Renan Souza

contact@renansouza.org • RenanSouza.org • renansouza1
ID: x9t36ewAAAAJ • renan-souza • 0000-0002-1794-808X
Updated on August 21, 2025.

Bio

Renan Souza holds a Ph.D., M.Sc., and B.Sc. in Computer Science (2009–2019) from the Federal University of Rio de Janeiro (UFRJ). Since 2022, he has been a staff research scientist at the Oak Ridge National Laboratory. From 2015 to 2022, he worked as a research scientist and software engineer at IBM Research. During his Ph.D., he was a visiting researcher at Inria, France. During his B.Sc., he spent one year at Missouri State University and interned at the SLAC National Laboratory at Stanford University. He has been active as a software engineer, researcher, and technical lead on multiple projects since 2010. His research and development focus on large-scale data management and AI to support the next generation of Edge–Cloud–HPC workflows.

Research Interests

Large-scale Data Science and Data Engineering • Edge–Cloud–HPC Workflows • Provenance Data • Big Data Analytics • Machine Learning Systems • Agentic AI

Education

- **Ph.D. in Computer Science**, Federal Univ. of Rio de Janeiro, Brazil — Sep 2015 – Dec 2019
  Supervised by Marta Mattoso (COPPE/UFRJ) and Patrick Valduriez (Inria).
  Title: Supporting User Steering in Large-scale Workflows with Provenance Data
- Visiting Ph.D. Student, Inria/Univ. Montpellier, France — Jan 2019 – Mar 2019
  Supervised by Patrick Valduriez (Inria).
- **M.Sc. in Computer Science**, Federal Univ. of Rio de Janeiro, Brazil — Jan 2013 – Jul 2015
  Supervised by Marta Mattoso (COPPE/UFRJ).
  Title: Controlling the Parallel Execution of Workflows Relying on a Distributed Database
- Computer Science exchange student, Missouri State University, U.S. — Jun 2011 – Jun 2012
- **B.Sc. in Computer Science**, Federal Univ. of Rio de Janeiro, Brazil — Jan 2009 – Dec 2012
  Supervised by Maria Luiza Machado Campos (DCC/UFRJ).
  Title: Linked Open Data Publication Strategies: An Application in Network Performance Data
- **Technical Degree in Information Systems**, Lemos de Castro — Jan 2005 – Dec 2007

Experience

- **Oak Ridge National Laboratory** — **Oct 2022 – Present**
  Research Scientist, HPC Workflows, Data & AI — **Knoxville, United States**
  As Principal Investigator and lead software developer, he leads the design and development of large-scale data and AI systems within the Workflows and Ecosystem Services group at ORNL, focusing on AI-driven observability, data integration, workflow provenance, and LLM-based agentic workflows to accelerate scientific discovery in Edge–Cloud–HPC environments.

- **IBM Research** **Apr 2015 – Oct 2022**
  Research Scientist and Software Eng., Cloud, Data & AI **Rio de Janeiro, Brazil**
  As a Staff Research Scientist (2021–2022), he was a lead researcher and developer on projects in large-scale data science and engineering to support AI systems in hybrid cloud and cluster environments with highly distributed and heterogeneous workloads for clients across energy, finance, physics, and cheminformatics domains. Although primarily focused on backend development, data engineering, cloud, and DevOps, he collaborated closely with front-end developers and HCI researchers to design domain-specific applications.
  As a Research Software Engineer (2015–2021), he worked with and led R&D projects on large-scale data integration for AI systems in the energy sector, targeting cluster and cloud platforms. He also led the Cloud DevOps team responsible for developing and deploying conversational AI systems.
  As a Software Engineering Intern (2015), he designed and implemented big data and machine learning solutions for real-time analysis of streaming social data.
- **SLAC National Accelerator Laboratory, Stanford Univ.** **May 2013 – Dec 2014**
  Research Software Engineering intern **Menlo Park, United States**
  Led the development of a cloud platform utilizing semantic web, big data, and data warehousing techniques. This platform is designed to store, retrieve, visualize, and publish structured data about internet performance worldwide, enabling understanding of global Internet quality.
- **COPPE-UFRJ** **Dec 2011 – Sep 2014**
  Software Engineer **Rio de Janeiro, Brazil**
  As a Lead Software Engineer (2013–2014), he led the development of a system that faciliated access of information about public services offered by the Brazilian Federal Government. Also lead the development of a platform to publish linked open data from the Brazilian Federal Register on the semantic web, applying agile practices, ontology-based data modeling, and natural language processing.
  As a Full-Stack Software Engineering Intern (2011–2013), he worked on the R&D of various web systems.
- **Federal Univ. of Rio de Janeiro** **Jan 2010 – Jul 2011**
  Software Engineering intern **Rio de Janeiro, Brazil**
  Developed a system that integrated data warehouse environments with both structured and unstructured data, enabling the generation of more intelligent and flexible information reports.
- **Petrobras** **May 2007 – May 2008**
  IT Intern **Rio de Janeiro, Brazil**
  Implemented features and provided ongoing maintenance for web systems supporting Petrobras employees.

Technical Knowledge

- **Languages**: Python, Java, C, C++, Shell scripting, NodeJS, Scala, Lua
- **Data Science/ML Technologies**: Pandas, Polars, Jupyter Notebooks, Numpy, Matplotlib, Seaborn, Plotly, Tensorflow, ScikitLearn, Keras, PyTorch, MLFlow, Airflow, Grafana
- **Agentic AI and LLMs**: MCP Agents, Crew AI (Multi-agent Framework), LangChain, Streamlit for AI Agents, RAG, Prompt Engineering Techniques
- **Big Data & Parallel Processing Frameworks**: Dask; Apache Spark: RDD, DataFrames, Streaming, MLib, GraphX, GraphFrames; Hadoop Ecosystem
- **Cloud and Cluster computing**: VMs, Dockers, Kubernetes, OpenShift, HPC (Slurm, LSF, PBS)
- **DevOps**: Containers, Kubernetes, OpenShift, CI/CD Pipelines, GitHub, GitHub Actions, Travis, Jenkins
- **Message Queueing Systems**: Kafka, RabbitMQ, Redis
- **GPU Programming and Profilling**: NVIDIA and AMD Python APIs for GPU performance analysis
- **Relational DBMS**: PostgreSQL/PostGIS, DB2, SQLite, MySQL, MySQL Cluster, MS SQL Server
- **NoSQL DBMS**: MongoDB, AllegroGraph, Jena, Blazegraph, Virtuoso, Sesame, Cloudant, CouchBase, Redis, Impala, Elasticsearch, HBase, Hive, Apache Ignite, LMDB
- **Heterogeneous Data Management**: Data Integration, Multi-database Queries, Polystores
- **Cluster Deployment**: YARN, Mesos, Standalone deployment
- **Business Intelligence**: MS SQL Server BI developer studio, Pentaho Solutions, Talend;
- **Semantic Web Tools/Languages**: OWL, RDF, SPARQL, Protege

- **Distributed and Concurrent Programming**: PubSub, MPI, OpenMP, CUDA, Data-centric distributed and parallel programming
- **Web Development**: Python Flask/UWSGI, Java EE, Tomcat/JBoss, Spring Boot

Selected Publications

For complete list, visit: RenanSouza.org/publications

[1] **R. Souza**, T. J. Skluzacek, S. R. Wilkinson, M. Ziatdinov, R. F. Silva, "Towards lightweight data integration using multi-workflow provenance and data observability," in *IEEE International Conference on e-Science*, 2023. DOI: 10.1109/e-Science58273.2023.10254822. [Online]. Available: https://doi.org/10.1109/e-Science58273.2023.10254822.

[2] **R. Souza**, L. G. Azevedo, V. Lourenço, E. Soares, R. Thiago, R. Brandão, D. Civitarese, E. Vital Brazil, M. Moreno, P. Valduriez, M. Mattoso, R. Cerqueira, M. A. S. Netto, "Workflow provenance in the lifecycle of scientific machine learning," *Concurrency and Computation: Practice and Experience*, vol. e6544, pp. 1–21, 2021. DOI: 10.1002/cpe.6544. [Online]. Available: https://doi.org/10.1002/cpe.6544.

[3] **R. Souza**, A. Gueroudji, S. DeWitt, D. Rosendo, T. Ghosal, R. Ross, P. Balaprakash, R. F. Silva, "Prov-agent: Unified provenance for tracking AI agent interactions in agentic workflows," in *IEEE International Conference on e-Science*, Chicago, U.S.A.: IEEE, 2025.

[4] **R. Souza**, S. Caino-Lores, M. Coletti, T. J. Skluzacek, A. Costan, F. Suter, M. Mattoso, R. F. Silva, "Workflow provenance in the computing continuum for responsible, trustworthy, and energy-efficient AI," in *IEEE International Conference on e-Science*, Osaka, Japan: IEEE, 2024. DOI: https://doi.org/10.1109/e-Science62913.2024.10678731.

Grants and Awards

- 2nd IBM Patent Plateau (8+ patents submitted to USPTO) 2021
- SBBD Honored Mention for the Best Ph.D. Thesis Award 2021
- 1st IBM Patent Plateau (4+ patents submitted to USPTO) 2020
- SBBD Best M.Sc. Thesis Award 2017
- SBBD Honored Mention on the paper *Spark Scalability Analysis in a Scientific Workflow* 2017
- CAPES M.Sc. Grant 2013 – 2014
- Brazil Science Mobility Grant - Missouri State University 2012 – 2013
- Scientific Initiation Grant - Federal Univ. of Rio de Janeiro 2010

Teaching and Supervisions

Teaching:

- Databases Laboratory, graduate, UFRJ 2017
  Teacher assistant to Prof. Marta Mattoso
- Logics for Computer Science, undergraduate, UFRJ 2012–2013
  Teacher assistant to Prof. Mario Benevides

Supervisions of final dissertations

- Pedro Paiva Miranda, undergraduate, UFRJ, Co-supervision with Prof. Marta Mattoso 2015
  Thesis title: *A Mechanism for Fault Tolerance in Parallel Executions of Workflows supported by a Database*
- Rachel Gonçalves de Castro, undergraduate, UFRJ, Co-supervision with Prof. Marta Mattoso 2015
  Thesis title: *Publication of Workflow Provenance Data in the Semantic Web*

Talks and Participation in Events

- **IEEE International Conference on e-Science** in Osaka, Japan (Virtual) 2024

- Workflow Provenance in the Computing Continuum for Responsible, Trustworthy, and Energy-Efficient AI, Oral presentation , link

- **IEEE/ACM Supercomputing (SC)** in Atlanta, GA 2024
  - Integrating Evolutionary Algorithms with Distributed Deep Learning for Optimizing Hyperparameters on HPC Systems, Oral presentation
- **IEEE/ACM Supercomputing (SC)** in Denver, CO 2023
- **IEEE International Conference on e-Science** in Limassol, Cyprus 2023
  - Towards Lightweight Data Integration using Multi-workflow Provenance and Data Observability, Oral presentation
- **Brazilian Symposium on Databases (SBBD)** in Rio de Janeiro, RJ (virtual) 2021
  - User Steering Support in Large-Scale Workflows, Oral presentation , link
- **Federal Fluminense University (UFF) Computer Science Seminars** in Rio de Janeiro, RJ (virtual) 2021
  - A Knowledge-centric Approach to Support Large-scale AI Systems, Invited talk (Portuguese) , link
- **SIAM Conference on Computational Science and Engineering** in Forth Worth, TX (virtual) 2021
  - AI4Seismic: An AI-Driven Platform to Accelerate Geological Discoveries, Invited talk, Highlighted by the SIAM press , link
  - Workflow Provenance in the Lifecycle of Scientific Machine Learning, Oral presentation , link
- **ACM International Conference on Management of Data (SIGMOD)** in Portland, OR (virtual) 2020
- **Brazilian Symposium on Databases (SBBD)** in Rio (virtual) 2020
- **High-Performance Data Science workshop** in Rio (virtual) 2020
- **Computational Science and Engineering Seminar at COPPE/UFRJ** in Rio (virtual) 2020
  - Workflow Provenance in the Lifecycle of Scientific Machine Learning, Invited talk
- **Open Subsurface Data Universe Development Workshop** in Houston, TX 2020
- **Open Subsurface Data Universe Development Workshop** in Houston, TX 2019
- **IEEE/ACM Supercomputing (SC)** in Denver, CO 2019
  Workflows in Support of Large-scale Science (WORKS)
  - Provenance Data in the Machine Learning Lifecycle in Computational Science and Engineering, Oral presentation
- **Scientific Data Analysis using Data-intensive Scalable Computing Workshop** in Rio de Janeiro, Brazil 2019
  - Provenance Data in the Machine Learning Lifecycle in Computational Science and Engineering, Invited talk
- **Open Subsurface Data Universe F2F Meeting** in Houston, TX 2019
- **IEEE International Conference on e-Science** in San Diego, CA 2019
  - Efficient Runtime Capture of Multiworkflow Data using Provenance, Oral presentation
- **Inria Talks** in Montpellier, France 2019
  - Providing Online Data Analytical Support for Humans in the Loop of Computational Science and Engineering Applications, Invited talk
- **IBM Regional Technical Exchange** in Rio de Janeiro, Brazil 2019
- **Provenance Week** in London, UK 2018
  International Provenance and Annotation Workshop (IPAW)
  - Provenance of Dynamic Adaptations in User-steered Dataflows, Oral presentation
  - Capturing Provenance for Runtime Data Analysis in Computational Science and Engineering Applications, Poster presentation

  Computational Reproducibility Workshop
  - Provenance of Dynamic Adaptations in User-steered Dataflows, Oral presentation

- **International Conference on Very Large Databases (VLDB)** in Rio de Janeiro, Brazil 2018
  Latin American Data Science Workshop
  - Tracking Hyperparameter Tuning in Deep Learning Training, Oral presentation
- **Brazilian Syposium on Databases (SBBD)** in Rio de Janeiro, Brazil 2018
- **Brazilian Syposium on Databases (SBBD)** in Uberlandia, Brazil 2017
  - Spark Scalability Analysis in a Scientific Workflow, Oral presentation
  - Controlling the Parallel Execution of Workflows Relying on a Distributed Database, Oral presentation
- **Federal University of Uberlandia, Brazil** in Uberlandia, Brazil 2017
  - Kubernetes, Invited talk
- **Smart City Cloud Hackathon OpenStack Rio** in Rio de Janeiro, Brazil 2017
- **Computer Science Week at UFRJ** in Rio de Janeiro, Brazil 2017
  - Kubernetes, Oral presentation
- **Brazilian Conference on Artificial Intelligence (BRACIS)** in Recife, Brazil 2017
  - Graph Analytics with Spark, Tutorial
- **IEEE/ACM Supercomputing (SC)** in Salt Lake City, UT 2016
  Workflows in Support of Large-scale Science (WORKS)
  - Online Input Data Reduction in Scientific Workflows, Oral presentation
- **ASE BigData/SocialCom/CyberSecurity** in Stanford University, Menlo Park, CA 2014
  - Linked open data publication strategies: Application in networking performance measurement data, poster presentation

Badges and Certifications

- **Machine Learning Specialist Professional** Course duration: 73h — 2022
  Exploratory Data Analysis, Regression, Classification, Deep Learning, Reinforcement Learning, Unsupervised Learning, Time Series and Survival Analysis, AI Ethics and Explainability
- **Trustworthy AI and AI Ethics** Course duration: 3.5h — 2022
- **Enterprise Design Thinking Practitioner** 2022
- **LinkedIn Skill Assessment: Python, MySQL, Linux, T-SQL, NoSQL**

Scientific Community Service

- IEEE Transactions on Parallel and Distributed Systems - Reviewer
- Future Generation Computer Systems - Reviewer
- Concurrency Computation Practice and Experience - Reviewer
- Journal of Parallel and Distributed Computing - Reviewer
- The Very Large Databases (VLDB) Journal - Reviewer
- IEEE Transactions on Big Data - Reviewer
- Journal of Cloud Computing - Reviewer
- Computer Physics Communications - Reviewer
- Discover Data - Reviewer
- Frontiers in High Performance Computing - Reviwer and Editorial Board
- International Workshop on AI Principles in Science Communication (AISC) - PC
- International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD'25) - PC
- Workflows in Distributed Environments (WiDE'24) - PC
- IEEE/ACM Supercomputing (SC'24) - PC
- IEEE International Conference on e-Science (eScience'23) - Session Chair, PC
- Workflows in Support of Large-Scale Science (WORKS'20, 21, 23, 24, 25) - PC
- Brazilian Workshop on Database and Artificial Intelligence Integration - PC

- Brazilian Symposium on Databases (SBBD'20, 23, 24, 25) - Session Chair, PC
- Innovation Summit on Information Systems (at SBSI'19,20) - PC

Languages

- **English** - Full proficiency
  - Missouri State University, U.S. Duration: 150h — Jun 2012 – Aug 2012
    Scientific English for Graduate Students
  - Cultura Inglesa (English Culture), Rio de Janeiro, Brazil 2001 – 2009
- **Portuguese** - Native
- **Spanish** - Fluent reading, intermediate speaking and understanding, limited writing

All Publications and Patents

Journal Articles

[J1] M. Dorier, A. Gueroudji, V. Hayot-Sasson, H. Nguyen, S. Ockerman, **R. Souza**, T. Bicer, H. Pan, P. Carns, K. Chard, "Toward a persistent event-streaming system for high-performance computing applications," *Frontiers in High Performance Computing*, vol. 3, 2025. DOI: 10.3389/fhpcp.2025.1638203. [Online]. Available: https://www.frontiersin.org/journals/high-performance-computing/articles/10.3389/fhpcp.2025.1638203/abstract.

[J2] D. Bard, K. Chard, S. Witt, I. T. Foster, C. Goble, W. Godoy, J. Gustafsson, U.-U. Haus, S. Hudson, L. Los, **R. Souza**, "Workflows community summit 2024: Future trends and challenges in scientific workflows," *Distributed, Parallel, and Cluster Computing (cs.DC)*, 2024. DOI: https://doi.org/10.48550/arXiv.2410.14943. [Online]. Available: https://arxiv.org/abs/2410.14943.

[J3] L. G. Azevedo, **R. Souza**, E. F. d. S. Soares, R. M. Thiago, J. C. C. Tesolin, A. C. Oliveira, M. F. Moreno, "A polystore architecture using knowledge graphs to support queries on heterogeneous data stores," *arXiv preprint Databases (cs.DB)*, 2023. DOI: 10.48550/arXiv.2308.03584. [Online]. Available: https://arxiv.org/abs/2308.03584.

[J4] **R. Souza**, L. G. Azevedo, V. Lourenço, E. Soares, R. Thiago, R. Brandão, D. Civitarese, E. Vital Brazil, M. Moreno, P. Valduriez, M. Mattoso, R. Cerqueira, M. A. S. Netto, "Workflow provenance in the lifecycle of scientific machine learning," *Concurrency and Computation: Practice and Experience*, vol. e6544, pp. 1–21, 2021. [Online]. Available: https://doi.org/10.1002/cpe.6544.

[J5] R. F. Silva, R. M. Badia, V. Bala, D. Bard, P.-T. Bremer, I. Buckley, S. Caino-Lores, K. Chard, C. Goble, S. Jha, ... **R. Souza**, et al. "Workflows community summit 2022: A roadmap revolution," *arXiv preprint Distributed, Parallel, and Cluster Computing (cs.DC)*, 2023. DOI: 10.48550/arXiv.2304.00019. [Online]. Available: https://arxiv.org/abs/2304.00019.

[J6] **R. Souza**, V. Silva, A. A. B. Lima, D. Oliveira, P. Valduriez, M. Mattoso, "Distributed in-memory data management for workflow executions," *PeerJ Computer Science*, vol. 7, pp. 1–30, 2021. DOI: 10.7717/peerj-cs.527. [Online]. Available: https://peerj.com/articles/cs-527/.

[J7] R. F. Silva, H. Casanova, K. Chard, ... **R. Souza**, et al. "Workflows community summit: Advancing the state-of-the-art of scientific workflows management systems research and development," 2021, pp. 1–24. [Online]. Available: https://arxiv.org/abs/2106.05177.

[J8] L. G. Azevedo, **R. Souza**, R. Brandão, V. N. Lourenço, M. Costalonga, M. Machado, M. Moreno, R. Cerqueira, "Adding hyperknowledge-enabled data lineage to a machine learning workflow management system for oil and gas," *First Break*, vol. 38, no. 7, pp. 89–93, 2020. DOI: 10.3997/1365-2397.fb2020055.

[J9] **R. Souza**, V. Silva, A. L. G. A. Coutinho, P. Valduriez, M. Mattoso, "Data reduction in scientific workflows using provenance monitoring and user steering," *Future Generation Computer Systems*, vol. 110, pp. 481–501, 2017, ISSN: 0167-739X. DOI: 10.1016/j.future.2017.11.028.

[J10] **R. Souza**, V. Silva, J. J. Camata, A. L. G. A. Coutinho, P. Valduriez, M. Mattoso, "Keeping track of user steering actions in dynamic workflows," *Future Generation Computer Systems*, vol. 99, pp. 624–643, 2019, ISSN: 0167-739X. DOI: 10.1016/j.future.2019.05.011. [Online]. Available: https://doi.org/10.1016/j.future.2019.05.011.

[J11] V. Silva, L. Neves, **R. Souza**, A. L. G. A. Coutinho, D. Oliveira, M. Mattoso, "Adding domain data to code profiling tools to debug workflow parallel execution," *Future Generation Computer Systems*, pp. 624–643, 2018, ISSN: 0167-739X. DOI: 10.1016/j.future.2018.05.078.

[J12] M. G. Bayser, P. Cavalin, **R. Souza**, A. Braz, H. Candello, C. Pinhanez, J.-P. Briot, "A hybrid architecture for multi-party conversational systems," *arXiv preprint Computation and Language (cs.CL)*, pp. 1–40, 2017. [Online]. Available: https://arxiv.org/abs/1705.01214.

Conference and Workshop Papers

[C1] **R. Souza**, A. Gueroudji, S. DeWitt, D. Rosendo, T. Ghosal, R. Ross, P. Balaprakash, R. F. Silva, "Prov-agent: Unified provenance for tracking AI agent interactions in agentic workflows," in *IEEE International Conference on e-Science*, Chicago, U.S.A.: IEEE, 2025.

[C2] A. Gueroudji, T. Mallick, **R. Souza**, R. F. Silva, R. Ross, M. Dorier, P. Carns, K. Chard, I. Foster, "Controla: Agentic workflow control mechanisms for reliable science," in *IEEE International Conference on e-Science*, Chicago, U.S.A.: IEEE, 2025.

[C3] **R. Souza**, S. Caino-Lores, M. Coletti, T. J. Skluzacek, A. Costan, F. Suter, M. Mattoso, R. F. Silva, "Workflow provenance in the computing continuum for responsible, trustworthy, and energy-efficient AI," in *IEEE International Conference on e-Science*, Osaka, Japan: IEEE, 2024. DOI: https://doi.org/10.1109/e-Science62913.2024.10678731.

[C4] M. Coletti, **R. Souza**, T. J. Skluzacek, F. Suter, R. F. Silva, "Integrating evolutionary algorithms with distributed deep learning for optimizing hyperparameters on HPC system," in *Workflows in Support of Large-Scale Science (WORKS) workshop co-located with the ACM/IEEE International Conference for High Performance Computing, Networking, Storage, and Analysis (SC)*, Atlanta, USA: IEEE, 2024.

[C5] R. F. Silva, W. Shin, F. Suter, A. Gainaru, **R. Souza**, D. Dietz, S. Jha, "Eco-Driven AI-HPC: Optimizing energy efficiency in distributed scientific workflows," in *Energy-Efficient Computing for Science Workshop*, Bethesda, MD, USA, 2024.

[C6] T. J. Skluzacek, **R. Souza**, M. Coletti, F. Suter, R. F. Silva, "Towards cross-facility workflows orchestration through distributed automation," in *Practice and Experience in Advanced Research Computing (PEARC 24)*, Providence, RI, USA: Association for Computing Machinery, 2024. DOI: 10.1145/3626203.3670606. [Online]. Available: https://doi.org/10.1145/3626203.3670606.

[C7] R. F. Silva, K. Maheshwari, T. Skluzacek, **R. Souza**, S. Wilkinson, "Advancing computational earth sciences: Innovations and challenges in scientific hpc workflows," in *European Geosciences Union (EGU)*, 2024.

[C8] L. G. Azevedo, **R. Souza**, E. Soares, R. M. Thiago, J. C. C. Tesolin, A. C. C. M. Oliveira, M. F. Moreno, "HKPoly: A polystore architecture to support data linkage and queries on distributed and heterogeneous data," in *Proceedings of the 20th Brazilian Symposium on Information Systems (SBSI)*, ser. SBSI '24, New York, NY, USA: Association for Computing Machinery, 2024, ISBN: 9798400709968. DOI: 10.1145/3658271.3658322. [Online]. Available: https://doi.org/10.1145/3658271.3658322.

[C9] **R. Souza**, T. J. Skluzacek, S. R. Wilkinson, M. Ziatdinov, R. F. Silva, "Towards lightweight data integration using multi-workflow provenance and data observability," in *IEEE International Conference on e-Science*, 2023. DOI: 10.1109/e-Science58273.2023.10254822. [Online]. Available: https://doi.org/10.1109/e-Science58273.2023.10254822.

[C10] D. Rosendo, M. Mattoso, A. Costan, **R. Souza**, D. Pina, P. Valduriez, G. Antoniu, "ProvLight: Efficient workflow provenance capture on the edge-to-cloud continuum," in *IEEE International Conference on Cluster Computing*, 2023. DOI: 10.1109/CLUSTER52292.2023.00026. [Online]. Available: https://www.computer.org/csdl/proceedings-article/cluster/2023/079200a221/1SfUrCnjgAM.

[C11] R. L. Cunha, L. V. Real, **R. Souza**, B. Silva, M. A. Netto, "Context-aware execution migration tool for data science jupyter notebooks on hybrid clouds," in *IEEE International Conference on e-Science*, 2021. DOI: 10.1109/eScience51609.2021.00013.

[C12] L. Azevedo, **R. Souza**, E. Soares, R. Thiago, A. Oliveira, M. Moreno, "Supporting polystore queries using provenance in a hyperknowledge graph," in *International Semantic Web Conference (ISWC)*, 2021, pp. 1–4.

[C13] **R. Souza**, "User steering support in large-scale workflows," in *PhD Thesis Contest: Brazilian Symposium on Databases (SBBD)*, 2021.

[C14] E. Soares, **R. Souza**, R. Thiago, M. Machado, L. Azevedo, "A recommender for choosing data systems based on application profiling and benchmarking," in *Brazilian Symposium on Databases (SBBD)*, 2021, pp. 265–270.

[C15] R. Brandão, V. Lourenço, M. Machado, L. Azevedo, M. Cardoso, **R. Souza**, G. Lima, R. Cerqueira, M. Moreno, "Cycle orchestrator: A knowledge-based approach for structuring cyclic ml pipelines in the o&g industry," in *International Semantic Web Conference (ISWC)*, 2020.

[C16] R. Brandão, V. Lourenço, M. Machado, L. Azevedo, M. Cardoso, **R. Souza**, G. Lima, R. Cerqueira, M. Moreno, "A knowledge-based approach for structuring cyclic workflows," in *International Semantic Web Conference (ISWC)*, 2020.

[C17] **R. Souza**, J. Camata, M. Mattoso, A. Coutinho, "Runtime steering of parallel cfd simulations," in *International Conference on Parallel Computational Fluid Dynamics*, 2020.

[C18] L. Azevedo, **R. Souza**, R. Thiago, E. Soares, M. Moreno, "Experiencing ProvLake to manage the data lineage of ai workflows," in *Innovation Summit on Information Systems (EISI) in Brazilian Symposium in Information Systems (SBSI)*, 2020.

[C19] L. Azevedo, **R. Souza**, E. Soares, M. Moreno, "Modern federated databases: An overview," in *International Conference on Enterprise Information Systems (ICEIS)*, 2020.

[C20] **R. Souza**, A. Codas, J. A. Nogueira Junior, M. P. Quinones, L. Azevedo, R. Thiago, E. Soares, M. Cardoso, L. Martins, "Supporting the training of physics informed neural networks for seismic inversion using provenance," in *American Association of Petroleum Geologists Annual Convention and Exhibition (AAPG)*, 2020.

[C21] R. Thiago, **R. Souza**, L. Azevedo, E. Soares, R. Santos, W. Santos, M. De Bayser, M. Cardoso, M. Moreno, R. Cerqueira, "Managing data lineage of O&G machine learning models: The sweet spot for shale use case," in *European Association of Geoscientists and Engineers (EAGE) Digitalization Conference and Exhibition*, 2020. DOI: 10.3997/2214-4609.202032075.

[C22] **R. Souza**, L. Azevedo, R. Thiago, E. Soares, M. Nery, M. Netto, E. V. Brazil, R. Cerqueira, P. Valduriez, M. Mattoso, "Efficient runtime capture of multiworkflow data using provenance," in *IEEE International Conference on e-Science*, 2019, pp. 1–10. DOI: 10.1109/eScience.2019.00047. [Online]. Available: https://doi.org/10.1109/eScience.2019.00047.

[C23] **R. Souza**, E. V. Brazil, L. Azevedo, D. Ferreira, E. Soares, R. Thiago, M. Nery, V. Torres, R. Cerqueira, "Managing data traceability in the data lifecycle for deep learning applied to seismic data," in *American Association of Petroleum Geologists Annual Convention and Exhibition (AAPG)*, 2019. [Online]. Available: https://www.searchanddiscovery.com/abstracts/html/2019/ace2019/abstracts/1718.html.

[C24] **R. Souza**, L. Azevedo, V. Lourenço, E. Soares, R. Thiago, R. Brandão, D. Civitarese, E. Vital Brazil, M. Moreno, P. Valduriez, M. Mattoso, R. Cerqueira, M. A. S. Netto, "Provenance data in the machine learning lifecycle in computational science and engineering," in *Workflows in Support of Large-Scale Science (WORKS) co-located with the ACM/IEEE International Conference for High Performance Computing, Networking, Storage, and Analysis (SC)*, 2019, pp. 1–10. DOI: 10.1109/WORKS49585.2019.00006.

[C25] **R. Souza**, L. Neves, L. Azeredo, R. Luiz, E. Tady, P. Cavalin, M. Mattoso, "Towards a human-in-the-loop library for tracking hyperparameter tuning in deep learning development," in *Latin American Data Science (LaDaS) workshop co-located with the Very Large Database (VLDB) conference*, Rio de Janeiro, Brazil, 2018, pp. 84–87.

[C26] V. Silva, **R. Souza**, J. Camata, D. Oliveira, P. Valduriez, A. L. G. A. Coutinho, M. Mattoso, "Capturing provenance for runtime data analysis in computational science and engineering applications," in *Provenance and Annotation of Data and Processes - International Provenance and Annotation Workshop (IPAW)*, ser. Lecture Notes in Computer Science (LNCS), Springer International Publishing, 2018, pp. 183–187, ISBN: 978-3-319-98379-0. DOI: 10.1007/978-3-319-98379-0_15.

[C27] **R. Souza** and M. Mattoso, "Provenance of dynamic adaptations in user-steered dataflows," in *Provenance and Annotation of Data and Processes - International Provenance and Annotation Workshop (IPAW)*, ser. Lecture Notes in Computer Science (LNCS), Springer International Publishing, 2018, pp. 16–29, ISBN: 978-3-319-98379-0. DOI: 10.1007/978-3-319-98379-0_2.

[C28] M. G. Bayser, C. Pinhanez, H. Candello, M. Affonso, M. P. Vasconcelos, M. A. Guerra, P. Cavalin, **R. Souza**, "Ravel: A mas orchestration platform for human-chatbots conversations," in *International Workshop on Engineering Multi-Agent Systems (EMAS@AAMAS 2018)*, 2018.

[C29] P. Valduriez, M. Mattoso, R. Akbarinia, H. Borges, J. Camata, A. L. G. A. Coutinho, D. Gaspar, N. Lemus, J. Liu, H. Lustosa, F. Masseglia, F. Nogueira Da Silva, V. Silva, **R. Souza**, K. Ocaña, E. Ogasawara, D. Oliveira, E. Pacitti, F. Porto, D. Shasha, "Scientific Data Analysis Using Data-Intensive Scalable Computing: the SciDISC Project," in *LADaS: Latin America Data Science Workshop*, vol. CEUR Workshop Proceedings, Rio de Janeiro, Brazil: CEUR-WS.org, 2018. [Online]. Available: https://hal-lirmm.ccsd.cnrs.fr/lirmm-01867804.

[C30] **R. Souza**, V. Silva, J. Camata, A. Coutinho, P. Valduriez, M. Mattoso, "Tracking of online parameter fine-tuning in scientific workflows," in *Workflows in Support of Large-Scale Science (WORKS) workshop co-located with the ACM/IEEE International Conference for High Performance Computing, Networking, Storage, and Analysis (SC)*, Denver, CO, 2017. [Online]. Available: https://hal-lirmm.ccsd.cnrs.fr/lirmm-01620974.

[C31] **R. Souza**, V. Silva, P. Miranda, A. A. B. Lima, P. Valduriez, M. Mattoso, "Spark scalability analysis in a scientific workflow," in *Brazilian Symposium on Databases (SBBD)*, 2017, pp. 288–293.

[C32] **R. Souza**, "Parallel execution of workflows driven by distributed database techniques," in *MSc Thesis Contest: Brazilian Symposium on Databases (SBBD)*, 2017.

[C33] **R. Souza**, V. Silva, A. Coutinho, P. Valduriez, M. Mattoso, "Online input data reduction in scientific workflows," in *Workflows in Support of Large-Scale Science (WORKS) workshop co-located with the ACM/IEEE International Conference for High Performance Computing, Networking, Storage, and Analysis (SC)*, 2016, pp. 1–10. [Online]. Available: https://hal.archives-ouvertes.fr/lirmm-01400538.

[C34] V. Silva, L. Neves, **R. Souza**, A. Coutinho, D. D. Oliveira, M. Mattoso, "Integrating domain-data steering with code-profiling tools to debug data-intensive workflows," in *Workflows in Support of Large-Scale Science (WORKS) workshop co-located with the ACM/IEEE International Conference for High Performance Computing, Networking, Storage, and Analysis (SC)*, Salt Lake City, USA, 2016.

[C35] J. J. Camata, J. M. Cela, D. Costa, A. L. G. A. Coutinho, D. Fernández-Galisteo, **R. Souza**, C. Jiménez, V. Kourdioumov, M. Mattoso, R. Mayo-García, T. Miras, J. A. Moríñigo, J. Navarro, D. d. Oliveira, M. Rodríguez-Pascual, V. Silva, P. Valduriez, "Applying future exascale HPC methodologies in the energy sector," pp. 9–19, 2016. [Online]. Available: https://upcommons.upc.edu/handle/2117/90905.

[C36] P. Cavalin, F. Figueiredo, M. Bayser, L. Moyano, H. Candello, A. Appel, **R. Souza**, "Building a question-answering corpus using social media and news articles," in *International Conference on Computational Processing of the Portuguese Language*, 2016, pp. 353–358.

[C37] J. Camata, J. M. Cela, D. Costa, A. L. G. A. Coutinho, D. Fernández-Galisteo, C. Jimenez, V. Kourdioumov, M. Mattoso, R. Mayo-García, T. Miras, J. A. Moríñigo, J. Navarro, P. O. A. Navaux, D. De Oliveira, M. Rodríguez-Pascual, V. Silva, **R. Souza**, P. Valduriez, "Enhancing Energy Production with Exascale HPC Methods," in *CARLA: Latin American High Performance Computing Conference*, vol. Communications in Computer and Information Science, Mexico City, Mexico: Springer, 2016, pp. 233–246. DOI: 10.1007/978-3-319-57972-6_17. [Online]. Available: https://hal-lirmm.ccsd.cnrs.fr/lirmm-01654914.

[C38] **R. Souza**, V. Silva, D. Oliveira, P. Valduriez, A. A. B. Lima, M. Mattoso, "Parallel execution of workflows driven by a distributed database management system," in *ACM/IEEE International Conference for High Performance Computing, Networking, Storage, and Analysis (SC)*, Salt Lake City, USA, 2015, pp. 1–3. [Online]. Available: `http://sc15.supercomputing.org/sites/all/themes/SC15images/tech_poster/tech_poster_pages/post284.html`.

[C39] R. Castro, **R. Souza**, V. Silva, K. Ocaña, D. Oliveira, M. Mattoso, "Uma abordagem para publicação de dados de proveniência de workflows científicos na web semântica," in *Brazilian Symposium on Databases (SBBD)*, 2015.

[C40] T. Barbosa, **R. Souza**, S. Cruz, M. Campos, R. L. Cottrell, "Applying data warehousing and big data techniques to analyze internet performance," in *International Conference on Internet Applications, Protocols, and Services (NETAPPS)*, 2015.

[C41] **R. Souza**, L. Cottrell, B. White, M. L. Campos, M. Mattoso, "Linked open data publication strategies: Application in networking performance measurement data," in *ASE BigData/SocialCom/CyberSecurity, Stanford, CA*, 2014.

Patents

[P1] M. A. S. Netto, L. C. V. Real, B. Silva, **R. Souza**, *Shortened narrative instruction generator for software code change*, US Patent App. 17/819,025, 2024. [Online]. Available: `https://patents.google.com/patent/US20240053980A1/en`.

[P2] L. C. V. Real, R. L. D. F. Cunha, **R. Souza**, M. A. S. Netto, *Data transformation for acceleration of context migration in interactive computing notebooks*, US Patent App. 17/683,279, 2023. [Online]. Available: `https://patents.google.com/patent/US20230012543A1/en`.

[P3] M. A. S. Netto, B. Silva, R. L. D. F. Cunha, **R. Souza**, L. C. V. Real, *Remotely healing crashed processes*, US Patent App. 17/480,087, 2023. [Online]. Available: `https://patents.google.com/patent/US20230088318A1/en`.

[P4] L. C. V. Real, R. L. D. F. Cunha, M. N. Santos, **R. Souza**, *Asset identification for collaborative projects in software development*, Granted, US Patent App. 17/118,646, 2022. [Online]. Available: `https://patents.google.com/patent/US11650812/en`.

[P5] L. C. V. Real, M. A. S. Netto, R. L. D. F. Cunha, **R. Souza**, A. Braz, *Program context migration*, US Patent App. 17/216,817, 2022. [Online]. Available: `https://patents.google.com/patent/US20220318049A1/en`.

[P6] A. P. Appel, C. R. L. De Freitas, **R. Souza**, C. R. D. A. Mendes, A. Vital, N. Dos, S. Marcelo, M. A. Stelmar Netto, P. B. Avegliano, C. Villas, *Model document creation in source code development environments using semantic-aware detectable action impacts*, US Patent App. 17/353,731, 2022. [Online]. Available: `https://patents.google.com/patent/US20220405065A1/en`.

[P7] L. C. V. Real, M. N. Santos, **R. Souza**, *Continuous storage of data in a system with limited storage capacity*, Granted, US Patent App. 16/678,375, 2021. [Online]. Available: `https://patents.google.com/patent/US11221925/en`.

[P8] **R. Souza**, R. Mozart, F. R. Da Silva, A. Vital, V. T. d. Silva, *Metadata-based scientific data characterization driven by a knowledge database at scale*, Granted, US Patent App. 16/527,546, 2021. [Online]. Available: `https://patents.google.com/patent/US11494611/en`.

[P9] M. G. De Bayser, A. Braz, P. R. Cavalin, F. Figueiredo, **R. Souza**, *Creating coordinated multi-chatbots using natural dialogues by means of knowledge base*, Granted, US Patent Application 15/217,660, 2018. [Online]. Available: `https://patents.google.com/patent/US20180025726A1/en`.

[P10] A. Braz, P. R. Cavalin, F. Figueiredo, M. G. De Bayser, **R. Souza**, *System and method for managing artificial conversational entities enhanced by social knowledge*, Granted, US Patent Application 15/265,615, 2018. [Online]. Available: `https://patents.google.com/patent/US10599644/en`.

[P11] A. P. Appel, A. Gama Leal, **R. Souza**, *Predicting user question in question and answer system*, Granted, US Patent Application 15/171,055, 2017. [Online]. Available: `https://patents.google.com/patent/US11687811B2/en`.