

Dr. Rosa Filgueira Vicente

Lecturer

Computer Science, University of St Andrews

February 28, 2022

Email: rf208@st-andrews.ac.uk

<http://www.rosafilgueira.com>

Professional Experience

- Lecturer, University of St Andrews, UK, since February 2022
- Assistant Professor, Heriot-Watt University, UK, March 2021 - February 2022
- Vice President, Applied Research, JP Morgan Chase & Co, UK, October 2020 - March 2021
- Research Fellow, University of Edinburgh, UK, June 2019 - October 2020
- Data Architect, University of Edinburgh, UK, March 2018 - May 2019
- Senior Data Scientist, British Geological Survey, Edinburgh, UK, October 2016 - February 2018
- Senior Research Associate, School of Informatics, University of Edinburgh, June 2011 - October 2016
- Research Assistant and Teaching Fellow, University Carlos III Madrid, Spain, Sept. 2005 - Sept. 2010
- Software Developer, Soluziona, Enterprise Consultancy, Spain, June 2004 - September 2004

Education

- PhD in Computer Science, University Carlos III Madrid, Spain, September 2010
- Master of Technology in Computer Science, University Carlos III, Madrid, Spain, July 2007
- Master in e-Commerce, University Carlos III Madrid, Spain, July 2004
- Master of Computer Science, University of Deusto, Spain, June 2003
- Bachelor of Computer Science, University of Deusto, Spain, June 2001

Visiting Positions and Consultancy Experience

- Honorary Research Fellow, School of Informatics, University of Edinburgh, since April 2021
- Research Visitor, ISI, University of Southern California, October 2017 - December 2017
- Honorary Research Fellow, School of Informatics, University of Edinburgh, January 2017 - March 2018
- Research Visitor, ISI, University of Southern California, Los Angeles, March 2016 - May 2016
- Consultancy Services for Accelogic, LOCATION, August 2014 - April 2016
- Research Visitor, AIST, Informatics Technology Research Institute, Tsukuba, Japan, July 2013 - July 2013
- Research Visitor, EPCC, University of Edinburgh, February 2011 - October 2011

Track Record of Research Funding

- Towards Large-scale Cultural Analytics in the Arts and Humanities. UKRI/AHRC, GBP 100,000
- Research Fellowship in Digital Scholarship, National Library of Scotland, 2021, GBP 7,500
- COVID-19 Response & Recovery Small Grant Funding, Data-Driven Innovation grant, 2020, GBP 10,000
- Text and Data Mining, Data-Driven Innovation grant, 2019, GBP 133,000
- Two Postdoctoral and Early Career Researcher Exchanges grant, SICSA, 2015/17, GBP 7,500/13,580
- Observation Imaging and Interpretation, UoE grant, 2016-2017, GBP 3000
- Combining knowledge to pioneer big-data solutions for SKA, CSIC grant, 2016-2018, EUR 26,000
- Driving UK HPC enabled science and innovation through US collaborations to SC14, 2014, GBP 2,500
- Two HPC-Europa2 PRACE grants, 2011, GBP 6,000

Awards and Scholarships

- International HPC Summer school student travel award, Ljubljana, Slovenia, 2016
- Best poster award, SACSIS 2012, Kobe, Japan, 2012
- Scholarship for Doctoral Training Programme, University Carlos III Madrid, Spain, 2005, EUR 12,000
- Student travel award, TCPP PhD Forum at IPDPS Conference, Miami, 2008, USD 500

Teaching, Supervision, Training and Examination Experience

- Course co-leader *Data Bases*, University of St Andrews, 2021/22
- Course marker for Data Structure and Algorithms, Heriot-Watt University, 2020/21
- Dissertation examiner for final year (four students), Masters (two students) and second year PhD thesis reviews (one student), Heriot-Watt University, 2020/21
- Course leader for *Apache Spark*, University of Edinburgh and BGS, from 2017 to 2020. My responsibilities included material preparation, delivering lectures and tutorials, laboratory demonstration, preparing the computing environment, and assisting students with course materials.
- Co-Supervision of PhD students (all successfully completed), since 2015
 - Ole Weidner, *A Telemetry Framework for High-Performance Computing Environments*, June 2021
 - Luca Trani, *Holistic Catalogues for Data-Intensive Science*, September 2019
 - Alessandro Spinuso, *Active Provenance for Data Intensive Research*, June 2018
- External PhD examination, since 2012
 - Giannis Evagorou, *Methods for Efficient Storage, Distribution and Scalable Analysis of Spatial Data*, Imperial College London University, UK, 2021
 - Pablo Cerro, *Modelling and validation of cloud systems using model driven engineering, metamorphic and mutation testing*, Complutense University of Madrid, Spain, 2020
 - Hector Ortega, *Parallel Approaches to Shortest-Path Problems for Multilevel Heterogeneous Computing*, Univeristy of Valladolid, Spain, 2015

- Idafen Santana Pérez, *Conservation of Computational Scientific Execution Environments for Workflow-based Experiments Using Ontologies*, Polytechnic University of Madrid, Spain, 2015
- Jordi Villaplana, *Management of Cloud systems applied to eHealth*, University of Lleida, Spain, 2015
- Miquel Orobitg, *High performance computing on biological sequence alignment*, University of Lleida, Spain, 2013
- Laura Prada, *New techniques to model energy-aware I/O architectures based on SSD and hard disk drives*, University Carlos III Madrid, Spain, 2012
- Supervision and examination of MSc students at the University of Edinburgh, from 2012 to 2020:
 - Supervised four and examined six MSc students at EPCC from 2018 to 2020
 - Supervised and examined six MSc students at the School of Informatics from 2012 to 2016
- Hosted five HPC-Europa3 visitors at the University of Edinburgh and Heriot-Watt University, since 2019
- Co-Supervision of two PRACE “Summer of HPC” students at University of Edinburgh, from 2017 to 2020
- Supervision of six Open Science Data Cloud-PIRE students, University of Edinburgh, from 2012 to 2015
- Coordination and preparation of seven training courses, since 2014: *Introduction to Machine Learning*, *Introduction to Data Science*, and *dispel4py*. My responsibilities also included to deliver training sessions in multi-day events attended by multidisciplinary audiences ranging from undergraduates and PhD students to senior academics
- Two invited lectures for the MSc in Informatics at the University of Lleida, Spain 2013-2014
- Invited lectures for Innovative Learning Week and The Data Lab Welcome for MSc students, University of Edinburgh, 2014 and 2016
- Teaching of Computer System Courses at University Carlos III of Madrid, from 2005 to 2010: *Operating Systems*, *Computer Architecture*, *Fault-tolerant Systems*, *Introduction to Programming*, *Computer Structure and Distributed Systems*. My responsibilities included material preparation, delivering lectures and tutorials, laboratory demonstration, grading homework and exams, and assisting students with course materials.
- Supervision of seven MSc students at University Carlos III Madrid, 2006 to 2010

Community Service

- Group Leader of the Systems Research Group, at University of St Andrews, since February 2022
- Reviewer for the H2020 *MAESTRO* project funded by the European Commission, February 2022
- Expert for the *Workflow Community Summit* track *Improving workflow systems’ interface with legacy and emerging HPC software and hardware stack*, April 2021
- Participation in the *Workflow Community Summit* events, since January 2021
- Expert external evaluator for *EU H2020-ICT* proposals, 2019
- Lead author of the *Data-Driven Geoscience Platform* research line for the British Geological Society, February 2017
- Participation in the *SC Mentor-Protege Program* as mentor, Super Computing, Salt Lake City, November 2016
- Reviewer for the H2020 *EXCAPE* project, May 2016 to September 2018
- Participation in the *European eXtreme Data and Computing Initiative* (EXDCI) as an expert in the field of *Weather, Climatology and Solid Earth Sciences*, since May 2016

Workshops Organizer, Programme Committee Memberships and Journal Reviewing

1. Programme Committee member for SC22, 2022
2. Co-Chair for the 16th Workshop on Workflows in Support of Large-Scale Science at SC21, 2021
3. Programme Committee member for SC21, 2021
4. Programme Committee member for IEEE Cluster, 2021
5. Programme Committee member for the International Workshop on Science Gateways, 2021
6. Programme Committee member for the IEEE IPDPS, 2021
7. Co-Chair for the 15th Workshop on Workflows in Support of Large-Scale Science at SC20, 2020
8. Invited Speakers Chair for Women in HPC at ISC20, 2020
9. Programme Committee member for the International Workshop on Science Gateways, 2020
10. Co-Organiser of the session Data-Oriented and Self-Optimizing Architectures and Workflows for Earth, Space and Climate Science Applications, AGU, 2019
11. Co-Chair of the Bridging from Concepts to Data and Computation (BC2DC) workshop co-located with IEEE eScience, 2019
12. Programme Committee member for the SC19 edition of the Women in HPC workshop, 2019
13. Programme Committee member for the IEEE eScience conference, 2019
14. Programme Committee member for the International Workshop on Science Gateways, 2019
15. Co-Organiser of the session Data-Driven Platforms and VREs for Complex Earth, Space, and Environmental Science Applications, AGU, 2018
16. Programme Committee member for the International Workshop on Science Gateways, 2018
17. Reviewer for the Journal of Frontiers of Information Technology & Electronic Engineering, 2017
18. Poster Committee Member for the IEEE Cluster conference, 2016
19. Reviewer for the IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid 2016) Conference, 2016
20. Programme Committee member for the TPD 2016 on Reproducible Open Science Workshop, 2016
21. Programme committee member for the first International Workshop on Interoperable infrastructures for Interdisciplinary Big Data Sciences (IT4RIs 15), 2015
22. Programme Committee member for the ACM Conference on Service-Oriented Architectures and Programming (SOAP 2015), 2015
23. Programme Committee member for the IEEE Conference on International Symposium on Parallel and Distributed Processing with Applications (ISPA 2013), 2012 and 2013
24. Reviewer for the Journal of Systems and Software, 2013
25. Reviewer for the Journal of Simulation Modelling Practice and Theory, 2013
26. Reviewer for the IEEE Symposium on Computers and Communications (ISCC 2010) Conference, 2010

Selected Invited Speaker

1. *Advanced data-driven frameworks to develop reproducible data science applications at scale*, Workshop on Re-envisioning Extreme-Scale I/O for Emerging Hybrid HPC Workloads, 2021
2. *Improving workflow systems' interface with legacy and emerging HPC software and hardware stack*, Workflows Community Summit, online, 2021
3. *Analysing Digital Historical Textual Data at Scale with Defoe Toolbox*, RLUK, Text and Data Mining Seminar, 2020
4. *defoe: A Spark-based Toolbox for Analysing Digital Historical Textual Data*. USC, Los Angeles, US, 2019
5. *Empowering domain experts with DARE, a new cloud-based platform and working environment*. RSEConUK, Birmingham, UK, 2019
6. *Delivering easy-to-use frameworks to empower data-driven research*. UK-France Bilateral International Meeting, Royal Society, Chicheley, UK, 2019
7. *Delivering easy-to-use frameworks to empower data-driven applications on HPC environments*, HPCKM'17, San Sebastian, Spain, 2017.
8. *ENVRiplus: Review of existing Research Infrastructures: Requirements, Technologies, Achievements and Gaps leading to Characterization*. USC, Los Angeles, US, 2016
9. *Data Science in Action*. Organised by The Data Lab, Edinburgh, UK, 2015
10. *Earthquake and Failure Forecasting in Real-Time: A Forecasting Model Testing Center*. Organised by INSPIRES Center at University of Lleida, Lleida, Spain, 2013

Selected Publications

• Journal Articles

1. Pablo C. Cañizares , Alberto Núñez, **Rosa Filgueira**, *et al.* *Parallel Mutation Testing for Large Scale Systems* . Future Generation Computer Systems (FGCS), 2021 [under review]
2. Liang Liang, **Rosa Filgueira**, *et al.* *Scalable Adaptive Optimizations for Stream-based Workflows in Multi-HPC-Clusters and Cloud Infrastructures*. Future Generation Computer Systems (FGCS), 2021
3. Johanna M. Scheidegger, ..., **Rosa Filgueira**. *Integration of 2D Lateral Groundwater Flow into the Variable Infiltration Capacity (VIC) Model and Effects on Simulated Fluxes for Different Grid Resolutions and Aquifer Diffusivities*. Water; 13 (5), March 2021.
4. Sandra Engwell, ... , **Rosa Filgueira**. *Near-real-time volcanic cloud monitoring: insights into global explosive volcanic eruptive activity through analysis of Volcanic Ash Advisories*. Bulletin of Volcanology 83 (2), 1-17, January 2021.
5. Iraklis A. Klampanos, Chrysoula Themeli, **Rosa Filgueira**, *et al.* *DARE Platform: a Developer-Friendly and Self-Optimising Workflows-as-a-Service Framework for e-Science on the Cloud*. The Journal of Open Software, October 2020.
6. Paolo Marangio, Vyrion Christodoulou, **Rosa Filgueira**, *et al.* *Highlights Automatic detection of Ionospheric Alfvén Resonances in magnetic spectrograms using U-net*. Computers & Geoscience, 2020
7. Rafael Ferreira da Silva, **Rosa Filgueira**, *et al.* *Exploring PID Controllers for Online Resilient Resource Management of Distributed Scientific Workflows*. Future Generation Computer Systems, vol. 95, pp. 615-628, 2019
8. Luca Trani, **Rosa Filgueira**, Rosa Filgueira *et al.* *Establishing Core Concepts for Information-Powered Collaborations*. Future Generation Computer Systems, vol. 89, pp. 421-437, 2018

9. Luca Trani, **Rosa Filgueira**, *et al.* *WFCatalog: a catalogue for seismological waveform data*. Computers and Geosciences, vol. 106, pp. 101-108, 2017
10. Rafael Ferreira da Silva, **Rosa Filgueira**, *et al.* *A characterization of workflow management systems for extreme-scale applications*. Future Generation Computer Systems, vol. 75, pp. 228-238, 2017
11. **Rosa Filgueira**, Amrey Krause, *et al.* *dispe14py: A Python Framework for Data-Intensive Scientific Computing*. Special issue on Data-Intensive High Performance Computing. International Journal of High Performance Computing Applications, vol. 31, no. 4, pp. 316-334, 2016
12. Jordi Vilaplana, **Rosa Filgueira**, *et al.* *The Cloud Paradigm applied to e-Health*. BMC Medical Informatics and Decision Making, vol. 35, no. 13, pp. 1-10, 2013
13. Alberto Nuñez, **Rosa Filgueira**, *et al.* *SIMCAN: A Flexible, Scalable and Expandable Simulation Platform for Modeling and Simulating HPC Architectures and Applications*. Simulation Modelling Practice and Theory, vol. 20, no. 1, pp. 12-32, 2012
14. **Rosa Filgueira**, David E. Singh, *et al.* *Adaptive CoMPI: Enhancing MPI-based Applications Performance and Scalability by Using Adaptive Compression*, The International Journal of High Performance Computing Applications, vol. 25, no. 1, pp. 93-114, 2012
15. **Rosa Filgueira**, Jesús Carretero, *et al.* *Dynamic-CoMPI: Dynamic Optimization Techniques for MPI Parallel Applications*. The Journal of Super Computing, vol. 59, no. 1, pp. 361-391, 2012

• **Refereed Conference and Workshop Papers**

1. Rosa Filgueira, Daniel Garijo *Inspect4py: A Knowledge Extraction Framework for Python CodeRepositories*. The 2022 Mining Software Repositories Conference. [Submitted, January 2022]
2. Rosa Filgueira, Claire Grover, *et al.* *Extending defoe for the Efficient Analysis of Historical Texts at Scale*. IEEE eScience, September 2021.
3. Liang Liang, **Rosa Filgueira**, *et al.* *Adaptive Optimizations for Stream-based workflows*. 15th Workshop on Workflows in Support of Large Scale, during SC20, 15th November 2020.
4. **Rosa Filgueira**, Beatrice Alex, *et al.* *Geoparsing the historical Gazetteers of Scotland: accurately computing location in mass digitised texts*. 8th Workshop on the Challenges in the Management of Large Corpora, June 2020.
5. **Rosa Filgueira**, Michael Jackson, *et al.* *defoe: A Spark-based Toolbox for Analysing Digital Historical Textual Data*. IEEE eScience, 2427 September 2019.
6. Malcolm Atkinson, **Rosa Filgueira**, *et al.* *Comprehensible Control for Researchers and Developers facing Data Challenges*. IEEE eScience, 2427 September 2019.
7. **Rosa Filgueira**. *DARE Platform: Enabling Easy Data-Intensive Workflow Composition and Deployment*. Research Objects (RO2019) during IEEE eScience, 2427 September 2019.
8. Vyron Christodoulou, **Rosa Filgueira**, *et al.* *Automatic classification of Aurora-related tweets using machine learning methods*, 2nd International Conference on Geoinformatics and Data Analysis (ICGDA), Prague, Czechia, 15-17 March 2019
9. **Rosa Filgueira**, Rafael Ferreira da Silva, *et al.* *IoT-Hub : New IoT Data-Platform for Virtual Research Environments*. 10th International Workshop on Science Gateways (IWSG 2018), 13-15 June 2018
10. Rafael Ferreira da Silva, **Rosa Filgueira**, *et al.* *Using Simple PID Controllers to Prevent and Mitigate Faults in Scientific Workflows*. WORKS'16 Workshop, SC16, 14th November 2016
11. **Rosa Filgueira**, Rafael Ferreira da Silva, *et al.* *Asterism: Pegasus and dispe14py Hybrid Workflows for Data-intensive Science*. DataCloud'16 Workshop, SC16 Computing, 13th November 2016
12. Ole Weidner, **Rosa Filgueira**, *et al.* *Rethinking High Performance Computing Platforms: Challenges, Opportunities and Recommendations*. DIDC16 Workshop, Kyoto, Japan, 1st June 2016
13. Rafael Ferreira da Silva, **Rosa Filgueira**, *et al.* *Automating Environmental Computing Applications with Scientific Workflows*. ECW Workshop, 12th IEEE eScience, Baltimore, US, 23rd October 2016

14. **Rosa Filgueira**, Amrey Krause, *et al.* *dispe14py: A User-friendly Framework for Describing eScience Applications*. 11th IEEE eScience, Munich, Germany, 1-4 September 2015
15. Malcolm Atkinson, **Rosa Filgueira**, *et al.* *VERCE Delivers a Productive e-Science Environment for Seismology Research*. 11th IEEE eScience, Munich, Germany, 1-4 September 2015
16. **Rosa Filgueira**, Malcolm Atkinson, *et al.* *VarPy: A Python Library for Volcanology and Rock Physics Data Analysis*. PyHPC'14 Worksop, Super Computing, New Orleans, US, 16 -21 November 2014
17. Sandra Gesing, **Rosa Filgueira**, *et al.* *Workflows in a Dashboard: A New Generation of Usability*. WORKS'14 Worksop, Super Computing, New Orleans, US, 16-21 November 2014
18. **Rosa Filgueira**, Malcolm Atkinson, *et al.* *eScience Gateway Stimulating Collaboration in Rock Physics and Volcanology*. 10th IEEE eScience, Guarujá, Brazil, 20-24 October 2014
19. **Rosa Filgueira**, Malcolm P. Atkinson, *et al.* *Applying Selectively Parallel I/O Compression to Parallel Storage Systems*. European Conference on Parallel and Distributed Computing (Euro-Par), Porto, Portugal, 27-29 August 2014
20. **Rosa Filgueira**, Iraklis Klampanos, *et al.* *FAST: Flexible Automated Synchronization transfer*. In proceedings of High-Performance Parallel and Distributed Computing. ACM International Workshop on Data Intensive Distributed Computing (DICT), Vancouver, Canada, 23-27 June 2014
21. Yusuke Tanimura, **Rosa Filgueira**, *et al.* *Two-Phase Collective I/O based on Advanced Reservations to Obtain Performance Guarantees from Shared Storage Systems*. Workshop on Interfaces and Architectures for Scientific Data Storage (IASDS), Indianapolis, IN, US, 1-9 September 2013
22. Yusuke Tanimura, **Rosa Filgueira**, *et al.* *MPI Collective I/O based on Advanced Reservations to Obtain Performance Guarantees from Shared Storage Systems*. IEEE International Conference on Cluster Computing (CLUSTER), Indianapolis, IN, US, 23-2 September 2013
23. Alberto Nunez, **Rosa Filgueira**, *et al.* *SANComSim: A Scalable, Adaptive and Non-intrusive Framework to Optimize Performance in Computational Science Applications*. International Conference on Computational Science (ICCS), Barcelona, Spain, 5-7 June 2013
24. **Rosa Filgueira**, Malcolm Atkinson, *et al.* *An Adaptive, Scalable, and Portable Technique for Speeding Up MPI-Based Applications*. International European Conference on Parallel and Distributed Computing (Euro-Par), Rhodes Island, Greece, August 27-31 2012
25. Gabriel G. Castañe, **Rosa Filgueira**, *et al.* *Dimensioning Scientific Computing Systems to Improve Performance of Map-Reduce based Applications*. International Conference on Computational Science (ICCS), pp.: 226-235, Omaha, Nebraska, 4-6 June 2012
26. **Rosa Filgueira**, David E. Singh, *et al.* *CoMPI: Enhancing MPI Based Applications Performance and Scalability Using Run-Time Compression*. EURO PVM/MPI, Espoo, Finland, 7-10 September 2009
27. **Rosa Filgueira**, David E. Singh, *et al.* *Exploiting Data Compression in Collective I/O Techniques*. IEEE International Conference on Cluster Computing (CLUSTER), Tsukuba, Japan, Sept. 29 - Oct. 1, 2008
28. **Rosa Filgueira**, David E. Singh, *et al.* *Data Locality Aware Strategy for Two-Phase Collective I/O*. In Proceedings of High Performance Computing for Computational Science (VECPAR), Toulouse, France, 24-27 June 2008
29. **Rosa Filgueira**, David E. Singh, *et al.* *Optimization and Evaluation of Parallel I/O in BIPS3D Parallel Irregular Application*. 21st IEEE International Parallel and Distributed Processing Symposium (IPDPS), Long Beach, California US, 26-30 March 2007

- **Books and Book chapters**

1. **Rosa Filgueira**, Malcolm Atkinson, *et al.* Chapter: *dispe14py: Agility and Scalability for Data-Intensive HPC*. Springer eBook on Conquering Big Data Using High Performance Computing, 2016
2. **Rosa Filgueira**, David E. Singh, *et al.* Book: *Optimization Techniques for Parallel Applications on Clusters*. Lambert Academic Publishing, 978-3-8433-70, 2010