

Sandro Fiore, Ph.D.
Senior Scientist

Senior Scientist and Head of the Data Science and Learning Research Group at the Advanced Scientific Computing Division of the Euro-Mediterranean Center on Climate Change Foundation, Italy.

I hold 15 years' experience on scientific data management topics with a comprehensive knowledge of databases, big data, high-performance data analytics, parallel and distributed computing.



Personal Information

Email

sandro.fiore1976@pec.it

Date of birth

02/06/1976

Nationality

Italian

Social media:

<https://www.linkedin.com/in/sandro-fiore-ph-d-080ba38/>

WORK EXPERIENCE

March 19, 2018 - Present

Visiting Scientist at the University of Chicago (Illinois, USA)

- Working at the University of Chicago (Department of Computer Science) on High Performance Data Analytics Workflows for eScience at scale (in particular for Climate Change use cases on HPC infrastructures) as well as analytics provenance and computational reproducibility (supervisor Prof. Ian Foster).

October 8, 2018 - Present

Head of the Data Science and Learning Research Group of the CMCC Advanced Scientific Computing Division

- Supervision and coordination of the CMCC Data Science and Learning Research Group
- Design and implementation of High-Performance Data Analytics and Machine Learning (HPDA&ML) solutions for eScience (in particular in the climate change domain)
- Principal Investigator of the **Data Science and Learning** CMCC Strategic Project
- Scientific Coordinator deputy of the **OFIDIA2** INTERREG Project
- Scientific officer for CMCC of the H2020 **ESIWACE2** and H2020 **EOSC-hub** projects.

Nov 2014 – Oct, 2018

Director of the Advanced Scientific Computing (ASC) Division at the Euro-Mediterranean Center on Climate Change (CMCC), Italy

- Supervision and coordination of the scientific & technical activities carried out at the Advanced Scientific Computing Division (4 Research Units, 16 people) of the Euro-Mediterranean Center on Climate Change (CMCC)
- Management, planning, financial and fund-raising tasks
- Engagement in outreach to local High Schools on climate change data management and high-performance computing.

Jan 2015 - Present

Senior Data Scientist, Advanced Scientific Computing Division (CMCC)

- Scientific officer for CMCC of the following research projects: OFIDIA2, EOSC-hub, EUBRAZILCC, CLIPC, INDIGO-DataCloud, EUBra-BIGSEA, BARRACUDA.
- Scientific Coordinator Deputy of the OFIDIA2 Project
- Research on high-performance data management, big data analytics and scientific computing.

Two key projects:

- EUBra-BIGSEA (www.eubra-bigsea.eu/): responsible for the design of a big and fast data analytics platform for smart cities based on Spark (and MLlib), Kafka/Storm, HDFS, and Ophidia. Implementation of descriptive and predictive models.
- INDIGO-DataCloud (<https://www.indigo-datacloud.eu/>): design of a cloud-based, big data analytics environment for parallel data analysis of multi-terabytes climate datasets based on Ophidia (<https://www.gitbook.com/book/indigo-dc/ophidia/>).
- Contribution to the scientific roadmap regarding extreme data and computing towards exascale in Europe within the European Extreme Data & Computing Initiative (EXDCI) project (<https://exdci.eu/>).
- Principal Investigator and main architect of the Ophidia big data analytics framework (www.ophidia.cmcc.it), a research effort on High-Performance Data Analytics and Mining for eScience (<https://github.com/OphidiaBigData> and <https://www.youtube.com/user/OphidiaBigData>).
- Member of the *Data Task Force* of the *European Network for Earth System Modelling* (ENES).

Jan 2012 – Mar 2015

Leader of the Scientific Data Management Research Group, Advanced Scientific Computing Division (CMCC)

- Supervision and coordination of the CMCC Scientific Data Management Research Unit (5 people)
- Design and implementation of open source solutions for the access, analysis & mining of scientific data.
- Main architect of the Ophidia big data analytics framework.

Nov 2010 – Dec 2014

Data Scientist, Advanced Scientific Computing Division (CMCC)

- Contribution to the design and implementation of large-scale data management solutions (access/analysis/processing), preparation of deliverables and presentation of results, in the following European projects: IS-ENES2, EUBRAZILCC, ORIENTGATE, and CLIPC.
- Use of parallel paradigms (MPI/OpenMP) to design scalable datawarehouse solutions for fast data analysis on top of relational (MySQL) and NoSQL (MongoDB) Databases.
- Lead developer of the ESGF-Dashboard module (<https://github.com/ESGF/esgf-dashboard>), based on a distributed datawarehouse (ROLAP on top of PostgreSQL) approach and several data marts to provide insights into hundreds of millions of data downloads across the Earth System Grid Federation.
- Contribution as *Scientific Data Management Expert* in the European Exascale Software Initiative (EESI - <http://www.eesi-project.eu/>) producing recommendations to lead research for efficient Exascale applications.

Jan 2012 – Nov 2014

Division Manager, Advanced Scientific Computing Division (CMCC)

- Support activity to the Head of the Advanced Scientific Computing Division concerning both the monitoring of the proper implementation of scientific activities and the coordination of scientific staff.
- Financial and organizational aspects, planning activities and human resources management within the Division.

Mar 2013 – May 2013

Visiting Scientist at the Lawrence Livermore National Laboratory (USA)

- Design and implementation of a distributed information system (datawarehouse, data marts, and collector component) on top of PostgreSQL and REST API for the analysis of large-scale data downloads across ESGF in the CMIP5 experiment (supervisor Dr. Dean Williams).

Apr 2012 – May 2012

Visiting Scientist at the Lawrence Livermore National Laboratory (USA)

- Design and implementation of the datawarehouse system of the ESGF Dashboard for the analysis of CMIP5 data download information (supervisor Dr. Dean Williams).

2004-2012

Contract Professor in Database Systems, Advanced Data Management & Parallel Computing II at the University of Salento, Lecce, Italy

Teaching of 8 Courses (Master's Degree) on Advanced Data Management (3), Database Systems

(3) and Parallel Computing (2) as Contract Professor at the Engineering Faculty of the University of Salento (Italy):

- **Parallel Computing II** (2004-2005 and 2008-2009)
Textbooks: *Parallel Programming in Openmp*, by Rohit Chandra, Ramesh Menon
Parallel Programming with MPI, by Peter S. Pacheco
- **Database Systems** (2005-2006, 2006-2007 and 2007-2008)
Textbook: *Fundamentals of Database Systems*, by Ramez Elmasri, Shamkant B. Navathe
- **Advanced Data Management** (2009-2010, 2010-2011 and 2011-2012)
Textbook: *Data Mining: Concepts and Techniques*, by Jiawei Han and Micheline Kamber

Oct 2011- Dec 2011

Visiting Scientist at the Lawrence Livermore National Laboratory (USA)

- Design and implementation of the first official release of the ESGF Dashboard to support distributed monitoring in the ESGF infrastructure (supervisor Dr. Dean Williams).

Jun 2006 – Oct 2010

Junior Scientist, Advanced Scientific Computing Division (CMCC)

- Implementation of scientific data management solutions at CMCC to enable climate data access, search & discovery, and metadata management.
- Development of Science Gateways to access to climate change data repositories at CMCC based on PostgreSQL/MySQL/SQLite and XML DBs and metadata standards: ISO19139 and ISO19115.

Jun 2004 – May 2006

Software Engineer at the University of Salento

- Contribution to the Enabling Grids for E-Science in Europe (EGEE) project regarding the deployment procedure of the LGC middleware on IA64 platforms.
- Active participation in the activities of sites certifications and sites monitoring at national level (Italian Grid Initiative).

EDUCATION AND TRAINING

July 2018

National Scientific Qualification (Associate Professor)

Qualification (awarded on July 26, 2018) as Associate Professor in Computer Engineering (ING-INF05) obtained by the Ministry of Education, University and Research (MIUR), Italy

October, 2004

Ph.D. degree in Innovative Materials & Technologies (Computer Science)

Ph.D. Thesis: "The GRelC Project: towards the definition of a Grid-DBMS" at the Istituto Superiore Universitario di Formazione Interdisciplinare (ISUFI), University of Lecce, Italy

Final Mark: Excellent

The thesis focused on the Grid Relational Catalog (GRelC) project, a reference implementation of a Grid-Database Management System joining scientific data management, database systems and grid computing applied to several eScience contexts (e.g. astrophysics, bioinformatics).

Jul 2002

International School for Computer Science Researchers

Participation in the 14th International School for Computer Science Researchers (Software Engineering), Lipari Island (Italy), June 30, 2002 - July 13, 2002

Jan 2001

Master's degree cum Laude in Computer Engineering

Thesis on the Parallelization of a QCD algorithm with Message Passing Interface at the Engineering Faculty of the University of Lecce, (Italy)

Final mark: 110/110 cum Laude (full marks with honors)

Key courses of my Master's Degree: Database Systems, Software Engineering, Operating Systems, High-Performance Computing, Information Systems, Computer Networks.

PERSONAL SKILLS

Mother Tongue

Italian

Other Languages

English (Full professional proficiency), **French** (Good working proficiency)

Technical Skills

Software engineering: data structures, algorithms, software development (from analysis of the problem to requirements gathering, design, implementation, test, validation, deployment), agile programming, UML documentation, DevOps approach, Lambda architecture.

More than 5 years' experience in the design, implementation and set-up of data analytics platforms for big data workflows. Experience of Hadoop, Hive, HDFS, Python Notebooks and libraries (i.e. NumPy, SciPy, pandas, scikit-learn, dplyr, ggplot2), JupyterLab/IPython, Matlab.

More than 10 years' experience in the design and implementation of OLTP (transactional DB) and OLAP systems (datawarehouses), data analytics, pre-processing (cleaning, integration, transformation, reduction), ETL systems for large scale (multi-Terabyte order) data, SQL, MySQL, PostgreSQL, SQLite and development of client/server applications in C and Java.

More than 10 years' experience on parallel programming with MPI, OpenMP.

Proficiency in the use of Office suites (word processors, spreadsheets, presentation software) as well as Google docs/spreadsheets, virtual machines, version control systems (e.g. Git), OSS (MacOS, Linux/UNIX, Windows)

More than 10 years' experience on spatial data infrastructure (e.g. services based on OGC standards (WMS, WPS, WCS) like GeoServer), NetCDF format, domain-specific services (OPeNDAP/THREDDS), ISO Geospatial standards (ISO19115/19139)).

ADDITIONAL INFO

GRANTS (a selection of the top ones with information on the funding and role in the project)

- **H2020 EOSC-Pillar** (Coordination and Harmonisation of National Initiatives, Infrastructures and Data services in Central and Western Europe), Proposal number 857650 (Accepted in February 2019, starting date to be defined). **(CMCC budget 164KEuro), Scientific Officer for CMCC**
- **H2020 ESIWACE2** (Excellence in Simulation of Weather and Climate in Europe, Phase 2), Proposal number SEP-210494560 (Accepted in August 2018, Planned start date March 2019). **(CMCC budget 436KEuro), WP Leader and Scientific Officer for CMCC**
- **H2020 IS-ENES3** (Infrastructure for the European Network for Earth System modelling - Phase 3), Proposal number SEP-210499769, Accepted in August 2018, Planned start date January 2019). **(CMCC budget 555KEuro), WP Leader**
- **H2020 IMMERSE** Improving Models for Marine Environment Services (Accepted in July 2018, Proposal number SEP- 821926, Planned start date December 2018), **(CMCC budget 255KEuro).**
- **OFIDIA2** (Operational Fire Danger prevention platform 2) Cooperation Programme INTERREG V/A Greece-Italy (EL-IT) 2014-2020 - MIS code: 5002571 **(CMCC budget 670KEuro). Scientific Coordinator Deputy**
- **H2020 EOSC-hub** (Integrating and managing services for the European Open Science Cloud), Proposal number: 777536, **(CMCC budget 210KEuro). Scientific Officer for CMCC**
- **H2020 EUBRA-BIGSEA** (Europe – Brazil Collaboration of BIG Data Scientific Research through Cloud-Centric Applications), Proposal number 690116. **(CMCC budget 280KEuro) Scientific Officer for CMCC and WP Leader**
- **H2020 ESIWACE** (Excellence in Simulation of Weather and Climate in Europe). H2020, Project Proposal number: 675191. **(CMCC budget 290KE). Scientific Officer for CMCC**
- **H2020 INDIGO-DataCloud** (Integrating Distributed data Infrastructures for Global Exploration). Proposal Number 653549 **(CMCC budget 340KEuro). Scientific Officer for CMCC.**
- **FP7 IS-ENES2** (Infrastructure for the European Network for Earth System modelling - Phase 2) **(CMCC budget 430KEuro).**
- **OFIDIA** (Operational Fire Danger prevention platform) - European Territorial Cooperation Operational Programme "Greece-Italy" (2007 – 2013). **(CMCC budget 460KEuro). Scientific Coordinator (Deputy).**
- **H2020 RDA Europe BARRACUDA** **(CMCC budget 15KEuro) Scientific Officer for CMCC.**
- **FP7 EUBRAZILCC** (EU-Brazil Cloud infrastructure Connecting Federated Resources for Scientific Advancement). Proposal Number 614048. **(CMCC budget 200KEuro). WP**

Leader.

- **UNIDATA Community Equipment Awards 2011** (sponsored by National Science Foundation (NSF). <http://www.unidata.ucar.edu/community/equipaward/index.htm#2011>)
- **FP7 CLIP-C** (*Climate Information Platform for Copernicus*) - (**CMCC budget 130KEuro**).
Project Manager for CMCC

**SCIENTIFIC, TECHNICAL,
PROGRAM, and
REVIEWING COMMITTEES**

- Participation in more than 30 *Program Committees* in Scientific Conferences.
- Organiser of more than 20 *Workshops / Sessions* on data management topics
- Member of the *ENES Data Task Force* (ENES - European Network for Earth System Modelling).
- Reviewer for (i) *Concurrency and Computation: Practice and Experience*, (ii) *Future Generation Computer Systems*, (iii) *Scientific Data (Nature)*, (iv) *IEEE Access*.
- Scientific data management expert in the Climate, Weather and solid Earth Science Working Groups within the following exascale initiatives/projects: IESP, BDEC, ESSI, ESSI2, EXDCI, EXDCI2.
- Guest editor of: (i) 4 Special Issues on High-Performance Computing and Simulations (*Concurrency and Computation: Practice and Experience*); 1 Special Issue on High-Performance Computing and Simulations (*Future Generation Computer Systems*); 1 Special Issue on Data management for eScience (*Future Generation Computer Systems*)

PUBLICATIONS

A comprehensive list of all publications can be found at:

DBLP <http://www.informatik.uni-trier.de/~ley/pers/hd/f/Fiore:Sandro.html>

Google Scholar: <https://scholar.google.com/citations?user=VmTh7icAAAAJ&hl=en>

**Ph.D., BACHELOR and
MASTER DEGREE
STUDENTS SUPERVISION**

1 Ph.D. Student supervised in 2009-2011 (Dr. Alessandro Negro)

1 Ph.D. student currently supervised (2018, 1st year) (Mr. Donatello Elia)

26 Master's and Bachelor's degree students supervised from 2004 to 2018.

AWARDS

Best Student Paper award for the paper: G. Aloisio, M. Cafaro, E. Blasi, M. Mirto, S. Fiore, D. Lezzi, "*Web Services for a Biomedical Imaging Portal*", Proceedings of ITCC2003, Volume IV, pp-432-437, Las Vegas, Nevada, 2003.

Earth System Grid Federation Team Award 2017, for the strong results of the ESGF Dashboard activity for the CMIP5 experiments, December 5-7, 2017, San Francisco, USA.

INTERESTS/HOBBIES

Reading, cooking, travelling, swimming and running outside.