# Sara Lafia, Ph.D.

**Research Fellow, ICPSR**                                slafia@umich.edu

**University of Michigan, Ann Arbor, MI**                         saralafia.com

## Education

**Ph.D. in Geography, UC Santa Barbara, CA**                    9/2014 - 7/2020
*Dissertation*: Designing for Serendipity: Research Data Curation in Topic Spaces
*Committee*: Werner Kuhn (Chair), James Frew, Kelly Caylor, and Daniel Montello
*Emphasis*: Information Technology and Society
*Certificate*: College and University Teaching (CCUT)

**B.S. in Urban Planning, Geography - GIS, Cal Poly Pomona, CA**          9/2009 - 6/2014
Kellogg Honors College (*Valedictorian*)

## Appointments

**Research Fellow**                                        8/2020 - present
ICPSR, University of Michigan, Ann Arbor, MI

**Engineering Intern**                                      6/2017 - 8/2017
Esri Research & Development, Arlington, VA

**Informatics Intern**                                      6/2016 - 9/2016
National Center for Ecological Analysis and Synthesis, Santa Barbara, CA

**Graduate Student Researcher**                               9/2014 - 7/2020
Center for Spatial Studies, UC Santa Barbara, CA

## Publications

### Peer Reviewed Articles

Fan, L., **Lafia, S.**, Wofford, M., Thomer, A., Yakel, E., & Hemphill, L. (accepted). Mining Semantic Relations in Data References to Understand the Roles of Research Data in Academic Literature. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries* (JCDL).

**Lafia, S.**, Thomer, A., Moss, E., Bleckley, D., and Hemphill, L. (2023). How and Why do Researchers Reference Data? A Study of Rhetorical Features and Functions of Data References in Academic Articles. *Data Science Journal*. doi: 10.5334/dsj-2023-010

**Lafia, S.**, Million, A. J., & Hemphill, L. (2023). Direct, Orienting, and Scenic Paths: How Users Navigate Search in a Research Data Archive. *Proceedings of the ACM on Human Information Interaction and Retrieval*. doi: 10.1145/3576840.3578275

**Lafia, S.**, Fan, L., Thomer, A., & Hemphill, L. (2022). Subdivisions and Crossroads: Identifying Hidden Community Structures in a Data Archive's Citation Network. *Quantitative Science Studies* (QSS). doi: 10.1162/qss_a_00209

**Lafia, S.**, Fan, L., & Hemphill, L. (2022). A Natural Language Processing Pipeline for Detecting Informal Data References in Academic Literature. *Proceedings of the Association for Information Science and Technology* (ASIS&T). doi: 10.1002/pra2.614

Thomer, A., Akmon, D., York, J., Tyler, A.R.B., Polasek, F., **Lafia, S.**, Hemphill, L., & Yakel, E. (2022). The craft and coordination of data curation: complicating "workflow" views of data science. *Proceedings of the ACM on Human Computer Interaction* (PACM HCI). doi: 10.1145/3555139.

Hemphill, L., Pienta, A., **Lafia, S.**, Akmon, D., & Bleckley, D. (2022). How do properties of data, their curation, and their funding relate to reuse?. *Journal of the Association for Information Science and Technology* (JASIST). doi: 110.1002/asi.24646

**Lafia, S.**, Thomer, A., Bleckley, D., Akmon, D., & Hemphill, L. (2021). Leveraging Machine Learning to Detect Data Curation Activities. In *Proceedings of 17th IEEE eScience 2021*, Innsbruck, Austria, September 20-23. Institute of Electrical and Electronics Engineers (IEEE). doi: 10.1109/eScience51609.2021.00025

**Lafia, S.**, Kuhn, W., Caylor, K., & Hemphill, L. (2021). Mapping Research Topics at Multiple Levels of Detail. *Patterns*, 100210. doi: 10.1016/j.patter.2021.100210

**Lafia, S.**, Zhu, R., Regalia, B., & Kuhn, W. (2021). Reimagining GIS Instruction through Concept-Based Learning. In *AGILE: GIScience Series* (AGILE 2021), 2, 1-7. doi: 10.5194/agile-giss-2-6-2021

Hervey, T., **Lafia, S.**, & Kuhn, W. (2020). Search Facets and Ranking in Geospatial Dataset Search. In *11th International Conference on Geographic Information Science* (GIScience 2021). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik. doi: 10.4230/LIPIcs.GIScience.2021.I.5

**Lafia, S.**, Last, C., & Kuhn, W. (2019). Enabling the Discovery of Thematically Related Research Objects with Systematic Spatializations. In *14th International Conference on Spatial Information Theory* (COSIT 2019). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik. doi: 10.4230/LIPIcs.COSIT.2019.18

**Lafia, S.**, Turner, A., & Kuhn, W. (2018). Improving Discovery of Open Civic Data. In *10th International Conference on Geographic Information Science* (GIScience 2018). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik. doi: 10.4230/LIPIcs.GISCIENCE.2018.9

**Lafia, S.**, & Kuhn, W. (2018). Spatial Discovery of Linked Research Datasets and Documents at a Spatially Enabled Research Library. *Journal of Map & Geography Libraries* (JMGL), 14(1), 21-39. doi: 10.1080/15420353.2018.1478923

Allen, C., Hervey, T., **Lafia, S.**, Phillips, D. W., Vahedi, B., & Kuhn, W. (2016). Exploring the Notion of Spatial Lenses. In *International Conference on Geographic Information Science* (GIScience 2016), pp. 259-274. Springer, Cham. doi: 10.1007/978-3-319-45738-3_17

**Lafia, S.**, Jablonski, J., Kuhn, W., Cooley, S., & Medrano, F. A. (2016). Spatial Discovery and the Research Library. *Transactions in GIS* (TGIS), *20*(3), 399-412. doi: 10.1111/tgis.12235 doi:10.1007/978-3-319-45738-3_17

## Conferences and Workshops

Fan, L., **Lafia, S.**, Bleckley, D., Moss, E., Thomer, A., Hemphill, L. (2022). Librarian-in-the-Loop: A Natural Language Processing Paradigm for Detecting Informal Mentions of Research Data in Academic Literature. *Workshop on Investigating Data Work Across Domains: New Perspectives on the Work of Creating Data* (CHI 2022).

**Lafia, S.**, Moss, E., Thomer, A. & Hemphill, L. (2021). Detecting Informal Data Use in Literature. *The Future of Research Communications and e-Scholarship* (FORCE11). doi: 10.5281/zenodo.5748382

**Lafia, S.,** Thomer, A., Bleckley, D., Akmon, D., and Hemphill, L. (2021). Analyzing Data Curation Work at a Social Science Archive with Machine Learning. *WebSci'21 Workshop on Research Infrastructure for Web Science* (RI4WebSci).

Akmon, D., **Lafia, S.**, Thomer, A., Hemphill, L., Pienta, A., Yakel, E., Bleckley, D., and Tyler, A.R.B. (2020). Measuring and improving the efficacy of curation activities in data archives. *Research Data Alliance* (RDA 2020).

**Lafia, S.**, Xiao, J., Kuhn, W., & Hervey, T. (2019). Talk of the Town: Discovering Government Data via Conversational Voice Assistants. In *14th International Conference on Spatial Information Theory* (COSIT 2019). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

Golubovic, N., Krintz, C., Wolski, R., **Lafia, S.**, Hervey, T., & Kuhn, W. (2016). Extracting Spatial Information from Social Media in Support of Agricultural Management Decisions. In *Proceedings of the 10th Workshop on Geographic Information Retrieval* (p. 4). ACM.

**Lafia, S.**, & Staehli, L. (2016). From Research Objects to Research Networks: Combining Spatial and Semantic Search. In *Workshop on Semantic Data on the Web* (SDW@GIScience) (pp. 35-40).

## News and Reports

**Lafia, S**., & Kuhn, W. (2019). Enabling the Spatial Discovery of Research Data in Libraries. In *US National Report* (US National Committee for the International Cartographic Association).

Coggins, B. L., **Lafia, S.**, & Torghabeh, B. V. (2018). Dramatic Change in North Korea: Instability and Human Flight Propensity. *North Korean Review*, 14(1), 49-70.

Seltmann, K., **Lafia, S.**, Paul, D., James, S., Bloom, D., Rios, N., ... & Davis, E. (2018). Georeferencing for Research Use (GRU): An Integrated Geospatial Training Paradigm for Biocollections Researchers and Data Providers. *Research Ideas and Outcomes* (RIO), *4*, e32449.

**Lafia, S.** & Kuhn, W. (2017). Opening Access to University Scholarship through Spatial Discovery. *ArcUser* (ESRI).

## Datasets and Code

**Lafia, S.** (2022). ICPSR Bibliography Citation Network. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2022-07-05. doi: 10.3886/E174361V1

**Lafia, S.** (2022). ICPSR Dataset References: v1.0.0 (v1.0.0). Zenodo. doi: 10.5281/zenodo.6798661

# Presentations

### Invited Talks

| | |
|---|---|
| "The Power and Perils of Spatial Metaphors: Storytelling and Decision-making with Maps." UMD iSchool | 2022 |
| "Computational Perspectives on Curating Data for Discovery and Reuse." University of Pittsburgh SCI | 2022 |
| "Measuring the Impacts of Curatorial Actions on Research Data Reuse." LIBER Webinar: Online | 2020 |
| "Designing for Serendipity: Supporting Data Discovery with Spatial Curation." ICPSR: Ann Arbor, Michigan | 2020 |
| "Discovering Spatial Data for Our Changing Earth." ESIP Summer Meeting: Tucson, Arizona | 2018 |

### Oral Presentations

| | |
|---|---|
| "A NLP Pipeline for Detecting Informal Data References in Academic Literature." ASIS&T: Pittsburgh, PA | 2022 |
| "Digitizing and Structuring Historical Records with Deep Learning." SIG AI at ASIS&T: Pittsburgh, PA | 2022 |
| "Surfacing Specimen Citations for Natural History Collections." AGU Annual Meeting: Chicago, IL | 2022 |
| "Expanding the ICPSR Bibliography of Data-related Literature." ICPSR Data Fair: Online | 2022 |
| "From Dataset Detection to Instructional Resources." ICPSR Biennial Meeting: Online | 2021 |
| "Analyzing Data Curation Work at a Social Science Archive with Machine Learning." RI4WebSci: Online | 2021 |
| "Designing for Serendipity: Research Data Curation in Topic Spaces." Dissertation Public Defense: UCSB | 2020 |
| "Enabling the Discovery of Thematically Related Research Objects." COSIT: Regensburg, Germany | 2019 |
| "Talk of the Town: Discovering Government Data." COSIT: Regensburg, Germany | 2019 |
| "Spatially Supported Thematic Discovery of Research Data." COSIT: Regensburg, Germany | 2019 |
| "Re-creating and Exploring Research Neighborhoods." FORCE11 Summer Institute: UCLA | 2019 |
| "Spatial Data Discovery at UCSB: Future Directions." Library Open Access Week: UCSB | 2018 |
| "Improving Discovery of Open Civic Data." GIScience: Melbourne, Australia | 2018 |
| "Discovering and Sharing Campus Scholarship." Center for Spatial Studies Lightning Talks: UCSB | 2017 |
| "Linking Geographic Footprints to Topic Spaces." Graduate Slam Competition: UCSB | 2017 |
| "From Research Objects to Research Networks." Western Association of Map Librarians: Stanford, CA | 2016 |
| "Spatial Discovery and the University Library." TGIS Session at ESRI User Conference: San Diego, CA | 2016 |
| "People and Places: Resource Discovery in a Library Context." AAG: San Francisco, CA | 2016 |
| "Degrees of Separation: Measuring Musicians and Places." Center for Spatial Studies: UCSB | 2016 |

**<u>Poster Presentations</u>**

| | |
|---|---|
| "Detecting Specimen Citations in Scientific Literature." ESIP Summer Meeting: Pittsburgh, PA. | 2022 |
| "Relating Data Reuse to Curation Decisions and Funding Models." RDAP Summit: Online. | 2022 |
| "Detecting Informal Data Use in Literature." FORCE11: Online | 2021 |
| "Detecting Informal Indicators of Research Data Reuse in Scientific Publications." Text as Data: Online | 2021 |
| "How Does Data Curation Impact Discovery?" ESIP Winter Meeting: Online | 2021 |
| "Measuring and Improving the Efficacy of Curation Activities in Data Archives." RDA 16th Plenary: Online | 2020 |
| "Supporting Interdisciplinary Research Reviews with Multi-Level Topic Maps." MIDAS Symposium: Online | 2020 |
| "Mapping Research Topics at Multiple Levels of Detail." AIDR Symposium: Online | 2020 |
| "Computing with Core Concepts of Spatial Information." COSIT: Santa Fe, New Mexico | 2015 |

# External Funding

**National Science Foundation**                                                      2021 - present
Co-Principal Investigator with Libby Hemphill (PI), & A.J. Million. "Transforming Data Discovery Through Behavior Modeling and Recommendation" (#2121789). SBE Social and Economic Sciences, Science of Science: Discovery, Communication, and Impact. $362,861.

**Michigan Institute for Data Science**                                              2021 - present
Co-Principal Investigator with J. Trent Alexander (PI), Katie R. Genadek, & David A. Bleckley. "Images to Integrated Data: Piloting new methods to digitize, parse, and link historical records". Propelling Original Data Science (PODS) Grants. $69,910.

# Awards and Competitions

| | |
|---|---|
| ESIP Community Fellow: Earth Science Information Partners (ESIP), Discovery Cluster | 2020 - 2022 |
| Kaggle Competitions Contributor: Top 45% in Coleridge Initiative - Show US the Data | 2021 |
| Best Paper of the Year Award: Journal of Map & Geography Libraries, Volume 14 | 2020 |
| Excellence in Teaching Award: Department of Geography, UC Santa Barbara | 2019 |
| Mellichamp Ph.D. Research Fellowship: 21st Century Global Dynamics, UC Santa Barbara | 2019 |
| Womxn/Hacks Hackathon Winner (Best Use of Google Cloud Platform): UC Santa Barbara | 2019 |
| SAP + Esri Spatial Hackathon Finalist (Geo-machine learning): Esri DevSummit | 2018 |
| Robert Raskin Scholarship: Earth Science Information Partners (ESIP) | 2018 |
| Jack & Laura Dangermond Graduate Fellowship: Department of Geography, UC Santa Barbara | 2017 - 2018 |

Greater Research Opportunities (GRO) Fellowship: U.S. Environmental Protection Agency    2012 - 2014

Prete Environmental Education Fellowship: Lyle Center for Regenerative Studies    2012 - 2013

## Teaching and Mentorship

**Lecturer**: School of Information, University of Michigan    2022 - 2023

- Co-instructed two remote, graduate-level data science courses in *Information Visualization* with 70+ students

**Graduate Research Mentor**: School of Information, University of Michigan    2021 - 2023

- Supervising two Information graduate students (a Master's study developing a Natural Language Processing model to detect data citations in articles and a PhD study analyzing a research data citation network)

**Undergraduate Research Mentor**: Center for Spatial Studies, UCSB    2015, 2017, 2019

- Supervised three Computer Science and Geography undergraduate students (a web map usability survey, a research publication visualization, and a spatial data ingest workflow with university librarians)

**Instructor of Record**: Department of Geography, UCSB    2017 - 2018

- Developed course materials and delivered lectures for *Introduction to Geographic Information Systems* (28 students) and *Geography of California* (60 students), receiving excellent student evaluations for both courses

**Workshop Co-Instructor**: iDigBio Georeferencing Workshop, NCEAS (with Dr. Katja Seltmann)    2016

- Developed training materials and sample data for a QGIS training workshop supporting collections managers

**Discovery Seminar Co-Instructor**: Department of Computer Science UCSB (with Dr. Chandra Krintz)    2015

- Introduced computer science students to GIS, designed learning activities, and facilitated guest lectures

**Teaching Assistant**: Department of Geography, UCSB    2014 - 2018

- Led sections for *Introduction to Geographic Information Systems*; *People, Place, and Environment; Analytical and Computer Cartography;* and *Geography of Surfing,* receiving excellent student evaluations for all courses

## Academic Service

**EarthArXiv Advisory Council and Moderation Crew**: Online    2023 - present

- Monthly team review of earth science submissions for compliance with preprint server's moderation policy

**Reviewer**: JASIST, Earth's Future, JGSA, JOSIS, IJDE, Semantic Web Journal, ASIS&T    2018 - present

- Reviewing manuscripts and conference submissions on data science and Semantic Web technology topics

**Institute for Social Research CoderSpaces Co-host**: University of Michigan    2021 - 2022

- Providing weekly help desk consultations for university researchers with computational questions

**Earth Science Information Partnership (ESIP) Community Fellow**: University of Michigan    2020 - 2022

- Organizing teleconference sessions for data discovery and semantic technology community working groups

**Antiracism Working Group Member**: Postdoctoral Association, University of Michigan    2020 - 2021

- Facilitating virtual discussion events and developing educational web resources for postdocs

**Faculty Search Committee Representative**: Environmental Data Science & GIS, UCSB          2017, 2019

- Reviewed candidate materials, supported invited candidate visits, and facilitated graduate student voting

**Meeting Co-organizer**: Spatial Discovery Expert Meetings, Center for Spatial Studies          2015, 2017, 2019

- Co-organized agenda and session topics on spatial search at the intersection of libraries and GIS services

**Graduate Student Association Representative**: Associated Student Government, UCSB          2018 - 2019

- Served as an elected department representative and campus liaison for Geography graduate students

**Workshop Co-organizer**: Core Computations on Spatial Information at GIScience in Melbourne, Australia          2018

- Reviewed invited papers and talks, contributed to agenda, and moderated discussion of a half-day workshop

**Graduate Student Representative**: Associated Students Campus Bicycle Committee, UCSB          2017 - 2018

- Inventoried bicycle infrastructure and deployed an interactive web application for mapping campus projects

**GIS Help Desk Consultant**: Center for Spatial Studies, UCSB          2017

- Provided GIS research services and consultations for students, staff, and faculty from campus research units