Ann Arbor is Already Recognized as a "Big Data Powerhouse"

Ann Arbor is one of top 5 US Cities for Big Data[1]

Ranking criteria:

- Data openness of city across 19 types of data
- Economic conditions for innovation in data science
  - *Volume*: Number companies specializing in data processing, hosting and related services, per 100,000 residents.
  - *Velocity*: Internet download speeds (megabits per second).
  - *Variety*: % of the workforce employed in computer and mathematical occupations.ations.

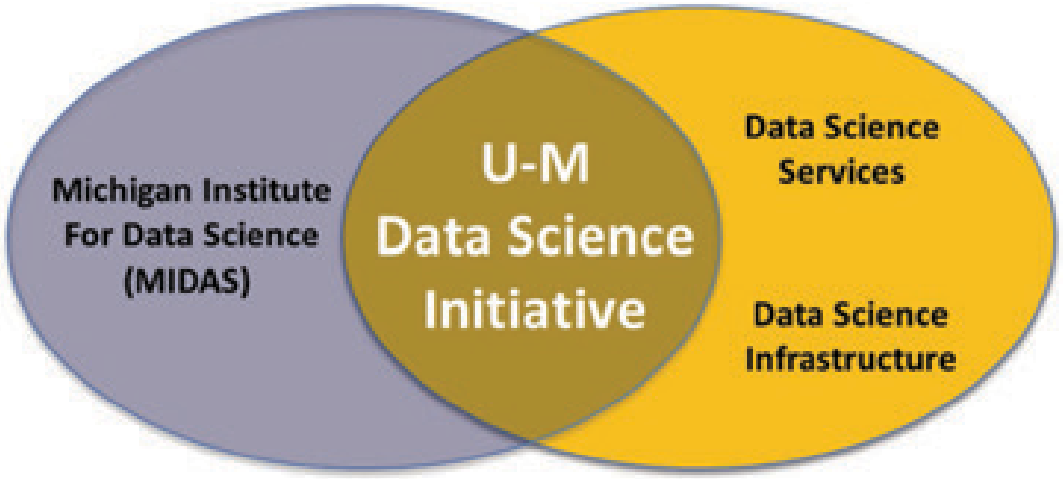
Ann Arbor is *the* most desirable small US city for Millenials[2]

- *Value*: Better quality of life and more satisfying careers for future workforce.

1. smartasset.com, Aug 16 2015
2. American Institute of Economic Research, May 15 2015

MIDAS MICHIGAN INSTITUTE FOR DATA SCIENCE UNIVERSITY OF MICHIGAN

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)

- 100+ U-M Faculty Affiliates (2015)
- 20-30 Existing U-M Faculty slots
- 10 New U-M Faculty slots
- 4 Data Science Grand Challenges
- Data Science Methodologies & Analytics
- Data Science Education & Training programs
- Industry Engagement

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

MIDAS is Positioning the U-M as a National Leader in Data Science

- **Research** - Will build from the existing strong Data Science research activities at U-M, bringing additional visibility and impact
- **Education** - Will result in training a new category of scientists and professionals to satisfy a growing market, by enhancing existing curricula and adding new content
- **Services** - Will place U-M in the forefront that will benefit the campus, the state and the nation
- **Industry Engagement** – Will engage local, regional and national partnerships with Big Data-oriented Companies

MIDAS MICHIGAN INSTITUTE FOR DATA SCIENCE UNIVERSITY OF MICHIGAN

Michigan Institute for Data Science (MIDAS)

http://midas.umich.edu/

- Currently have 100+ U-M Faculty Affiliates (Fall 2015)
- Launching Data Science Education & Training programs
- Launching Data Science Services (with CSCAR)
- Launching 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

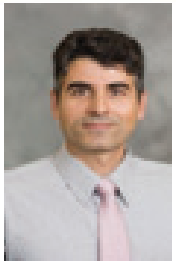
MIDAS MICHIGAN INSTITUTE FOR DATA SCIENCE UNIVERSITY OF MICHIGAN

Selected MIDAS Faculty Leaders

H. V. Jagadish, PhD, is the Bernard A. Galler Collegiate Professor of Electrical Engineering and Computer Science.

Vijay Nair, PhD, is the Donald A. Darling Professor of Statistics and Professor of Industrial and Operations Engineering.

MIDAS Distinguished Scientists



Ivo D. Dinov, PhD, is an Associate Professor of Nursing; Faculty Affiliate, Center for Computational Medicine and Bioinformatics

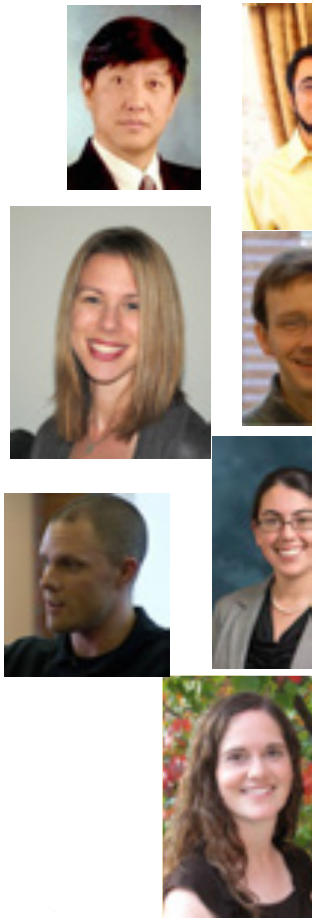
Associate Director, MIDAS Education & Training

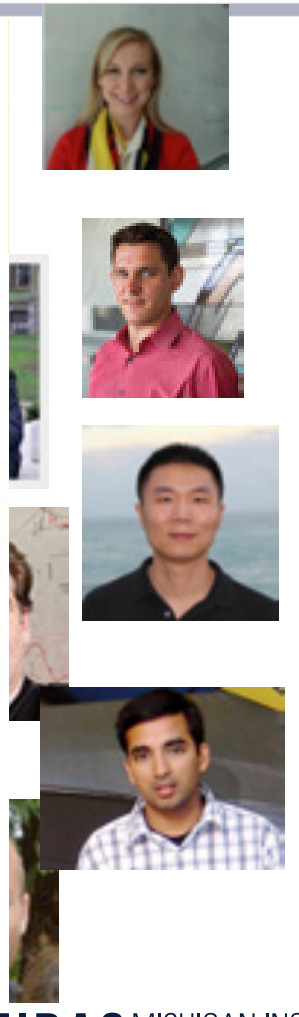
MIDAS Management Committee

George Alter, Institute for Social Research
Brian Athey, Medical School
Satinder Baveja, College of Engineering
Ivo Dinov, School of Nursing, Medical School
Anna Gilbert, Literature, Science, and the Arts
Margaret Hedstrom, School of Information
Al Hero III, College of Engineering
HV "Jag" Jagadish, College of Engineering
Timothy McKay, Literature, Science, and the Arts
Eric Michielssen, College of Engineering
Vijayan Nair, Literature, Science, and the Arts
Kerby Shedden, Center for Statistical Consulting and Research
Kevin Smith, Medical School
Peter Sweatman, Mobility Transformation Center
Jeremy Taylor, School of Public Health
Jieping Ye, Medical School

MICHIGAN INSTITUTE FOR DATA SCIENCE
UNIVERSITY OF MICHIGAN

Current MIDAS Jr. Faculty Members and the Future of Data Science





MIDAS MICHIGAN INSTITUTE FOR DATA SCIENCE UNIVERSITY OF MICHIGAN

Data Science Training and Education

Focus session on MIDAS Education
- Ivo Dinov (MIDAS Assoc. Dir)
- Erin Shellman
- Patrick Harrington
- Nandit Soparkar

Data Scie

New U-M
- Rackham Data Science Certificate
- Undergraduate Major in Data Science

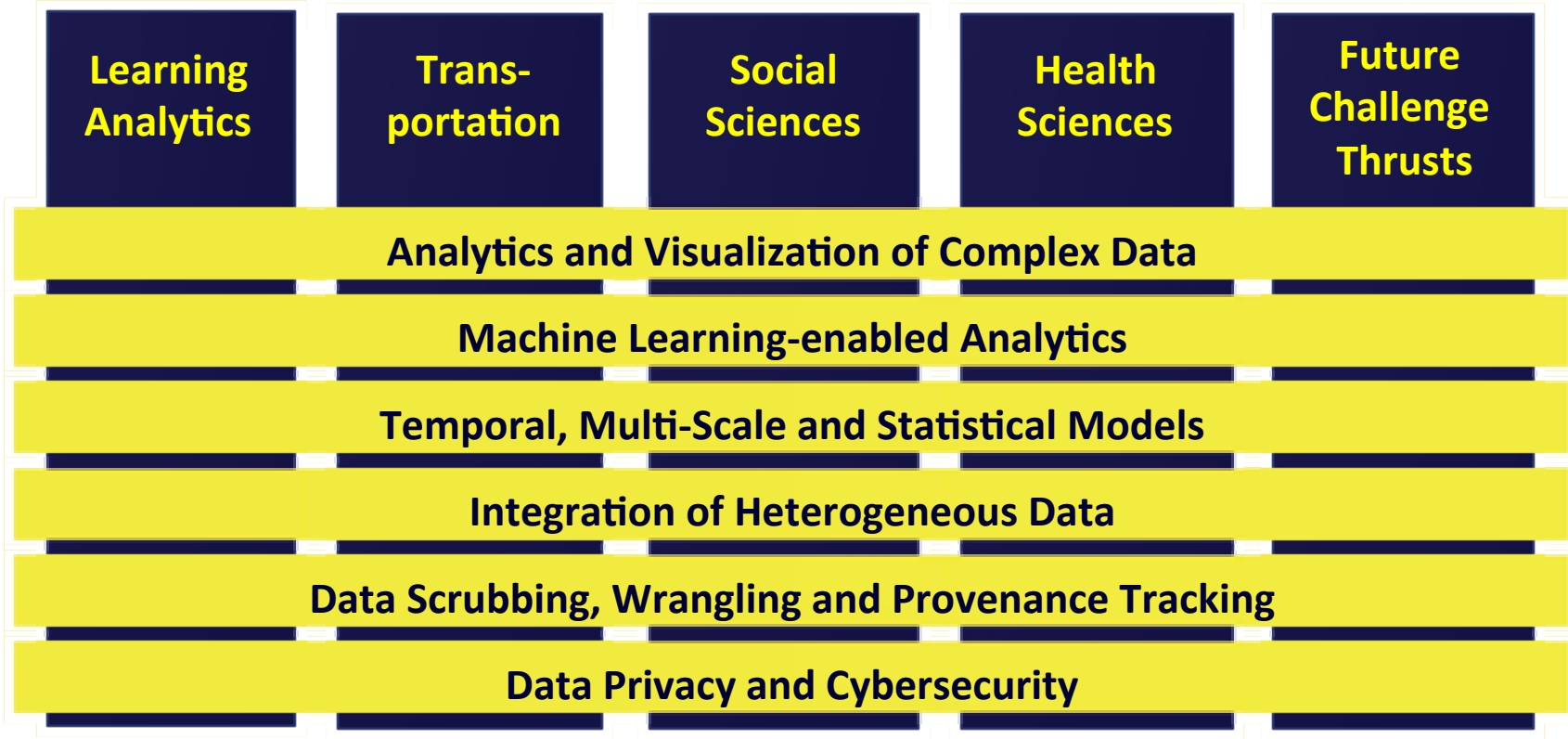
Other activities
- Data Science Bootcamps
- Summer Schools, Online Training

Understand principles of:
data collection, access, security, preparation,
analysis

of
lved –
cs,
ural
nd life
science, information science, social and political science; business, economics and law

MIDAS MICHIGAN INSTITUTE FOR DATA SCIENCE UNIVERSITY OF MICHIGAN

U-M Data Science Challenge Thrusts



Leveraging Data Science Services & Infrastructure

MIDAS MICHIGAN INSTITUTE FOR DATA SCIENCE UNIVERSITY OF MICHIGAN

U-M Data Science Challenge Thrusts: Crosscutting methodologies

Analytics and Visualization of Complex Data —networked single-user and collaborative visualization of massive multi- Carley 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 Nowak, Murphy, McKeown

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

Integration of Heterogeneous Data —Integration of numerical data, symbolic data, structured data, and streaming data at various stages of the analysis p Carley, Murphy

Data Scrubbing, Wrangling and Provenance Tracking – Automation of data preparation steps such as normalization, calibration, outlier treatmer Soparkar 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., medicin Goroff portation, or business analytics, throughout the data storage, management, and analysis pipeline.

MIDAS MICHIGAN INSTITUTE FOR DATA SCIENCE UNIVERSITY OF MICHIGAN

MIDAS Health Challenge

Integrated personal omics profiling

Bio-behavioral Outcomes

Murphy

Pervasive wearable health sensors

Environment

Demographics

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

Predictive analytics for personalized health and medicine

Poste

Security & Privacy Expertise
(EECS. ISR)

MIDAS

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

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

Data de-identification and privacy

Goroff

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

Saxberg

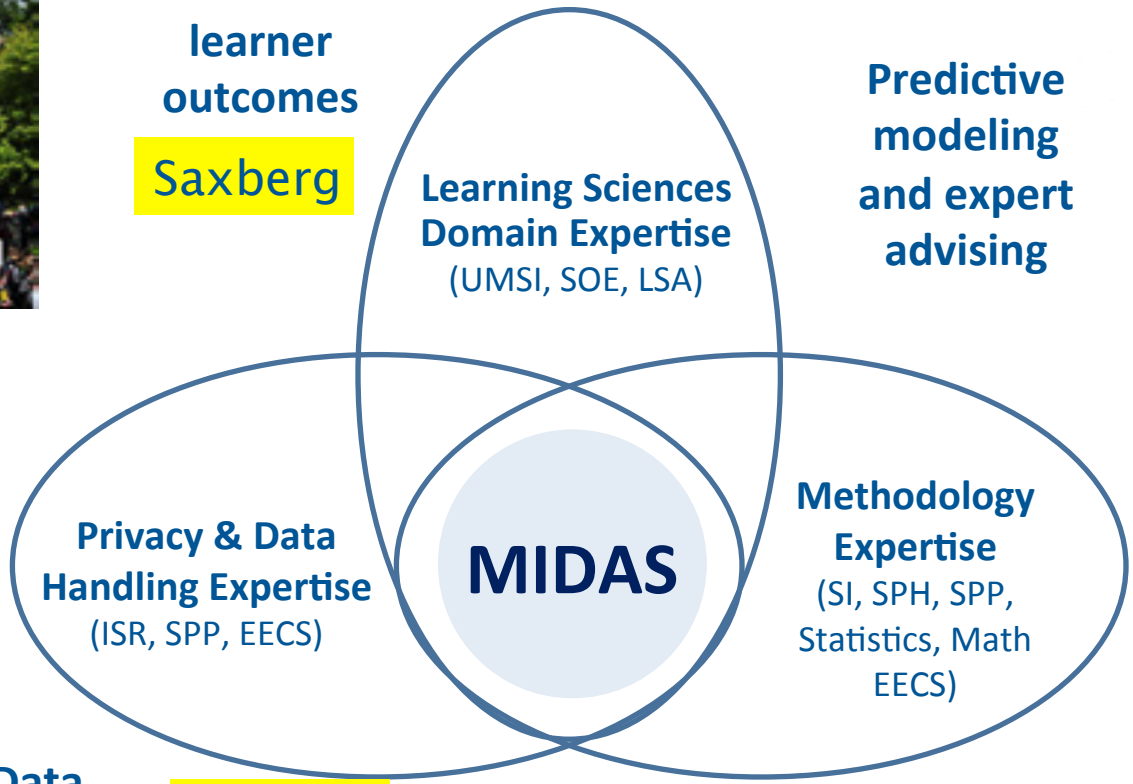
Multimodal assessment of learner outcomes

Saxberg

Personalized education at scale

Saxberg

Predictive modeling and expert advising



MIDAS MICHIGAN INSTITUTE FOR DATA SCIENCE UNIVERSITY OF MICHIGAN

MIDAS Social Science Challenge

Institute for Social Research

Media-driven socio-economic prediction Carley

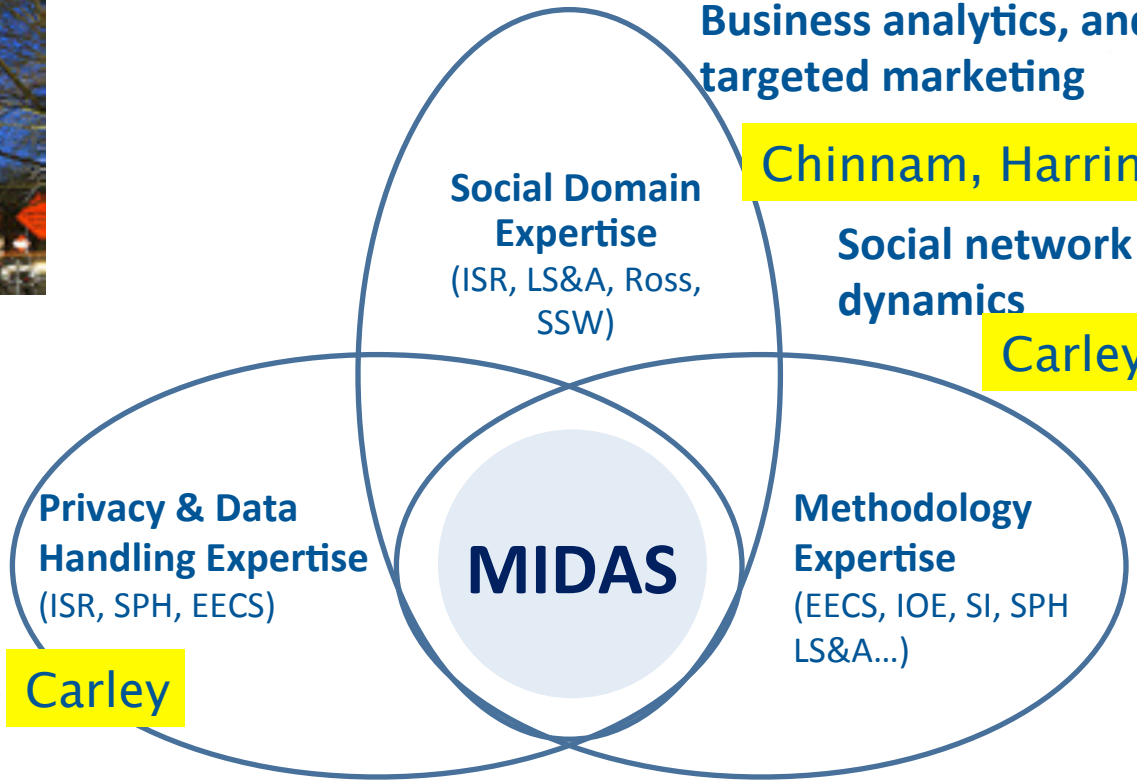
Data aggregation: postings, social, economic, demographic Carley

Social-media survey analytics

Information security and privacy Goroff

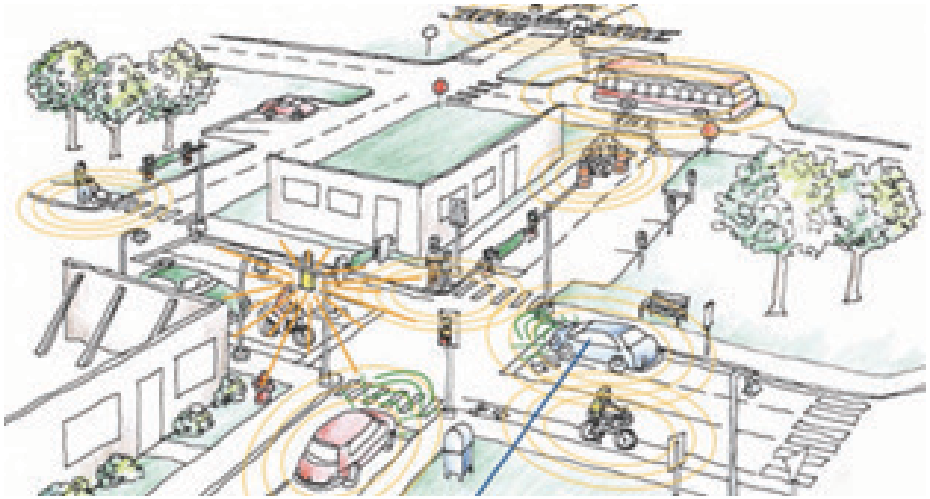
Business analytics, and targeted marketing Chinnam, Harringto

Social network dynamics Carley



MIDAS MICHIGAN INSTITUTE FOR DATA SCIENCE UNIVERSITY OF MICHIGAN

MIDAS Transportation Challenge

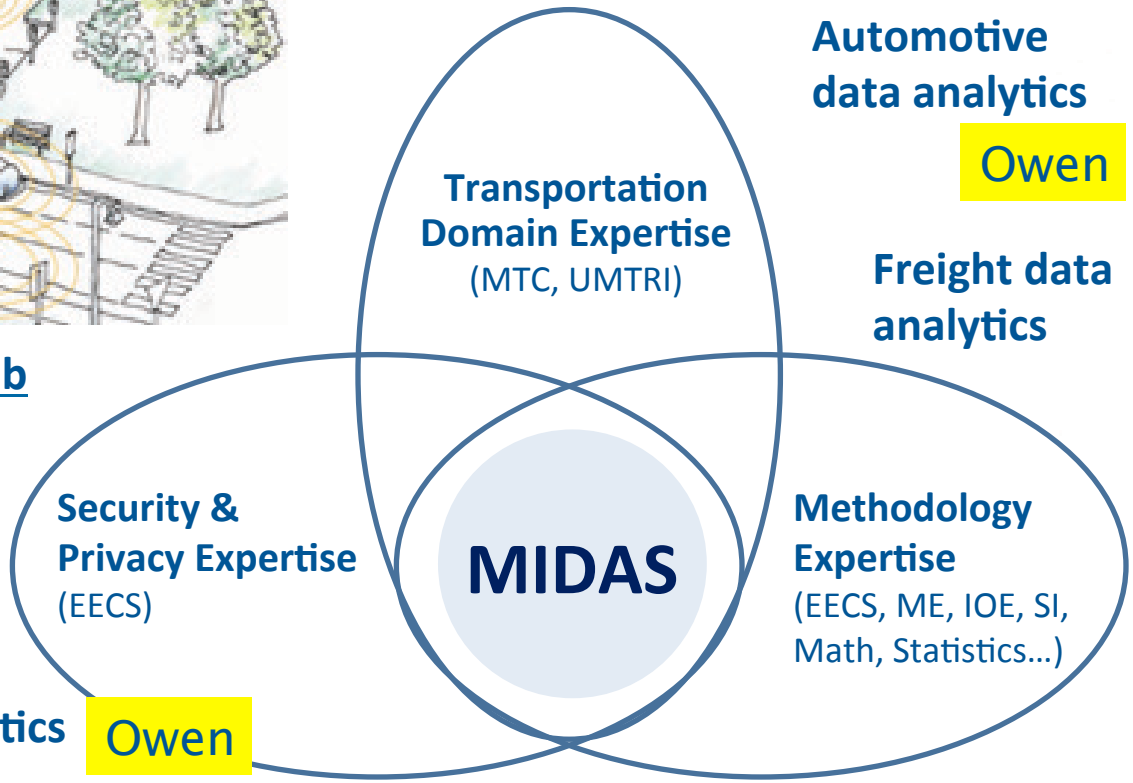


Mcity: A 32-Acre Outdoor Lab

Transportation data ecosystems for connected vehicles Owen

Automotive data analytics Owen

Freight data analytics



Automotive cybersecurity for connected vehicles

Accident and safety data analytics Owen

Data-analysis for mass transit

MIDAS MICHIGAN INSTITUTE FOR DATA SCIENCE UNIVERSITY OF MICHIGAN

Staging of Challenge Thrust Proposals

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

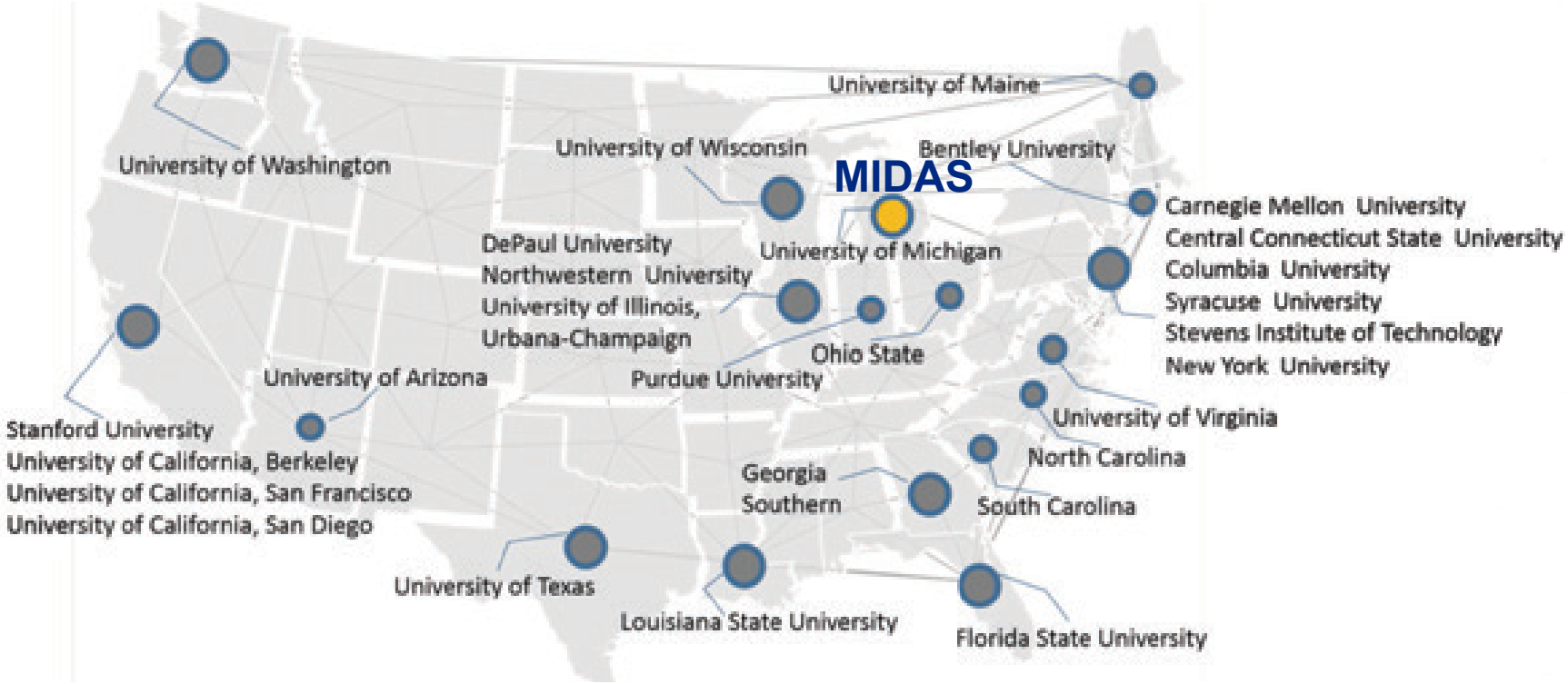
MIDAS plans to fund a total of 8 proposals in the next 2 years

- Evenly split over the 4 challenge thrusts
- Multi-disciplinary teams
- Funded over 3 years

New Challenge Thrusts will be launched In Year 3 and beyond

MIDAS MICHIGAN INSTITUTE FOR DATA SCIENCE UNIVERSITY OF MICHIGAN

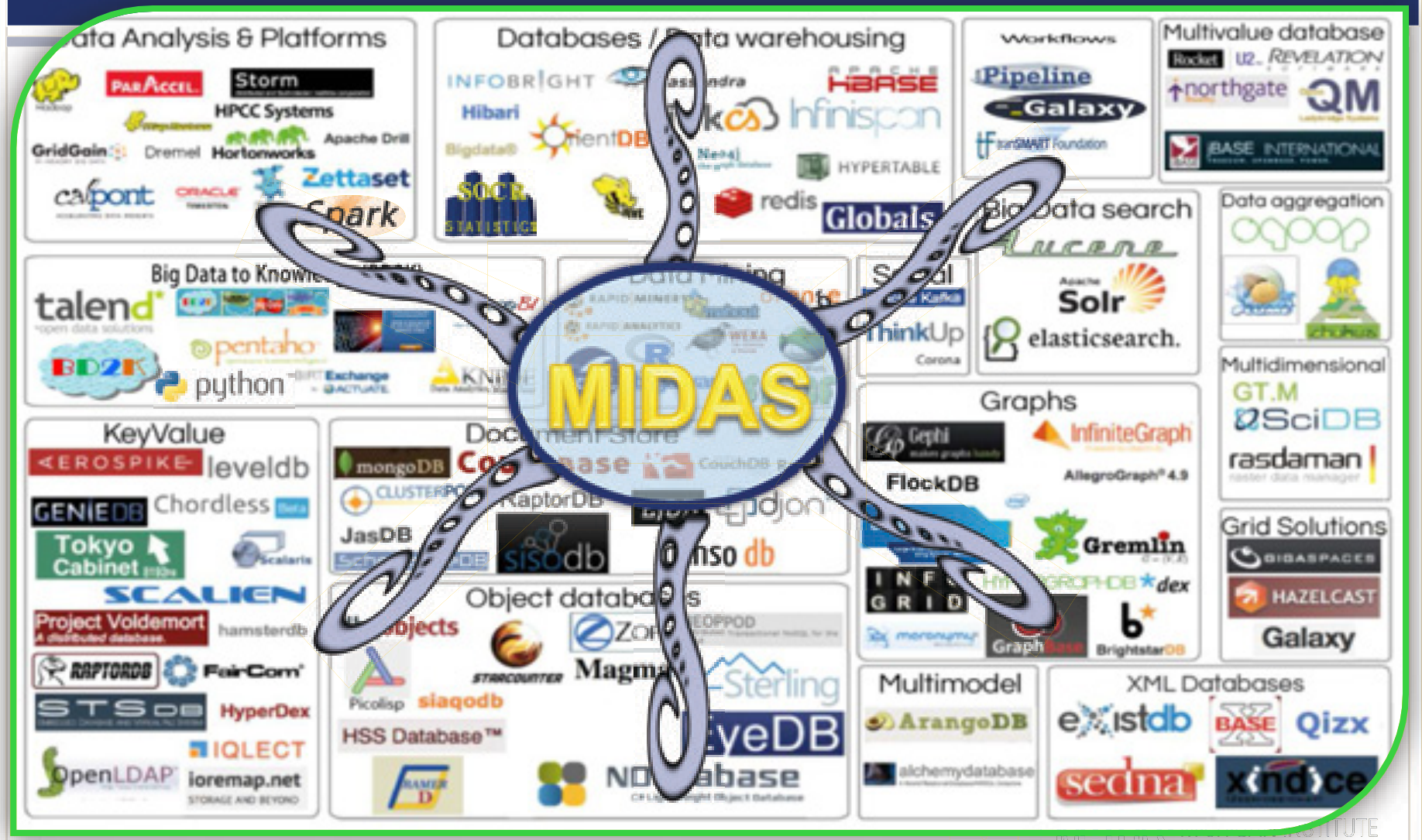
University Data Science Efforts are Launching Widely



How can MIDAS Efforts Stand Apart?

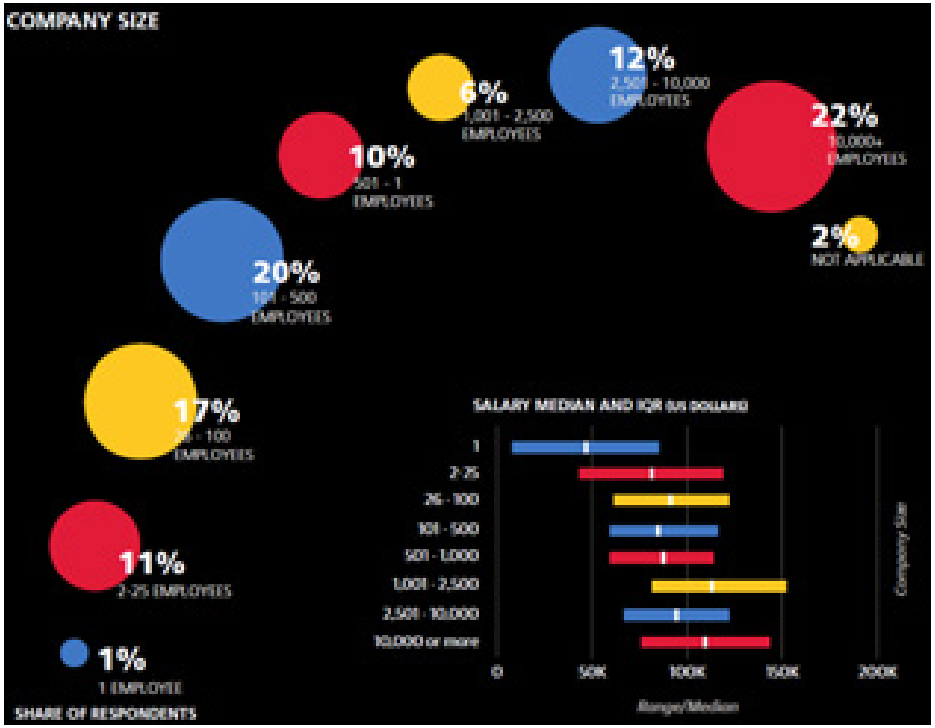
MIDAS MICHIGAN INSTITUTE FOR DATA SCIENCE UNIVERSITY OF MICHIGAN

The Opportunities for Partnerships are Limitless



http://socr.umich.edu/docs/BD2K/BigDataResourceome.html

Data About Data Scientists



O'Reilly 2015 Data Science Salary Survey

https://www.oreilly.com/ideas/2015-data-science-salary-survey

MIDAS MICHIGAN INSTITUTE FOR DATA SCIENCE UNIVERSITY OF MICHIGAN

MIDAS' Value Proposition for Industry Partners

Contact MIDAS.umich.edu

1. Education and recruiting
   - Help shape UM data science education/training programs
   - Provide employee continuing education: degree, non-degree, short course
   - Industry-student interaction via DS Certificate project sponsorships, internships
   - Student scholarships (Northrup-Grumman has endowed one of these already)
2. Access to DS experts
   - Machine learning, statistics, signal processing, data mining, database management, security, etc.
3. R&D partnerships
   - Involvement in challenge thrusts as sponsor, data provider, etc.
   - Industry partnering on MIDAS extramural grant proposals
4. Broader visibility
   - Participation or sponsorship of seminar series, workshop or short course

MIDAS MICHIGAN INSTITUTE FOR DATA SCIENCE UNIVERSITY OF MICHIGAN

Henry Kelly—MIDAS Industry Partnership Leader

- Chief Scientist, Energy Policy and Systems Analysis (EPSA), US Department of Energy
- Senior Advisor to the Director of the White House, Office of Science and Technology Policy (7/2012–8/2013)
- Principal Deputy Assistant Secretary for Energy Efficiency and Renewable Resources, US Department of Energy (7/2008–7/2012)
- President, Federation of American Scientists (FAS) (6/2000–6/2008)
- Assistant Director for Technology, Office of Science and Technology Policy (1993–2000)
- Acting Associate Director for Technology, Office of Science and Technology Policy (10/96–11/97)

Dr. Kelly will manage NSF MWBH Activities and support building novel industry partnerships on behalf of MIDAS

MIDAS MICHIGAN INSTITUTE FOR DATA SCIENCE UNIVERSITY OF MICHIGAN

Data Science Services and IT Infrastructure

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

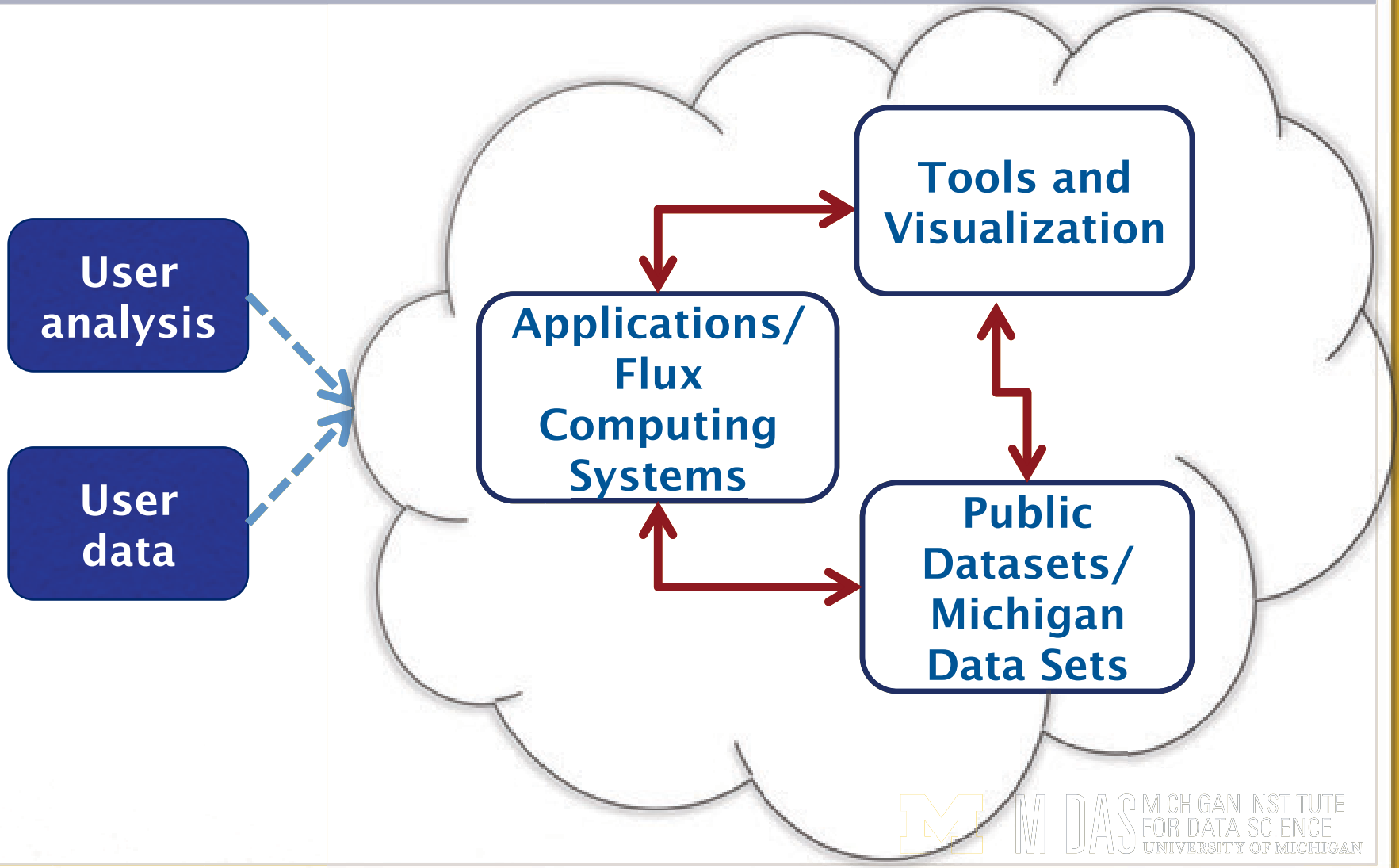
Consulting for

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

Data Science Infrastructure (ARC-TS)

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

MIDAS ARC Cloud Computing and Data Infrastructure Vision



UM Data Science Initiative Working Group (N= 32)

- **"Hosted"** by UMOR Office of Advanced Research Computing (ARC); Michielssen
- **Co-Chaired** by H.V. Jagadish and Brian Athey
- **Engineering (n=7):** Jagadish, (CSE), Singh (CSE), Hero (ECE), Jin (IOE), Powell (Aero)
- **Med School (n=4):** Athey (DCMB/Psych), Ayanian (IHPI/Int. Med), Friedman (LHS)
- **LS&A (n=8):** Nair (Statistics), Miller (Astronomy), Gilbert (Math), Smith (EEB), Evrard (Physics)
- **School of Information (n=6):** Finholt, Mei, Lagoze, van Houweling
- **School of Nursing (n = 1):** Dinov
- **School of Public Health (n=2):** (SPH): Abecasis, Little
- **ISR (n=4):** Alter (ISR/LSA), Gonzales (ISR/LSA), Pasek (ISR/LSA)

MIDAS MICHIGAN INSTITUTE FOR DATA SCIENCE UNIVERSITY OF MICHIGAN