

Ioana Manolescu

Informations personnelles

E-mail ioana.manolescu@inria.fr, web: <https://pages.saclay.inria.fr/ioana.manolescu>

Expérience professionnelle

Recherche

- 2016- **Responsable de l'équipe-projet CEDAR** (Rich Data Analytics at Cloud Scale), équipe-projet du centre Inria de l'Institut Polytechnique de Paris. L'équipe mène des recherches sur la gestion des données complexes à grande échelle, l'analyse des réseaux sociaux et le traitement du langage naturel.
Directrice de recherche Inria 1re classe (2017-)
- 2025- **Titulaire d'une Fellowship du Cluster IA Hi!Paris**, PREDIAL: AI Data Dialogs for the Press
- 2019-2023 **Directrice scientifique de LabIA**, visant à déployer des méthodes d'intelligence artificielle en réponse à des besoins de l'administration publique française.
- 2010-2015 **Responsable de l'équipe LEO puis OAK** (Distributed and Heterogeneous Data and Knowledge), équipe-projet du centre Inria de Saclay, équipe commune avec l'Université de Paris XI. L'équipe a travaillé sur la gestion de grands volumes de données distribuées, notamment dans des architectures *cloud*, et sur l'intégration de données à l'aide des ontologies.
Directrice de recherche Inria (2010), précédemment Chargée de Recherche (CR) Inria (2002)
- 2002 **Post-doctorante**, *Politecnico di Milano*, Milan, Italie
Modélisation d'applications Web s'appuyant sur des bases de données (AWBD), afin d'y intégrer des services Web, ainsi que sur la compilation de processus structurés (workflows) vers des AWBD. Ces travaux ont été intégrés et exploités dans l'outil de l'entreprise WebRatio, issue des recherches de l'équipe.
- 1999 **Stagiaire**, *Lucent - Bell Labs*, Murray Hill, New Jersey
Algorithmes pour la sélection automatique de schémas de stockage efficace des données XML

Enseignement

- 2017-2026 **Professeure Chargée de Cours** temps partiel 50%, *Ecole Polytechnique*, Palaiseau, France
Enseignements de 3e année (M1, en anglais: cours "Database Management Systems", co-responsable du parcours IA et Data Science, des stages de recherche de ce parcours, et d'un cours-projet de recherche de 6 mois à temps partiel, permettant aux élèves polytechniciens d'avoir une première expérience de recherche.
- 2016 **Professeure invitée**, *Aalborg University*, Aalborg, Danemark
Cours M2/PhD "Scalable Tools for RDF Data Analytics" (16h)
- 2011-2024 **Professeure (vacataire)**, *Institut Mines Télécom*, Paris puis Palaiseau, France
Cours M2 "Big Data Architectures" (15-21h/an).
- 2010-2014 **Professeure (vacataire)**, *Université de Paris Dauphine*, Paris, France
Cours M2 "Services Web" (18h/an)
- 2008 **Professeure invitée**, *Politecnico di Milano*, Milan, Italie
Cours M2 "Novel Architectures for Data Integration" (24h)
- 2002-2010 **Enseignante (vacataire)**, *Université Paris Sud*, Orsay, France
Cours M2 "Semistructured Data", avec S. Abiteboul (12h/an)

Diplômes et études

- 2009 **Habilitation à diriger les recherches en informatique**, *Université Paris Sud*, Orsay, France
"Optimization techniques for XML Query Processing"
Parrain: Serge Abiteboul (Inria).
Rapporteurs: Donald Kossmann (ETH Zurich), Torsten Grust (Universität Tübingen), Philippe Pucheral (Université de Versailles Saint-Quentin et Inria).
Autres membres du jury: Véronique Benzaken (Université de Paris Sud), Patrick Valduriez (Inria)
- 2001 **Thèse en informatique**, *Université de Versailles Saint-Quentin et Inria Rocquencourt*, Versailles, France
"Optimization techniques for querying distributed heterogeneous data sources"
Directeurs de thèse: Daniela Florescu et Patrick Valduriez (Inria).
Rapporteurs: Serge Abiteboul (Inria) et Tamer Özsu (Université de Waterloo, Canada).
Autres membres du jury: Eric Simon (Business Objects), Philippe Pucheral (Inria)
- 1999 **Magistère en Mathématiques Fondamentales et Appliquées et Informatique**, *ENS Paris et Université Paris 6*, Paris, France
"Query optimization with binding patterns"
- 1998 **DEA Systèmes Distribués**, *Université Paris 6*, Paris, France
- 1996-1997 **Elève étrangère (concours G/S 1996)**, *Ecole Normale Supérieure*, Paris, France

Prix et distinctions

- 2024 ACM SIGMOD Distinguished PC Member
- 2020,2022 PVLDB Outstanding Reviewer
- 2021 **ACM Senior Member**
- 2020 **ACM SIGMOD Contribution Award** pour avoir initié (en 2008) et présidé (en 2008, 2009, et 2010) l'initiative de Reproductibilité Expérimentale en conjonction avec la conférence ACM SIGMOD. Ces efforts dans la communauté de gestion de données ont été parmi les premiers de ce genre en informatique.
- 2015 **Google Computational Journalism Award**, un financement de 70K€ pour la recherche sur des usages innovants des données dans le journalisme

Responsabilités liées à la recherche

Sociétés savantes

- 2023- **Présidente du comité de pilotage BDA** (élue pour 4 ans), membre du comité depuis 2013. Ce comité de 20 personnes organise la conférence annuelle en gestion de bases de données (BDA), et une école d'été tous les deux ans. En 2019, j'ai été à l'initiative de la création du prix de thèse BDA, pour la meilleure thèse soutenue en France dans le domaine chaque année
- 2022- Membre de **ELLIS European Network of Excellency in AI**
- 2016-2022 Membre élue du **PVLDB Endowment Board**, l'association internationale (non-profit) qui organise la conférence annuelle VLDB et gère le journal PVLDB
- 2016-2018 Membre, puis co-présidente du **ACM Jim Gray Dissertation Committee**, qui décerne chaque année un prix à la meilleure thèse dans la communauté internationale de recherche sur la gestion de données
- 2013-2016 Membre du **IEEE Technical Committee of Data Engineering (TCDE) Award Committee** qui décerne des prix annuels (prix de carrière, meilleure contribution, etc.)
- 2010-2014 Membre du **ACM SIGMOD Executive Committee**, en charge de l'organisation de la conférence SIGMOD

Responsabilités institutionnelles

- 2023- Représentante d'Inria Saclay dans le **Comité Opérationnel de Hi!Paris**, un pôle d'excellence puis (depuis 2025) Cluster IA réunissant l'Institut Polytechnique de Paris (Ecole polytechnique, ENSTA, Institut Mines Télécom, etc.) ainsi que HEC. J'ai contribué à la rédaction du projet de Cluster, qui a reçu en 2025 un financement de 85M€
- 2023- et Membre élue dans la **Commission d'Evaluation Inria**, qui organise les concours et les campagnes
2014-2019 de promotions des chercheurs
- 2023- Membre élue du **Conseil Académique de l'Institut Polytechnique de Paris**
- 2019-2024 Responsable du **pôle "Data Analytics and Machine Learning" (DAML)** au sein du **LIX (UMR 7161, CNRS/Ecole polytechnique)**, membre du conseil de direction
- 2016-2022 Membre du **Bureau du Comité des Equipes-Projets Inria d'Inria Saclay**, un comité d'une dizaine de personne qui instruit et prépare en amont les évolutions scientifiques du centre Inria (nouvelles équipes, grands projets, évolutions institutionnelles etc.)

Activité éditoriale

J'ai été **présidente** des journées BDA 2022, **co-présidente et organisatrice** de la conférence IEEE ICDE (2018), **présidente du comité de programme** des conférences ICWE (2021), SSDBM (2016), BDA (2012) et EDBT (2010). J'ai été à plus de 20 reprises **présidente de section (track)** des conférences internationales ACM SIGMOD, VLDB, EDBT, ACM CIKM, WWW (The Web Conference). J'ai fait partie de plus de 70 comités de programmes de conférences internationales de rang **A** ou **A*** (classement CORE), notamment ACM SIGMOD, IEEE ICDE, EDBT, ACM CIKM, The Web Conference, ACL (American Computational Linguistics) etc.

J'ai été **éditrice** de journaux tels que VLDB Journal et PVLDB, et **éditrice en chef** de ACM SIGMOD Record. Plus de détails sur mon implication dans des conférences et journaux sont disponibles online.

Principaux projets de recherche financés

- 2025- **Coordinatrice** du projet **ANR TOPOL**, avec LISN (U. Paris Saclay), LAMSADE (U. Dauphine), LIFAT (U. Tours), LIFO (U.Orléans) ainsi que l'International Consortium of Investigative Journalism (ICIJ) (partenaire associé). Le projet vise à développer des méthodes de gestion de données, extraction d'information et traitement du langage naturel pour faciliter l'interrogation et l'exploration des lacs de données hétérogènes.
- 2022-2027 **Responsable de site** dans le projet ELIAS (European Lighthouse of AI for Sustainability), projet Horizon Europe. Mon équipe y étudie des méthodes de fact-checking et interrogation de données ouvertes sur des sujets de débat dans les sociétés européennes
- 2020-2024 **Titulaire de la Chaire ANR IA SourcesSay**, un projet qui a développé des méthodes et logiciels pour analyser des "arènes" (ou lacs) de données hétérogènes, structurées, semi structurées, et non structurées, interconnectées à l'aide de l'extraction d'information (entités nommées et relations). Les recherches de SourcesSay ont été inspirées par des problèmes soulevés par Le Monde et Radio France, partenaires associées. De cette chaire est résulté le système StatCheck, transféré à RadioFrance, qui a été mentionné par la Ministre de la Culture comme participant à la protection de l'intégrité de l'information.
- 2018-2023 **Responsable de site** dans le projet ANR CQFD (Requêtes Ontologiques Complexes sur des Fédérations de Données Hétérogènes)
- 2015-2020 **Coordinatrice** du projet ANR **ContentCheck**, avec LIMSI (U. Paris Saclay), LIRIS (U. Lyon 1), U. Rennes et le quotidien Le Monde. Le projet visait à étudier des modèles, algorithmes et outils basés sur les données pour le journalisme de données et la vérification d'informations. Dans ce projet, j'ai démarré la recherche sur le fact-checking statistique (première ébauche de StatCheck). Nous avons collaboré avec Les Décodeurs (équipe de fact-checking du Monde) puis avec Stéphane Horel, journaliste d'investigation.

- 2011-2012 **Coordinatrice** d'une activité KIC EIT ICT Labs (action financées par l'EU sur des durées d'un) "DataBridges: Data Integration for Digital Cities", avec DFKI (Allemagne), U. Aalto (Danemark), TU Delft (Pays-Bas) et KTH (Suède).
- 2009-2012 **Coordinatrice** du projet ANR CODEX (Coordination, données dynamiques et performance pour des données XML), avec Inria Grenoble, Inria Lille, Inria Saclay, Univ. Tours, Univ. Paris 7, Univ. Paris Sud et la start-up Innovimax.

Encadrement

Doctorants: thèses en cours (5)

- 2026- Asuncion Carolina Gomez Colomeri, co-encadrée avec Madhulika Mohanty (CRCN Inria, équipe CEDAR): *"Towards Effective Data Discovery in a Heterogeneous Data Lake"*
- 2025- Przemyslaw Dominikowski, co-encadré avec Madhulika Mohanty (CRCN Inria, équipe CEDAR): *"Indexing and Querying Heterogeneous Data Lakes"*
- 2025- Marijan Soric, co-encadré avec Pierre Senellart (ENS Paris): *"Exploitation and Structuring of Heterogeneous Geological Data and Knowledge"*
- 2024- Hritika Kathuria, co-encadrée avec Maxime Buron (U. Clermont Auvergne): *"Decentralized semantic data sharing with access control"*
- 2023- Antoine Gauquier, co-encadré avec Pierre Senellart (ENS Paris): *"Intelligent construction of a multimodal and heterogeneous data warehouse"*

Doctorants: thèses soutenues (18) et emplois actuels

- 2022-2025 Ghufuran Khan, co-encadré avec Angelos Anadiotis (professeur assistant à l'Ecole polytechnique dans l'équipe CEDAR, puis chercheur à Oracle Zurich): *"Scalable Analytics on Multi-Streams Dynamic Graphs"*. Chercheur post-doctorant à INP Grenoble
- 2022-2025 Kun Efimov-Zhang, co-encadré avec Oana Balalau (Inria Starting Faculty dans CEDAR): *"Contributions to evaluating and improving the faithfulness for text generation"*. Chercheur post-doctorant au LORIA
- 2021-2024 Nelly Barret, **Accessit au prix de thèse BDA**: *"User-oriented exploration of semi-structured datasets"*. Maîtresse de Conférences à l'INSA Lyon
- 2018-2021 Paweł Guzewicz, co-encadré with Yanlei Diao (professeur à l'Ecole polytechnique, équipe CEDAR): *"Expressive and Efficient Analytics for RDF Graphs"*. Ingénieur de recherche chez Huawei Edinburgh, Royaume-Uni
- 2017-2020 Maxime Buron, co-encadré with François Goasdoué (U. Rennes 1) and Marie-Laure Mugnier (LIRMM): *"Efficient reasoning on large and heterogeneous graphs"*, **Prix de thèse BDA**. Maître de conférences à l'Université Clermont Auvergne.
- 2016-2019 Tien-Duc Cao, co-encadré with Xavier Tannier (U. Paris Sud puis U. Paris Descartes): *Toward Automatic Fact-Checking of Statistic Claims*. Ingénieur ML chez MadKudu, France
- 2013-2016 Damian Bursztyn, co-encadré avec François Goasdoué: *Efficient Big Data query answering in the presence of constraints*. Director of engineering chez Payfit, une "licorne française"
- 2012-2016 Raphaël Bonaque, co-encadré with Bogdan Cautiș (U. Paris Sud) and François Goasdoué (U. Paris Sud): *"Structured, social and semantic search"*. VP ingénierie chez Monk AI, France
- 2012-2015 Stamatis Zampetakis, co-encadré avec François Goasdoué (U. Paris Sud): *"Massively parallel processing for Web data"*. Ingénieur sénior chez Cloudera
- 2011-2014 Jesús Camacho-Rodríguez, co-encadré avec Dario Colazzo (U. Paris Sud): *"Optimization Techniques for Semistructured Semantic Web data"* Principal engineering manager chez Microsoft Research, USA.

- 2011-2014 Alexandra Roatiş, co-encadrée avec François Goasdoué (U. Paris Sud) and Dario Colazzo (U. Paris Sud): *"Data Warehousing for the Semantic Web"*. Ingénieure chez Google, Canada
- 2010-2013 Julien Leblay, co-encadré with François Goasdoué (U. Paris Sud): *"Database techniques for semantics-rich semi-structured Web data"*. Ingénieur chez Rakuten, Japon
- 2010-2013 Asterios Katsifodimos: *"Scalable View-based Techniques for Web Data: Algorithms and Systems"*. Professeur associé à Technical University Delft, Pays-Bas
- 2009-2012 Konstantinos Karanasos, co-encadré avec François Goasdoué. *"View-based Technniques for the Efficient Management of Web Data"*. Ingénieur de recherche chez Meta, USA
- 2008-2011 Wael Khemiri, co-encadré avec Jean-Daniel Fekete (Inria) et Véronique Benzaken (U. Paris Sud): *"Data-intensive Interactive Workflows for Visual Analytics"*. Ingénieur chez Lemo, Suisse
- 2006-2009 Spyros Zoupanos, co-encadré avec Serge Abiteboul: *"Efficient peer-to-peer data management"*. Ingénieur sénior chez Deloitte, Grèce
- 2004-2008 Nicoleta Preda, co-encadrée avec Serge Abiteboul: *"Efficient Web resource management in structured peer-to-peer networks"*. Maîtresse de Conférences à l'Université de Paris Saclay
- 2004-2007 Andrei Arion, co-encadré avec Véronique Benzaken: *"Structured Materialized Views for XML Queries"*. Ingénieur chez Younited

Post-doctorants (2 en cours, 8 passés) et leurs emplois actuels

- 2024- Georgios Siachamis (PhD TU Delft, Pays-Bas): *"Efficient Sharing and Querying of Graph Data"*
- 2023- Garima Gaur (PhD IIT Delhi, Inde): *"NLP and Data Management for Digital Democracy"*
- 2021-2023 Madhulika Mohanty (PhD IIT Delhi, Inde): *"Graph Querying and Heterogeneous Data Exploration"*, Chargée de recherche dans l'équipe CEDAR
- 2021-2022 Prajna Upadhyay (PhD IIT Delhi, Inde): *"Abstracting Heterogeneous Data and Open Information Extraction"*, Professeure Assistante à IIT Hyderabad
- 2018-2021 Mirjana Mazuran (PhD Politecnico di Milano, Italy): *"Issue-Driven European Arena Analytics"*, Senior Data Scientist chez Enablon, Paris, France
- 2014-2015 Francesca Bugiotti (PhD Università de Roma, Italy): *"Logical Data Independence for the Cloud"*, Professeure Assistante à Centrale-Supélec, Gif-sur-Yvette, France
- 2014-2015 Ioana Ileana (PhD Télécom Paris): *"Rewriting Queries over Heterogeneous Data Models"*, Maîtresse de Conférences, Université Paris Descartes
- 2013-2014 Soudip Roy Chowdhury (PhD Università di Trento): *"Task Factorization in Massively Parallel Semistructured Data Processing"*, Senior Product Executive chez Asper.AI, USA
- 2012-2014 Zoi Kaoudi (PhD University of Athens): *"Massively Parallel Evaluation of RDF Queries"*, Professeure Assistante, IT University de Copenhague, Danemark
- 2006-2007 Daniele Braga (PhD Politecnico di Milano): *"Incremental maintenance of XML views"*, Maître de Conférences, Politecnico di Milano, Italie

Ingénieurs et stagiaires

J'ai encadré ou co-encadré une douzaine de jeunes ingénieurs, impliqués dans nos différents projets de recherche et de transfert. Certains ont poursuivi par des thèses que j'ai encadrées (J. Leblay, J. Camacho Rodríguez), d'autres ont créé des start-ups (S. Ebel et T. Galizzi) ou poursuivent des carrières dans l'industrie.

J'ai (co-)encadré plus de trente stagiaires au niveaux M1 et M2. Une grande partie de ces stages a conduit à des publications: S. D'Aguanno, A. Arion, N. Barret, P. Bertaud-Velten, I. Burger, C. Chaniel (**Prix de stage de l'Ecole polytechnique 2018**), R. Dziri, A. Gauquier, J.-J. Law, Q. Massonnat (**Prix de stage 3A des Laboratoires de l'Ecole polytechnique 2022**), G. Thiry et R. Vijay. J'ai collaboré et publié aussi avec des doctorants étrangers en stage ou visite dans mon équipe: R. Al-Otaibi de l'Université de Californie à San Diego, M. Goodfellow de l'Université de Strathclyde, D. Sileo de l'Université de Calabre, M. Herschel (maintenant M. Weis) du Hasso-Plattner Institute.

Dissemination et vulgarisation

Tutoriels

- 2002- Dès la première année après ma thèse, j'ai conçu et présenté, seule ou avec un ou des collègues, des tutoriels de trois heures ([T152] à [T172] ci-dessous), systématisant et organisant les connaissances d'un domaine de mes recherches, et les disséminant dans la communauté. Mes tutoriels sur le fact-checking [T153, T160, T167], en 2018-2019, ont contribué à le faire reconnaître comme un domaine d'application d'importance cruciale; selon du feedback que j'ai reçu, ils ont inspiré d'autres chercheurs à s'y intéresser.

Présence dans les médias

- 2021- J'ai répondu à des sollicitations de journalistes sur le rôle de l'**IA** dans les **deep fakes**, la manipulation des utilisateurs par la désinformation, et plus récemment sur les perspectives et impacts réels mais aussi fantasmés de ces technologies. Ces contributions sont apparues notamment dans Les Echos (2018, 2024, 2025), Libération, L'Humanité, La Croix, TF1, Le Figaro, France Culture, France Inter, etc.
- 2015- Mes travaux sur le **fact-checking** ont attiré de l'intérêt dans la presse généraliste. Après un premier article dans Le Monde en 2015, suivi par Ouest France, Rue89 et d'autres, j'ai publié une contribution dans le blog Binaire, présentant les enjeux du fact-checking et les défis que les chercheurs peuvent soulever dans ce domaine. D'autres articles de presse issus de mon travail dans ce domaine sont parus dans Le Monde (2016, 2018), Sciences et Avenir, 20 minutes, La revue des médias (INA) etc.

Rencontres d'autres communautés

- 2019- J'ai répondu à des invitations de parler de ma recherche dans des rencontres des **journalistes**: Journées Sciences et Médias 2025, Médias en Seine 2023, SciCar (Science Meets Computer-Assisted Reporting) 2023, Médias en Seine 2022, DataHarvest 2020. Ponctuellement, j'ai aussi décrit mes travaux dans des conférences en **économie** (Les Journées de l'Economie 2019, Journées Rencontres de l'Association Française du Marketing 2021), et **sciences sociales** (Colloque Post-vérité et Infox 2019 organisé par la Fondation Maison des Sciences de l'Homme).

Création d'une communauté interdisciplinaire

- 2023- Avec François Bancelhon, je suis à l'initiative de la création du workshop **Infox sur Seine**, qui réunit informaticiens, chercheurs en sciences sociales et sciences sociales computationnelles, en géopolitique, dans les études des médias, ainsi que des personnels des services de l'Etat comme Viginum, autour des thématiques liées à la désinformation. J'ai présidé le comité de programme au lancement, j'en fais partie chaque année, et j'y ai fait une keynote. Le workshop a réuni 100 participants en 2024, et grandit depuis (125 présents en 2026).