

FERNANDO SEABRA CHIRIGATI

NYU Tandon School of Engineering • CSE Department
2 MetroTech Center, 10th Floor, 10.053P • Brooklyn, NY • 11201
fchirigati@nyu.edu • fernando.chirigati@gmail.com
<http://fchirigati.com>

EDUCATION

- **Ph.D. in Computer Science** *Jan 2012 – Present*
NYU Tandon School of Engineering • Brooklyn, NY • USA
Advisors: Juliana Freire, Ph.D. and Cláudio T. Silva, Ph.D.
 - **B.E. in Computer and Information Engineering** *Jan 2007 – Dec 2011*
Federal University of Rio de Janeiro • Rio de Janeiro, RJ • Brazil
Advisor: Marta Mattoso, D.Sc.
 - One-semester exchange program *Aug 2010 – Dec 2010*
University of Central Florida • Orlando, FL • USA
Scholarship granted by Brazilian government
-

RESEARCH INTERESTS

Data Management • Large-Scale Data Analysis • Provenance Management and Analytics
Reproducibility • Data Visualization

RESEARCH EXPERIENCE

- *Research Assistant* *Jan 2012 – Present*
Supervised by Juliana Freire, Ph.D.
NYU Tandon School of Engineering • Brooklyn, NY • USA
- *Summer Research Intern* *May 2016 – Aug 2016*
Supervised by Flip Korn and Cong Yu
New York Structured Data Research Group
Google Research • New York City, NY • USA
- *Summer Research Intern* *May 2015 – Aug 2015*
Supervised by Flip Korn and Cong Yu
New York Structured Data Research Group
Google Research • New York City, NY • USA
- *DataONE Working Group Member* *Jun 2012 – Jan 2015*
Scientific Workflows and Provenance Working Group
- *Summer Research Intern* *Jun 2013 – Aug 2013*
Supervised by Jérôme Siméon and Martin Hirzel
IBM T. J. Watson Research Center • Yorktown Heights, NY • USA

- *Research Intern* *Jan 2009 – Feb 2009*
Supervised by Juliana Freire, Ph.D. and Cláudio T. Silva, Ph.D.
University of Utah • Salt Lake City, UT • USA
 - *Research Assistant* *Aug 2007 – Aug 2010*
Supervised by Marta Mattoso, D.Sc.
Federal University of Rio de Janeiro • Rio de Janeiro, RJ • Brazil
-

RESEARCH PROJECTS

- *Urban Data Management and Analytics* *Aug 2014 – Present*
Project focused on understanding the urban environment by efficiently analyzing the numerous interactions among large spatio-temporal urban data.
NYU Tandon School of Engineering • Brooklyn, NY • USA
 - *Reproducibility in Science* *Jan 2012 – Present*
Project focused on developing an infrastructure to facilitate the creation, reuse, and sharing of reproducible computational experiments. ReProZip and noWorkflow are examples of outcomes of the project.
NYU Tandon School of Engineering • Brooklyn, NY • USA
 - *Computational Infrastructure to Support Scientific Experiments* *Jul 2009 – Aug 2010*
Project funded by the Brazilian oil company Petrobras focused on developing efficient solutions for deep-sea oil exploitation experiments.
Federal University of Rio de Janeiro • Rio de Janeiro, RJ • Brazil
 - *Data Management: from Requirements to Distribution* *Aug 2007 – Jul 2009*
Project focused on development infrastructure for running and collecting provenance from scientific experiments.
Federal University of Rio de Janeiro • Rio de Janeiro, RJ • Brazil
-

AWARDS AND HONORS

- *Honorable Mention, SIGMOD 2017 Best Demonstration Award*
Querying and Exploring Polygamous Relationships in Urban Spatio-Temporal Data Sets
Chicago, IL • USA • 2017
- *SIGMOD 2017 Most Reproducible Paper Award*
Data Polygamy: The Many-Many Relationships among Urban Spatio-Temporal Data Sets
Chicago, IL • USA • 2017
- *SIGMOD 2017 Student Travel Award*
Chicago, IL • USA • 2017
- *Pearl Brownstein Doctoral Research Award*
Doctoral research that shows the greatest promise
NYU Tandon School of Engineering • Brooklyn, NY • USA • 2016
- *2nd Place – SIGMOD 2014 Programming Contest*
ViDA Team, together with Tuan-Anh Hoang-Vu, Kien Pham, and Huy T. Vo
Snowbird, UT • USA • 2014

- *Deborah Rosenthal, MD Award*
Outstanding performance on the Ph.D. qualifying examination
NYU Tandon School of Engineering • Brooklyn, NY • USA • 2014
- *A3P Special Honor*
Outstanding performance achieved at the Federal University of Rio de Janeiro
Alumni Association of the Polytechnic School (A3P)
Federal University of Rio de Janeiro • Rio de Janeiro, RJ • Brazil • 2013
- *Magna Cum Laude Honor*
Outstanding performance achieved in Computer and Information Engineering
Federal University of Rio de Janeiro • Rio de Janeiro, RJ • Brazil • 2013
- *Research Honor*
Academic Deliberative Council of Graduate Department of Engineering (COPPE)
Federal University of Rio de Janeiro • Rio de Janeiro, RJ • Brazil • 2010
- *Best Poster Award*
Development of Explicit Control Structures for SWfMS VisTrails (in Portuguese)
XXIV Brazilian Symposium on Databases
Fortaleza, Ceará • Brazil • 2009
- *Honorable Mention*
Top 10 presentations among more than 500 presentations
XXXI Conference on Young Research Assistant
Federal University of Rio de Janeiro • Rio de Janeiro, RJ • Brazil • 2009

PUBLICATIONS

Journals

- *ReproZip: The Reproducibility Packer*, R. Rampin, **F. Chirigati**, D. Shasha, J. Freire, and V. Steeves. In *Journal of Open Source Software (JOSS)*, 2016
- *Knowledge Exploration Using Tables on the Web*, **F. Chirigati**, J. Liu, F. Korn, Y. Wu, C. Yu, and H. Zhang. In *Proceedings of the VLDB Endowment (PVLDB)*, 10(3), pp. 193-204, 2016
- *Exploring What not to Clean in Urban Data: A Study Using New York City Taxi Trips*, J. Freire, A. Bessa, **F. Chirigati**, H. T. Vo, and K. Zhao. In *IEEE Data Engineering Bulletin*, 39(2), pp. 63-77, 2016
- *YesWorkflow: A User-Oriented, Language-Independent Tool for Recovering Workflow Information from Scripts*, T. McPhillips, T. Song, T. Kolisnik, S. Aulenbach, K. Belhajjame, R. Kyle Bocinsky, Y. Cao, J. Cheney, **F. Chirigati**, S. Dey, J. Freire, C. Jones, J. Hanken, K. W. Kintigh, T. A. Kohler, D. Koop, J. A. Macklin, P. Missier, M. Schildhauer, C. Schwalm, Y. Wei, M. Bieda, B. Ludäscher. In *International Journal of Digital Curation (IJDC)*, 10(1), pp. 298-313, 2015
- *The More the Merrier: Efficient Multi-Source Graph Traversal*, M. Then, M. Kaufmann, **F. Chirigati**, T. Hoang-Vu, K. Pham, A. Kemper, T. Neumann, and H. T. Vo. In *Proceedings of the VLDB Endowment (PVLDB)*, 8(4), pp. 449-460, 2014
- *The PBase Scientific Workflow Provenance Repository*, V. Cuevas-Vicenttín, P. Kianmajd, B. Ludäscher, P. Missier, **F. Chirigati**, Y. Wei, D. Koop, and S. Dey. In *International Journal of Digital Curation (IJDC)*, 9(2), pp. 28-38, 2014

- *A Computational Reproducibility Benchmark*, **F. Chirigati**, M. Troyer, D. Shasha, and J. Freire. In IEEE Data Engineering Bulletin, 36(4), pp. 54-59, 2013
- *Chiron: A Parallel Engine for Algebraic Scientific Workflows*, E. Ogasawara, J. Dias, V. Souza, **F. Chirigati**, D. Oliveira, F. Porto, P. Valduriez, and M. Mattoso. In Journal of Concurrency and Computation: Practice and Experience, 25(16), pp. 2327-2341, 2013
- *Similarity-Based Workflow Clustering*, V. Souza, **F. Chirigati**, K. Maia, E. Ogasawara, D. Oliveira, V. Braganholo, L. Murta, and M. Mattoso. In Journal of Computational Interdisciplinary Sciences, vol. 2, pp. 23-35, 2011

Conferences and Workshops

- *Querying and Exploring Polygamous Relationships in Urban Spatio-Temporal Data Sets*, Y. Chan, **F. Chirigati**, H. Doraiswamy, C. Silva and J. Freire. In Proceedings of the 2017 ACM SIGMOD International Conference on Management of Data (**SIGMOD**), pp. 1643-1646, 2017
Honorable Mention, SIGMOD Best Demonstration Award
- *Data Polygamy: The Many-Many Relationships among Urban Spatio-Temporal Data Sets*, **F. Chirigati**, H. Doraiswamy, T. Damoulas, and J. Freire. In Proceedings of the 2016 ACM SIGMOD International Conference on Management of Data (**SIGMOD**), pp. 1011-1025, 2016
SIGMOD Most Reproducible Paper Award
- *ReproZip: Computational Reproducibility With Ease*, **F. Chirigati**, R. Rampin, D. Shasha, and J. Freire. In Proceedings of the 2016 ACM SIGMOD International Conference on Management of Data (**SIGMOD**), pp. 2085-2088, 2016
- *Virtual Lightweight Snapshots for Consistent Analytics in NoSQL Stores*, **F. Chirigati**, J. Siméon, M. Hirzel, and J. Freire. In Proceedings of the 32nd International Conference on Data Engineering (**ICDE**), pp. 1310-1321, 2016
- *noWorkflow: Capturing and Analyzing Provenance of Scripts*, L. Murta, V. Braganholo, **F. Chirigati**, D. Koop, and J. Freire. In Provenance and Annotation of Data and Processes, vol. 8628, Lecture Notes in Computer Science (**LNCS**), pp. 71-83, Springer International Publishing, 2015
- *Packing Experiments for Sharing and Publication*, **F. Chirigati**, D. Shasha, and J. Freire. In Proceedings of the 2013 International Conference on Management of Data (**SIGMOD**), pp. 977-980, 2013
- *ReproZip: Using Provenance to Support Computational Reproducibility*, **F. Chirigati**, D. Shasha, and J. Freire. In Proceedings of the 5th USENIX Conference on Theory and Practice of Provenance (**TaPP**), 2013
- *VisTrails Provenance Traces for Benchmarking*, **F. Chirigati**, D. Koop, J. Freire, and C. Silva. In Proceedings of the 2013 Joint **EDBT/ICDT** Workshops, pp. 323-324, 2013
- *Towards Integrating Workflow and Database Provenance*, **F. Chirigati** and J. Freire. In Provenance and Annotation of Data and Processes, vol. 7525, Lecture Notes in Computer Science (**LNCS**), pp. 11-23, Springer Berlin / Heidelberg, 2012
- *Evaluating Parameter Sweep Workflows in High Performance Computing*, **F. Chirigati**, V. Souza, E. Ogasawara, D. Oliveira, J. Dias, F. Porto, P. Valduriez, and M. Mattoso. In Proceedings of the 1st International Workshop on Scalable Workflow Enactment Engines and Technologies (**SWEET**), article 2, 2012

- *An Evaluation of the Distribution of Dynamic and Static Activities in Parallel Environments using Hydra*, V. Souza, **F. Chirigati**, E. Ogasawara, J. Dias, D. Oliveira, F. Porto, P. Valduriez, and M. Mattoso. In Proceedings of the XXXI Congress of the Brazilian Computer Society, 2011
In Portuguese
- *SimiFlow: An Architecture for Clustering Workflows by Similarity*, V. Souza, **F. Chirigati**, K. Maia, E. Ogasawara, D. Oliveira, V. Braganholo, L. Murta, and M. Mattoso. In Proceedings of the XXX Congress of the Brazilian Computer Society, 2010
In Portuguese
- *GExpLine: A Tool for Supporting Experiment Composition*, D. Oliveira, E. Ogasawara, **F. Chirigati**, V. Souza, L. Murta, and M. Mattoso. In Provenance and Annotation of Data and Processes, vol. 6378, Lecture Notes in Computer Science (**LNCS**), pp. 251-259, Springer Berlin / Heidelberg, 2010
- *A Semantic Approach for Scientific Experiment Lines using Ontologies*, D. Oliveira, E. Ogasawara, **F. Chirigati**, V. Souza, L. Murta, C. Werner, and M. Mattoso. In Proceedings of the III e-Science Workshop, XXIV Brazilian Symposium on Databases, 2009
In Portuguese
- *Scientific Workflow Management System Applied to Uncertainty Quantification in Large Eddy Simulation*, G. Guerra, F. Rochinha, R. Elias, A. Coutinho, V. Braganholo, D. Oliveira, E. Ogasawara, **F. Chirigati**, and M. Mattoso. In Proceedings of the 30th Iberian-Latin-American Congress on Computational Methods in Engineering (CILAMCE), 2009
- *Exploring Many Task Computing in Scientific Workflows*, E. Ogasawara, D. Oliveira, **F. Chirigati**, C. E. Barbosa, R. Elias, V. Braganholo, A. Coutinho, and M. Mattoso. In Proceedings of the 2nd Workshop on Many-Task Computing on Grids and Supercomputers, International Conference for High Performance, Networking, Storage and Analysis (SC), 2009
- *Expliciting Control Flow in Scientific Workflows*, S. M. S. Cruz, F. Chirigati, R. Dahis, M. L. M. Campos, and M. Mattoso. In Proceeding of the II e-Science Workshop, XXIII Brazilian Symposium on Databases, 2008
- *Using Explicit Control Processes in Distributed Workflows to Gather Provenance*, S. M. S. Cruz, **F. Chirigati**, R. Dahis, M. L. M. Campos, and M. Mattoso. In Provenance and Annotation of Data and Processes, vol. 5272, Lecture Notes in Computer Science (**LNCS**), pp. 186-199, Springer Berlin / Heidelberg, 2008

Book Chapters

- *Glossary*, A. Rokem and **F. Chirigati**. In J. Kitzes, D. Turek, and F. Deniz (Eds.), *The Practice of Reproducible Research: Case Studies and Lessons from the Data-Intensive Sciences*, 2017
To Be Published
- *Provenance and Reproducibility*, **F. Chirigati** and J. Freire. In L. Liu and M. T. Özsu (Eds.), *Encyclopedia of Database Systems*, 2017
- *Reproducibility Using VisTrails*, J. Freire, D. Koop, **F. Chirigati**, and C. Silva. In V. Stodden, F. Leisch, and R. Peng (Eds.), *Implementing Reproducible Computational Research (**The R Series**)*, 2014

Editorials

- *A Collaborative Approach to Computational Reproducibility*, **F. Chirigati**, R. Capone, R. Rampin, J. Freire, and D. Shasha. In *Information Systems*, vol. 59, pp. 95-97, 2016

Posters

- *Data Polygamy: The Many-Many Relationships among Urban Spatio-Temporal Data Sets*, **F. Chirigati**, H. Doraiswamy, T. Damoulas, and J. Freire. International Conference on Management of Data (**SIGMOD**), 2016
- *ReproZip: Computational Reproducibility With Ease*, **F. Chirigati**, R. Rampin, D. Shasha, and J. Freire. International Conference on Management of Data (**SIGMOD**), 2016
- *Enhancing Scholarly Communication with ReproZip*, **F. Chirigati**, R. Rampin, V. Steeves, D. Shasha, and J. Freire. FORCE2016 Conference, 2016
- *Constructing a Social Network Analysis System for SIGMOD 2014 Programming Contest*, **F. Chirigati**, K. Pham, T. Hoang-Vu, and H. T. Vo. SIGMOD 2014 Programming Contest, 2014
- *Provenance Storage, Querying, and Visualization in PBase*, V. Cuevas-Vicenttín, P. Kianmajd, B. Ludäscher, P. Missier, **F. Chirigati**, Y. Wei, D. Koop, and S. Dey. In Proceedings of the International Provenance and Annotation Workshop (**IPAW**), 2014
- *Packing Experiments for Sharing and Publication*, **F. Chirigati**, D. Shasha, and J. Freire. International Conference on Management of Data (**SIGMOD**), 2013
- *ReproZip: Packing Experiments for Sharing and Publication*, **F. Chirigati**, D. Shasha, and J. Freire. Beyond the PDF 2 Conference, 2013
- *Procedure to Build Scientific Workflows*, M. P. Rodrigues, J. C. C. Fernandez, **F. Chirigati**, S. M. S. Cruz, and M. C. R. Cavalcanti. XXIV Brazilian Symposium on Databases, 2009
In Portuguese
- *Development of Explicit Control Structures for SWfMS VisTrails*, **F. Chirigati**, R. Dahis, S. M. S. Cruz, J. Freire, C. Silva, and M. Mattoso. XXIV Brazilian Symposium on Databases, 2009
Best Poster Award
In Portuguese
- *A Conception Process for Abstract Workflows: An Example on Deep Water Oil Exploitation Domain*, W. Martinho, E. Ogasawara, D. Oliveira, **F. Chirigati**, F. Correa, B. Jacob, I. Santos, G. H. Travassos, and M. Mattoso. 5th IEEE International Conference on e-Science, 2009

Reproducibility Papers

- *HESML: A Scalable Ontology-based Semantic Similarity Measures Library with a Set of Reproducible Experiments and a Replication Dataset*, J. Lastra-Díaz, A. García-Serrano, M. Batet, M. Fernández, and **F. Chirigati**. In Information Systems, vol. 66, pp. 97-118, 2017
- *Reproducible Experiments on Dynamic Resource Allocation in Cloud Data Centers*, A. Wolke, M. Bichler, **F. Chirigati**, and V. Steeves. In Information Systems, vol. 59, pp. 98-101, 2016

PRESENTATIONS

- *Preserving and Reproducing Research with ReproZip*
Brainhack NYC
New York City, NY • USA • 2017

- *Data Polygamy: The Many-Many Relationships among Urban Spatio-Temporal Datasets*
AWS re:Invent 2016
Las Vegas, NV • USA • 2016
- *Preserving and Reproducing Research with ReProZip*
Preservation and Archiving Special Interest Group (PASIG), Fall 2016 Meeting
New York City, NY • USA • 2016
- *Data Polygamy: The Many-Many Relationships among Urban Spatio-Temporal Data Sets*
International Conference on Management of Data (SIGMOD)
San Francisco, CA • USA • 2016
- *Virtual Lightweight Snapshots for Consistent Analytics in NoSQL Stores*
32nd International Conference on Data Engineering (ICDE)
Helsinki • Finland • 2016
- *ReProZip: Computational Reproducibility with Ease*
Dagstuhl Seminar 16041, Reproducibility of Data-Oriented Experiments in e-Science
Wadern • Germany • 2016
- *Achieving Reproducibility with ReProZip*
CS Colloquium/Tutorial, Columbia University
New York City, NY • USA • 2015
- *Facilitating Reproducibility After the Fact*
BIDS Reproducibility Conference, University of California, Berkeley
Berkeley, CA • USA • 2015
- *Facilitating Reproducibility After the Fact*
Reproducibility Seminar, University of Washington
Seattle, WA • USA • 2015
- *Constructing a Social Network Analysis System for SIGMOD 2014 Programming Contest*
SIGMOD Programming Contest
Snowbird, UT • USA • 2014
- *ReProZip: Packing Experiments for Sharing and Publication*
Workshop on Software Infrastructure for Reproducibility in Science
Brooklyn, NY • USA • 2013
- *ReProZip: Packing Experiments for Sharing and Publication*
Beyond the PDF 2 Conference – Visions of the Future Session
Amsterdam • Netherlands • 2013
- *ReProZip: Packing Experiments for Sharing and Publication*
ICERM Workshop on Reproducibility in Computational and Experimental Mathematics
Lightning Talk
Providence, RI • USA • 2012
- *Towards Integrating Workflow and Database Provenance*
4th International Provenance and Annotation Workshop
Santa Barbara, CA • USA • 2012

- *Evaluating Parameter Sweep Workflows in High Performance Computing*
1st International Workshop on Scalable Workflow Enactment Engines and Technologies
Scottsdale, AZ • USA • 2012

PROFESSIONAL ACTIVITIES

- *Reproducibility Editor* 2015 – Present
Information Systems Journal • Elsevier North-Holland
- *Program Committee Member – Research Track* 2018
Very Large Data Bases (VLDB) Conference
- *Program Committee Member – Demo Track* 2018
ACM International Conference on Management of Data (SIGMOD)
- *Program Committee Member* 2017
SciPy Conference
- *Reproducibility Committee Member* 2016
ACM International Conference on Management of Data (SIGMOD)
- *Artifact Evaluation Committee Member* 2016
European Conference on Object-Oriented Programming (ECOOP)
- *Reproducibility Committee Member* 2015
ACM International Conference on Management of Data (SIGMOD)
- *Artifact Evaluation Committee Member* 2015
European Conference on Object-Oriented Programming (ECOOP)
- *Student Volunteer* 2013
ACM International Conference on Management of Data (SIGMOD)
- *Co-Organizer* 2013
Workshop on Software Infrastructure for Reproducibility in Science
Brooklyn, NY • USA
- *Co-Organizer* 2013
Workshop on Reproducibility in Science
Brooklyn, NY • USA

FREELY-AVAILABLE SOFTWARE SYSTEMS

- **ReproZip**: ReproZip is a tool that automatically captures provenance of experiments and packs all the necessary files, library dependencies, and variables to reproduce the results. Reviewers can then unpack and run the experiments without having to install any additional software.
<http://vida-nyu.github.io/reprozip/>
- **noWorkflow**: noWorkflow is a tool that can transparently capture detailed provenance information from Python scripts. It is non-intrusive, does not require users to change the way they work, and provides different ways to analyze the captured provenance.
<https://github.com/gems-uff/noworkflow/>