Ana Trisovic

anatrisovic.com | github.com/atrisovic | twitter.com/atrisovic
Department of Biostatistics, 655 Huntington Ave, Building II, Room 403, Boston, MA 02115
anatrisovic@g.harvard.edu

Research interests

My primary research interests are computational reproducibility, software sustainability, data engineering and workflows, research dissemination and reuse. I am enthusiastic about data science, machine learning and open science.

Education

2014 - 2018	**University of Cambridge, Newnham College**, United Kingdom PhD in Computer Science, *Thesis title: Data preservation and reproducibility at the LHCb experiment at CERN* The CERN experiments collect vast amounts of data from particle collisions, inducing a large number of publications. This thesis presents approaches to facilitate reproducible research, advance open data efforts and assist in transparently running scientific software in the future using novel technologies like *Neo4j* graph database, *Docker*, *OpenStack* and *DevOps* tools.
2010 - 2014	**Union University, School of Computing**, Belgrade, Serbia BSc in Computer Science
2010 - 2013	**University of Belgrade, Faculty of Mechanical Engineering** BSc in Mechanical Engineering

Work Experience

Feb 2022 - Present	**Research Associate** at Harvard Biostatistics Undertake research, software development, and integration of health and various geospatial data toward advancing the quality, reproducibility, and reuse of statistical analysis on air pollution and public health.
Sept 2019 - Feb 2022	**Sloan Postdoctoral Fellow** at Harvard University Advancing reproducibility and reuse at the Dataverse project · work on various projects to improve the quality and reproducibility of the shared replication packages (research data & code).
Sept 2018 - Sept 2019	**CLIR Postdoctoral Fellow** at the University of Chicago In collaboration with the Energy Policy Institute at the University of Chicago (EPIC), facilitate reproducibility and openness of energy, environmental and climate research.
Sept 2017 - Sept 2018	**Project Associate** at CERN CERN Analysis Preservation and CERN Open Data · back-end web development for the preservation and open data services for the LHC collaborations.
July 2013 – July 2014	**Technical Student** at CERN Educational LHCb particle collision display · development of the application, which was later used by hundreds of high-school students through the event International Masterclass in Physics.

Updated 01/2023

MARCH 2013 – JULY 2013	**Data Science Associate** at MICROSOFT *Development Center Serbia*
	Blacklister Efficiency · evaluation of the load balancer and blacklister using data from the Azure cloud and creating a proposal for improvements.

PUBLICATIONS

1. Daniel Garijo, Hervé Ménager, Lorraine Hwang, Ana Trisovic, Michael Hucka, Thomas Morrell, Alice Allen, Task Force on Best Practices for Software Registries, and SciCodes Consortium. Nine Best Practices for Research Software Registries and Repositories. *PeerJ Computer Science*, 2022

2. Ana Trisovic, Matthew K Lau, Thomas Pasquier, and Mercè Crosas. A Large-scale Study on Research Code Quality and Execution. *Nature Scientific Data*, 2022

3. Stian Soiland-Reyes, Peter Sefton, Mercè Crosas, Leyla Jael Castro, Frederik Coppens, José M Fernández, Daniel Garijo, Björn Grüning, Marco La Rosa, Simone Leo, Eoghan Carragáin, Marc Portier, Ana Trisovic, RO-Crate Community, Paul Groth, and Carole Goble. Packaging research artefacts with RO-Crate. *Data Science*, 2021

4. Nadica Miljkovic, Ana Trisovic, and Limor Peer. Towards FAIR principles for open hardware. *Conference on Application of Free Software and Open Hardware (PSSOH)*, 2021

5. Kelly Blumenthal, Aleksandrina Goeva, Sara Stoudt, Ana Trisovic, and Pavle Trisovic. Why Do We Plot Data? *Harvard Data Science Review*, 2021

6. Ana Trisovic, Katherine Mika, Ceilyn Boyd, Sebastian Feger, and Mercè Crosas. Repository Approaches to Improving the Quality of Shared Data and Code. *Data*, 2021

7. Aleksandrina Goeva, Peyton Jones, Sara Stoudt, and Ana Trisovic. Recipes for Connector Courses From the Early-Career Board Kitchen. *Harvard Data Science Review*, 2021

8. Shuang Frost, Aleksandrina Goeva, Javin Pombra, Will Seaton, Sara Stoudt, Ana Trisovic, Chris Wang, and Catherine Zucker. Kaleidoscopic Perspectives on Practicum-Based Data Science Education. *Harvard Data Science Review*, 2021

9. Ana Trisovic, Philip Durbin, Tania Schlatter, Gustavo Durand, Sonia Barbosa, Danny Brooke, and Mercè Crosas. Advancing Computational Reproducibility in the Dataverse Data Repository Platform. *3rd International Workshop on Practical Reproducible Evaluation of Computer Systems (P-RECS)*, 2020

10. Aleksandrina Goeva, Sara Stoudt, and Ana Trisovic. Toward Reproducible and Extensible Research: from Values to Action. *Harvard Data Science Review*, 2020

11. Shuang Frost, Aleksandrina Goeva, William Seaton, Sara Stoudt, and Ana Trisovic. Early-Career View on Data Science Challenges: Responsibility, Rigor, and Accessibility. *Harvard Data Science Review*, 2020

12. Anna E. Woodard, Ana Trisovic, Zhuozhao Li, Yadu Babuji, Ryan Chard, Tyler Skluzacek, Ben Blaiszik, Daniel S. Katz, Ian Foster, and Kyle Chard. Real-Time HEP Analysis With FuncX – a High-Performance Platform for Function as a Service. *24th International Conference on Computing in High Energy & Nuclear Physics (CHEP)*, 2020

13. Ana Trisovic, Chris R. Jones, Ben Couturier, and Marco Clemencic. Provenance Tracking in the LHCb Software. *Computing in Science & Engineering (CISE)*, 2020

14. Xiaoli Chen, Sünje Dallmeier-Tiessen, Robin Dasler, Sebastian Feger, Pamfilos Fokianos, Jose Benito Gonzalez, Harri Hirvonsalo, Dinos Kousidis, Artemis Lavasa, Salvatore Mele, Diego R. Rodriguez, Tibor Simko, Tim Smith, Ana Trisovic, Anna Trzcinska, Ioannis Tsanaktsidis, Markus Zimmermann, Kyle Cranmer, Lukas Heinrich, Gordon Watts, Michael Hildreth, Lara Lloret Iglesias, Kati Lassila-Perini, and Sebastian Neubert. Open Is Not Enough. *Nature Physics*, 2019

15. Ana Trisovic. Graph Mining at the High-Energy Physics Experiment LHCb. *7th International Symposium on Industrial Engineering*, 2018

16. Ana Trisovic, Ben Couturier, Val Gibson, and Chris R. Jones. Recording the LHCb Data and Software Dependencies. *22th International Conference on Computing in High Energy & Nuclear Physics (CHEP)*, 2017

17. Thomas Pasquier, Matthew K. Lau, Ana Trisovic, Emery R. Boose, Ben Couturier, Mercè Crosas, Aaron M Ellison, Valerie Gibson, Chris R. Jones, and Margo Seltzer. If These Data Could Talk. *Nature Scientific Data*, 2017

18. Ana Trisovic. Measuring the D^0 Lifetime at the LHCb Masterclass. *37th International Conference on High Energy Physics (ICHEP)*, 2016

Upcoming Publications

1. Whanhee Lee, Wu Xiao, Heo Seulkee, Kim Joyce Mary, Fong Kelvin C., Son Ji-Young, Sabath Matthew Benjamin, Trisovic Ana, Braun Danielle, Park Jae Yoon, Kim Yong Chul, Lee Jung Pyo, Schwartz Joel, Kim Ho, Dominici Francesca, Al-Aly Ziyad, and L. Bell Michelle. Air Pollution and Acute Kidney Injury in the US Medicare Population: A Longitudinal Cohort Study. *Accepted in Environmental Health Perspectives*, 2023

2. Ana Trisovic. Cluster Analysis of Open Research Data: A Case for Replication Metadata. *To appear in International Journal of Digital Curation*, 2023

3. Daina Bouquin, Oliver Bertuch, Elena Colon-Marrero, and Ana Trisovic. Advancing Software Citation Implementation. 2023

I am a coauthor of many publications as a member of the LHCb collaboration at CERN.

Notable Conferences & Workshops

| 2022 | **Statistical Software Conference** |
| | Invited talk on "Evidence-based practices for better research software". |

| 2022 | **IEEE eScience Conference** |
| | Poster presentation "Cluster Analysis on Open Research Data". |

| 2022 | **International Conference on Applied Sciences (ICAS)** |
| | Talk "Conducting reusable big data analyses on air pollution and public health". |

| 2022 | **Journal Editors Discussion Interface (JEDI) workshop** |
| | Invited talk. |

2022	**Serbian Academy of Sciences and Arts** Invited talk.
2022	**Pew Research Center** Invited talk "The Landscape of Data Sharing and Computational Reproducibility for Social Research".
2021	**Gathering for Open-Source Hardware (GOSH) meeting** Invited talk.
2021	**SciCodes meeting** Talk on research code review.
2021	**HERMES (Helmholtz Rich Metadata Software Publication) workshop** Invited talk on Dataverse Project.
2021	**Metascience Conference** Talk "Evidence-based steps toward a culture for replicability and reproducibility".
2021	**Toward Open, Reproducible and Reusable Research** Organiser, co-chair with Qian Zhang and presenter.
2020	**SORSE – International Series of Online Research Software Events** Talk "Improving FAIRness with Containers".
2020	**Portage Network Webinar** Talk "Enabling painless reuse of shared research data and code".
2020	**Women in Data Science (WiDS)** Poster presentation "Reproducibility in Data Science".
2019	**14th Research Data Alliance (RDA) Plenary** Poster presentation	Awarded the RDA/US Early-Career Fellowship.
2019	**Building Reproducible Workflows for Earth Sciences at ECMWF** Keynote "Responding to reproducibility challenges from physics to social sciences".
2019	**Open-Source Software Health Index workshop** Invited attendee.
2019	**University of Texas at Austin** Talk "Responding to reproducibility challenges"
2019	**CNI Spring 2019 Membership Meeting** Invited attendee.
2019	**Education and Training for Reproducible Research with Whole Tale** Invited attendee.
2017	**10th LHCb Computing Workshop** Talk on the development of the CERN Analysis Preservation portal.
2016	**22nd Conference on Computing in High Energy Physics (CHEP)** Oral presentation	Awarded the CHEP 2016 Student Scholarships.

2016	**7th LHCb Computing Workshop at CERN**
	Talk on capturing and serving LHCb data provenance.

2016	**Multidisciplinary Perspectives on Data Preservation and Access**
	Talk on the development of the provenance database at LHCb.

2016	**3rd Annual Oxbridge Women in CS Conference**
	Lightning talk on the LHCb provenance database.

2016	**6th LHCb Computing Workshop**
	Talk on data preservation and provenance capture efforts.

2015	**2nd Annual Oxbridge Women in CS Conference**
	Poster presentation on reproducibility at LHCb.

2015	**Newnham College Graduate Conference**
	Poster presentation	Best poster award.

2014	**37th International Conference in High-Energy Physics (ICHEP)**
	Oral presentation "Measuring the D^0 lifetime at the LHCb Masterclass."

2011	**10th Annual Petnica Students' Science Conference**
	Talk "Design and optimization of rectangular optical waveguides".

Grants, Scholarships and Awards

2022 - 2023	**National Institutes of Health (NIH) Supplement Grant**
	for supporting dissemination of geospatial and climate datasets in data repository Dataverse

2022 - 2023	**National Institutes of Health (NIH) Supplement Grant**
	for approaches for AI/ML Readiness for Wildfire Exposures

2020 - 2023	**Alfred P. Sloan Foundation Grant**
	for improving dissemination of computational research in data repository Dataverse

2019 - 2021	**CLIR Micro-Grant for Research**
	for improving computational reproducibility with gamification

2019 - 2020	**Amazon AWS Cloud Credits for Research**
	for Evaluation of reproducibility in data repository Dataverse

2015 - 2017	**Muir Wood Scholarship**
	for Mathematics, Engineering and Science awarded by Newnham College

2014 - 2017	**CERN Doctoral Studentship**
	to work with the LHCb computing group and undertake the PhD research

2015	**Scholarship to attend the Summer School on Machine Learning**
	in High Energy Physics in Saint-Petersburg, Russia awarded by Yandex

2014	**Google Anita Borg Memorial Scholarship (Women Techmakers)**
	for outstanding achievement, leadership and community involvement

2013	**Dositeja Award**
	for the best students in senior year of study, awarded by the Republic of Serbia

2012	**Scholarship by the German Academic Exchange Service DAAD**
	for attending international summer school of German in Munich, Germany

2011 - 2012	**Praise for outstanding academic achievements**
	awarded by the Faculty of Mechanical Engineering, University of Belgrade

2010	**International Competition First Step to Nobel prize in Physics**
	Honorable Mention

2006	**Best Student Award**
	in secondary school Despot S. Lazarevic, Belgrade, Serbia

Computer Skills

Programming languages:	Python, R
Cloud & Distributed technologies:	Docker, OpenStack, AWS
Databases & query languages:	Neo4j graph database, CYPHER, SQL
Web development & design:	Flask, HTML/CSS, Adobe Photoshop
Online training:	R (coursera.org)
Other:	Slurm, Vim, Git, LaTeX

Teaching Experience

2021	Xidian University, Xi'an China
	Lecturer in Introduction to Machine Learning summer course

2019	Workshop for students and staff at UChicago Library
	Instructor on version control with git

2017	LHCb starterkit · training course for PhD students
	Instructor

2017	Django girls Geneva · programming for high-school students
	Coach on Python, Django and SQL

2015	Newnham College, University of Cambridge
	Teaching assistant (supervisor) in Algorithms

2015	Cavendish Laboratory, University of Cambridge
	Demonstrator in Computational Physics and Physics Practical

Professional and outreach activities

2023	**IEEE eScience Conference**
	Poster Session Co-Chair.

2023 –	**Journal of Open-Source Software (JOSS)** Editor.
2021 –	**Consortium of Scientific Software Registries and Repositories (SciCodes)** Committee Member.
2021 –	**Journal of Systems Research (JSys)** Review Board Member in Data Science and Reproducibility.
2020 –	**Harvard Data Science Review (HDSR)** Member of the Early-Career Board.
2021 –	**GDCC Software and Container Metadata Working Group** Working Group Chair.
2022	**The Scientist** Interview on challenges with research code.
2022	**The RSE Stories podcast** Conversation on data engineering.
2020 - 2021	**Annual Dataverse Community Meeting** Member of the Planning Committee and a Session Co-Chair.
2020	**SciPy Scientific Computing Conference** Program Committee Member of SciPy 2020 Talks and Posters.
2020	**The Theory and Practice of Provenance (TaPP) Workshop** Program Committee Member.
2017	**Outreach for Researchers' Night at CERN**
2015	**Volunteer at TEDxCERN Conference**
2013 – 2016	**Moderator at the International Masterclass in Physics Events**

Languages

ENGLISH:	Fluent
SERBIAN:	Mother tongue
SPANISH, GERMAN:	Basic Knowledge