Data Science Initiative Advisory Council

15 January 2016

Mladen A. Vouk
Director

Alyson Wilson
Associate Director

Trey Overman
Program Manager

Research, Innovation + Economic Development

Agenda

- Winter Symposium
- Data Science Survey Results
- Education Initiatives
- Other Updates

Winter Symposium

Friday, February 5th, 9:00 – 5:00

Institute of Advanced Analytics

Reception to Follow

Winter Symposium Goals

- Highlight internal resources
- Highlight external partnerships
- Foster internal collaboration within NC State
  - Research Lightning Talks:
    - Short presentations by faculty members and researchers on individual topics, organized by college
      - Need to ensure a broad representation across disciplines
      - Please sign up to present

Winter Symposium Plenary Talks

- Opening Remarks – Dr. Alan Rebar
- Data Science Initiative – Dr. Mladen Vouk
- Institute for Advanced Analytics – Dr. Michael Rappa
- NCSU Institutional Data Science Efforts – Dr. Mary Lelik
- National Consortium for Data Sciences – RENCI
- NCSU Libraries Research Support – Steve Morris/ Hilary Davis
- Office of Information Technology – Eric Sills

Full Agenda at dsi.ncsu.edu

Data Science Survey Results

November, 2015
55 respondents

Data Science Initiative Surveys

Education

- Are there data science educational efforts already underway in your department?
- Should data science be institutionalized at the university?
- What do you most care about in data science education?

Infrastructure

- Is your current infrastructure sufficient?
- Do you use existing resources?
- If no, why not?

Education Survey Results

- 75% of respondents from CALS, Engineering and College of Sciences
- 71% have data science related activities on-going within the college or department
- Only 48% have data science as a part of any curriculum
- 64% agree that data science needs to be institutionalized as an education requirement at the University
  - 61% believe it should be a basic education requirement

What are the key things you care about data science?

- How to collect data and interpret
- General architecture of data collection, analysis and interpretation ecosystem
- Data management, storage, analysis, big data concepts, visualization
- All students should understand the difference between data and information
- Data is a major part of every organization. Today's students will be the professionals that will have to work with data in a regular basis beyond their title or expertise. The better we prepare them to face the challenge the more successful we will be

Infrastructure Survey Results

- 81% of respondents from CALS, Engineering and College of Sciences
- 79% say existing hardware infrastructure is not sufficient to support existing research. Top three insufficient areas:
  - 21% Storage Capacity
  - 17% Computer Processing
  - 14% Servers
- 55% say the current size of their data is 1TB – 100TB, however 23% say the current size of their data is <1TB
- Only 43% indicate they currently use the HPC, yet 70% indicate they would use it if upgraded by the University

Infrastructure Survey Selected Comments

If you do not use the High Performance Computing Services, why not?

- Domain specific suitability
- Awareness
- Support
- Software
- Inability to transfer large datasets (1TB+)

Education Initiatives

2016 Summer Short Course Series

- Week long professional development series
- Exploring leveraging Data Matters series by the National Consortium for Data Science and host at NC State
  - Intro to Data Science - Data Curation
  - Intro to Data Visualization - Data Mining and
  - Intro to Data Science using R Machine Learning
  - Etc…
- Additional Internally Curated Professional Development Series being examined by Dr. Alyson Wilson (Statistics) and Dan McGurrin (Poole College of Management)

Data Science in General Education

61% or survey respondents agree that data science should be a basic education requirement

Best Path Forward?

1. Establish Data Science Thematic Track in general education requirements
   a) 12 credit hours
   b) Courses listed should be offered every year and must be on one of the approved GEP course lists
2. New undergraduate course offering on Data Science
3. New undergraduate course on Data Science that is a general education requirement

Other Updates

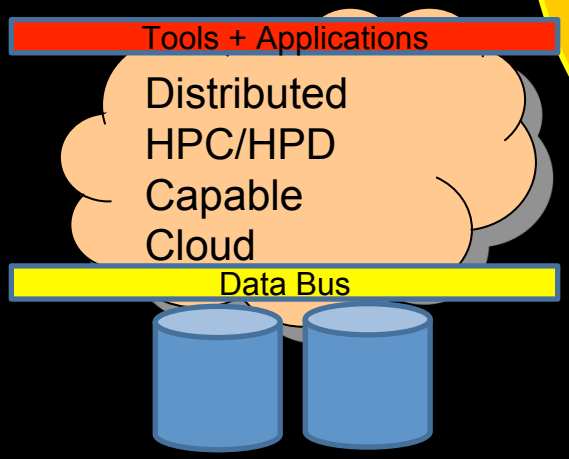
Research

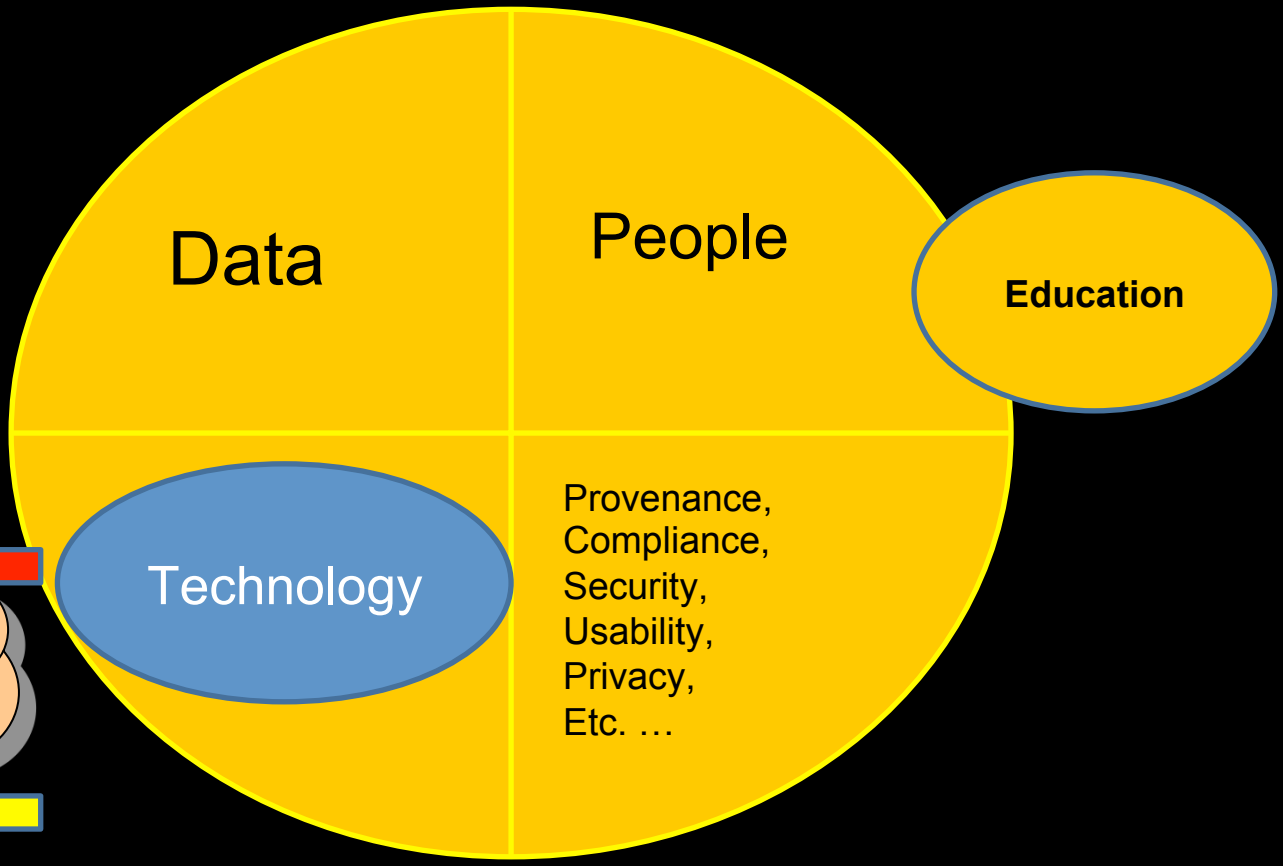
- NCSU Data Science Initiative (NCSU)
  Institutionalization of DS&A
- NC Data Science and Analytics Initiative (State)
- NSF Big Data Hub South East (NSF, US)

Framework and Infrastructure

Data Interaction Models:

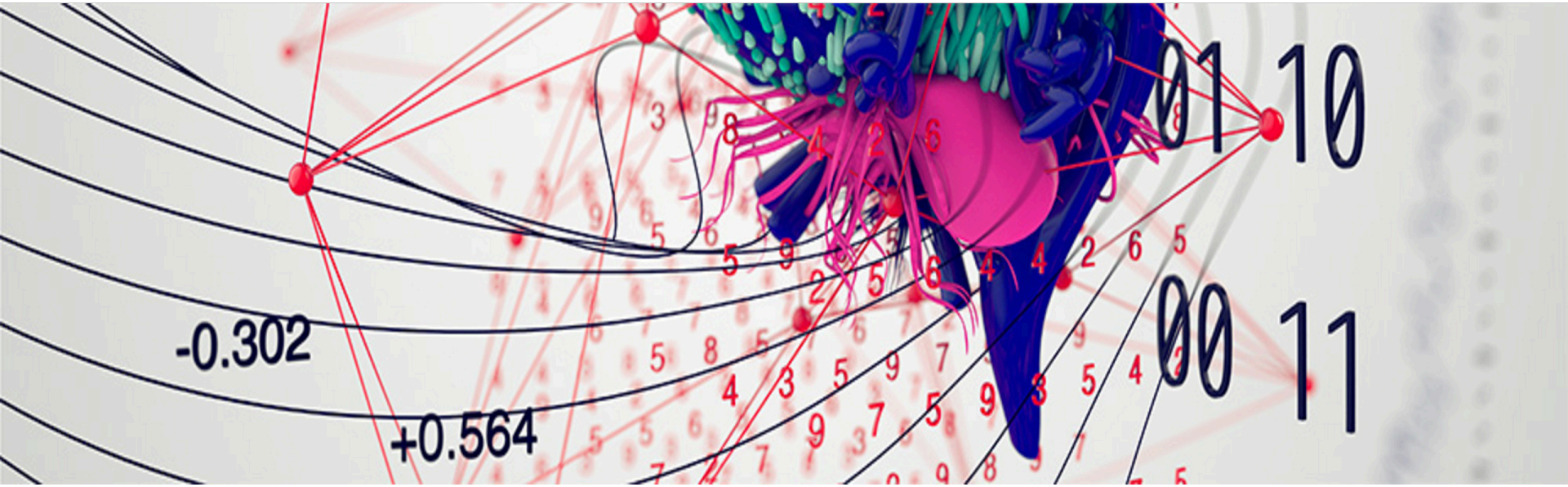
- Total Isolation
- Compute-to-Data
- Data-to-Compute
- Open model





NC STATE UNIVERSITY

Data Science Initiative (DSI)



2016 Winter Symposium, February 5, 2016, see details

Welcome to the Data Science Initiative

Understanding, managing, and using data — often large amounts of unstructured data — is becoming increasingly important in nearly every industry, government sector, and academic domain. Indeed, not having the skills and infrastructure to apply data science reliably has become a major risk in itself. The UNC System has recognized North Carolina's capacity to provide national leadership in this important sector, and NC State is uniquely positioned to address this goal.

NC State's Data Science Initiative will expand our research, teaching and outreach efforts to create a nationally recognized hub of excellence in data science and analytics by:

- integrating our existing capabilities;
- enhancing human and physical resources; and
- facilitating external partnerships.

Questions?