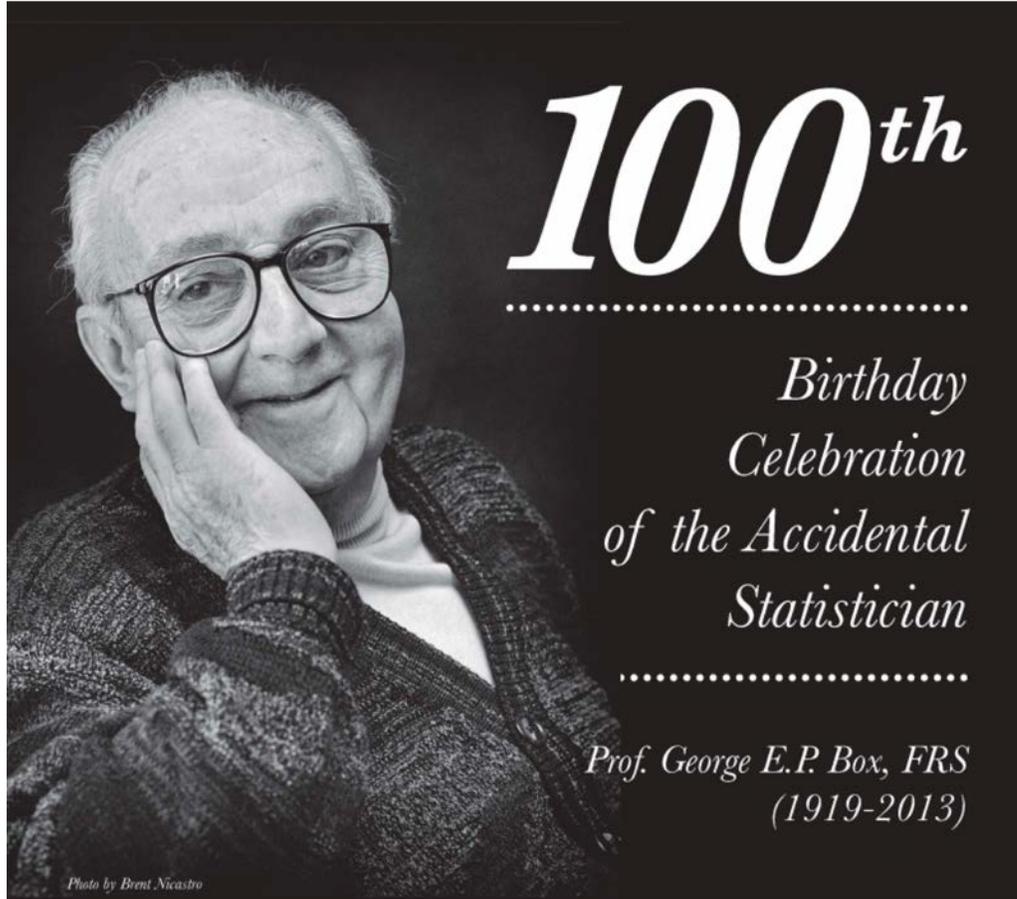# American Family Insurance Data Science Institute Progress Update
Brian S. Yandell, Interim Director

George Box at 100, 18 October 2019

All models are wrong but some are useful.

- Founded Department of Statistics.
- Co-founded Center for Quality & Productivity Improvement.
- Profound contributor in several areas, particularly experimental design, time series and quality improvement.
- Mentored several generations of grateful students.
- Influenced researchers far and wide.
- Known for his sense of humor, genial nature and compelling storytelling abilities.

https://old-www.stat.wisc.edu/people/george_box

American Family Insurance
Data Science Institute
UNIVERSITY OF WISCONSIN–MADISON

Box: multiple books, >250 papers, >30 students from 1954 to 2005

- *The Design and Analysis of Industrial Experiments* (ed. by O. L. Davies; authors Box et al. 1954)
- *Time Series Analysis: Forecasting and Control* (Box, G. M. Jenkins, and G. C. Reinsel. 1970, 1976, 1994, 2007)
- *Bayesian Inference in Statistical Analysis* (Box and G. C. Tiao. 1973)
- *Statistics for Experimenters: an introduction to design, data analysis, and model building* (Box, J. S. Hunter, and W. G. Hunter. 1977, 2005)
- *Evolutionary Operation – A Statistical Method for Process Improvement* (Box and N. R. Draper. 1969)
- *Empirical Model-Building and Response Surfaces* (Box and N. R. Draper. 1987)
- *Response Surfaces, Mixtures and Ridge Analyses* (Box and N. R. Draper. 2007)
- *Statistical Control by Monitoring and Feedback Adjustment* (Box, A. Luceño, and M. C. Paniagua-Quiñones. 1997, 2009)
- *The Collected Works of George E.P. Box* (Edited by G. C. Tiao. Volumes 1&2, 1985)
- *Box on quality and discovery – With design, control, and robustness* (Edited by G C Tiao, S Bisgaard, W J Hill, D Pena and S M Stigler. John Wiley & Sons, 2000)
- *Improving Almost Anything: Ideas and Essays* (Box and Friends. 2006)
- *An Accidental Statistician* (Box. 2013)

Statistics Department Updates

- new School of Computer, Data & Information Sciences
  - houses Statistics, Computer Sciences, iSchool
  - potential for substantial growth in coming years
- new Undergraduate Major in Data Science
  - managed by Statistics Department
  - collaboration with Math, CS, iSchool and many others
- new Data Science Institute
  - directed by member of Statistics Department
- growth in degree programs
  - Undergraduate Major in Statistics: ~500 (from ~20 ten years ago)
 - MS Statistics Option Data Science: ~65 (from 0 five years ago)

What are data? What is data science?

- **What are data?**
  - Data include numbers, text, images, graphs, sounds, code, and metadata. Anything you can measure.
- **What is data science?**
  - Data science is the study, development, and application of methods that reveal new insights from data.
- **What about Big Data?**
  - Big Data to some are small data to others.
  - Most data-rich problems are of modest size.
  - Size is relative, and keeps changing.

American Family Insurance
Data Science Institute
UNIVERSITY OF WISCONSIN–MADISON

What is Data Science Institute business motivation?

- Mission:
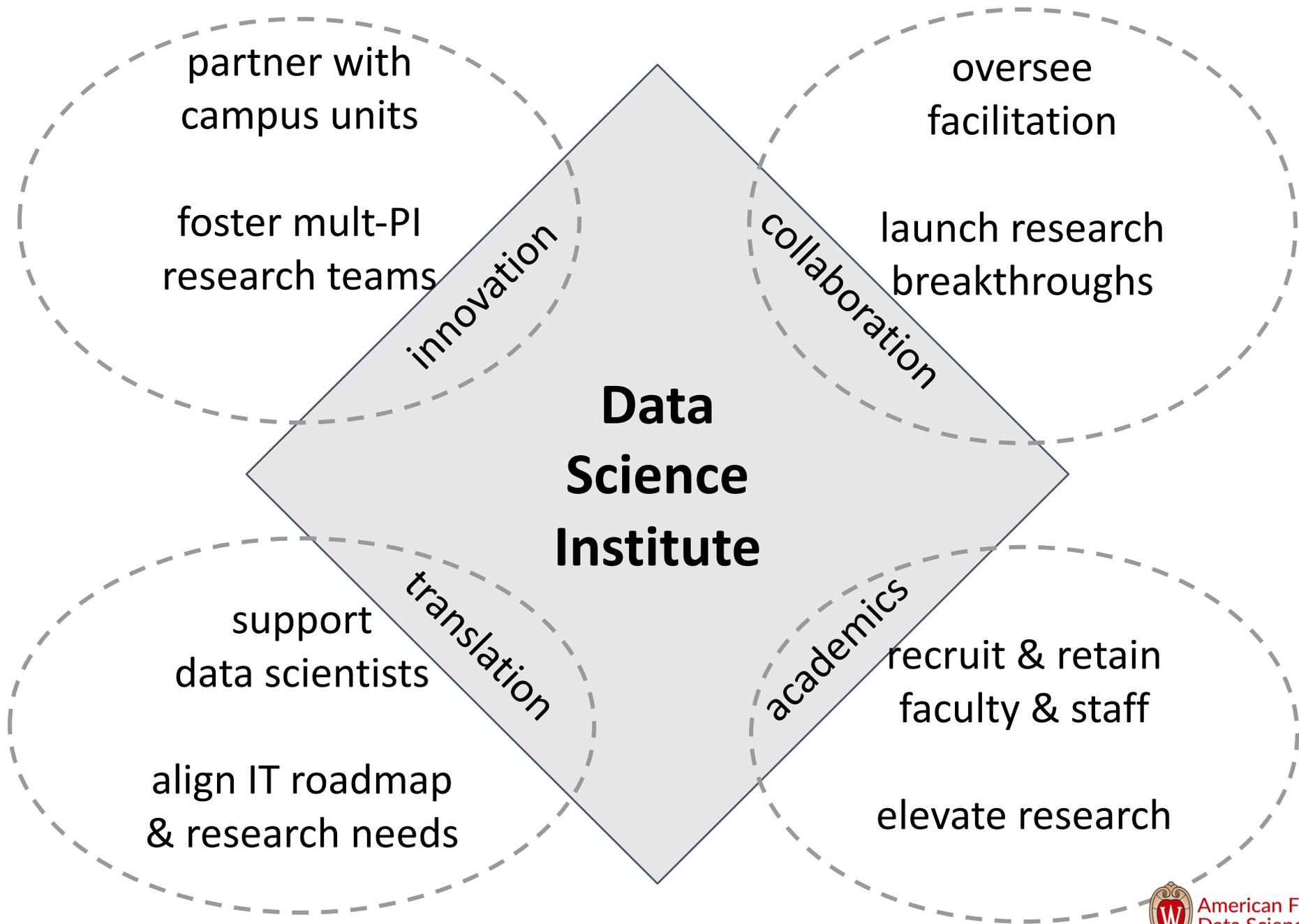 - **innovate** data science methods and approaches
  - **translate** research into practice
  - **collaborate** to advance scientific discovery
- Motivation:
  - create exciting collaborative research space
  - leverage expertise in teams tackling data-rich projects
  - build momentum toward large-scale data science funding

American Family Insurance
Data Science Institute
UNIVERSITY OF WISCONSIN–MADISON
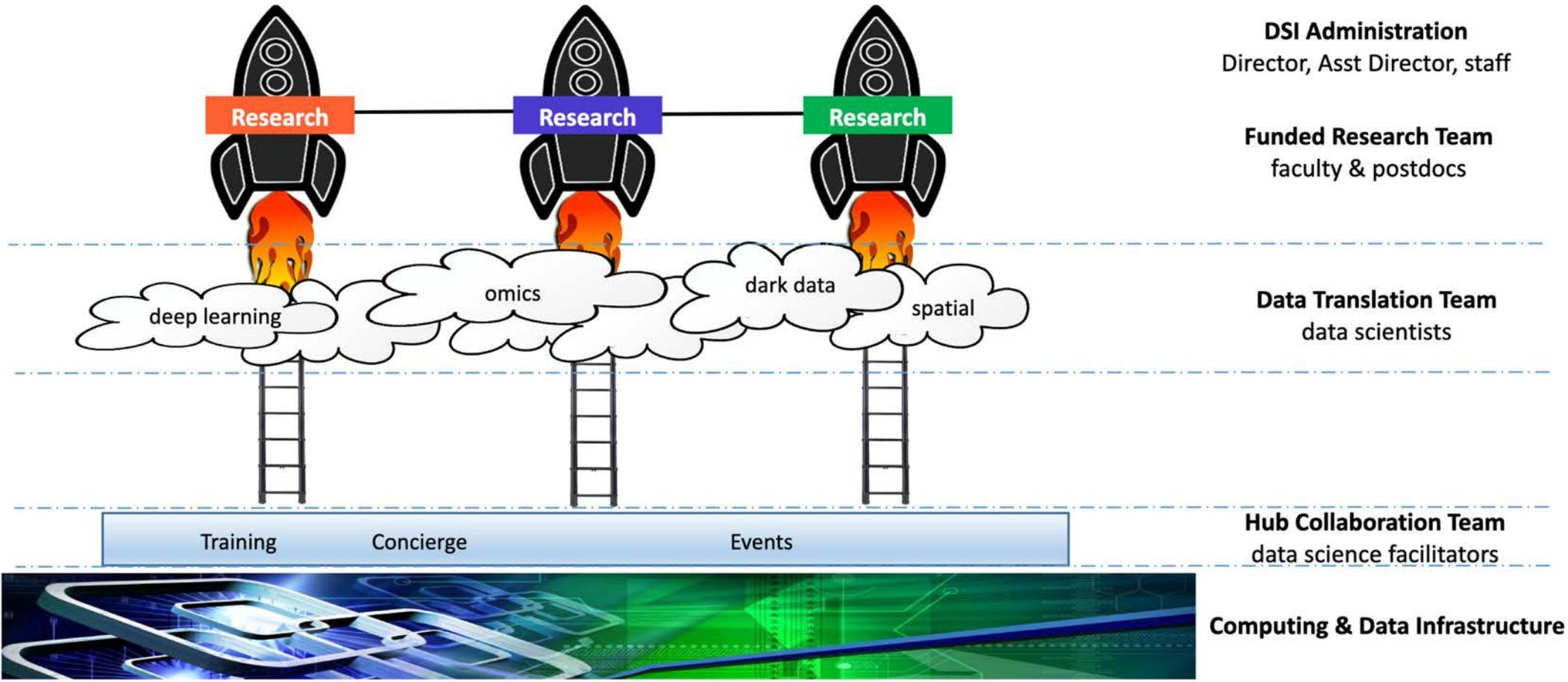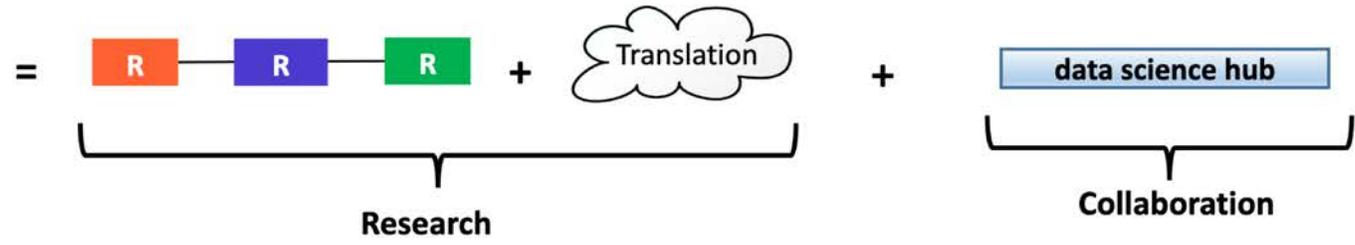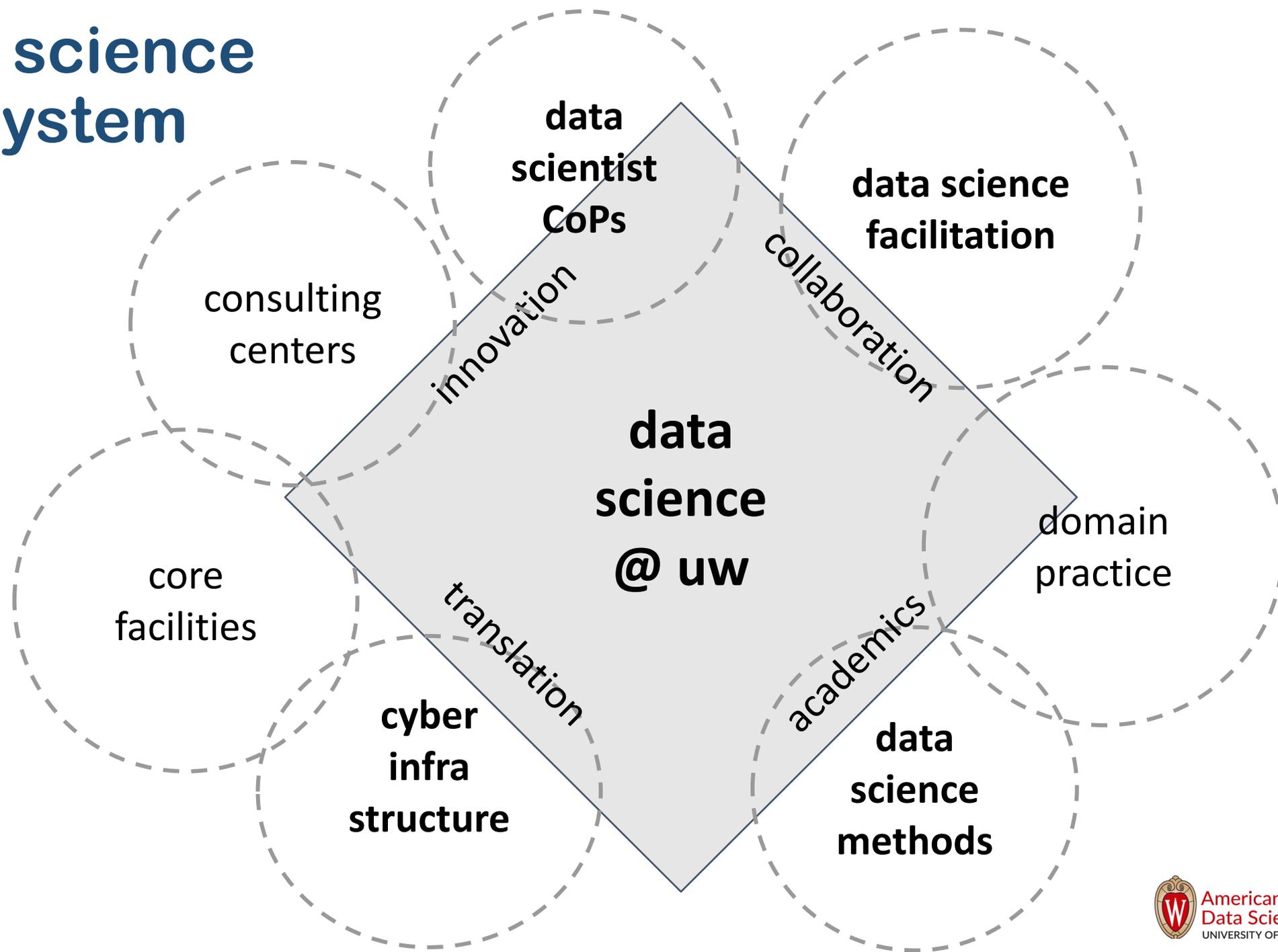
Some Practical Data Science Challenges

- How to improve data quality in research domains?
  - bottlenecks of data curation
- How to best integrate diverse data into systems-level models?
  - multiple data sources, models at scale, policy challenges
- Help nonprofits, governments, companies make better decisions?
  - prepare managers to use data to make better decisions
- What are key evolving patterns in human literature?
  - text analytics on 17M-volume Hathi Trust
- What chemotherapy works best for this cancer patient?
  - data lakes, HIPAA, biomed visualization, and real-time team decisions
- How does genetics influence metabolic disease?
  - central dogma DNA-RNA-protein at scale and with microbiomes

American Family Insurance
Data Science Institute
UNIVERSITY OF WISCONSIN–MADISON

partner with campus units

foster mult-PI research teams

oversee facilitation

launch research breakthroughs

innovation

collaboration

Data Science Institute

translation

academics

support data scientists

align IT roadmap & research needs

recruit & retain faculty & staff

elevate research

American Family Insurance
Data Science Institute
UNIVERSITY OF WISCONSIN–MADISON

DSI workshop space

American Family Insurance
Data Science Institute
UNIVERSITY OF WISCONSIN–MADISON

Data science ecosystem

data scientist CoPs

data science facilitation

consulting centers

innovation

collaboration

data science @ uw

core facilities

translation

academics

domain practice

cyber infra structure

data science methods

American Family Insurance
Data Science Institute
UNIVERSITY OF WISCONSIN–MADISON

11

What are the anticipated benefits?

- Central data science presence with enhanced visibility
  - One stop shopping for UW collaborations
  - Coordinated thinking about data-rich research needs
  - Talent acquisition and development
- New design for partnerships around data-rich research
  - Integration of virtual/cloud tools into research
  - Expanded projects, new connections
  - New possibilities for industry and public engagement

American Family Insurance
Data Science Institute
UNIVERSITY OF WISCONSIN–MADISON