

Summer Institute in Computational Social Science

June 18 to July 1, 2017

Sponsored by the Russell Sage Foundation

From the evening of Sunday, June 18 to the morning of Saturday, July 1, 2017, the Russell Sage Foundation will sponsor the first Summer Institute in Computational Social Science, to be held at Princeton University. The purpose of the Summer Institute is to introduce graduate students and beginning faculty in the social and data sciences (broadly conceived) to computational social science – the use of digital-age data sources and methods to conduct social research. The program will highlight issues about access, privacy, and confidentiality that are raised by the emergence of computational data and methods. The institute will also introduce participants to a network of scholars across disciplines with similar interests in these new data and methods. The co-organizers and principal faculty of the Summer Institute are Christopher Bail (Duke University) and Matthew Salganik (Princeton University).

The instructional program will involve lectures, group problem sets, and student-led research projects. There will also be outside speakers with relevant expertise from academia, industry, and government. Topics covered will include text as data, website scraping, digital field experiments, non-probability sampling, mass collaboration, and ethics. There will be ample opportunities for students to discuss their ideas and research with the organizers, other participants, and visiting speakers. Because we are committed to open and reproducible research, all materials created by faculty and students for the Summer Institute will be released open source.

Participation is restricted to Ph.D. students, postdoctoral researchers, and untenured faculty within 7 years of receiving their Ph.D. Most participant costs during the workshop, including housing and most meals, will be covered, and most travel expenses will be reimbursed up to a set cap. We expect that about 30-35 participants will be invited. Participants with less experience with social science research will be expected to complete additional readings in advance of the Institute, and participants with less experience coding will be expected to complete a set of online learning modules on the R programming language. Students doing this preparatory work will be supported by a teaching assistant who will hold online office hours during the two months before the Institute. Accepted applicants must agree to attend all sessions of the two-week program.

There is no application form or program information beyond what is in this announcement. To apply, interested applicants must send (i) a curriculum vitae, (ii) a statement (maximum three pages) describing your interest and any current research in computational social science (especially as it relates to RSF research programs in [Social Inequality](#), [Behavioral Economics](#), [Future of Work](#), and [Race, Ethnicity, and Immigration](#)), (iii) one writing sample (no more than 35 pages), (iv) for graduate students, an unofficial course transcript, and (v) for graduate students and postdoctoral researchers, two letters of recommendation from faculty advisors. Letters of recommendation should be as informative as possible about your standing in your program, general ability, research potential, and (if applicable) special interest in computational social science. All applications must include an e-mail and an alternative means of contact (e.g., phone number). Send your CV, statement, and writing sample (all collapsed into a single pdf file) to rsfcompsocsci@gmail.com. Ask your recommenders to send their letters to the same email address, with the following subject line: **RSF recommendation letter for APPLICANT NAME**.

We anticipate a large pool of highly qualified applicants and plan to make final decisions quickly, so applications and letters must be received by the deadline – no extensions will be granted. **Complete applications, including letters of recommendation, must be received by Friday, February 17.** We will notify applicants solely through e-mail, by Friday, March 10 and will ask participants to confirm their participation very soon thereafter. Inquiries can be sent to rsfcompsocsci@gmail.com.

Russell Sage Foundation
Summer Institute in Computational Social Science, 2017

Location: Princeton University

Organizing Faculty: Christopher Bail and Matthew Salganik

Sunday, June 18

20.30 Opening Dinner

Monday, June 19 *Introduction and Ethics*

- What is Computational Social Science?
- Ethics: Principles-based approach
- Four areas of difficulty: informed consent, informational risk, privacy, and making decisions in the face of uncertainty
- Problem set
- Dinner & discussion

Tuesday, June 20 *Collecting “Big” Data*

- What is big data?
- Strengths and weakness of big data
- Screen-Scraping
- Application Programming Interfaces
- Problem Set
- Dinner & Discussion

Wednesday, June 21 *Text as data*

- Dictionary-Based Approaches
- Cluster Analysis
- Topic Models
- Network-Based Text Analysis
- Problem set
- Dinner & discussion

Thursday, June 22 *Surveys*

- Introduction to total survey error

- Probability and non-probability sampling
- New approaches to measurement
- Linking surveys and big data
- Developing apps for survey research
- Problem set
- Dinner & discussion

Friday, June 22 *Experiments*

- Why experiments
- Moving beyond simple experiments
- Four strategies for experiments
- Problem set
- Dinner & discussion

Saturday, June 23 *Mass Collaboration*

- Why mass collaboration?
- Human computation
- Open call
- Distributed data collection
- Design advice
- Problem set
- Dinner & discussion

Sunday: Go to New York

Monday, June 26-Thursday, June 30

Work on projects

At the end of each day groups will present their progress and receive feedback

Friday, June 30

Present final projects

Make plans for next steps on projects

Faculty Bios

Christopher A. Bail is the Douglas and Ellen Lowey Associate Professor of Sociology and Public Policy at Duke University. He studies how non-profit organizations and other political actors shape public discourse by analyzing large groups of texts from newspapers, television, public opinion surveys, and social media sites such as Facebook and Twitter. His research has been published by Princeton University Press, *Proceedings of the National Academy of Sciences*, *American Sociological Review*, the *American Journal of Public Health*, and other leading journals. His work has been recognized by awards from the American Sociological Association, the Association for Research on Non Profit Organizations and Voluntary Action, the Society for the Scientific Study of Religion, and the Society for Study of Social Problems, and supported by the National Science Foundation and the Robert Wood Johnson Foundation. His research has also been covered by major media outlets such as *NBC News*, *National Public Radio*, and the *Washington Post*. Bail earned his Ph.D. from Harvard University in 2011. Dr. Bail is currently conducting a study of how non-profit organizations reach new audiences on the Internet using a unique sample of groups recruited via Facebook.

Matthew J. Salganik is Professor of Sociology at Princeton University, and he is affiliated with several of Princeton's interdisciplinary research centers: the Office for Population Research, the Center for Information Technology Policy, the Center for Health and Wellbeing, and the Center for Statistics and Machine Learning. His research interests include social networks and computational social science. He is the author of the forthcoming book *Bit by Bit: Social Research in the Digital Age*. Salganik's research has been published in journals such as *Science*, *Proceedings of the National Academy of Sciences*, *Sociological Methodology*, and *Journal of the American Statistical Association*. His papers have won the Outstanding Article Award from the Mathematical Sociology Section of the American Sociological Association and the Outstanding Statistical Application Award from the American Statistical Association. Popular accounts of his work have appeared in the *New York Times*, *Wall Street Journal*, *Economist*, and *New Yorker*. Salganik's research is funded by the National Science Foundation, National Institutes of Health, Joint United Nations Program for HIV/AIDS (UNAIDS), Facebook, and Google. During sabbaticals from Princeton, he has been a Visiting Professor at Cornell Tech and a Senior Research at Microsoft Research.