

Ioana Manolescu-Goujot

Current position

Senior researcher (*directeur de recherche 1re classe*) at Inria Saclay-Île-de-France (SIF)
Head of the CEDAR project-team of Inria Saclay-Île-de-France and LIX (CNRS and Ecole Polytechnique)
Professeuse chargée de cours (part-time) at Ecole Polytechnique
Email: ioana.manolescu@inria.fr, web: <http://pages.saclay.inria.fr/ioana.manolescu>

Research interests

I am interested in models, algorithms and techniques for efficiently managing large volumes of data with complex structure and semantics. The core scientific areas of my research pertain to query optimization through algebraic techniques, and to storage optimization through indexes and materialized views. A central aspect of my research is distributed data management, in particular in massively parallel and cloud environments.

According to Google Scholar (<https://scholar.google.com/citations?user=q6Ft35wAAAAJ>), my H-index is 37.

Career summary

01/2016 -	Lead of the CEDAR team with Ecole Polytechnique (http://team.inria.fr/cedar); <i>directeur de recherche 1re classe</i> since 2017.
01/2010-12/2015	Lead of the LEO team, then OAK project, with Université de Paris-Sud; <i>directeur de recherche 2e classe</i> since 2010.
09/2002-12/2009	Junior Inria researcher (<i>chargée de recherche</i>).
03/2002-08/2002	Post-doc at Politecnico di Milano, Italy.
12/1998-12/2001	PhD at Université de Versailles and Inria Rocquencourt.

Diplomas

Habilitation à Diriger des Recherches: “Optimization Techniques for XML Query Processing”, Université de Paris Sud-XI, 2009. My HDR defense committee included: Serge Abiteboul (Inria Saclay-Île de France), Véronique Benzaken (Université de Paris Sud-11) Torsten Grust (University of Tuebingen), Donald Kossmann (ETH Zurich) Philippe Pucheral (Université de Versailles) and Patrick Valduriez (Inria Sophia Antipolis – Méditerranée).

PhD thesis: “Optimization techniques for querying heterogeneous distributed data”, 2001, Université de Versailles. Supervised by Patrick Valduriez and Daniela Florescu.

Magistère de Mathématiques Fondamentales et Appliquées et d’Informatique en juillet 1999, Ecole Normale Supérieure Paris and Université Paris VI.

Master at Université de Paris VI and Ecole Normale Supérieure Paris, 1998.

Undergraduate at Ecole Normale Supérieure Paris and Université de Paris-Sud (1996-1997), and at Universitatea Politehnica Bucuresti, Romania (1993-1996).

Teaching experience

Ecole Polytechnique (2016-) Since 2017, I am a part-time professor (*professeuse chargée de cours*) at Ecole polytechnique. I have taught in 2016 and 2017 Database Management Systems (Inf553) to third year students as part of the Data Science track (90 registered students). Since 2018, I also teach an advanced undergraduate project-based course on graph data management.

Aalborg University, Denmark (2016) I have taught a two-days course on Massive Open Data Analytics. The students were Master and PhD students from Aalborg and a few other Denmark universities; the course was evaluated through a coding project on which the students worked for six more weeks.

Université de Paris Saclay (2015-2016) I have designed and taught a course (including lab projects- on Architectures for Big Data Management in the 2nd year Master on Data and Knowledge (D&K).

Université de Paris Sud (2002-2014) I have taught a 2nd year Research Master course covering Big Data technologies around Map Reduce, peer-to-peer data management, Web data management, XML query processing, Web services etc. Some of the material contributed to the “Web Data Management” book I co-authored¹.

Politecnico di Milano, Italy (2008) As a visiting professor, I have taught a “Web Data Management” course to 1st year Master students for three months. .

Université de Paris Dauphine (2006-2012) I have designed and taught a 2nd year Research Master optional course on Web services and workflows.

Earlier In 1999-2005, I have taught courses on Databases, Web data management, and XML databases at Master level (first and second year) at Université d’Orsay (MIAGE), Université de Versailles and Télécom ParisTech.

Tutorials I have had a longstanding interest in systematizing and disseminating knowledge on data management, Web and knowledge representation research topics. Since 2002, I had the opportunity to devise and present 14 such tutorials in summer schools and international conferences, typically selected through peer review. The full list of tutorials appears in the separate Publications document submitted with this part-time PCC application.

Conference and editorial responsibilities

- **General chair** of IEEE International Confernece on Data Engineering 2018
- **Associate editor** of PVLB (2017 and 2018), member of the **PVLDB Endowment Board of Trustees** (since 2016).
- **Editor-in-chief** of the quarterly ACM SIGMOD Record journal from 2010 to 2014 (www.sigmod.org/record).
- **Associate editor** of the ACM Transactions on Internet Technology (TOIT) from 2010 to 2014, of the ACM Transactions on the Web (2015-2016), **member of the editorial board** of the Proceedings of VLDB (PVLDB) in 2008-2011.
- **Program chair** of the SSDBM 2016 and EDBT 2010 conferences, and **track (co-)chair** for major conferences on databases and the Web such as PVLDB, SIGMOD, ICDE, EDBT, WWW, and ACM CIKM.
- **Co-chair of the international workshops** Programming Languages for XML (PLAN-X) 2008, in cooperation with ACM SIGPLAN, and Web and Databases (WebDB) 2007, in cooperation with ACM SIGMOD.
- **Program committee member** of more than **70** international conferences such as SIGMOD, VLDB, ICDE etc. and workshops.

Collaborative projects

My research contributed to many projects, the main ones being listed below. Those that I coordinated appear in bold.

International: “**WebClaimExplain**” associated team between Inria and AIST Japan (2017-2019); “**DataBridges: Data Integration for Digital Cities**”, a KIC EIT ICT Labs activity (2011-2012); “**OakSad**” associated team between Inria and UC San Diego (2013-2015); “Europa: cloud-based architectures for scalable data management”, KIC EIT ICT Labs activity (2012-2013); “**XClean: Declarative Data Cleaning Techniques for XML**”, bilateral project with Portugal (2005-2007); EU FP6 project eDOS on the efficient distribution of open-source software (2004-2007).

National: “iCODA: Knowledge-mediated Content and Data Interactive Analytics, The case of data journalism” Inria Project Lab (IPL), 2017-2020; “**ContentCheck: Models, Algorithms and Tools for Journalistic Fact-Checking**”, national project (2016-2019); “**Scalable and secure tools for large-scale data management**” task of the DigiCosme LABEX (IT network of excellency, 2012-2016); “Business Models for the Cloud” national ANR project (2013-2017); Datalyse *AAP Cloud & Big Data 2012* (2012-2016) project on Big Data analytics in the cloud; “**Coordination, dynamicity and efficiency for XML**” national ANR project (2009-2012); “**EdiFlow: data-intensive workflows for visual analytics**” regional Digiteo project (2008-2011). “**Peer-to-peer**” work package of the WebContent RNTL project with Thales, Bongrain, EADS and other academics on generic platforms for warehousing Web data (2005-2008).

¹<http://www.cambridge.org/aus/catalogue/catalogue.asp?isbn=9781107012431>

Software My research is systematically validated through the development of software prototypes, some of which grow quite complex and mature, and may involve up to 10 person \times year of development by PhD students, interns, engineers, and myself. Some of these have been distributed in open-source:

RDFSsummary is a tool for summarizing large RDF graphs, while reflecting both explicit data and implicit data that results from the graph through reasoning; a summary can be used as a static analysis tool for query optimization, or as a first GUI to help users formulate queries (<https://team.inria.fr/cedar/projects/rdfquotient/>)

CliqueSquare is a Hadoop-based platform for large-scale RDF data management, approx. 40.000 lines of Java code (<http://sourceforge.net/projects/cliquesquare/>).

Amada is a platform for storing Web data (in particular, XML documents and RDF graphs) based on the Amazon Web Services (AWS) cloud infrastructure (<https://team.inria.fr/oak/projects/amada/download-amada/>)

Awards

ACM Recognition of Service Award in Appreciation for Contributions to ACM, as a workshop Co-Chair of the 1st Int. Workshop on the Performance and Evaluation of Data Management Systems (ExpDB), in June 2006.

Supervision

I have supervised or co-supervised **12 graduated PhD students**: 6 of them hold senior positions in technology companies in France, 2 are academics (in France and Japan), 3 work abroad in Microsoft Research, SAP, and respectively HortonWorks and one is still a post-doc in Canada. I have also advised **6 post-docs** (4 of which are now faculty or researchers in France, Italy, and Qatar), and **17 interns** which were 2nd year CS Master students or engineering students in their final year of study.