The Michigan Data Science Initiative

Challenge thrust information and teaming meetings
October 21 and 22, 2015

Al Hero and Brian Athey
Co-Directors, MIDAS
Eric Michielssen, AVP ARC
✓ Introduction to Data Science Initiative and MIDAS
✓ MIDAS Challenge Initiatives
✓ RFP and review process
✓ Spotlight talks
**U-M Data Science Initiative (DSI)**

**Michigan Institute for Data Science (MIDAS)**
- 150+ U-M Faculty Affiliates (2015)
- 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

**U-M Data Science Initiative**

**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

**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

http://midas.umich.edu/

• Currently have 150+ U-M Faculty Affiliates (FALL 2015)
• 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
Data Science Services and IT Infrastructure

Data Science Services
(CSCAR—Center for Statistical Consulting and Research)

*Consulting for*

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

Data Science Infrastructure

(ARC-TS)

- Hadoop, SPARK
- SQL, NoSQL databases
- Analytics Platforms
- Integration with the Flux HPC Platform
U-M Data Science Challenge Initiatives

- Learning Analytics
- Transportation
- Social Sciences
- Health Sciences
- Future Challenge Thrusts

- Analytics and Visualization of Complex Data
- Machine Learning-enabled Analytics
- Temporal, Multi-Scale and Statistical Models
- Integration of Heterogeneous Data
- Data Scrubbing, Wrangling and Provenance Tracking
- Data Privacy and Cybersecurity

Leveraging Data Science Services & Infrastructure
U-M Data Science Challenge Thrusts: Crosscutting methodologies

Analytics and Visualization of Complex Data — networked single-user and collaborative visualization of massive multi-modality datasets.

Machine Learning-enabled Analytics — Machine learning methods such as anomaly detection, dictionary learning, reinforcement learning, similarity learning, and transfer learning must be scalable to massive data scales.

Temporal, Multi-Scale and Statistical Models — Mathematical, computational and statistical models are needed to integrate multimodal data collected at many different time and length scales.

Integration of Heterogeneous Data — Integration of numerical data, symbolic data, structured data, and streaming data at various stages of the analysis pipeline.

Data Scrubbing, Wrangling and Provenance Tracking — Automation of data preparation steps such as normalization, calibration, outlier treatment, and annotation.

Data Privacy and Cybersecurity — The tradeoffs between data privacy/security and data utility must be understood in the context of the specific application, e.g., medicine, transportation, or business analytics, throughout the data storage, management, and analysis pipeline.
MIDAS Transportation Challenge

Mcity: A 32-Acre Outdoor Lab

Transportation data ecosystems for connected vehicles

Automotive data analytics

Transportation Domain Expertise (MTC, UMTRI)

Freight data analytics

Methodology Expertise (EECS, ME, IOE, SI, Math, Statistics...)

Security & Privacy Expertise (EECS)

MIDAS

Automotive cybersecurity for connected vehicles

Accident and safety data analytics

Data-analysis for mass transit
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
Privacy & Data Handling Expertise (ISR, SPP, EECS)

Learning Sciences Domain Expertise (UMSI, SOE, LSA)

Personalized education at scale
Predictive modeling and expert advising

Methodology Expertise (SI, SPH, SPP, Statistics, Math EECS)

MIDAS
MIDAS Health Challenge

Health Domain Expertise
(MED, SPH, SoN, Pharmacy, Dentistry, LS&A, LSI, CoE)

Methodology Expertise
(EECS, SPH, DCMB, IOE, SI, Math, Statistics...)

Security & Privacy Expertise
(EECS. ISR)

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 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

Social Domain Expertise
(ISR, LS&A, Ross, SSW)

Privacy & Data Handling Expertise
(ISR, SPH, EECS)

Methodology Expertise
(EECS, IOE, SI, SPH LS&A...)

MIDAS
### Staging of challenge RFPs

<table>
<thead>
<tr>
<th>Timeline</th>
<th>Challenge Thrust</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall 2015</td>
<td>Transportation, Learning Analytics</td>
</tr>
<tr>
<td>Winter 2016</td>
<td>Personalized Medicine and Health, Social Sciences</td>
</tr>
<tr>
<td>Fall 2016</td>
<td>Transportation, Learning Analytics</td>
</tr>
<tr>
<td>Winter 2017</td>
<td>Personalized Medicine and Health, Social Sciences</td>
</tr>
</tbody>
</table>

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
## Fall 2015 RFP Timeline

<table>
<thead>
<tr>
<th>Date</th>
<th>Challenge Thrust</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oct 6 2015</td>
<td>RFPs disseminated</td>
</tr>
<tr>
<td>Oct 6 – Nov 30</td>
<td>Hold townhall information sessions</td>
</tr>
<tr>
<td>Nov 30</td>
<td>White papers due with 2 week down selection</td>
</tr>
<tr>
<td>Jan 18 2016</td>
<td>Full proposal due</td>
</tr>
<tr>
<td>Feb 15</td>
<td>Awards announced</td>
</tr>
</tbody>
</table>
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.
The Associate Deans for Research (ADR) of all colleges or schools in which the coPIs 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.
## Potential Industry Partners for Transportation

<table>
<thead>
<tr>
<th>Partner</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delphi Automotive PLC</td>
<td></td>
</tr>
<tr>
<td>DENSO Corporation</td>
<td></td>
</tr>
<tr>
<td>Econolite Group, Inc</td>
<td></td>
</tr>
<tr>
<td>Ford Motor Company (*)</td>
<td></td>
</tr>
<tr>
<td>General Motors Company (*)</td>
<td></td>
</tr>
<tr>
<td>Honda Motor Co., Ltd</td>
<td></td>
</tr>
<tr>
<td>Iteris, Inc.</td>
<td></td>
</tr>
<tr>
<td>Navistar, Inc.</td>
<td></td>
</tr>
<tr>
<td>Nissan Motor Co., Ltd</td>
<td></td>
</tr>
<tr>
<td>Qualcomm Technologies, Inc.</td>
<td></td>
</tr>
<tr>
<td>Robert Bosch LLC</td>
<td></td>
</tr>
<tr>
<td>State Farm Mutual Automobile Insurance Company</td>
<td></td>
</tr>
<tr>
<td>TRW automotive</td>
<td></td>
</tr>
<tr>
<td>IBM (*)</td>
<td></td>
</tr>
<tr>
<td>Google (**)</td>
<td></td>
</tr>
<tr>
<td>UPS (*)</td>
<td></td>
</tr>
<tr>
<td>Toyota (**)</td>
<td></td>
</tr>
<tr>
<td>Dymler (**)</td>
<td></td>
</tr>
<tr>
<td>ZipCar Inc.</td>
<td></td>
</tr>
<tr>
<td>Microsoft (α)</td>
<td></td>
</tr>
<tr>
<td>Xerox (α)</td>
<td></td>
</tr>
<tr>
<td>Cisco (α)</td>
<td></td>
</tr>
<tr>
<td>Siemens (α)</td>
<td></td>
</tr>
</tbody>
</table>

Mcity partners, (*) registered for kickoff, (**) aspirational, (α)=demonstrated interest in urban data
## Potential Industry Partners for Education

### Interested in Hiring Data Specialists
- Dow (*)
- Ford (*)
- GM (*)
- IBM (*)
- Microsoft (*)
- Northrop Grumman (*)
- Oracle (*)
- Rockwell Automation (*)
- Samsung (*)
- Stanley Black and Decker (*)
- Steelcase (*)
- UPS (*)
- Verizon (*)

### Interested in Design and Supply of Learning Tools
- IBM (*)
- Pearson (Kaplan) (*)
- Knewton (*)
- Microsoft (*)
- Red Hat (*)
- Oracle (*)
- Coursera (**) 
- Kahn Academy (**) 

(*) participated in kickoff, (**) aspirational
Virtual Teaming

http://myumi.ch/aMneW
Spotlight talks

• Several short presentations from faculty on opportunities