Renan Souza

✉ contact@renansouza.org   •   🌐 RenanSouza.org   •   in renansouza1

🆔 ID: x9t36ewAAAAJ   •   ⌗ renan-souza

Generated on August 11, 2021

Bio

Renan Francisco Santos Souza holds a Ph.D. (2019) and an M.Sc. (2015) in Computer Science from COPPE/Federal University of Rio de Janeiro (UFRJ), and a B.Sc. in Computer Science from UFRJ (2009-2013). Since 2015, he works at IBM Research Brazil, where he is a Research Scientist in the Intelligent Cloud Technologies group. He has been working both as a software engineer and a researcher in several projects since 2010 and has been actively publishing scientific papers in refereed international conferences and journals since 2014. During his B.Sc., he spent a school year at the computer science department at Missouri State University and did a summer internship at Stanford University in the SLAC National Laboratory. During his Ph.D., he was a visiting researcher with the Scientific Data Management team at Inria/Univ. Montpellier in France in 2019. In 2017, he won the best M.Sc. thesis award from SBBD, the main conference on data management in Latin America. He researches large-scale data science and data engineering techniques for the support of Artificial Intelligence systems.

Research Interests

Large-scale Data Science and Data Engineering ▪ Parallel Workflows ▪ Data Provenance ▪ Big Data Analytics ▪ High Performance Computing in Clusters and Clouds ▪ Machine Learning ▪

Education

○ Ph.D. in Computer Science, Federal Univ. of Rio de Janeiro, Brazil   Sep 2015 – Dec 2019
  Supervised by Marta Mattoso and Patrick Valduriez
  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
○ M.Sc. in Computer Science, Federal Univ. of Rio de Janeiro, Brazil   Jan 2013 – Jul 2015
  Supervised by Marta Mattoso
  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
  Title: Linked Open Data Publication Strategies: An Application in Network Performance Data (in pt)
○ Technical Degree in Information Systems, Lemos de Castro   Jan 2005 – Dec 2007

Experience

○ **IBM Research, Research Scientist**   **Aug 2019 – present**
  Leading research and development projects in large-scale data science and data engineering for the support of Artificial Intelligence systems running on hybrid cloud and cluster environments with highly distributed and heterogeneous applications, data, data management systems, and users. My work is often applied to solving industrial data and AI problems.
○ **IBM Research, Research Engineer**   **Sep 2015 – Aug 2019**
  Participated in several research and development projects with clients in the Oil and Gas industry. Developed techniques and implemented systems for large-scale data integration of AI systems running

on clusters and clouds. Also, led Cloud DevOps initiatives to support the design and development of conversational Artificial Intelligence systems.

- o **IBM Research, Software Engineer Intern**   **Apr 2015 – Sep 2015**
  Designed and implemented big data and machine learning solutions to analyze streaming social data.
- o **Stanford University, SLAC, Research Collaborator**   **Aug 2013 – Dec 2014**
  Research collaboration of systems for big data processing, data management, semantic web, and data visualization of PingER project, which contains data about internet performance worldwide.
- o **Stanford University, SLAC, Research Intern**   **May 2013 – Aug 2013**
  Led a project to design and develop a cloud platform that uses semantic web, big data, and data warehousing techniques to store, retrieve, visualize, and publish structured data about internet performance worldwide, giving information and comparison of internet quality.
- o **CAPGov COPPE/UFRJ, Software Eng. Lead**   **Dec 2013 – Sep 2014**
  Led the development of a system that helped the Brazilian population to have easy access to information about public services provided by the Federal Government.
- o **CAPGov COPPE/UFRJ, Software Eng.**   **Sep 2013 – Jan 2014**
  Development of projects to publish linked open data of the Brazilian Federal Register ("Diário Oficial da União") on the semantic web using agile methodology, ontology data modeling, and natural language processing.
- o **CAPGov COPPE/UFRJ, Software Eng. Intern**   **Jan 2011 – Jun 2012**
  Development of a web content management system to facilitate business e-negotiations.
- o **Federal Univ. of Rio de Janeiro, Software Eng. Intern**   **Jan 2010 – Jul 2011**
  Developed a system to integrate data warehouse environments with structured and unstructured data to enable more intelligent and flexible information reports.
- o **Petrobras, IT Intern**   **May 2007 – May 2008**
  Helped to implement features and provided maintenance for web systems to support Petrobras employees.

Selected Publications   For complete list, visit: RenanSouza.org/publications

[1] **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/.

[2] **R. Souza**, L. G. Azevedo, V. Lourenço, E. Soares, R. Thiago, R. Brandão, D. Civitarese, E. V. Brazil, M. Moreno, P. Valduriez, M. Mattoso, R. Cerqueira, M. A. S. Netto, "Workflow provenance in the lifecycle of scientific machine learning," *accepted at Concurrency and Computation: Practice and Experience*, pp. 1–21, 2020. [Online]. Available: https://arxiv.org/abs/2010.00330.

[3] **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 (eScience)*, 2019, pp. 1–10. DOI: 10.1109/eScience.2019.00047.

[4] **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.

Technical Knowledge

- o **Languages**: Python, Java, C, C++, Shell scripting, NodeJS, Scala, Lua
- o **Relational DBMS**: PostgreSQL/PostGIS, DB2, DashDB, MySQL, MySQL Cluster, MS SQL Server
- o **NoSQL DBMS**: MongoDB, AllegroGraph, Jena, Virtuoso, Sesame, Cloudant, CouchBase, Redis, Impala, Elasticsearch, HBase, Hive, Apache Ignite
- o **Heterogeneous Data Management**: Data Integration, Multi-database Queries, Polystores, Foreign Data Wrappers

- **Big Data Frameworks**: Apache Spark: RDD, DataFrames, Streaming, MLib, GraphX, GraphFrames; Hadoop Ecosystem
- **Message Brokers**: Kafka, RabbitMQ
- **Data Science/ML Technologies**: Pandas DataFrames, Jupyter Notebooks, Numpy, SciPy, Tensorflow, and PyTorch
- **Big Data 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**: MPI, OpenMP, CUDA, Data-centric distributed and parallel programming
- **Cloud and Cluster computing**: VMs, Dockers, Kubernetes, OpenShift, HPC Clusters
- **DevOps**: Containers, Kubernetes, OpenShift, CI/CD Pipelines, GitHub, Travis, Jenkins
- **Web Development**: Python Flask/UWSGI, Java EE: JSP, JSF, JPA, Hibernate, Tomcat/JBoss, Spring Boot

Languages

- **English** - Full proficiency
  - Missouri State University, U.S.   Jun 2012 – Aug 2012 (150h)
    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

Grants & Awards

- SBBD Best M.Sc. Thesis Award   2017
- Honored Mention at SBBD 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 Events Participation

- **SIAM Conference on Computational Science and Engineering** in Forth Worth, TX (virtual) 2021

- AI4Seismic: An AI-Driven Platform to Accelerate Geological Discoveries, Oral presentation , link
- Workflow Provenance in the Lifecycle of Scientific Machine Learning, Oral presentation , link

- ○ **Open Subsurface Data Universe Development Workshop** in

- ○ **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

- ○ **Seminarios de Engenharia e Ciencia Computacional** in Rio (virtual)   2020
  - Workflow Provenance in the Lifecycle of Scientific Machine Learning, Oral presentation

- ○ **Open Subsurface Data Universe Development Workshop** in Houston, TX   2020

- ○ **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

- ○ **SciDISC Workshop** in Rio de Janeiro, Brazil   2019
  - Provenance Data in the Machine Learning Lifecycle in Computational Science and Engineering, Oral presentation

- ○ **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, Oral presentation

- ○ **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

- ○ **SBC Brazilian Syposium on Databases (SBBD)** in Rio de Janeiro, Brazil   2018

- ○ **SBC 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

- ○ **Hacker at the 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

- ○ **SBC 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

All Publications and Patents

Journal Articles

[J1] **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/.

[J2] L. G. Azevedo, **R. Souza**, R. Brandão, V. N. Lourenço, M. Costalonga, M. OC 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.

[J3] **R. Souza**, L. G. Azevedo, V. Lourenço, E. Soares, R. Thiago, R. Brandão, D. Civitarese, E. V. Brazil, M. Moreno, P. Valduriez, M. Mattoso, R. Cerqueira, M. A. S. Netto, "Workflow provenance in the lifecycle of scientific machine learning," *accepted at Concurrency and Computation: Practice and Experience*, pp. 1–21, 2020. [Online]. Available: https://arxiv.org/abs/2010.00330.

[J4] **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.

[J5] 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.

[J6] 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.

[J7] **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.

Conference and Workshop Papers

[C1] L. Azevedo, **R. Souza**, E. Soares, R. Thiago, A. Oliveira, M. Moreno, "Supporting polystore queries using provenancein a hyperknowledge graph," in *accepted at International Semantic Web Conference (ISWC)*, 2021, pp. 1–4.

[C2] L. Azevedo, **R. Souza**, E. Soares, M. Moreno, "Modern federated databases: An overview," in *International Conference on Enterprise Information Systems (ICEIS)*, 2020.

[C3] L. Azevedo, **R. Souza**, R. Thiago, E. Soares, M. Moreno, "Experiencing provlake to manage the data lineage of ai workflows," in *Meeting in Innovation in Information Systems (EISI) in Brazilian Symposium in Information Systems (SBSI)*, 2020.

[C4] 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.

[C5] ——, "Cycle orchestrator: A knowledge-based approach for structuring cyclic ml pipelines in the o&g industry," in *International Semantic Web Conference (ISWC)*, 2020.

[C6] **R. Souza**, J. Camata, M. Mattoso, A. Coutinho, "Runtime steering of parallel cfd simulations," in *International Conference on Parallel Computational Fluid Dynamics*, 2020.

[C7] **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.

[C8] 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.

[C9] **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.

[C10] **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 (eScience)*, 2019, pp. 1–10. DOI: 10.1109/eScience.2019.00047.

[C11] **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.

[C12] 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.

[C13] 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.

[C14] **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.

[C15] **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.

[C16] 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.

[C17] **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.

[C18] **R. Souza**, V. Silva, P. Miranda, A. A. B. Lima, P. Valduriez, M. Mattoso, "Spark scalability analysis in a scientific workflow," in *Simpósio Brasileiro de Banco de Dados (SBBD)*, 2017, pp. 288–293.

[C19] T. Barbosa, **R. Souza**, S. Cruz, M. Campos, R. L. Cottrell, "Applying data warehousing and big data techniques to analyze internet performance," SLAC National Accelerator Lab., Menlo Park, CA (United States), Tech. Rep., 2016.

[C20] 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.

[C21] J. J. Camata, J. M. Cela, D. Costa, A. L. G. A. Coutinho, D. Fernández-Galisteo, C. Jiménez, V. Kourdioumov, M. Mattoso, R. Mayo-García, T. Miras, J. A. Moríñigo, J. Navarro, D. d. Oliveira, M. Rodríguez-Pascual, V. Silva, **R. Souza**, P. Valduriez, "Applying future exascale HPC methodologies in the energy sector," pp. 9–19, 2016. [Online]. Available: https://upcommons.upc.edu/handle/2117/90905.

[C22] 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.

[C23] 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.

[C24] **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.

[C25] 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 *Simpósio Brasileiro de Banco de Dados (SBBD)*, 2015.

[C26] **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.

[C27] **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] L. C. V. Real, M. N. Santos, **R. F. S. Souza**, *Continuous storage of data in a system with limited storage capacity*, US Patent App. 16/678,375, 2021.

[P2] **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*, US Patent App. 16/527,546, 2021.

[P3] 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.

[P4] 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.

[P5] A. P. Appel, A. Gama Leal, **R. Souza**, *Predicting user question in question and answer system*, Granted, US Patent Application 15/171,055, 2017.