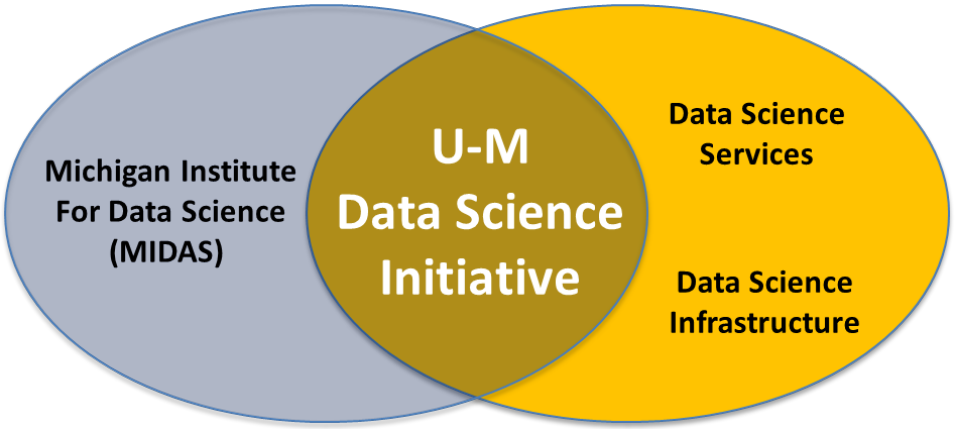
The Michigan Data Science Initiative



Data Science for Social Science Challenge
Town Hall Meeting
March 10, 2016

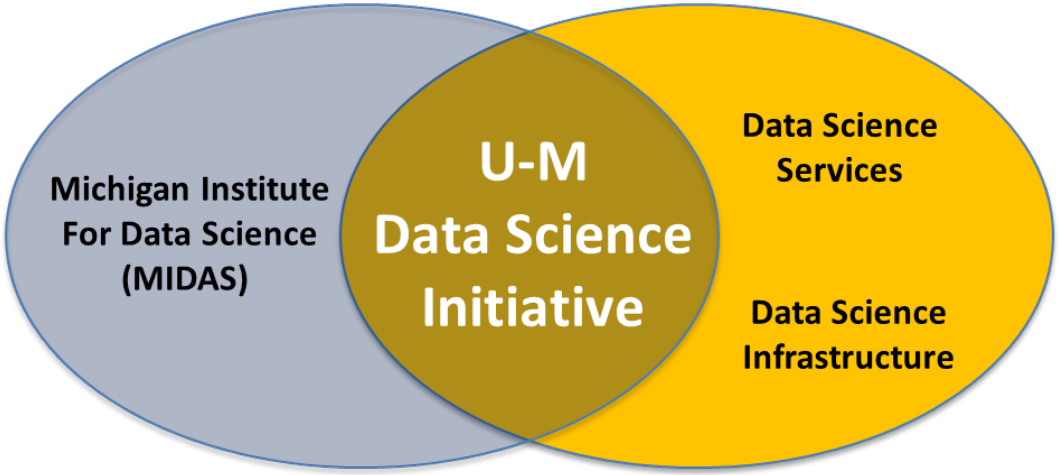
Al Hero and Brian Athey
Co-Directors, MIDAS
Eric Michielssen, AVP ARC

MIDAS MICHIGAN INSTITUTE FOR DATA SCIENCE UNIVERSITY OF MICHIGAN

Outline

- Introduction to the Data Science Initiative and MIDAS
- MIDAS Challenge Initiatives
- RFP and review process
- Data Science Resources
  - CSCAR (Data Science Services)
  - ARC Technology Services (Computing Infrastructure)
- Social Science Example Project Topics
- ICPSR Data Sets

U-M Data Science Initiative (DSI)



UM Collaborating Units

Academic Leadership & Engagement

COE, UMMS, LS&A, SI, SPH, SON, ISR, UMBS, others

Services & Infrastructure

ARC-TS, CSCAR, others

Michigan Institute for Data Science (MIDAS)

- 178 U-M Faculty Affiliates
- Cross-cutting Data Science Methodologies & Analytics
- Data Science Education & Training programs
- Industry Engagement
- 4 Data Science Grand Challenges
- 20-30 Existing U-M Faculty slots
- 10 New U-M Faculty slots

Data Science Services (CSCAR)

Consulting for

- Database Creation, Preparation & Ingestion
- Data Visualization
- Data Access
- Data Analytics

Data Science Infrastructure (ARC-TS)

- Hadoop, SPARK
- SQL, NoSQL databases
- Analytics Platforms
- Integration with HPC Flux Platform

MIDAS MICHIGAN INSTITUTE FOR DATA SCIENCE UNIVERSITY OF MICHIGAN

Michigan Institute for Data Science

http://midas.umich.edu/

- 178 U-M Faculty Affiliates (Ann Arbor, Dearborn, Flint)
- Launching Data Science Education & Training programs
- Involved in growing the Data Science Services component
- Actively involved in industry engagement activities
- Will fund 4 Data Science Grand Challenges in 2015-2016
- Will grow to 30+ core faculty over the next two years
  - 20 slots for existing U-M faculty
  - 10 slots for recruiting external faculty

MIDAS MICHIGAN INSTITUTE FOR DATA SCIENCE UNIVERSITY OF MICHIGAN

Leadership and Core Faculty

Management Committee

- Al Hero, COE-ECE
- Brian Athey, MED-DCMB
- H.V. Jagadish, COE-CSE
- Vijay Nair, LS&A-Statistics
- Ivo Dinov, School of Nursing
- George Alter, ISR
- Satinder Bajeva, COE-CSE
- Anna Gilbert, LS&A-Math
- Margaret Hedstrom, SI
- Tim Mckay, LS&A – Physics
- Eric Michielssen COE-ECE
- Kerby Shedden, LS&A-Statistics
- Jeremy Taylor, SPH-Biostatistics
- Kevin Ward, MED
- Jiepeng Ye, MED-DCMB

Core Faculty (20 slots)

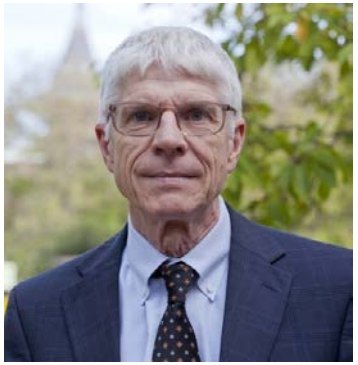
- Anna Gilbert, LS&A-Math
- Raj Rao Nadakuditi, COE-ECE
- Dragomir Radev, SI
- Jeremy Taylor, SPH-Biostatistics
- Pascal Van Hentenryck, COE-IOE
- Ji Zhu, LS&A-Statistics

Core Faculty Recruiting (12 slots)

- 109 Applications Received
- 31 Applications Recommended
- 19 Candidates Interviewing
- 1 Offer (in process)

MIDAS MICHIGAN INSTITUTE FOR DATA SCIENCE UNIVERSITY OF MICHIGAN

MIDAS Industry Engagement Program



Henry Kelly - MIDAS Senior Scientist & Industry Partnership Leader

- Recently retired from federal service (DOE, OSTP)
- Manages NSF Midwest BDHub activities and supports building novel industry partnerships
- **Available to meet with MIDAS Faculty**

Company Exploratory Meetings (16+ completed)

- Agilent, AT&T, Barracuda Networks, Booz Allen Hamilton, Delta Dental, Ford, Goldman Sachs, JPMC, Konica-Minolta, Magna International, Microsoft, MTC Leadership Circle, Naval Research Labs, Northrup Grumman, Proquest, Sandia National Labs, Taubman Institute SAB

Developing Industry Engagement Model

- Working with Business Engagement Center (BEC)
- Exploring partnership with CoE Multidisciplinary Design Program

Data Science Services and IT Infrastructure

Consulting for Statistics Computing & Analytics Research (CSCAR)

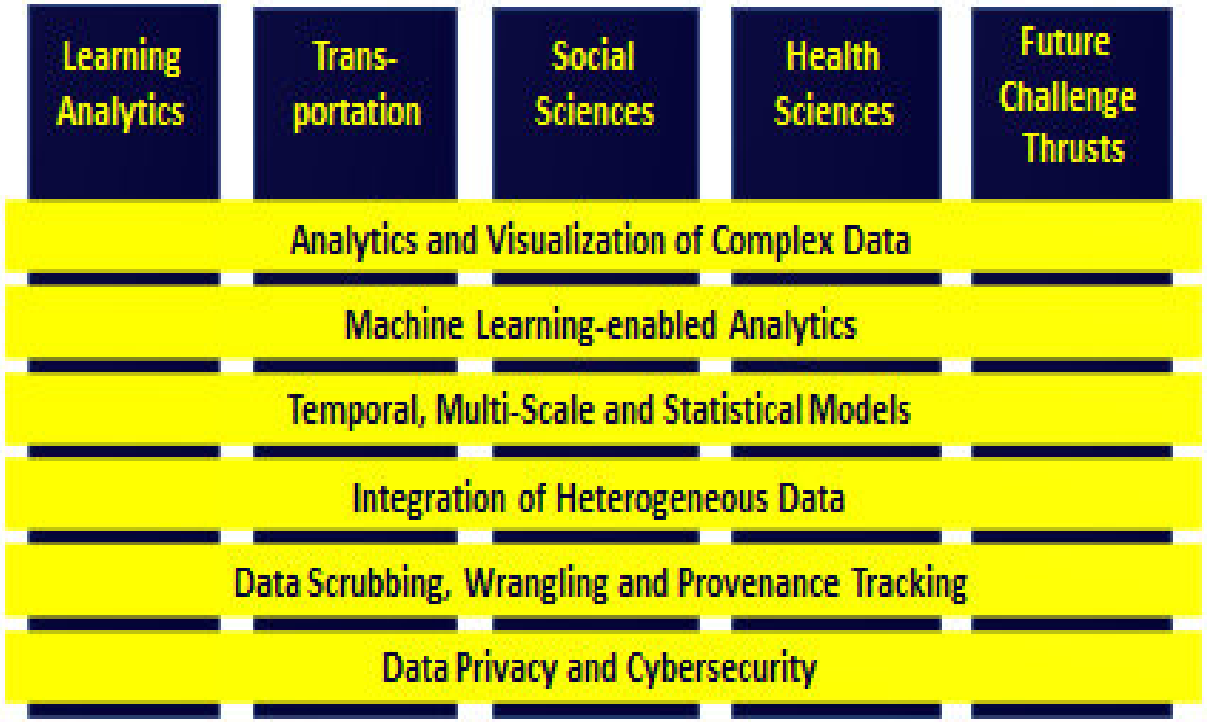
Consulting for

- Database Creation, Preparation & Ingestion
- Data Visualization
- Data Access
- Data Analytics
- Advanced Geographic Information Systems (GIS+)

Advanced Research Computing - Technology Services (ARC-TS)

- Hadoop, SPARK
- SQL, NoSQL databases
- Analytics Platforms
- Integration with the Flux HPC Platform

MIDAS Challenge Initiatives Program



Leveraging Data Science Services & Infrastructure

MIDAS plans to fund a total of 8 proposals

- Evenly split over the 4 challenge thrusts
- Multi-disciplinary teams
- Funded at approximately $1.25M over 3 years
- 50% cost sharing between UMOR and units

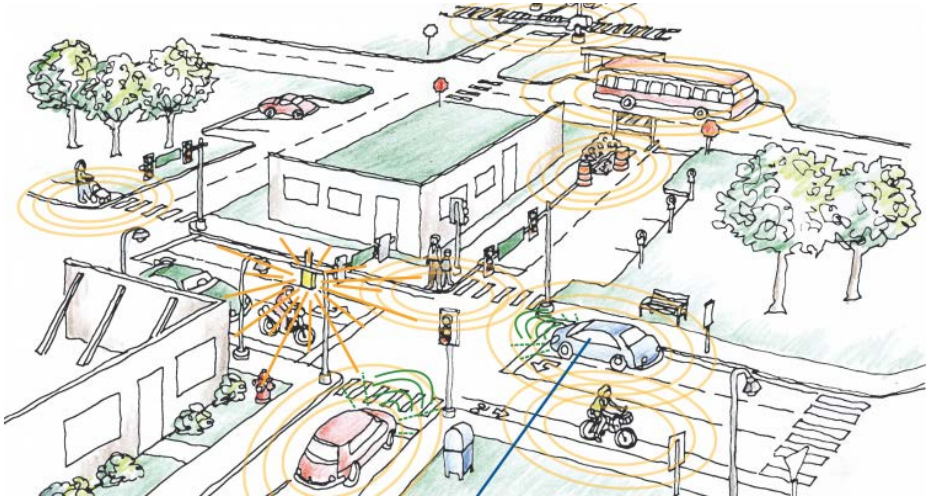
MIDAS MICHIGAN INSTITUTE FOR DATA SCIENCE UNIVERSITY OF MICHIGAN

Methodological Approach for Successful Proposal

A successful proposal must develop methodological approaches and might apply to one or more of the following.

1. Massive integration and harmonization of complex heterogeneous data.
2. Scalable active learning and causal inference.
3. Human-in-the-loop learning and analytics.
4. Adaptive anomaly detection.
5. Dimensionality reduction for visualization of complex data.
6. Embedded systems for data mining and statistical inference.
7. Distributed, cloud-enabled and interoperable algorithms.
8. Bayesian approaches for big data.

MIDAS Transportation Challenge



Mcity: A 32-Acre Outdoor Lab

Automotive cybersecurity for connected vehicles

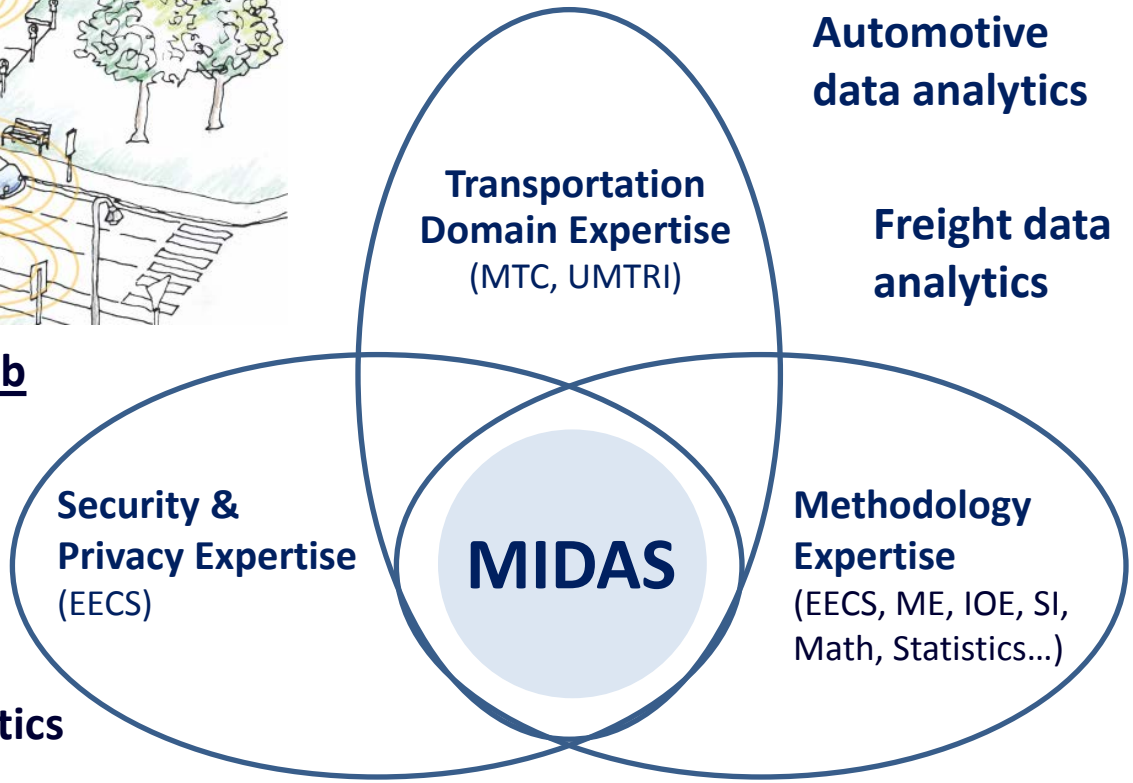
Accident and safety data analytics

Data-analysis for mass transit

Transportation data ecosystems for connected vehicles

Automotive data analytics

Freight data analytics



MIDAS MICHIGAN INSTITUTE FOR DATA SCIENCE UNIVERSITY OF MICHIGAN

MIDAS Learning Analytics Challenge

UM: Education at Scale

Multimodal capture of learning behavior

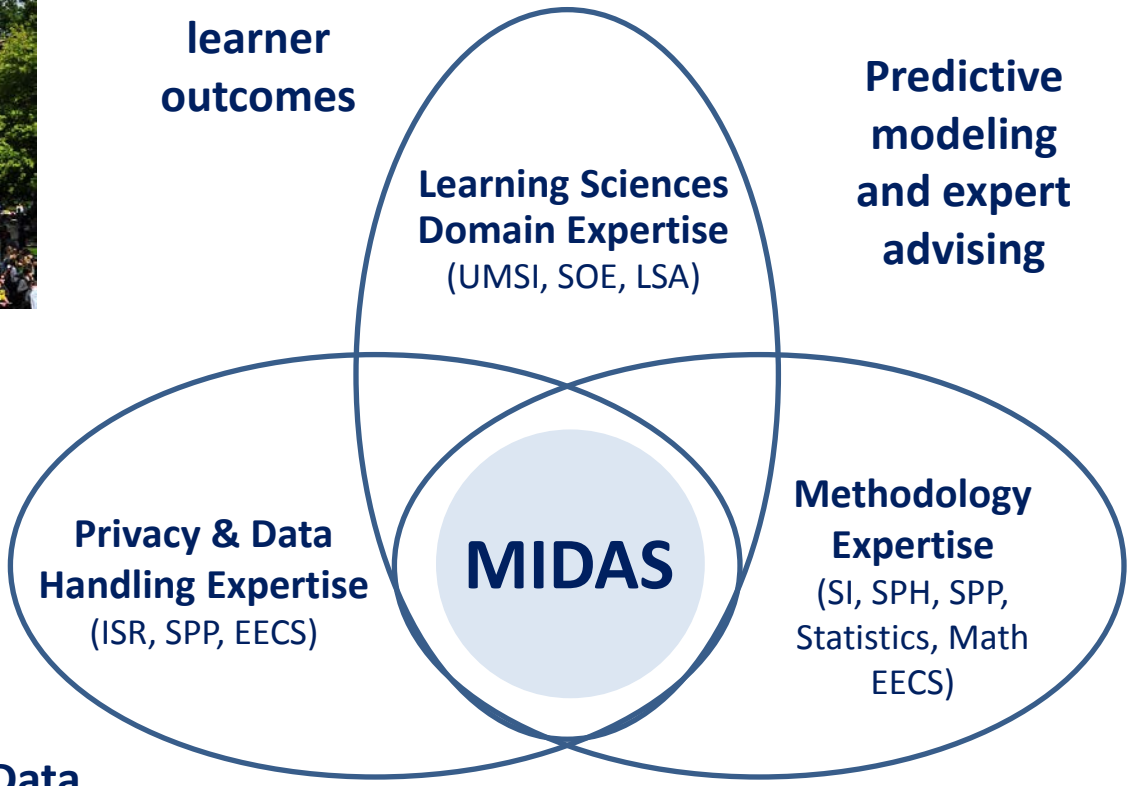
Social network characterization and intervention

Development of Big Data enabled teachers and learners

Multimodal assessment of learner outcomes

Personalized education at scale

Predictive modeling and expert advising



MIDAS MICHIGAN INSTITUTE FOR DATA SCIENCE UNIVERSITY OF MICHIGAN

MIDAS Health Science Challenge

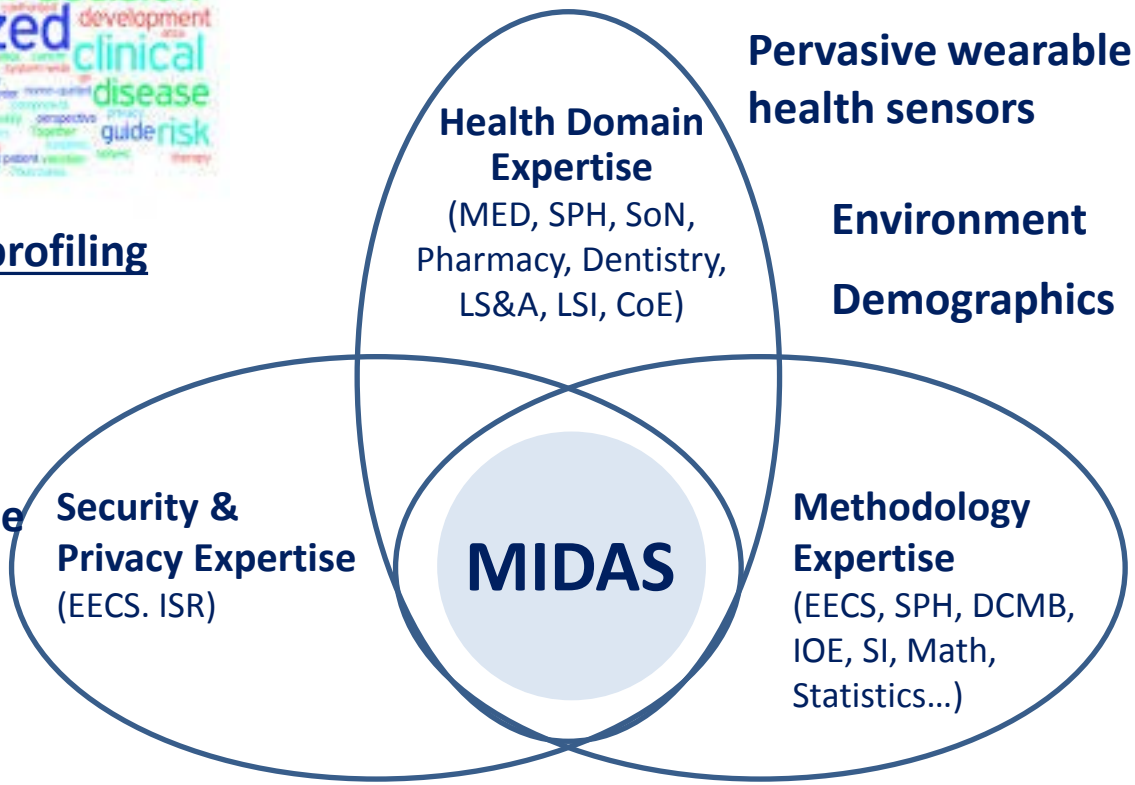
Integrated personal omics profiling

Bio-behavioral Outcomes

Pervasive wearable health sensors

Environment

Demographics



Predictive analytics for personalized health and medicine

Cancer, Obesity, Diabetes, Alzheimer's Disease, …

Data de-identification and privacy

MIDAS MICHIGAN INSTITUTE FOR DATA SCIENCE UNIVERSITY OF MICHIGAN

MIDAS Social Science Challenge

Institute for Social Research

Media-driven socio-economic prediction

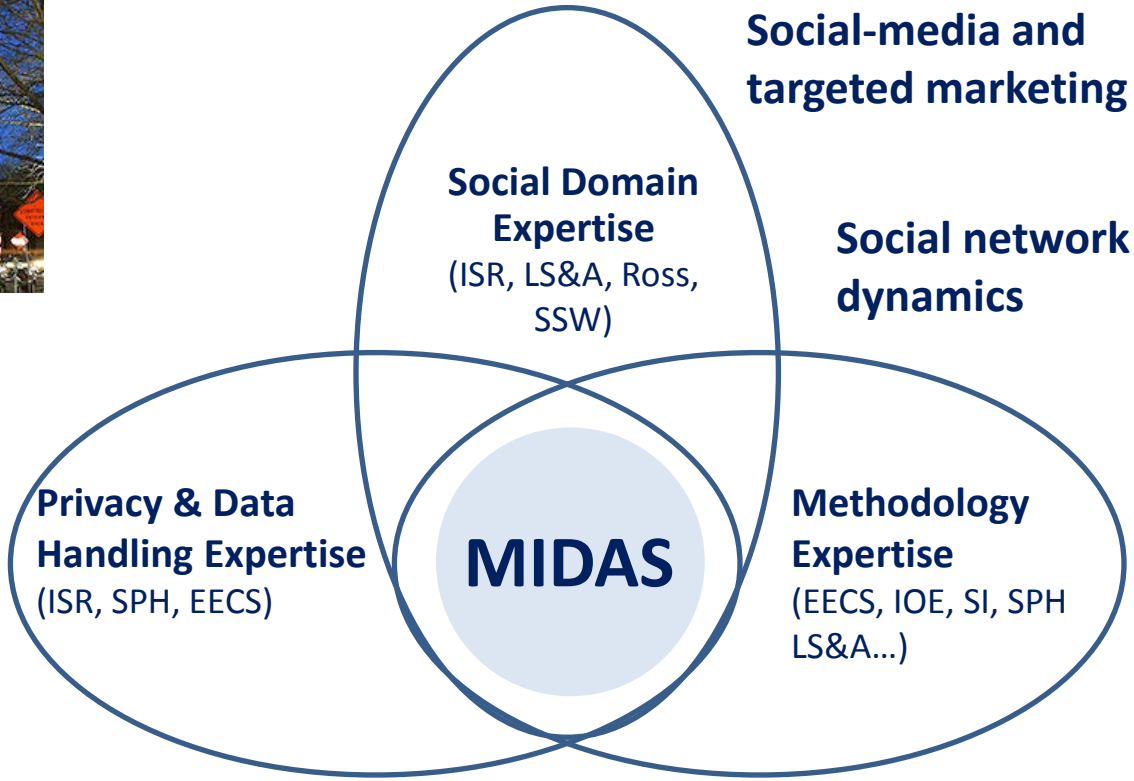
Data aggregation: postings, social, economic, demographic

Social-media survey analytics

Information security and privacy

Social-media and targeted marketing

Social network dynamics



MIDAS MICHIGAN INSTITUTE FOR DATA SCIENCE UNIVERSITY OF MICHIGAN

Example Social Topics for Successful Proposal

Example topics for a successful Data Science for Social Science Challenge proposal might include:

- Media-driven socio-economic prediction
- Integration of numeric, geo-spatial, sensor, social media and administrative data
- Cross-validation of survey data, social media, and passive data collections
- Privacy, disclosure risks, and informed consent
- Data visualization for business analytics and targeted marketing

Example Social Topics for Successful Proposal

Example topics for a successful Data Science for Social Science Challenge proposal might include:

- Social network dynamics
- Analysis of space-time trajectories and ecological momentary assessment
- Extracting meaning from large collections of text, video, audio, and other digital objects
- Data and methods for integrative assessment and prediction
- Novel techniques for data collection

Challenge RFPs - White Paper Requirements

No longer than **5 pages** (excluding budget and bios)

- P1. Title page with proposed project title, DSI Thrust designation, project abstract, names of co-PI's and contact information for the lead PI.
- P2-P5. Technical description. Problem to be addressed and technical approach to solve problem. Nature of data to be collected/analyzed/managed. Methodology to be applied and analytical tools to be used or developed. Data Science Services and computational infrastructure to be used. Description and justification of team, including partners from industry or other institutions (cannot be part of budget). Expected impact of research resulting from the project.
- Draft budget of approximately$1.25M total over three years broken down yearly.
- One page bios of each co-PI.

MIDAS MICHIGAN INSTITUTE FOR DATA SCIENCE UNIVERSITY OF MICHIGAN

Staging of Challenge RFPs

Timeline	Challenge Thrust
Fall 2015	Transportation, Learning Analytics
Winter 2016	Personalized Medicine and Health, Social Sciences
Fall 2016	Transportation, Learning Analytics
Winter 2017	Personalized Medicine and Health, Social Sciences

Challenge Awards for
Data Science in Transportation & Data Science in Learning Analytics
will be announced on April 22

MIDAS MICHIGAN INSTITUTE FOR DATA SCIENCE UNIVERSITY OF MICHIGAN

Health & Social Sciences Challenge Timeline

Date	Challenge Thrust
February 16	RFPs disseminated
March 9	Health Sciences Town Hall Information Session
March 10	Social Sciences Town Hall Information Session
March 30	Health Sciences Town Hall Information Session
May 2	Social Sciences Town Hall Information Session
June 30	White papers due with 2+ week down selection
July 22	Full proposal solicitations communicated
October 17	Full proposals due
November 18	Awards announced

http://midas.umich.edu/rfp/

MIDAS MICHIGAN INSTITUTE FOR DATA SCIENCE UNIVERSITY OF MICHIGAN

Challenge RFPs - White Paper Requirements

Real-time Monitoring and Data Visualization of Infectious Disease Outbreaks

PI/Co-PI Name	Department	School/College	Budget Year 1	Budget Year 2	Budget Year 3	Total Funding
Principal Investigator	Information	Information	$ 70,000	$ 80,000	$ 97,000	$ 247,000
Co-Principal Investigator #1	EECS - CSE	CoE	$ 45,000	$ 55,000	$ 50,000	$ 150,000
Co-Principal Investigator #2	Inf Diseases	Medicine	$ 85,000	$ 97,000	$ 82,000	$ 264,000
Co-Principal Investigator #3	DCMB	Medicine	$ 45,000	$ 55,000	$ 50,000	$ 150,000
Co-Principal Investigator #4	Mathematics	LS&A	$ 73,000	$ 77,000	$ 120,000	$ 270,000
Co-Principal Investigator #5	Biostatistics	Public Health	$ 45,000	$ 55,000	$ 69,000	$ 169,000
		TOTAL	**$ 363,000**	**$ 419,000**	**$ 468,000**	**$ 1,250,000**

Schools/Colleges	
CoE	**150,000**
EECS - CSE	150,000
Information	**247,000**
LS&A	**270,000**
Mathematics	270,000
Medical School	**414,000**
Inf Diseases	264,000
DCMB	150,000
Public Health	**169,000**
Biostatistics	169,000

In addition to a detailed budget, broken down yearly and including cumulative totals, a budget summary that shows the distribution of the budget by faculty member is required.

This information will be used to determine unit (school/college) cost-share.

MIDAS MICHIGAN INSTITUTE FOR DATA SCIENCE UNIVERSITY OF MICHIGAN

Challenge RFPs - White Paper Requirements

The Associate Deans for Research (ADR) of all colleges or schools in which the co-PIs and senior investigators hold their primary appointments should be sent a copy of the white paper.

Challenge RFPs - Full Proposal Requirements

No longer than **10 pages** (excluding title page, budget, bios, letters)

- P1-P10. **Sec. 1. Technical description**. Sec. 1.2 Problem to be addressed and challenges faced. Sec 1.3 Nature of data to be collected/managed/analyzed. Sec. 1.3 Technical approach proposed to solve problem, including methodology to be applied and analytical tools to be used or developed. Sec. 1.4 Expected impact on technology, science and society. **Sec 2. Resources**. Sec. 2.1 Databases or data collections, including IRB and HIPPA issues if applicable. Sec 2.2 Computational and data services and infrastructure resources to be used, including UM flux or cloud resources. **Sec 3. Data management and dissemination plan**. **Sec. 4 Description and justification of team**, including partners from industry or other institutions (cannot be part of budget).
- A draft budget (up to $1.25M for three years), broken down yearly and showing 50% cost sharing.
- One page bios of each co-PI.
- Letters from ADRs confirming 50% cost sharing of Ann Arbor component

Challenge RFPs - Review Process and Criteria

- Evaluation will be done by a panel of experts.
- The panel will review each proposal according to the following criteria:
    1. relevance to the stated thrust area(s);
    2. likelihood of the project to result in innovative creation and/or application of data science methodology for the stated thrust area(s);
    3. complementarity to existing projects at UM;
    4. multi-disciplinary coherence of team;
    5. likelihood that proposed work will lead to competitive major extramural grant proposals within 3 years.
    6. substantial involvement of students
- The decision to solicit a full proposal from a white paper or to fund a full proposal will be made by the MIDAS co-Directors.

Challenge RFPs - Post-selection Expectations

- All co-PIs are expected to become active affiliate members of MIDAS.
- All teams will be expected to:
  1. submit yearly reports on progress towards the aims of their grant;
  2. participate in a yearly review, organized as a workshop for all co-PI's on all projects funded by the DSI intramural funding program;
  3. maintain an active project website;
  4. actively work with MIDAS to enhance data science at UM, e.g., through hosting DS student interns, sharing resources like software, and participating in targeted industry outreach.

Data Science Services

CSCAR CONSULTING FOR STATISTICS, COMPUTING & ANALYTICS RESEARCH UNIVERSITY OF MICHIGAN

- Free consulting for U-M researchers in all aspects of data management and analysis.
- Support for programming, software and advanced computing infrastructure: Python, R, Matlab, and many more packages on desktops and on FLUX.
- Experienced consultants with diverse research backgrounds.
- Send your students to talk to us - we will help them be more productive!
- Data Science Skills Series: Wednesdays 3:30-5:00 pm.
- Call 764-7828 to schedule an appointment, or drop in to talk to the GSRAs.
- We are hiring 6 new consultants to build capacity and expand our scope.
- Add a CSCAR consultant to your team, recharge by percent effort. Limited time -- no recharge arrangements available (thanks to MIDAS/DSI).

cscar.research.umich.edu

MIDAS MICHIGAN INSTITUTE FOR DATA SCIENCE UNIVERSITY OF MICHIGAN

Data Science Computing Infrastructure

Advanced Research Computing – Technology Services

- Infrastructure for Sensitive Data

Examples of Social Science Projects & ICPSR Data Sets

Questions, Discussion, and Mixer