008 - January 2010*)
Part-time research activity on data mining applied to clinical data.

Consip [2]
(*December 2008 - January 2010*)
I collaborated with Consip as a consultant on the quality of a data migration for the new information system of the Italian State General Accounting Department (Ragioneria Generale dello Stato Italiano).

Education

Università Roma Tre
(*November 2008 - April 2012*)
PhD in Computer Science - Computer Science and Automation department

- Thesis:"A model oriented approach to heterogeneity".

[1] ISA s.r.l. http://www.isa.it/, is an Italian enterprise that provides software for small and medium companies. It is focused on ERP services and business intelligence.

[2] Consip (Concessionaria Servizi Informativi Pubblici) s.p.a. is a public stock company owned by Italy's Ministry of the Economy and Finance that operates in behalf of the State.

- Advisor: Prof. Paolo Atzeni[3]

Università Roma Tre - IBM - Formit [4]
(*January 2008 - March 2009*)
Post lauream degree in IT governance: development, management and monitoring (*Governo dei sistemi informativi: gestione, sviluppo e monitoraggio*).

Università Roma Tre
(*October 2008*)
Qualifying examination to exercise the engineering profession.

Università Roma Tre
(*October 2005 - December 2007*)
Master degree in Computer Engineering ("Laurea Specialistica in Ingegneria informatica").
- Title of the thesis:"Tools and methodology for model management problems".
- Advisor: prof. Paolo Atzeni
- Final grade: 110/110 lode (maximum honors)

Università Roma Tre (*October 2002 - July 2005*)
First level degree in Computer Engineering ("Laurea in Ingegneria informatica").
- Thesis:"Datalog rules management for data and schema translation".
- Advisor: Prof. Paolo Atzeni
- Final grade: 110/110 lode (maximum honors)

Awards

Accenture - Università Roma Tre
(*March 2009*)
"Accenture" Outstanding Engineering Graduate Award.

IBM
(*July 2007*)
Participant in "IBM EMEA Best Student Recognition Event", Nice.

Industry Collaborations

Transvalor
(*2021 - present*)
Participation to a research-contract for the development of a chaire de recherche.

Schlumberger
(*2017 - present*)
Co-supervision of multiple projects, focusing on data analysis and integration, developed with students of CentraleSupélec. Co-Supervision of a PhD thesis.

Vires - Msc Sowtware
(*2018 - present*)
Co-supervision of projects developed with students of CentraleSupélec.

Teaching Responsibilities

BSc in Artificial Intelligence, Data & Management Sciences
(*2021 - present*)
Scientific responsible of the bachelor. Participation to the development of the scientific program. Program academical director since September 2022.

[3] `http://www.dia.uniroma3.it/~atzeni/`

[4] FORMIT is a Foundation that performs activities of scientific research, technical support, analysis and industrial, financial and socio-economic evaluation to sustain migration processes and integration of technological systems in every field of society.

Class	Type	Level	Kind	Hours	Students
CentraleSupélec					
BDDM	C/TD/TP	CS (M2)	INI	25	20
Algoritmi for distributed systems	C/TP/TD	CS (M2)	INI	24	20
Modern infrastructures and cloud	C/TP/TD	CS (M2)	INI	24	40
Cloud Computing	C/TP/TD	CS (M1)	INI	25	100
Big Data	C/TP/TD	CS (M1)	INI		80
Software Engineering	C/TP/TD	CS (M1)	INI	15	100
Object Oriented Programming	TP/TD	CS (M1)	INI	15	25
Algorithms	TP/TD	CS (L3)	INI	25	60
Information Systems and Programming	TP/TD	CS (L3)	INI	15	35
Models and Systems for Big Data Management	TP/TD	CS (M1)	INI	15	30
Hardware Architecure	TP/TD	CS (M1)	INI	15	25
Computer Engineering	TP/TD	CS (M1)	INI	15	25
EXED/CentraleSupélec					
Relational Databases and NoSQL	C/TP/TD	MS	APP	25	30
CentraleSupélec/Instutut Villebon Charpak					
Data Analysis	C/TP/TD	MS	APP	25	15
Essec/CentraleSupélec					
Big Data: algorithms, techniques & platforms	C/TP/TD	M2	INI	25	120
CentraleSupélec/Erasmus Mundus					
Big Data Research Project	C/TP/TD	M2	INI	25	20
TU-Berlin					
Data Analytics in Energy Sector Applications	C/TP/TD	M2	INI	25	15
Advanced Database Design, Data Management & Integration	C/TP/TD	M2	INI	25	15
Computer science and programming methods for Energy engineering	C/TP/TD	M1	INI	25	15
Universitá Roma Tre					
Databases	TP/TD	L3	INI	25	60
Object Oriented Analysis and Design	TP/TD	M2	INI	25	15
IT Governance	C/TP/TD	M2	INI	25	40
Java programming and Algorithms	TP/TD	L3	INI	25	80
Remedial Mathematics	TP/TD	L2	INI	25	30
Universitá della Tuscia					
Big Data	C/TP/TD	M2	INI	25	15

TEACHING ACTIVITY

My teaching activities started in 2007 with a class on Java programming and algorithms and continues with a year volume of 192H/TD.

CentraleSupélec

BDDM *(2020-2022)* - lectures, exercises, test grading, office hours. The class focusses on the Big-Data Storage Management Systems and the different query execution processes: (i) How the relational DBMS and the NoSQL systems store data (ii) How the queries are translated in low-level operations that are executed on data physically stored according to the physical pattern of the system under analysis (iii) How to write and re-write queries taking into account this physical organization for improving performances.

Algorithms for distributed systems *(2020-2022)* - lectures, exercises, test grading, office hours. The class illustrates how to treat heterogeneous, complex, and massive data using distributed methods, algorithms, and platforms. Each optimization challenge and theory aspect will be analyzed with reference to a standard Big Data technology and a framework of reference such as Docker, Kubernetes, and Argo Workflow.

Modern infrastructures and cloud *(2020-2022)* - lectures, exercises, test grading, office hours. At the end of this course students have a precise knowledge of: the fundamental components used in the design of infrastructures and IT solutions, of the concepts,

methods and tools applicable to infrastructures and Cloud services, and the different approaches to databases, NoSQL, graphs, and their main uses.

Cloud Computing *(2019-2022)* - lectures, exercises, test grading, office hours. The class covers Big Data introduction, to Hadoop, Spark, Kubernetes, and data analysis using standard algorithms (k-means, page rank, etc.).

Object Oriented programming *(2015-2022)* - exercises, test grading, office hours. The course covers Java Programming, Object Oriented Analysis and Design, introduces to UML and concurrent programming.

Algorithms *(2015-2021)* - exercises, test grading, office hours. The course introduces comprises a practical introduction to computer science basic algorithms (indexing, searching, sorting, recursion, etc.) and more advanced algorithms (compute Voronoi diagram, graph coloring, etc.).

Information systems and programming *(2018-2021)* - exercises, test grading, office hours. The SIP course is intended to educate future engineers to understand and master the IT tools and information systems they will be confronted with in their career, as well as build their skills to properly design and write code. The course is composed of two main parts: Information Systems and Programming, the second focusing on the Python language.

Big Data *(2016, 2017, 2018)* - lectures, exercises, test grading, office hours. The course covers Big Data introduction, NoSQL Databases, to Hadoop, Spark, and data analysis using standard algorithms (k-means, page rank, etc.).

Models and Systems for Big Data management *(2019)* - exercises, test grading, and office hours. The course covers introduction to NoSQL Databases.

Software Engineering *(2015)* - lectures, exercises, test marking, office hours. The course covers Java programming, object-oriented analysis and design, introduction to UML and concurrent programming.

Hardware Architecture *(2016, 2017, 2018)* - exercises. Introduction to the computer logic, to the physical architecture of a computer, and to the programming languages.

Computer Engineering *(2014, 2015, 2016)* - exercises, office hours. The course explains the rules and methods that describe the functionality, organization, and implementation of computer systems.

EXED CentraleSupélec - Mastère Spècialisé en Architecte des Systèmes d'Information

My teaching activities have also provided for interventions in a specialized master's degree: Specialized Master in Information Systems Architect.

Relational databases and NoSQL *(2019, 2021, 2022)* - the class introduce the database management systems and compares the classical relational databases with the NoSQL databases.

CentraleSupélec - Institut Villebon Charpak

My teaching activities also included interventions in the Villebon Charpak Institute[5]. In this context I developed and used innovative teaching techniques.

Data Analysis *(2021-2022)* - 30H/TD. The objective of this course is to make students autonomous in the study and exploration of data using digital tools with the focus on machine learning and thanks to the discussion of a representative set of examples.

[5]

CentraleSupélec/ESSEC - Master in Data sciences & Business analytics

My teaching activities have also included a participation in an international master with a class of 120 students.

Big Data Algorithms, Techniques & Platforms *(2020-2022)* - 30H équiv. TD, lectures, exercises, test grading, office hours. The class introduces the main characteristics of Big Data. The first part is focused on NoSQL datastores showing how their characteristics allow to store a big and heterogeneous amount of data. The second part analyzes the basic programming methods enabling users to deal with these large amounts of data. Finally the class presents the map-reduce programming paradigm and shows how it can be used in cloud-based computing frameworks (like Hadoop) and natively in some NoSQL datastores (like MongoDB).

CentraleSupélec/Erasmus Mundus - Master Big Data Management & Analytics

My teaching activities also included interventions in an Erasmus Mundus master's degree: BDMA (Big Data Management & Analytics).

Big Data Research Project *(2015-2023)* - the class introduces to research methodology. The class objective is in the development of a research project on that the analysis of an heterogenous and massive data.

TU - Berlin

My teaching activities have also foreseen interventions in the international master with a collaboration with the TU-Berlin *(2018-2021)* within " IT for Energy " program.

Data Analytics in Energy Sector Applications *(2018, 2019, 2020)* - lectures, exercises, test grading, office hours. This class explains the best practices in NoSQL database design. This class will also provides a strong understanding of data integrity concepts and a practical demonstration of how those concepts are handled in Relational and NoSQL databases.

Advanced Database Design, Data Management & Integration *(2019)* - lectures, exercises, test grading, office hours. This class focus on the design of Relational databases and show how they can be deployed, populated, and queried using SQL query language and MySQL DBMS.

Computer science and programming methods for energy engineering - campus El-Goona *(2019)* - lectures, exercises, test grading, office hours. It presents the fundamental principles of computer science, object oriented programming and Java language.

Università della Tuscia

My teaching activities at Università della Tuscia with the objective of the internationalization of their program.

Big Data *(2018)* - lectures, exercises, test grading, office hours. The class explains the fundamentals of data science: Big Data, data partitioning and distribution, data clustering.

Alternanza scuola-lavoro *(13-24 April 2018).* Supervision of 23 students for the definition of a Java programming project for Under graduate students. The project focused on the developing of an NP algorithm and on the usage of a database management system.

Università Roma Tre

My teaching activities at Roma Tre University (2008-2012).

Database *(2008, 2009, 2011)* - course for **2nd year Bachelor students** with prof. Paolo Atzeni, exercises, test grading, office hours. The course introduces databases design and data base management systems (DBMS). I curated the practical part that requires to perform experiments in some DBMss (IMB DB2, PostgresSQL, Oracle, SimpleDB) and in some environments for data warehousing (Pentaho).

Object-Oriented Analysis and Design *(2009, 2011, 2012)* - course for **3rd year Bachelor students** with prof. Luca Cabibbo, exercises, test grading, office hours. The course covers Object Oriented Analysis and Design, introduces to agile object oriented programming and to the iterative and incremental software developing approach.

IT Governance *(2012)* - course for **2nd year Master students**, with prof. Paolo Atzeni, lectures, exercise sessions, consultation hours, grading of the exam. The course introduces the principles governing the lifecycle of an IT systems describing the methodological issues, the economic rules and the juridical aspects.

Java Programming and Algorithms *(2007)* - course for **3rd year Bachelor students**, with prof. Luca Cabibbo, laboratory exercise lessons and test grading for the exam. The course introduces the Java programming methodologies.

Remedial Mathematics *(2009)* - course for **1st year Bachelor students**. Math lessons, lectures and exercise sessions. The course introduces math concepts to students. It covers linear algebra, geometric, trigonometry and function analysis.

Contribution to a database book, "Databases - Models and query languages" ("Basi di dati - Modelli e linguaggi di interrogazione"), McGraw Hill 2009, under contract with the publisher. I wrote 100 exercises accompanying the book chapters and I provided the solutions (accessible on-line at `http://www.ateneonline.it/atzeni3e/areastudenti.asp`).

PhD. Thesis Defenses

Thursday, October 29th 2020

- Candidate: Amine GHRAB
- Title: Graph data warehousing
- Director: Dr. Oscar Romero Moral (Universitat Politècnica de Catalunya) Co-director: Dr. Esteban Zimanyi (Université libre de Bruxelles)
- Jury: Prof. Stijn Vansummeren (Université libre de Bruxelles), Dr. Hannes Voigt (Empreses d'Alemanya), Prof. Francesca Bugiotti (CentraleSupélec)

Research Supervision

Thesis supervision

I supervised several students (Master 2 and Bachelor) for their thesis project in computer engineering with my research supervisors prof.Paolo Atzeni (Università Roma Tre - 50 % of supervision), prof. Tetiana Morozyuk (TU-Berlin - 50 % of supervision), and prof. Nacèra Seghouani (CentraleSupélec 50% of supervision).

1. Konstantino Mira: "Energy analysis techniques from literature a full and systematic classification", 2021/22. Konstantinos is going to provide an automatic analysis of energy-literature for providing a classification methodology of thechniques.
2. Lin Siying: "Graph Neural Networks analysis", 2021/22. Lin is going to analyze limits and challenging around neural network analysis in distributed environment.

3. Hem Bhatt: "Comparative DER Data Analysis for Architecture for Energy Consumption Optimization and Control", 2020/21. Hem evaluated techniques for energy consumption optimisation and analysis.
4. Antony Joseph: "A classification of BigData techniques applied in energy sector for the development of new research approaches", 2020/21. Antony studied how BigData are collected, stored and analyzed in energy field.
5. Akshay Tayde: "Evaluation of the IT methodology applied for energy sector and management systems", 2020/21. Akshay evaluated different methodologies used in IT management systems comparing their impact and usage.
6. Shinji Kaneko: "The forecast and impact of day-ahead electricity price in Germany", 2019/20. Shinji studied the elecritcity data-consumpion in Berlin area giving a complete related work study about the methodology developed for energy-consumtpion predictions.
7. Pallavi Katihalli-Manjegowda:"Effective data integration in smart cities for energy analysis", 2019/20. Pallavi studied methodologies for integrating energy data from multiple sources in smart-cities contex.
8. Moditha Hewasinghage: "Modeling Strategy for Storing Data in Distributed Heterogeneous NoSQL Databases", 2016/2017. *Moditha developed a modhel for storing NoSQL data in distributed environments.*
9. Daniele Calabresi: "Integration of Oracle NoSQL into a Platform for the Management of Non-Relational Data Stores", 2011/2012. *Daniele worked on the integration of Oracle NoSQL into the* ***SOS*** *platform integrating multiple, heterogeneous Non-Relational systems. In this context, he also developed NoSQL data indexing strategies.*
10. Tommaso D'Amora: "Integration of Amazon DynamoDB into a Platform for the Management of Non-Relational Data Stores", 2011/2012. *Tommaso worked on the integration of DynamoDB into the SOS platform. He also proposed the introduction of an indexing module that improves query performances.*
11. Marco De Leonardis: "Statistical Databases Management: an Approach Based on Translation Rules", 2011/2012. *Marco worked at Bank of Italy implementing a module that translates high level statistical data manipulation rules into the concrete language supported by the statistical tools used by the Bank. The approach is based on schema mappings and is presented in greater details in.*
12. Luca Rossi: "Heterogeneous Data Management on Innovative Database Systems", 2010/2011. *Luca helped develop a platform that interfaces with multiple heterogeneous NoSQL data stores. The approach is described in.*
13. Luca Tracuzzi: "Methodologies for Data Translation between Heterogeneous Data Models", 2009/2010. *Luca worked on a new internal representation for instances of data organized according to different native data models. This internal (pivot) model is relational, however its organization and the rules describing it vary based on the native data model being thus encoded. Data instances mapped into this internal model can then be stored within a model-independent schema and data translation platform (***MIDST***).*
14. Simone Folino: "Definition of Operators into a Model Management System", 2009/2010. *Simone implemented a subset of the schema evolution operators (project, rename, join, nest etc.) which are used to transform schemas at the level of the metamodel.*
15. Marianna Ciminiello: "Object-Relational Mappings using MIDST", 2009/2010. *Marianna implemented a methodology for expressing object-relational schema mappings within the MIDST high-level data model.*
16. Stefano Mazzoni and Raimondo Tanariva: "Object to Relational Mapping: a Metamodeling Approach", 2008/2009. *Stefano and Raimondo worked together to devise a methodology for implementing object-relational mappings as a particular instance of the model-independent data mapping methodology developed as part of my Ph.D. thesis.*
17. Fabrizio Celli: "Model Independent Data and Schema Translation: a Runtime Approach", 2008/2009. *Fabrizio implemented a model independent data and schema*

translation approach based on the generation of (virtual) views, which enables an efficient run-time access to base data (which may be relational, object-oriented or XML) under a common target model.

18. Andrea Gozzi: "Import and Export of Schema and Data into a Model Management Platform", 2008/2009. *Andrea studied the "Object Relational - XML" model supported by Microsoft SQL Server and IBM DB2, and devised an approach for expressing this model within the common metamodel internal to the MIDST platform. Moreover, this approach also supports importing and exporting data to/from the above mentioned systems and the integration platform (MIDST).*

PhD candidate co-supervision

Molood Arman *(March 2019 - December 2022)*, direction and co-supervision with Nacéra Seghouani (Professor at CentraleSupélec and member of LISN) and Sylvain Wlodarczyk (Schlumberger). *T*hesis subject: Weakly unsupervised approaches for building knowledge bases from geological and petrophysics heterogeneous data sources. *F*unding: Thesis CIFRE.

TECHNICAL SKILLS

- *Languages*: Italian (native), French, English.
- *Programming languages*: Python, Java, C/C++, Prolog, OCaml, Assembly.
- *Databases*: Relational Databases, SQL, No-SQL data stores (MongoDB, Redis, Oracle NoSQL, HBase, DynamoDB).
- *Operating Systems*: Linux, Mac Os, Windows.

COMMUNICATIONS SCIENTIFIQUES

April 22nd 2016 - Seminar - Roma Tre - Database research group, "Flexible Stores and Data".

December 4th 2017 - Invited tutorial - TU-Berlin - research team DIMA / DFKI, Title: "Database Design for NoSQL Systems".

December 5th 2017 - Seminar - TU-Berlin - research team DIMA / DFKI, "Modeling Methodology for a uniform access to NoSQL systems".

July 9th 2018 - Seminar - TU-Berlin - research group DIMA / DFKI, "Interpreting Reputation through Frequent Named Entities in Twitter".

July 4th 2021 - Seminar - CentraleSupélec, Title: "The cartography of the Artificial Intelligence Research in CentraleSupélec".

December 12th 2022 - Seminar - CEA - In the art - DataIA, Title: "Weakly supervised Named Entity Recognition using Deep Neural Networks".

SCIENTIFIC RESPONSIBILITIES

Correspondante HUB AI CentraleSupélec - LISN *(2019 - present)* I am part of the Copil of the HUB AI of CentraleSupélec and I am a correspondent between the HUB and the LISN research laboratory.

The AI Hub was launched in 2020, with the support of the general management and the CentraleSupélec Foundation. At the crossroads of teaching, research and innovation, the HUB as for the purpose the spreading of AI "made in CS" externally but also internally. The HUB wants to create an ecosystem of students, doctoral and post-doctoral students, researchers, professors, and companies through partnerships and actions around entrepreneurship.

Co-responsible for the organization of seminars for the LaHDAK team and the Data Science Department - *(2020 - present)*

Strategic and collaborative activities for the organization of weekly (LaHDAK team) and monthly (Data Science Department) seminars.

Member of the PhD board "Engineering for Energy and Environment" - of Università della Tuscia - *(2018 -present)*

The main objectives of the doctoral board are to plan the whole of the strategic activities of the research doctorate and to verify the status of the planned activities.

Program Committee and Reviews

European Conference on Advances in Databases and Information Systems (ADBIS) 2023, workshop-track chair.

Data-driven Smart Cities (DASC) 2023, ICDE Workshops, member of the program commitee.

Multi-Armed Bandits for Knowledge Discovery (MAB-KG) 2023, ICDM Workshops, member of the program commitee.

Handiversite 2023, member of the program commitee.

Gestion de Donnèes - Principes, Technologies et Applications DBA (2022), member of the DEMO program commitee.

Member of the program committee of the 26th International Conference on Scientific and Statistical Database Management, 2014.

External reviewer for the International Conference on Extending Database Technology (EDBT) in 2012, for the Data & Knowledge Engineering (DKE) Journal in 2013, for the Proceedings of Very Large Data Bases in 2014, and for the ACM SIGMOD Conference in 2014, Journal of Information Systems in 2022, MENACIS 2022.

Research dissemination

Mentorat Solinum *(2020/21)* - Mentoring of Solinum company for data analysis and the application of AI algorithms. Action in collaboration with the school's Entrepreneurship cell. The objective of this collaboration was to provide an environment of scientific expertise to consolidate this project born during the COVID crisis. In my action I participated in the improvement of Soliguide, a tool from the company Solinum, which makes it possible to guide social action thanks to AI.

"A volte ritornano" *(2018-2023)* - Series of scientific seminars held by ex-students of the Liceo Paolo Ruffini in Viterbo. The aim of the seminars is to make research accessible to all and to communicate to people the passion for science. For high school students, the initiative aims to provide useful tools in the choice of their future career, and to inform on current lines of scientific research.

Dalkia: Women's Energy In Transition *(2019 - 2020 - 2022 - 2023)* - Member of the jury - *Dalkia challenge* aims to promote the place of women in the field of energy, to highlight rich and exemplary careers and to encourage young women to join these professions, by participating in its promotion among students. I contributed to this project taking part in the evaluation of the candidates files and the jury.

Project "Summer school for girls" *(2021)* - The University of Paris-Saclay aims to introduce students to disciplines traditionally neglected by girls, in a vast recruitment pool that includes priority areas. I participated in two weeks of initiatives for high school and middle school students through a seminar: "Big Data: the incredible opportunities for storage and interpretation" and the participation in several newsgroups.

European Project

VRAILEXIA European Erasmus+ Project - *(2020-2023)*
Le projet VRAILEXIA is an European Erasmus+ project, prized by Unesco, which aims to change perception and develop a tool model to overcome the main difficulties of dyslexics by strengthening their motivation and self-esteem. The project main objective is to develop a digital platform to support dyslexic students based on AI. My role is in the integration of data that comes from several tests, in several languages, for the evaluation of the profile of dyslexia and the effects of the use of the platform on the psychological aspects.

Research Projects

My research activities mainly focus on the efficient on the analysis and the integration of heterogeneous databases, which has also been the core of my thesis. My research interests are still focused on these problems and have also spread in related-research challenges on: (*i*) independent transformation of schema and data models, (*ii*) uniform access to NoSQL databases, (*iii*) management of data from the Semantic Web in a cloud architecture, (*iv*) efficient integration of "big data" independent of the model, and in the last years (*v*) on the application if Artificial Intelligence techniques on such integrated data.

1. LISN

PROCLAIM - *(2020-2023)*
The objective of this project is to explore and define information extraction approaches and to build learning models to obtain a knowledge base from documents drilling (cuttings), laboratory reports of core data analysis and geological studies in order to automatically provide the a priori information necessary to interpreting logs using the available knowledge and the business rules.

B-GRAP - *(2018-2020)*
The definition of effective strategies for graph partitioning is a major challenge in distributed environments since an effective graph partitioning allows to considerably improve the performance of large graph data analytics computations. In this project we studied and defined a multi-objective and scalable Balanced GRAph Partitioning (B-GRAP) algorithm to produce balanced graph partitions. B-GRAP is based on Label Propagation (LP) approach and defines different objective functions to deal with either vertex or edge balance constraints while considering edge direction in graphs. The experiments are performed on various graphs while varying the number of partitions.

SATT DataForYou - *(2018-2020)*
Participation in the SATT DataForYou project aimed at supporting the creation of the start-up DataForYou, which aims to build tools to support local authorities (for example, town halls, departmental administrations) in France. The objective of the project was to integrate data for the optimization of services provided to citizens by relying on behavioral analysis tools. In this project, I was involved in coaching an engineer on the data integration batch.

APIQA - *(2016-2017)* - [budget : 10KE]
This project had as objective to define a methodology that provides complete answers to queries over data accessible via Web APIs. The project focused on Twitter graph data for the beginning. The project defined a query engine that integrates real-time (or online) queries over the Twitter API with a local (or offline) data source. It was possible to build and maintain the data using a NoSQL graph datastore. Moreover, we focused on different ways of organizing data on the offline datastore, in order to improve the performance of queries and the completeness of the results.

2. Università Roma Tre

SOS - *(2011-2012)*
Save Our Systems (SOS) is a common programming interface [D10] to NoSQL systems. Its goal is to support application development by hiding the specific details of the various systems.
I contributed to the definition to the architecture of the platform, the operations it exposes and the query strategies it implements. I have been involved in the definition of the strategies for integrating the NoSQL data stores into the system. I also participated in the definition of the data storage techniques that are used in each datastore in order to perform operations the interface exposes [IC7,IJ4].

NOAM and ONDM - *(2013-2017)*
NoSQL Abstract Model (NOAM) is a logical approach to the NoSQL database design problem [N15]and aims at exploiting the commonalities of various NoSQL systems. It is based on an intermediate, abstract data model where aggregates are units of distribution (to support

scalability) and consistency (to the extent it is needed). Some intermediate representations can be implemented in target NoSQL datastores, taking into account their specific features and providing an effective support for scalability, consistency, and performance [M16].ONDM (Object-NoSQL Datastore Mapper), is the framework [M17]that supports NOAM approach. It provides application developers with a uniform programming interface, as well as the ability to map application data to different data representations and can be used, in an effective way, for performing the experiments during the design of a NoSQL database.

MIDST - *(2005-2012)*
Model-Independent Schema and Data Translation (MIDST) is a platform for model-independent schema and data translation based on a meta-level approach over a wide range of data models (Relational, OR, OO, ER, XML).
I contributed to the extension of MIDST *supermodel* (a general model handled by the platform that describes the various data models in terms of a small set of *basic constructs*) and I also implemented some core software components like the Datalog-SQL translation engine giving some ideas about the evolution of the platform [N12].

MIDST-RT - *(2008-2012)*
Model-Independent Schema and Data Translation-RunTime (MIDST-RT) is a platform based on MIDST but that implements a runtime approach.
I contributed to the definition, design and implementation of the MIDST-RT algorithm that, given the schema of the source database and the model of the target one, generates views on the operational system that expose the underlying data according to the corresponding schema in the target model. The implemented approach generates views in an automatic way, on the basis of the Datalog rules for schema translation [IJ2,IC6,N14].

MISM - *(2009-2012)*
Model Independent Schema Management (MISM) is a platform for model management that offers a set of operators to manipulate schemas.
I designed and implemented the algorithm that gives one solution to the round-trip engineering problem considering the main model management operators (merge, diff, and modelgen) implemented according to a model-independent and model-aware approaches based on MIDST supermodel [IJ3,IN13].

MATRIX and EXL - *(2012)*
I collaborated with the Bank of Italy, supporting the implementation of EXLEngine, a tool that manipulates statistical data at high level, in terms of entities of statistical models such as time series. We proposed (*i*) a language, EXL, has been defined for the declarative specification of statistical programs, (*ii*) an approach for the translation of EXL code into executables in various target systems has been developed, and (*iii*) a concrete implementation, EXLEngine. The approach leverages on schema mappings as an intermediate specification step, in order to facilitate the translation from EXL towards several target systems [IC5].

GENDATA - *(2013)*
The work regarding data models continues within the GENDATA European project `http://gendata.weebly.com/` that aims at building the abstractions, models, and protocols for supporting a network of genomic data, making them available for genome servers located in the major biologist laboratories in the world. I started to collaborate to the project within the working packages involving Università Roma Tre investigating about the model design, the query language and the model standardization.

3. INRIA

AMADA - *(2011)*
During my internship I contributed to the AMADA project: a platform [D9,N11]for storing Web data (XML documents and RDF graphs) based on the Amazon Web Services (AWS) cloud infrastructure.
I provided a solution for the problem of indexing RDF datasets by using SimpleDB, a key-value store provided by AWS. I contributed to the definition and development of four

indexing strategies [B1,IC8].

ESTOCADA - *(2013-2016)*
A novel system capable of exploiting side-by-side a practically unbound variety of DMSs, all the while guaranteeing the soundness and completeness of the store, and striving to extract the best performance out of the various DMSs. Our system leverages recent advances in the area of query rewriting under constraints, which we use to capture the various data models and describe the fragments each DMS stores. [N15].

PUBLICATIONS

All my publications can be found at:

`http://www.bugiotti.it/publications.html`

The general practice within my scientific community is to *list all authors in alphabetic order* unless one author has contributed very significantly more than her or his share.

Book Chapters

[1] Francesca Bugiotti, Jesús Camacho-Rodríguez, François Goasdoué, Zoi Kaoudi, Ioana Manolescu, and Stamatis Zampetakis. SPARQL query processing in the cloud. In *Linked Data Management.*, pages 165–192. CRC Press, 2014.

Articles in international journals

[IJ2] Paolo Atzeni, Luigi Bellomarini, Francesca Bugiotti, Fabrizio Celli, and Giorgio Gianforme. A runtime approach to model-generic translation of schema and data. *Information Systems*, 37(3):269–287, 2012.

[IJ3] Paolo Atzeni, Luigi Bellomarini, Francesca Bugiotti, and Giorgio Gianforme. MISM: A Platform for Model-Independent Solutions to Model Management Problems. *Journal of Data Semantics*, 14:133–161, 2009.

[IJ4] Paolo Atzeni, Luigi Bellomarini, Francesca Bugiotti, and Marco De Leonardis. Executable schema mappings for statistical data processing. *Distributed Parallel Databases*, 36(2):265–300, 2018.

[IJ5] Paolo Atzeni, Francesca Bugiotti, Luca Cabibbo, and Riccardo Torlone. Data modeling in the nosql world. *Comput. Stand. Interfaces*, 67, 2020.

[IJ6] Paolo Atzeni, Francesca Bugiotti, and Luca Rossi. Uniform access to nosql systems. *Inf. Syst.*, 43:117–133, 2014.

[IJ7] Adnan El Moussawi, Nacéra Bennacer Seghouani, and Francesca Bugiotti. BGRAP: balanced graph partitioning algorithm for large graphs. *J. Data Intell.*, 2(2):116–135, 2021.

[IJ8] Nacéra Bennacer Seghouani, Francesca Bugiotti, Moditha Hewasinghage, Suela Isaj, and Gianluca Quercini. A frequent named entities-based approach for interpreting reputation in twitter. *Data Sci. Eng.*, 3(2):86–100, 2018.

Full articles in international conferences and workshops

[IC9] Molood Arman, Sylvain Wlodarczyk, Nacéra Bennacer Seghouani, and Francesca Bugiotti. PROCLAIM: an unsupervised approach to discover domain-specific attribute matchings from heterogeneous sources. In Nicolas Herbaut and Marcello La Rosa, editors, *International Conference on Advanced Information Systems Engineering (CAiSE)*, volume 386, pages 14–28. Springer, 2020.

[IC10] Paolo Atzeni, Luigi Bellomarini, and Francesca Bugiotti. EXLEngine: executable schema mappings for statistical data processing. In *International Conference on Extending Database Technology (EDBT)*, pages 672–682, 2013.

[IC11] Paolo Atzeni, Luigi Bellomarini, Francesca Bugiotti, and Giorgio Gianforme. A runtime approach to model-independent schema and data translation. In *International Conference on Extending Database Technology (EDBT)*, pages 275–286, 2009.

[IC12] Paolo Atzeni, Francesca Bugiotti, and Luca Rossi. Uniform Access to Non-relational Database Systems: The SOS Platform. In *International Conference on Advanced Information Systems Engineering (CAiSE)*, pages 160–174, 2012.

[IC13] Nacéra Bennacer, Francesca Bugiotti, Jorge Galicia, Mariana Patricio, and Gianluca Quercini. Eliminating incorrect cross-language links in wikipedia. In *Web Information Systems Engineering (WISE)*, pages 109–116, 2017.

[IC14] Nacéra Bennacer, Francesca Bugiotti, Moditha Hewasinghage, Suela Isaj, and Gianluca Quercini. Interpreting reputation through frequent named entities in twitter. In *Web Information Systems Engineering (WISE)*, pages 49–56, 2017.

[IC15] Francesca Bugiotti, Damian Bursztyn, Alin Deutsch, Ioana Ileana, and Ioana Manolescu. Invisible glue: Scalable self-tunning multi-stores. In *Conference on Innovative Data Systems Research (CIDR)*, 2015.

[IC16] Francesca Bugiotti, Luca Cabibbo, Paolo Atzeni, and Riccardo Torlone. Database Design for NoSQL Systems. In *International Conference on Conceptual Modeling (ER)*, pages 1–7, 2014.

[IC17] Francesca Bugiotti, François Goasdoué, Zoi Kaoudi, and Ioana Manolescu. RDF Data Management in the Amazon Cloud. In *Workshop on Data analytics in the Cloud (DanaC 2012)*, 2012.

[IC18] Moditha Hewasinghage, Nacéra Bennacer Seghouani, and Francesca Bugiotti. Modeling strategies for storing data in distributed heterogeneous nosql databases. In Juan Trujillo, Karen C. Davis, Xiaoyong Du, Zhanhuai Li, Tok Wang Ling, Guoliang Li, and Mong-Li Lee, editors, *International Conference on Conceptual Modeling (ER)*, volume 11157 of *Lecture Notes in Computer Science*, pages 488–496. Springer, 2018.

[IC19] Ahmad Khazaie, Nacéra Bennacer Seghouani, and Francesca Bugiotti. Smart crawling: A new approach toward focus crawling from twitter. volume abs/2110.06022, 2021.

[IC20] Siying Li, José Alves, Francesca Bugiotti, and Frédéric Magoulès. A comparison study of graph neural network and support vector machine. In *Distributed Computing and Applications for Business Engineering and Science (DCABES)*, 2022.

[IC21] René Gómez Londono, Sylvain Wlodarczyk, Molood Arman, Francesca Bugiotti, and Nacéra Bennacer Seghouani. Weakly supervised named entity recognition for carbon storage using deep neural networks. In *International Conference on Discovery Science (DS)*, 2022.

[IC22] Venkata Shivaditya Meduri, José Alves, Francesca Bugiotti, and Frédéric Magoulès. Point-cloud-based deep learning models for finite element analysis. In *Distributed Computing and Applications for Business Engineering and Science (DCABES)*, 2022.

[IC23] Venkata Shivaditya Meduri, Francesca Bugiotti, and Frédéric Magoulès. Point-cloud-based deep learning models for finite element analysis. In *Distributed Computing and Applications for Business Engineering and Science (DCABES)*, 2022.

[IC24] Adnan El Moussawi, Nacéra Bennacer Seghouani, and Francesca Bugiotti. A graph partitioning algorithm for edge or vertex balance. In Sven Hartmann, Josef Küng, Gabriele Kotsis, A Min Tjoa, and Ismail Khalil, editors, *Database and Expert Systems Applications - 31st International Conference, DEXA 2020, Bratislava, Slovakia, September 14-17, 2020, Proceedings, Part I*, volume 12391 of *Lecture Notes in Computer Science*, pages 23–37. Springer, 2020.

Demonstrations in international conferences

[D25] Andrés Aranda-Andújar, Francesca Bugiotti, Jesús Camacho-Rodríguez, Dario Colazzo, François Goasdoué, Zoi Kaoudi, and Ioana Manolescu. AMADA: Web Data Repositories in the Amazon Cloud. In *International Conference on Information and Knowledge Management (ACM CIKM)*, 2012.

[D26] Paolo Atzeni, Francesca Bugiotti, and Luca Rossi. SOS (Save Our Systems): a uniform programming interface. for non-relational systems. In *International Conference on Extending Database Technology (EDBT)*, pages 582–585, 2012.

[D27] Francesca Bugiotti, Damian Bursztyn, Alin Deutsch, Ioana Manolescu, and Stamatis Zampetakis. Flexible hybrid stores: Constraint-based rewriting to the rescue. In *International Conference on Data Engineering (ICDE)*, pages 1394–1397, 2016.

Articles and demos at national database conférences

[N28] Rana B. AL-Otaibi, Francesca Bugiotti, Damian Bursztyn, Alin Deutsch, Ioana Manolescu, and Stamatis Zampetakis. Estocada : Stockage hybride et ré-écriture sous contraintes d'intégrité. In *Journées des Bases de Données Avancées (BDA)*, 2016.

[N29] Andrés Aranda-Andújar, Francesca Bugiotti, Jesús Camacho-Rodríguez, and Zoi Kaoudi. AMADA: Web Data Repositories in the Amazon Cloud. In *Journées des Bases de Données Avancées (BDA)*, 2012.

[N30] Paolo Atzeni, Luigi Bellomarini, Francesca Bugiotti, and Giorgio Gianforme. From Schema and Model Translation to a Model Management System. In *British National Conference on Databases (BNCOD)*, pages 227–240, 2008.

[N31] Paolo Atzeni, Luigi Bellomarini, Francesca Bugiotti, and Giorgio Gianforme. A platform for model-independent solutions to model management problems. In *Italian Symposium on Advanced Database Systems (SEBD)*, pages 310–317, 2008.

[N32] Paolo Atzeni, Luigi Bellomarini, Francesca Bugiotti, and Giorgio Gianforme. A runtime approach to model-independent schema and data translation. In *Italian Symposium on Advanced Database Systems (SEBD)*, pages 245–252, 2009.

[N33] Quinio Bernard, Antoine Harfouche, and Francesca Bugiotti. Human-centric ai to mitigate ai biases: The advent of augmented intelligence. In AIM, editor, *Conférence de l'Association Information et Management (AIM)*, 2022.

[N34] Francesca Bugiotti, Damian Bursztyn, Alin Deutsch, Ioana Ileana, and Ioana Manolescu. Toward Scalable Hybrid Stores. In *Italian Symposium on Advanced Database Systems (SEBD)*, pages 312–319, 2015.

[N35] Francesca Bugiotti and Luca Cabibbo. A Comparison of Data Models and APIs of NoSQL Datastores. In *Italian Symposium on Advanced Database Systems (SEBD)*, pages 63–74, 2013.

[N36] Francesca Bugiotti, Luca Cabibbo, Paolo Atzeni, and Riccardo Torlone. How I Learned to Stop Worrying and Love NoSQL Databases. In *Italian Symposium on Advanced Database Systems (SEBD)*, pages 216–223, 2015.

[N37] Nacéra Bennacer Seghouani, Francesca Bugiotti, Jorge Galicia, Mariana Patricio, and Gianluca Quercini. Élimination des liens inter-langues erronés dans wikipédia. In Mustapha Lebbah, Christine Largeron, and Hanane Azzag, editors, *Conférence francophone sur l'Extraction et la Gestion des Connaissances (EGC)*, volume E-34 of *RNTI*, pages 397–402. Éditions RNTI, 2018.

Manuscripts

[M38] Molood Arman, Sylvain Wlodarczyk, Nacéra Bennacer Seghouani, and Francesca Bugiotti. Ocrana: Optical character recognition analytics. 2022.

[M39] René Gómez Londono, Sylvain Wlodarczyk, Molood Arman, Francesca Bugiotti, and Nacéra Bennacer Seghouani. Weakly supervised named entity recognition for carbon storage using deep neural network. 2022.