Social scientists have long relied on parliamentary votes, surveys, in depth interviews, or aggregate election results to understand elite and mass social behavior. But now, there are vast new (or newly accessible) datasets of mass and elite behavior. Voter files, legislative debates, tweets, Facebook posts, Instagram photos, and geolocation records from mobile phones provide immense datasets of human behavior that require different analytical tools as well as new theoretical models of understanding. Scholars have studied and continue to study topics including censorship in China, Supreme Court amicus brief networks, development of party through parliamentary rhetoric in the UK, and racial context on racial attitudes using big data and the associated methodological toolkits. These advances occur when scholars apply substantive theories using advanced methodological toolkits to model and make causal inference from big data. New data sources and increased computing power are merely necessary conditions for scientific advancement in the era of big data.

As Washington University in St. Louis, there are a variety of people developing advanced methodological toolkits to analyze, describe, and draw both prediction and causal inference from big data. Others begin with social science questions about human behavior and utilize big data sources to test theories and make causal inferences. We propose convening a group of scholars in order to develop a proposal for a campus wide master's degree in data science and a program for data science.

Organizers
Roman Garnett, Computer Science
Andrew Reeves, Political Science

Invitees
Sanmay Das, Computer Science
Lee Epstein, Political Science and the Law School
Patrick Fowler, Brown School
Jenine Harris, Brown School
David Law, Law School
Jacob Montgomery, Political Science
Robert Pless, Computer Science
Joshua Swamidass, Medical School
Betsy Sinclair, Political Science