

# Massive Data Institute

Michael Bailey

GEORGETOWN  
UNIVERSITY | McCourt School  
of Public Policy



# Mission

MDI harnesses modern computing power to improve public policy.



## Organizing principles

- Interdisciplinary: MDI brings together social scientists, computer scientists and others from all schools at Georgetown.
- Entrepreneurial: MDI seeks out and supports projects that involve novel uses of data and computing power.
- Partnership-driven: MDI leverages partnerships and external funding.



Cutting-Edge Research

MDI builds research capacity that places Georgetown at the leading edge of data science research.

Cloud Computing

Spatial Data

Text-as-Data

General Research Support

Policy Impacts

MDI uses data to help government and other partners improve public policies.

Partnerships

Elections

Law

Harnessing Data

MDI connects researchers and partners to data and tools that help them unlock insights on important societal problems.

Social Media Data

Census Data

Data Curation

# Cutting-Edge Research

GEORGETOWN  
UNIVERSITY | McCourt School  
of Public Policy



# Computing Resources



## Achievements

Georgetown scholars have used cloud computing to connect new forms of large-scale data to societal scale issues ranging from corruption in public-private partnerships to climate change to understanding civil service to global biodiversity. Some projects have created databases that will become shared repositories for researchers while others involve large-scale simulations.

## Services

We provide access to cloud computing for faculty who need large-scale and fast computing resources. Our services include compute clusters, large-scale databases including Postgres and BigQuery and access to statistical packages for large-scale data and private cloud-based storage.

We will be increasing our services next year to include creation of data portals and data labeling services using Mechanical Turk.

## Contacts

- Lisa Singh (MDI Research Professor)
- Virinche Marwadi (MDI System Administrator)

# Research support



## Achievements

The **MDI Scholars** program provides undergraduate and graduate students from a broad range of disciplines with research experience using large-scale data. The Scholars have participated in over 10 different research projects led by faculty across campus and have been involved in research publications.

Funded by a substantial NSF grant, MDI Research Professor Sing directs a interdisciplinary group of scholars who develop research tools for scholars using social media and other modern data in cross-disciplinary research.

MDI's seed grants have funded Georgetown researchers in Public Policy, Biology, Business, Government, the School of Foreign Service and elsewhere. Scholars receiving these grants have gone on to raise over \$3.5 million in grants.

## Services

MDI can connect researchers to faculty and student data science experts. MDI can also provide support for researchers applying for grants.

## Contacts

- Jonathan Beam (MDI Program Coordinator)
- Tricia Waller (Director, Grants Administration for MDI)

# Advanced Data Science Research



## Achievements

MDI scholars have used text-as-data tools to analyze social media and newspaper data the understand the 2020 US election, online social movements ([#MeToo](#) and [#BlackLivesMatter](#)), [forced migration](#) in Iraq, [Covid misinformation](#) and [gun ownership](#).

MDI also convenes events designed to build the text-as-data and spatial data communities at Georgetown.

## Services

We build and support a community of scholars doing research on text as data. This includes programming for faculty and a Phd-oriented working group. We hope to offer a web-scraping consulting service via our MDI Fellow.

## Contacts

- Jaren Haber (MDI Fellow)
- Michael Bailey (MDI Director)



# Policy Impacts

GEORGETOWN  
UNIVERSITY | McCourt School  
of Public Policy





# Partners

## Achievements

MDI has worked with multiple partners on policy-projects. We have worked with a **government agency** in Washington and have a Fellow at the Lab@DC. Working with interdisciplinary partners via Georgetown's Tech & Society, MDI has sponsored a Fellows with working on racial justice in the law and ethics in data science.

## Services

MDI can work with local governments, public policy organizations and university researchers to create a team that can provide high end data services in a data-driven analysis of a public policy question.

## Contact

- Amy O'Hara (MDI Research Professor)
- Michael Bailey (MDI Director)



# Democracy

## Achievements

MDI has undertaken projects assessing the nature and content of election information. This work is joint with the University of Michigan. This work includes high-profile, current-use [data](#) on election dynamics at [CNN](#).

## Services

MDI is eager to collaborate with scholars interested in using our data or building off our experience analyzing campaigns and elections.

## Contact

- Lisa Singh (MDI Research Professor)
- Michael Bailey (MDI Director)



# Data Access and Policy



## Achievements

MDI has engaged directly with local governments and nonprofit organizations to provide expertise and support on the ethical and effective sharing and reuse of data. Our ADRI and SPE projects have received generous support from the Sloan Foundation to promote the use and linkage of administrative data.

MDI Research Professor Amy O'Hara has also had a high profile in public conversations about [Census](#) and [other government](#) data.

## Services

We have put together [guides of best practices and a directory of intermediaries](#) that can assist in such efforts.

We are developing a data use agreement (DUA) generator through a partnership with the National Head Start Association to assist in the streamlining of creating and approving DUAs.

Our Facebook project helped assess their efforts to provide social scientists access to some of their data.

A Fellow hired under the auspices of Georgetown's Tech & Society program assesses bias and other ethical issues associated with algorithmic analysis of social media data.

## Contact

# Legal system

## Achievements

MDI is engaged in multiple projects related to the U.S. legal system.

- Working with the Georgetown University Law Center, we are creating a [civil justice data commons](#) that will help researchers understand racial and other implications of civil justice in the United States.
- Working with the [Justice Innovation Lab](#), MDI researchers are trying to make the legal process more fair at the prosecutorial level.
- MDI researchers developed innovative measures of the influence of law and politics on the [Supreme Court](#).

## Services

We are eager to join with researchers interested in these and other projects at the intersection of law and data.

## Contacts

- Amy O'Hara (MDI Research Professor)
- Michael Bailey (MDI Director)

# Harnessing Data

GEORGETOWN  
UNIVERSITY | McCourt School  
of Public Policy



# Federal Statistical Research Data Center



## Achievements

The FSRDC is open again! Thanks to the Restarting Research protocols and collaboration between the McCourt School and Census Bureau, our approved researchers can safely and securely access restricted data. The FSRDC has generated interest across area non-profits and institutions. Scholars using our U.S. Census Bureau Research Data Center have produced multiple working papers and journal articles.

## Services

MDI provides access to non-public U.S. Census data at our secure on-sight U.S. Census Research Data Center. We support researchers by helping them learn about what data is available, apply for access and, if approved, to access the data. Executive Director Amy O'Hara serves on the FSRDC Technical Committee and on the FSRDC Budget Committee. She also works with the Census Bureau to streamline data request processes and improve linkage methods.

## Contact

- Amy O'Hara (FSRDC Director)



# Social Media Data Portals



We have developed four portals to date: election 2020, covid-19, and parenting. These portals are controlled by research teams, but allow for construction of variables from social media and newspaper data through easy to use forms. It is particularly valuable for those with limited large-scale data programming skills.

## Services

MDI is developing a service that will give researchers access to variables constructed from the Twitter decahose.

## Contact

- Lisa Singh (MDI Research Professor)

# Data curation



## Achievements

We have developed efficient approaches for storing and processing large-scale data sets. Currently, we have eight projects that are curating different social media and newspaper data sets.

## Services

MDI can provide scholars with access to certain specific data sets. One is a large scale database of more than 700 million publicly available, open-source media articles and blog posts actively compiled since 2006 from a wide range of non-English sources. Another is a very large data set of anonymized credit information.

MDI also makes available **administrative data metadata** obtained through data use agreements with federal and state agencies which show the specific data elements shared with the census in order to improve research designs and increase the specificity of proposals.

We are also exploring whether we can assist in other data access projects.

## Contacts

- Amy O'Hara (FSRDC Director)
- Lisa Singh (MDI Research Professor)
- Michael Bailey (MDI Director)

