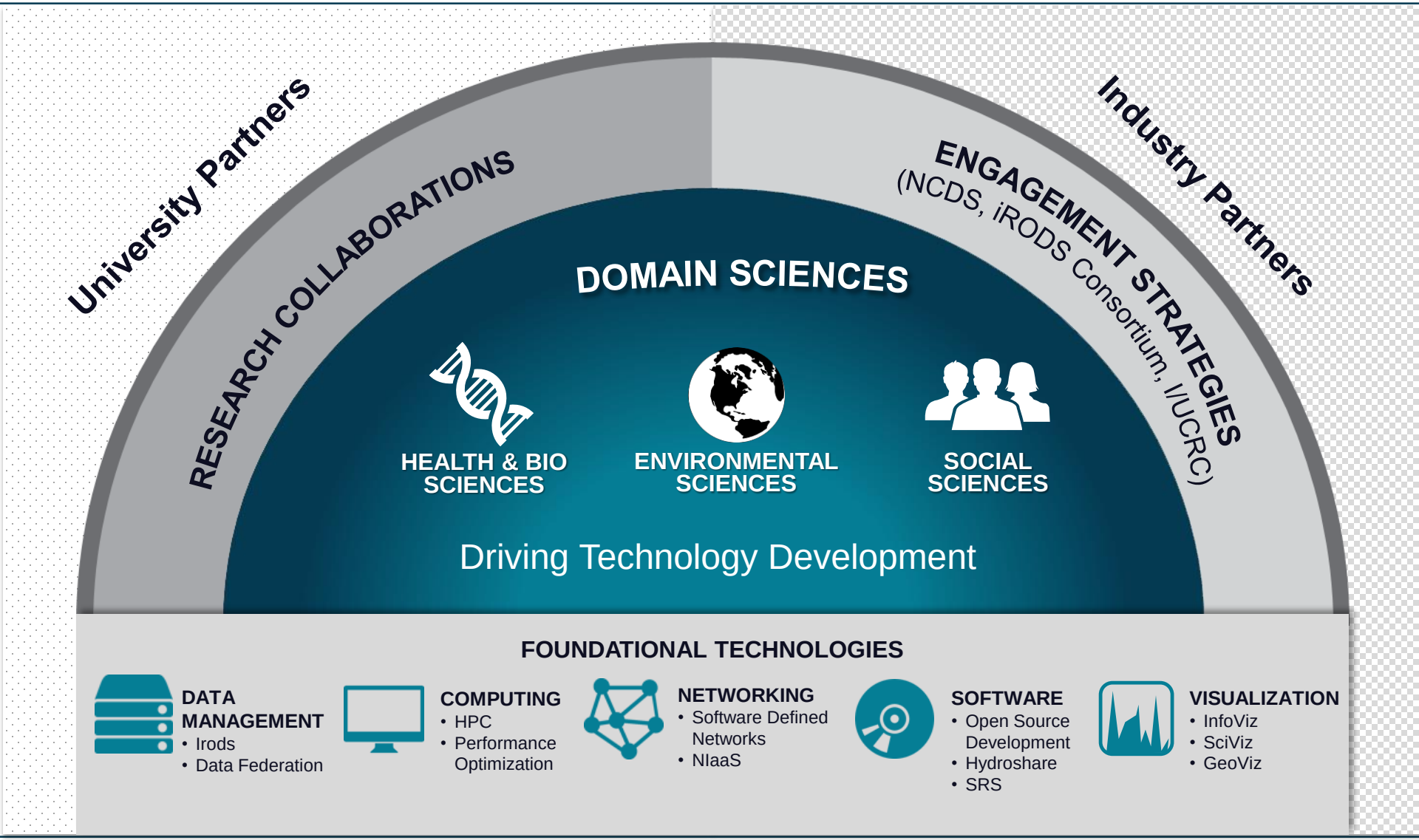
The National Consortium for Data Science

A public-private collaboration to advance data science

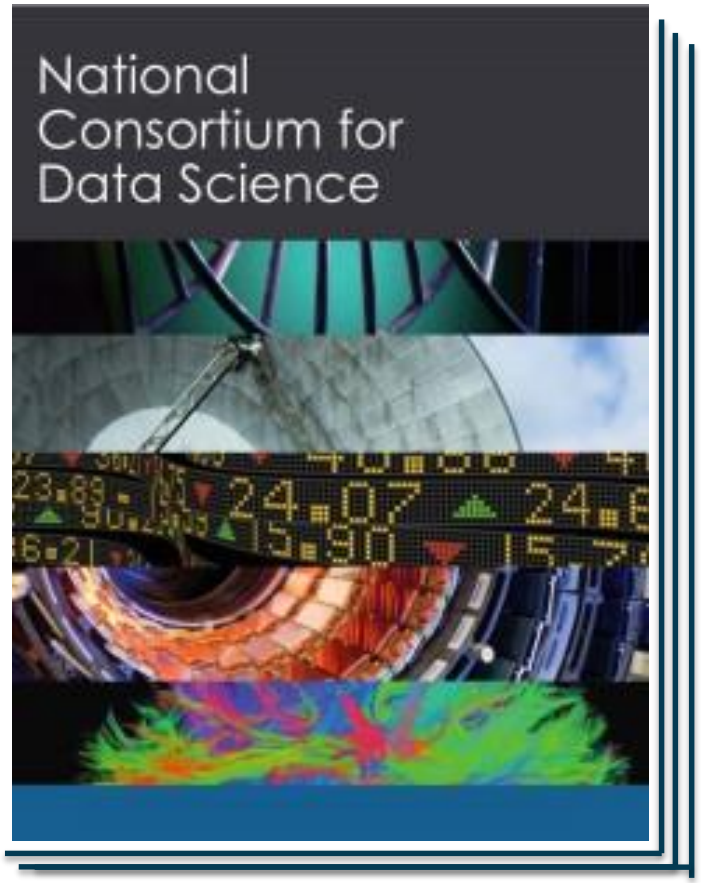
Renaissance Computing Institute (RENCI)
University of North Carolina, Chapel Hill

ncds | THE NATIONAL CONSORTIUM for DATA SCIENCE

RENCI FOCUS: DATA SCIENCE TO ENABLE RESEARCH, INNOVATION, AND ECONOMIC DEVELOPMENT



Foundational Concepts



- Data science needs to emerge as a complete domain science on par with other domains
- Public-private partnerships are necessary to access the requisite talent and perspectives
- Long-term thinking is essential
  - focus on research, education, and commerce

The National Consortium for Data Science

Mission

Help secure US role as leaders in data science research & education; position US industry to use the power of data to drive economic growth

Vision

Focused multi-sector, multidisciplinary data science community to solve big data challenges and drive the field forward

Goals

- **Engage** broad communities of data experts
- **Coordinate** data science research priorities that span disciplines and industries
- **Facilitate** development education & training programs
- **Apply** NCDS expertise to data challenges in science, business and government

ncds is a **strategic** approach to data science and big data opportunities.

NCDS Members

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

NC STATE UNIVERSITY

CISCO

EMC²

RTI INTERNATIONAL

Deloitte Consulting

IBM

THE UNIVERSITY of NORTH CAROLINA GREENSBORO

renci

Consortium Highlights

Research

- Data Fellows program
- Partnerships with NCBSTI & the DSA Initiative
- Focus on three overarching research themes

Education

- *Data Matters* Summer Short Course Series
- Data Science Career Panel Events

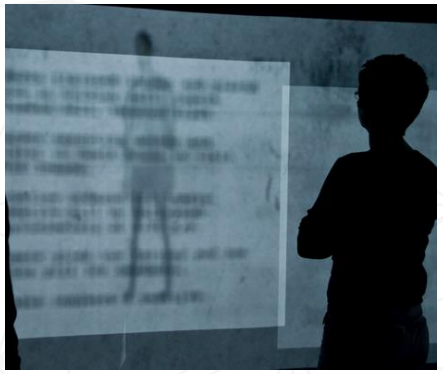
Engagement

- *DataBytes* Lunchtime Webinar Series
- Support for multiple data science conferences, hackathons
- NSF Big Data Hub

Current Research Themes

Innovative Infrastructure & the IOT

Workforce Development



Anonymizing Data

Purpose: Focus NCDS activities on data challenges of interest to members in order allow members to have a voice, increase the value of NCDS membership, produce products, and establish thought leadership.

Outcomes: White papers, position papers, best practice documents, lectures, panels, special events, etc.

Data Fellows Program

Fosters private-public relationships; **engages** future data scientists; **bridges gaps** between research and practice; **creates** NCDS-sponsored scholarship

$50,000 seed grants for early career faculty

Seed grant approach to fund initial cadre of Fellows from NCDS academic member campuses

Teaming with an NCDS member encouraged; industry-led selection process

Funds used for **course buy-outs, summer salary, graduate student support**, conference travel and modest infrastructure costs

Target: **3 awards, $50K each, *focus on early career faculty***

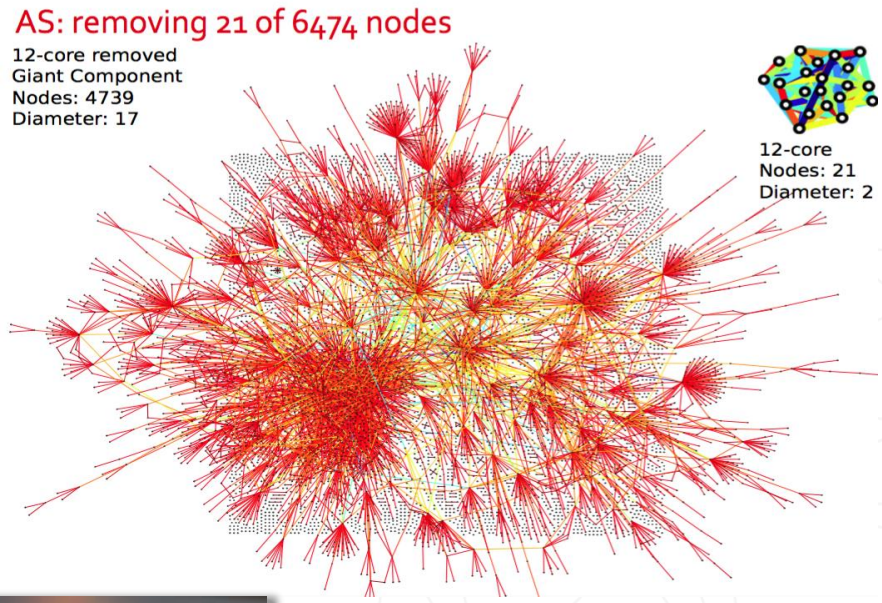
Additional support provided by UNC General Administration to **offer fellowships to UNC System campuses**.

NCSU Data Fellows Highlight

- Dr. Blair Sullivan selected for $1.5 million Moore Investigator Award as part of its Data-Driven Discovery Initiative

- Sullivan's work is based on a field of study called parameterized complexity. These algorithms leverage a graph's structure to solve time-consuming problems much more quickly.

NC STATE UNIVERSITY



"The NCDS congratulates Dr. Sullivan on this great opportunity to expand her research and turn theory into practice. This is exactly what we hope NCDS Data Fellowships will enable for talented faculty—new opportunities to move their research forward and answer important data science questions."

- Stan Ahalt, director of the NCDS Data Fellows Steering Committee

NC Data Science and Analytics Initiative

- Shared, multi-institutional data environment for collaboration & discovery
- Collaborative data store and discovery - iRods V4, DataVerse Network
- On-demand cloud computing for analytics workflows – NCSU's Virtual Computing Lab (VCL)
- Includes proposal process and financial support for promising data-related research across campuses
- Web interface & connectors to "conventional" HPC and Hadoop
- Distributed authentication managed by each campus
- 3 year & $2.2M investment by the State of North Carolina

UNC CHARLOTTE

NC STATE UNIVERSITY

renci

Partnership with NC Dept. of Commerce

NCBSTI identified data science as one of six "Grand Opportunities" for the NC economy

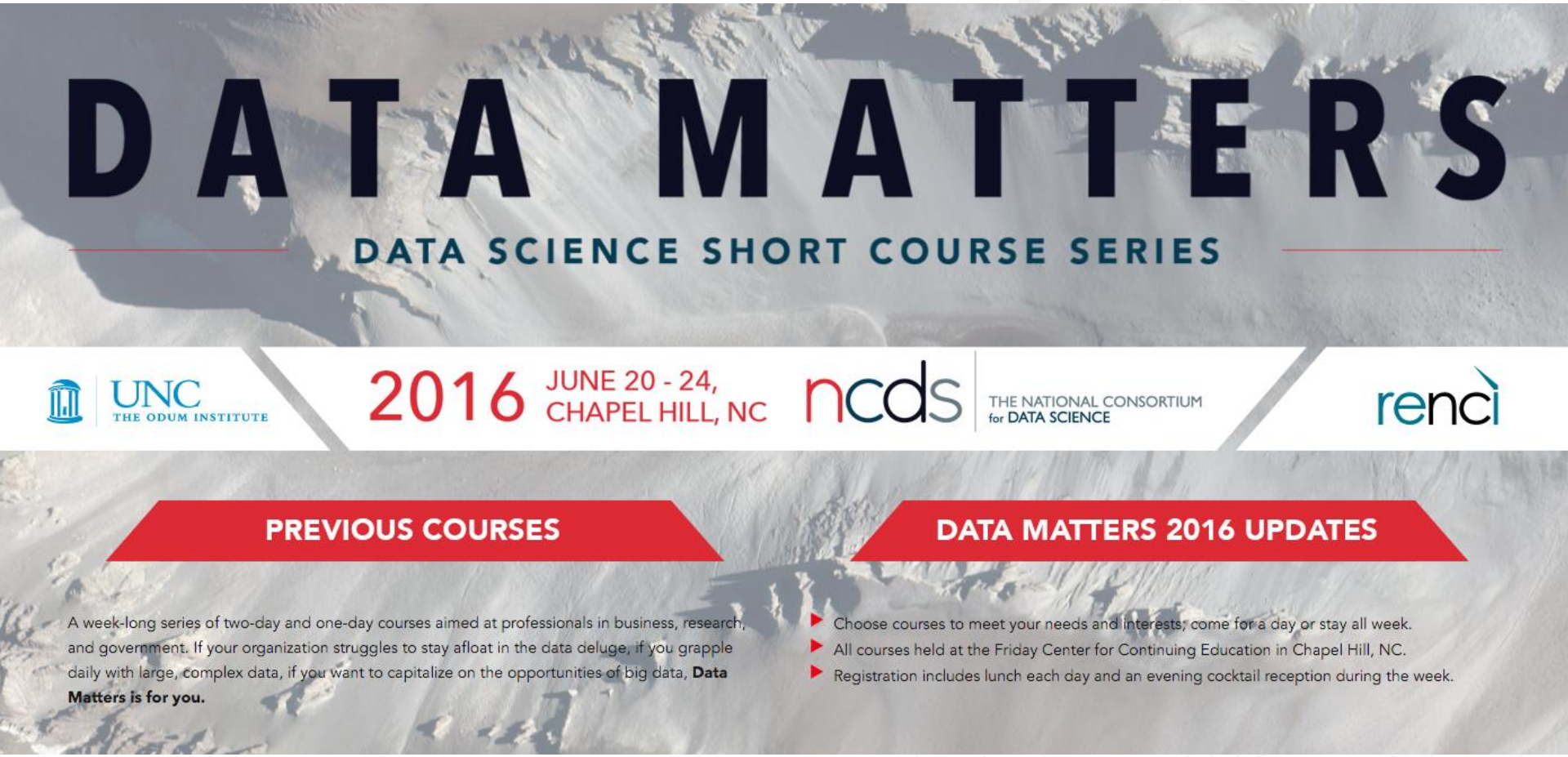
Project underway to survey NC data science leaders on identification of the most valuable strategy:

- Survey/research of data science assets in the state
- Develop "Data Sciences Index" and other methodology to assess the state's relative standing nationally
- Survey employer skill needs
- Engagement with KFBS "STAR" Program student team

North Carolina Board of Science, Technology & Innovation

Key question: ***How important is data science to North Carolina's economy now, and how important could it be?***

Workforce Training



Monday & Tuesday June 20 & 21	Wednesday June 22	Thursday & Friday June 23 & 24
Intro to Data Science	Conceptual Diagrams in Information Visualization	Intro to Data Mining & Machine Learning
Intro to Information Visualization	Programming in R	Health Informatics in the Age of Interoperability
Intro to Data Science Using R	Open(ing) Data	Collecting, Classifying, & Analyzing Textual Data
Data Curation: Managing Data throughout the Research Lifecycle	Creating Surveys in Qualtrics	Simulation Strategies in Data Science: System Dynamics & Agent-based Modeling
Writing Questions for Surveys	Intro to Big Data & Machine Learning for Survey Researchers & Social Scientists	Conducting & Analyzing Cognitive Interviews
Intro to Survey Sampling	Geospatial Data Science	Analysis with Complex Sample Survey Data

Career Networking Events

Student-focused, cross campus events designed to connect industry with future data science talent

SAS RTI INTERNATIONAL IBM CISCO EMC²

UNC KENAN-FLAGLER BUSINESS SCHOOL Deloitte. MetLife

DataBytes Webinar Series

Objective: Present novel and compelling data science concepts and projects

Member and non-member presenters

First Wednesday of the month

2015

Geospatially Explicit Synthetic Populations
Bill Wheaton

The Internet of Everything: The Next Big Wave
Russ Gyurek

Leveraging AI for Big Industrial Data Science
Steve Gustafson

Big Data, Small Data, & Behavioral Insights
James Guszcza

Analytics for Large-Scale Temporal Event Data
David Gotz

Toward Machine Oblivious Graph Analysis
Erik Saule

Assessing the Impact of Data & Software on Science
Erjia Yan

2016

JAN — *Semantics: Revolutionary Breakthrough or Same Old Thing*
Andy Crapo

FEB — *Simulating Storm Surge for Coastal Hazards and Risk Assessment*
Brian Blanton

MAR — *The Computing Universe*
Tony Hey

APR — *NOAA's Collaborative Big Data Project*
Jeff de La Beaujardiere

MAY — *Title TBD*
Stephen W. Edwards

JUN — ***TBD***

AUG — ***TBD***

SEP — ***TBD***

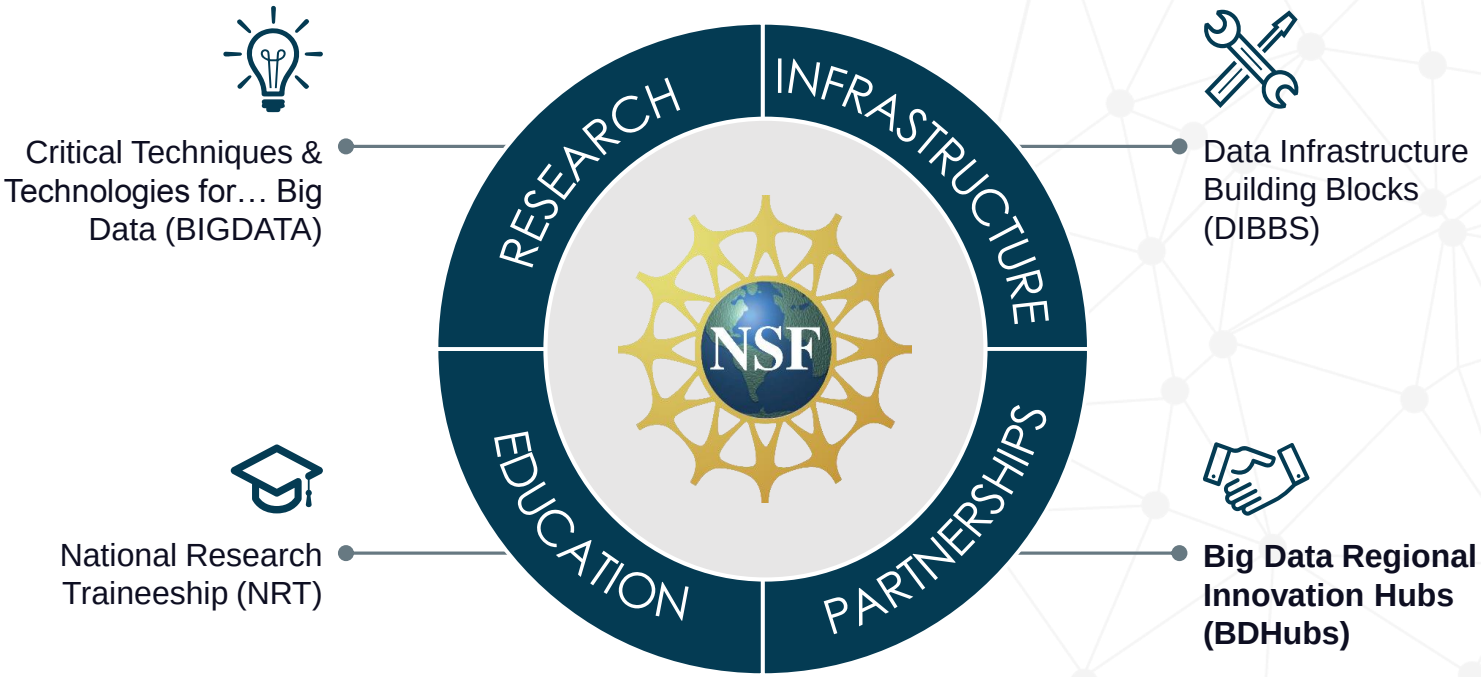
OCT — ***TBD***

NOV — ***TBD***

Partnership with BD Hub

SOUTHBDHUB

NSF Big Data Portfolio of Programs

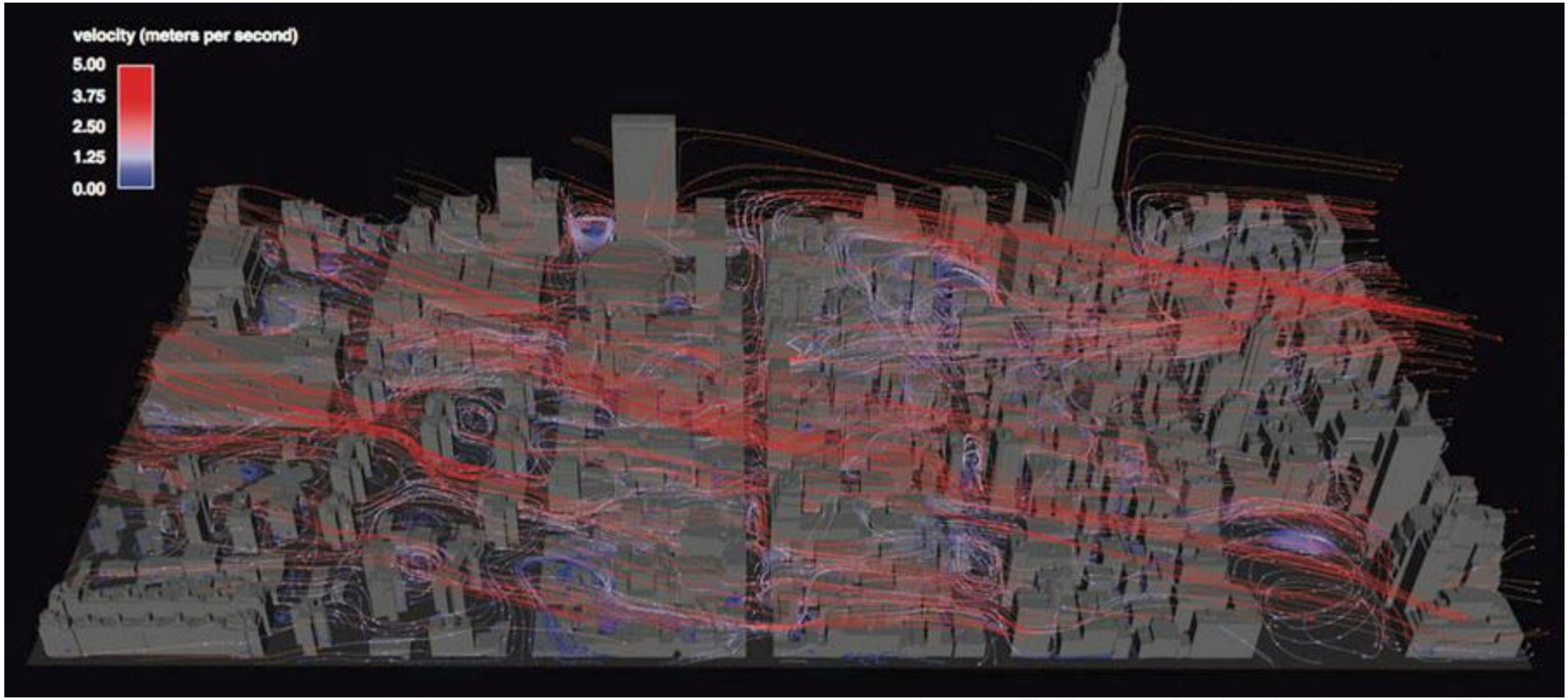


GOAL: To stimulate regional and grassroots partnerships involving public and private organizations of all types in order to establish a national Big Data innovation ecosystem

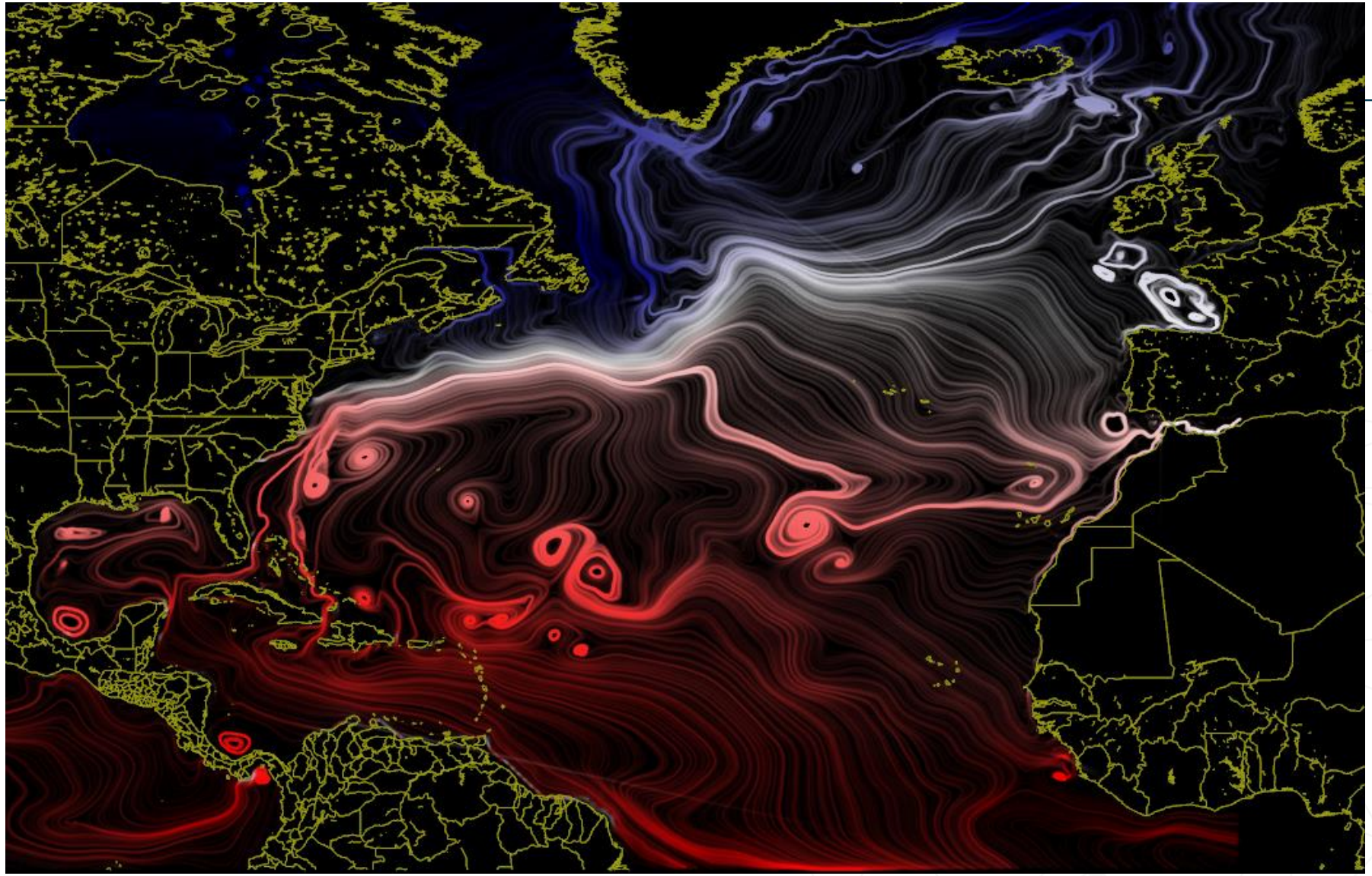
RENCI Engagement Center at NCSU

- Established in 2007 at Centennial Campus to foster collaborations with NCSU research community
- Visualization-focused; Two full-time resident visualization researchers (PhD)
- Project focus spanning scientific visualization and information visualization
- Establishment of **Visualization Studio** (RENCI's Social Computing Room) at DH Hill Library
- RENCI-sponsored Vis Call for Projects focusing on Applied Visualization Research
  - $12,000 (each) for two projects
  - 50% dedicated personnel time
  - Running two years in a row (2012-13)

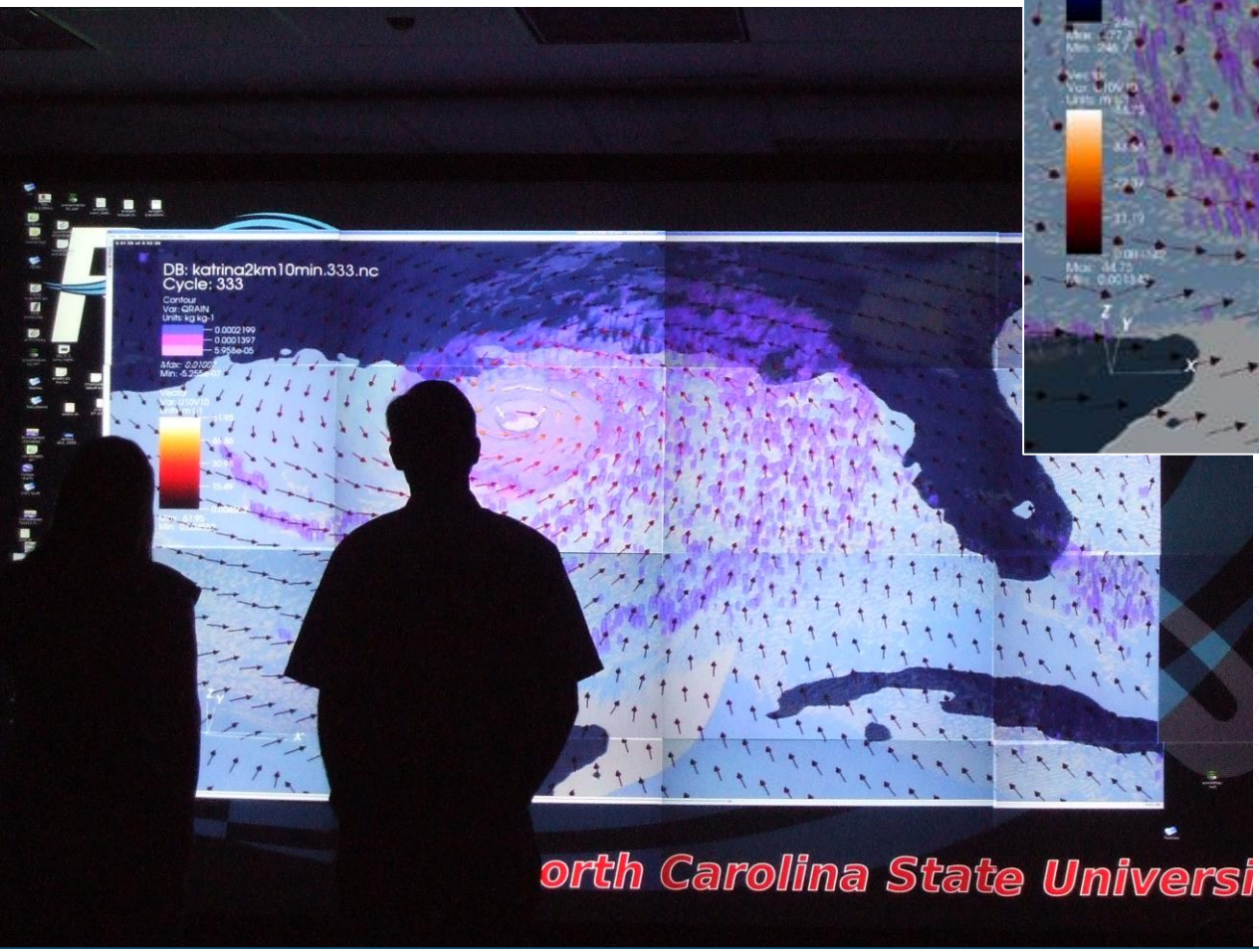
RENCI Visualization Examples

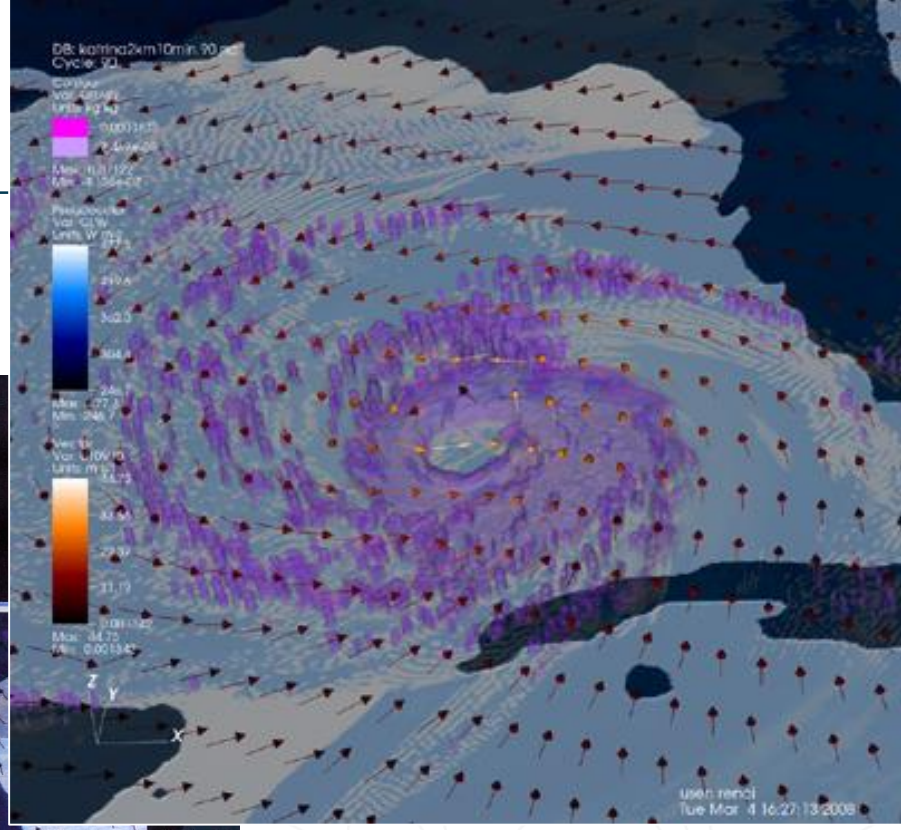


Visualization: Dr. David Borland, RENCI

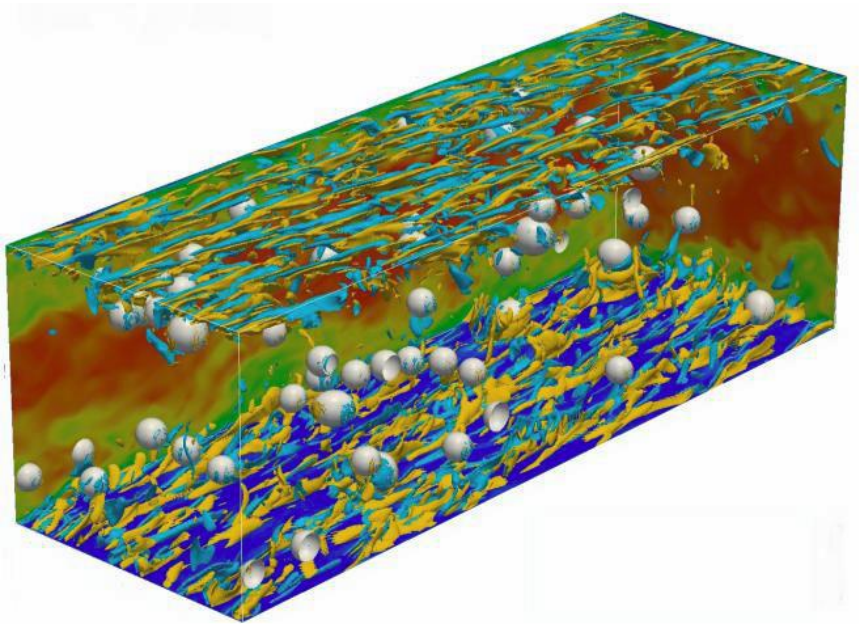


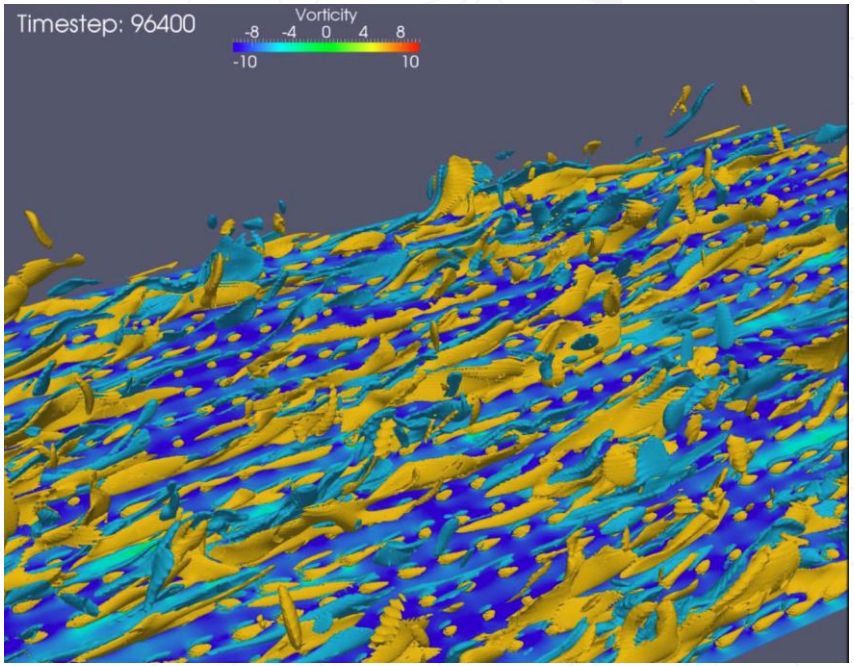
Visualization: Dr. David Borland, RENCI



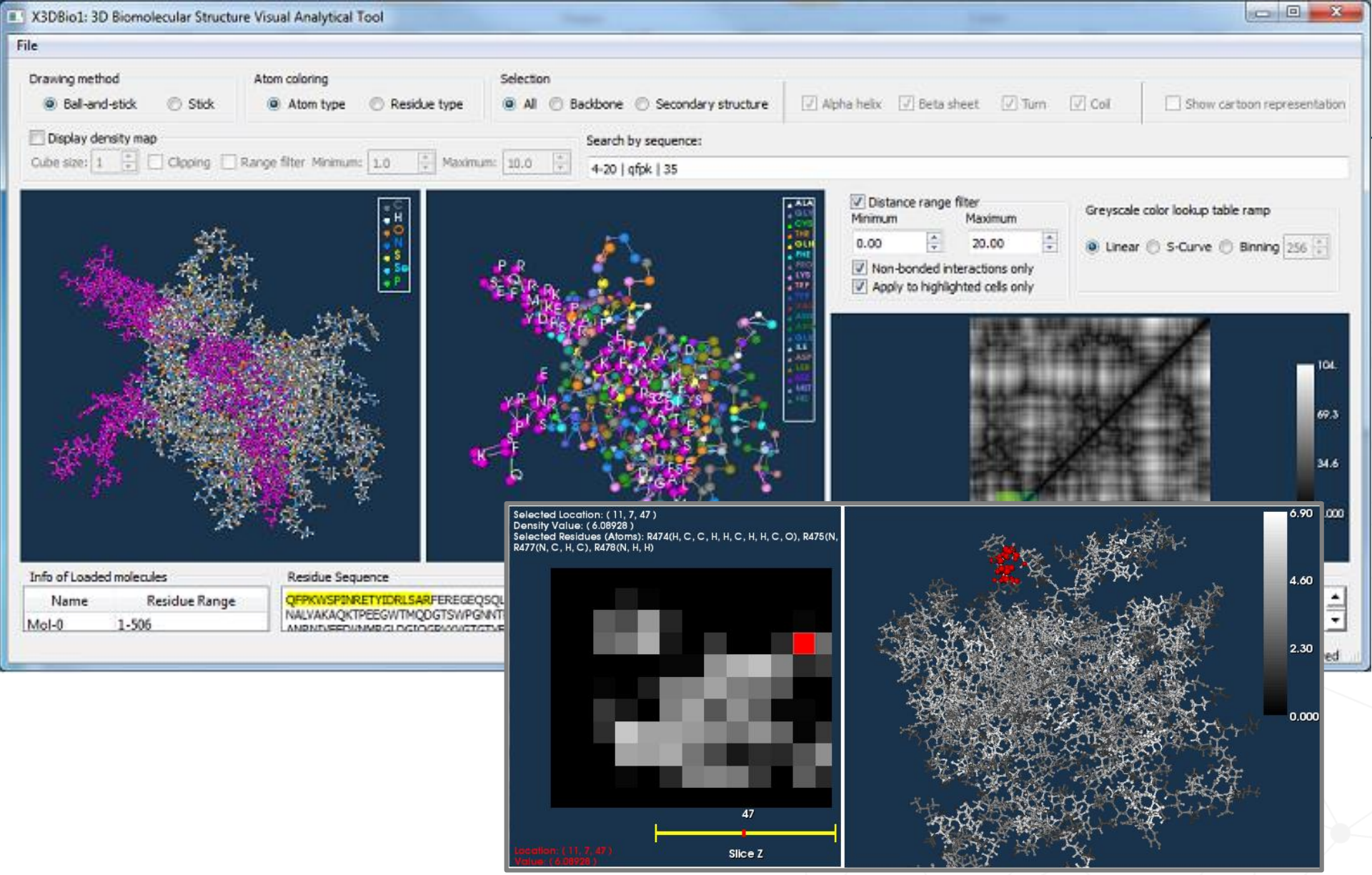


Visualizations: Steve Chall, Theresa-Marie Rhyne

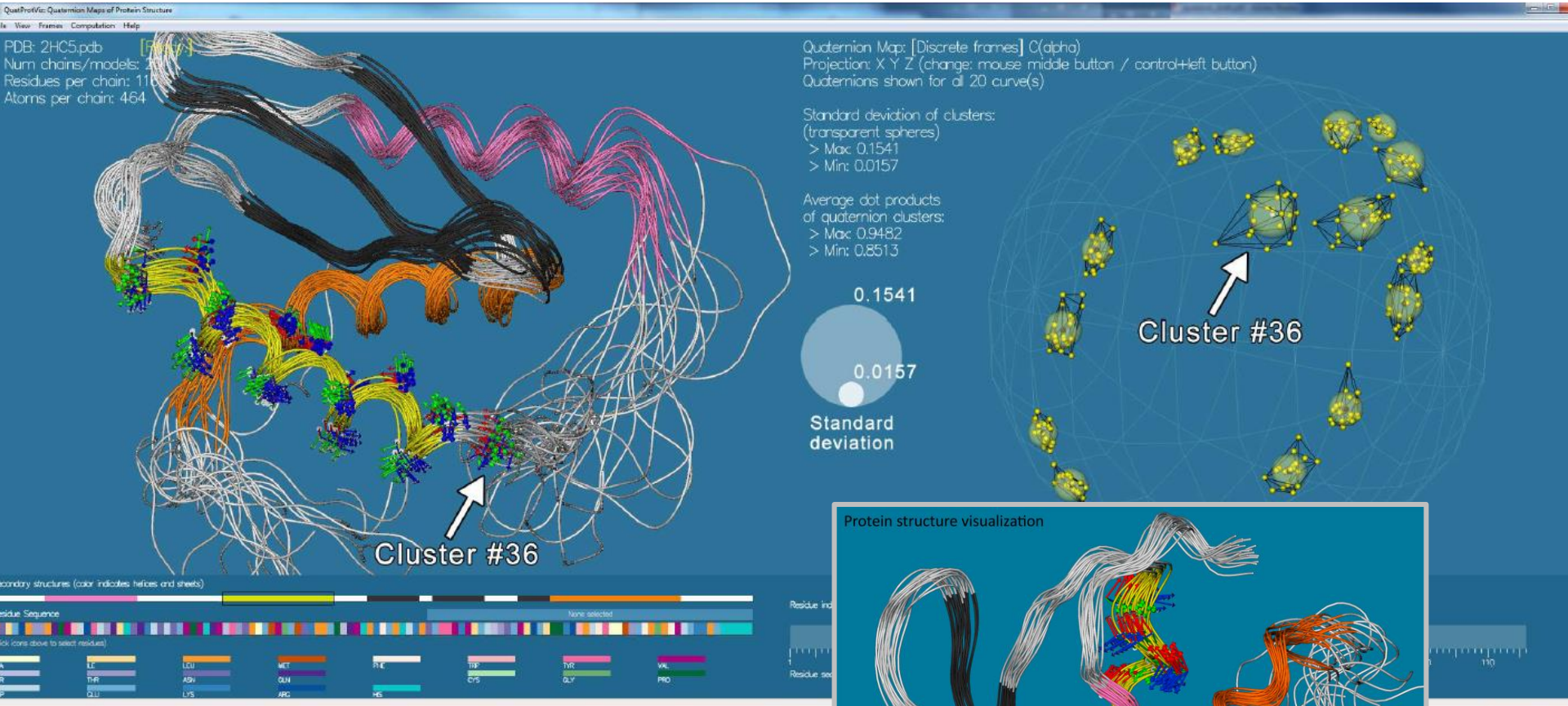


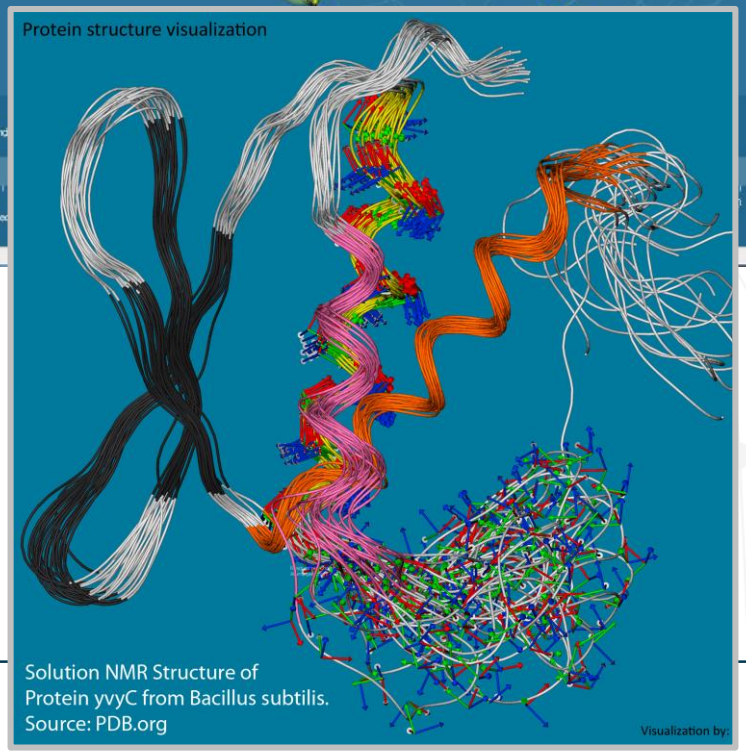


Visualizations: Dr. Igor Bolotnov's group,
Dr. Hong Yi (RENCI)

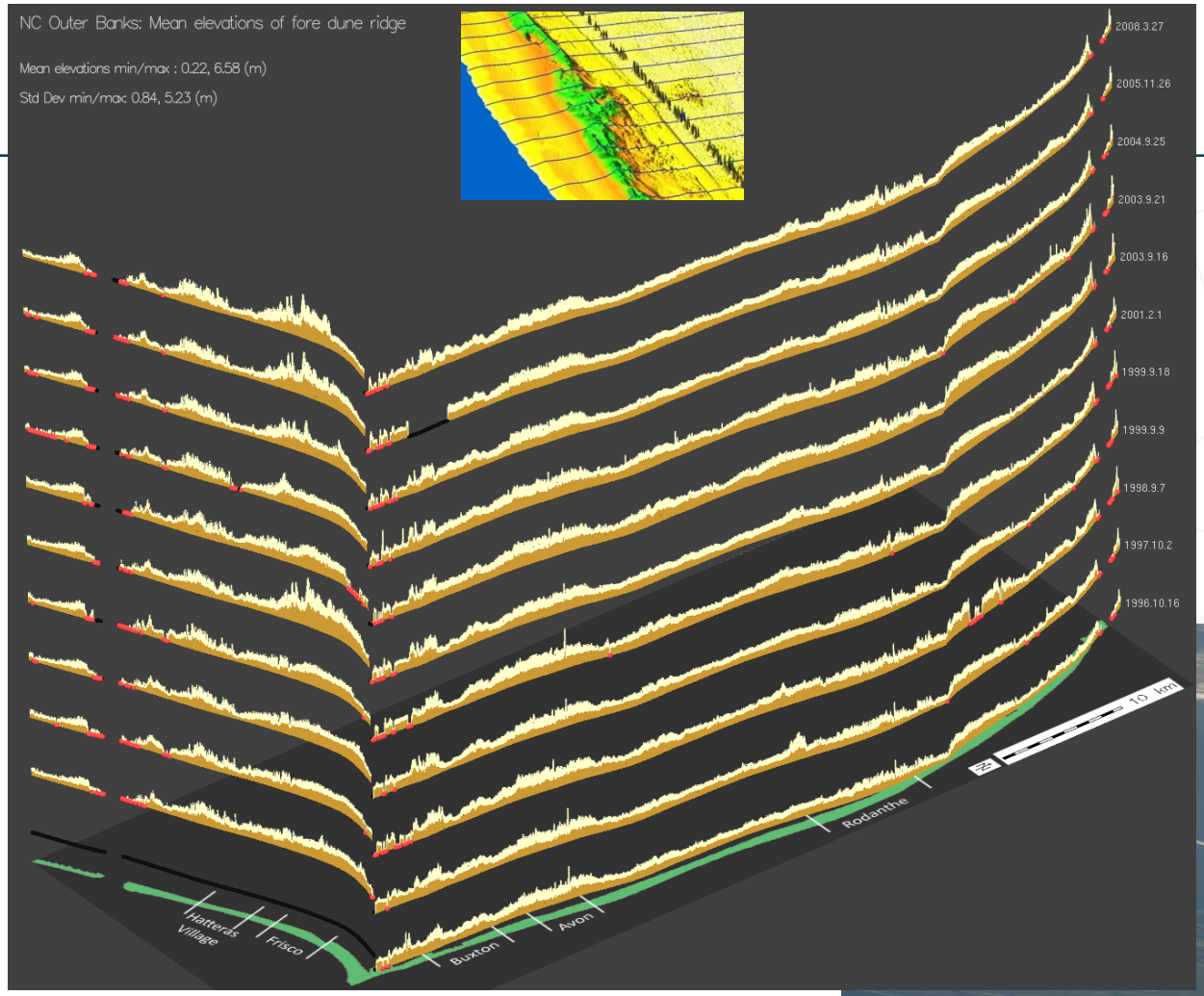


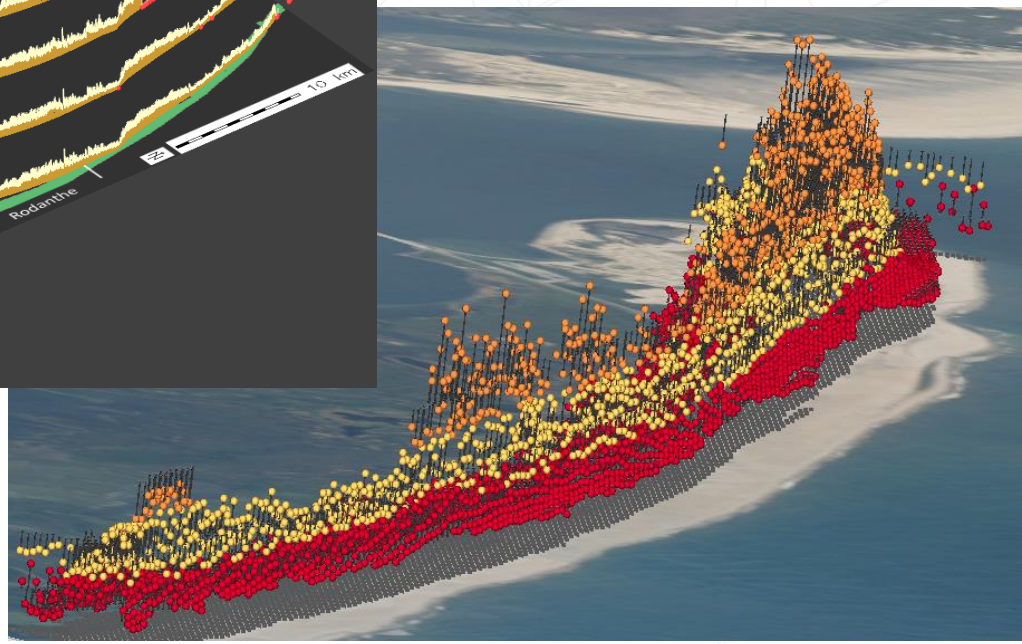
Visualization: Dr. Hong Yi, RENCI



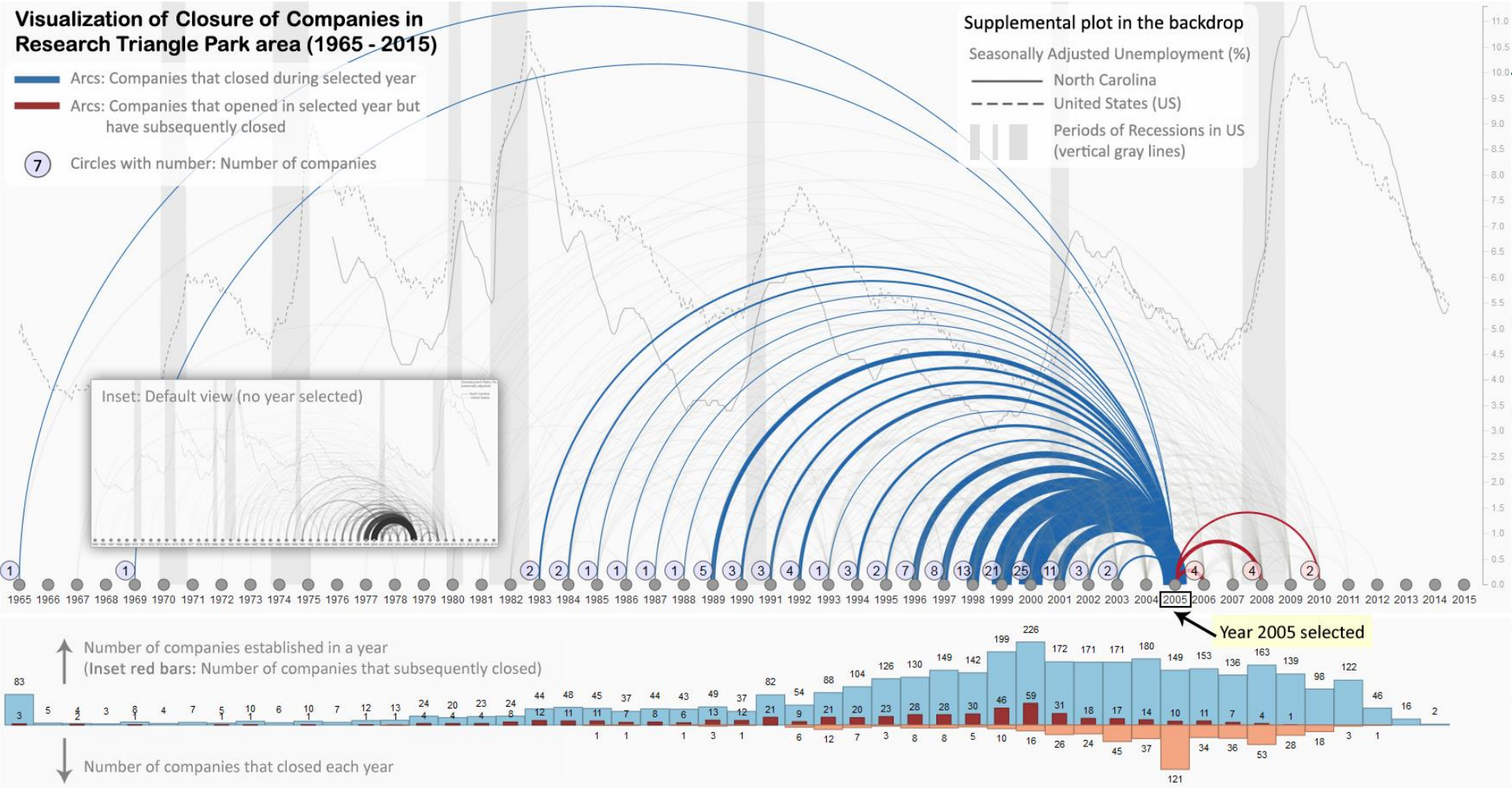


Visualizations: Dr. Sid Thakur, RENCI



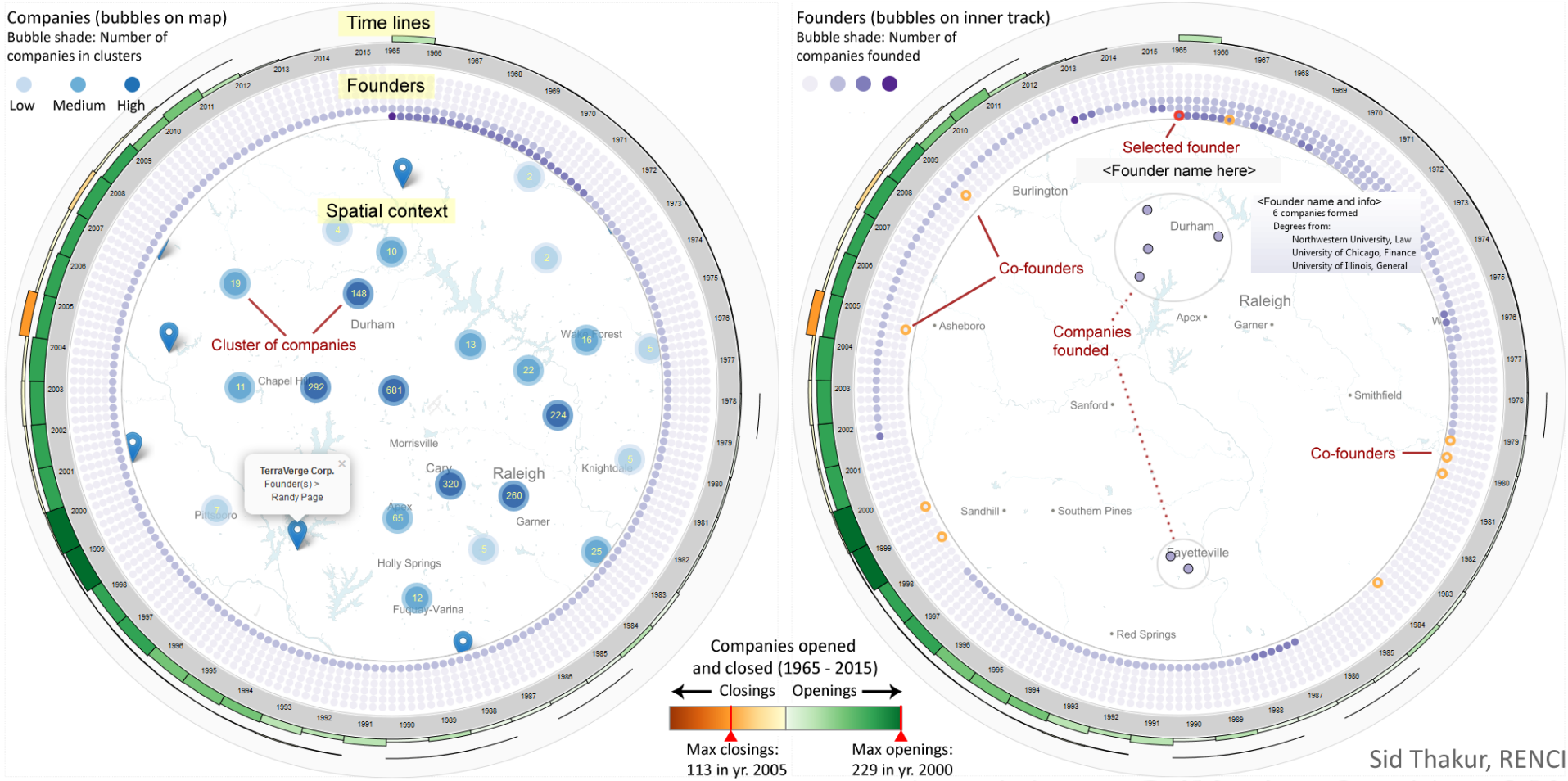


Visualizations: Dr. Sid Thakur, RENCI

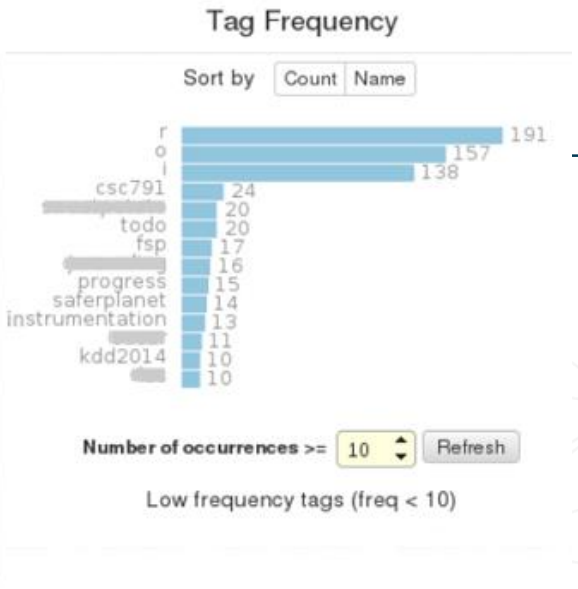


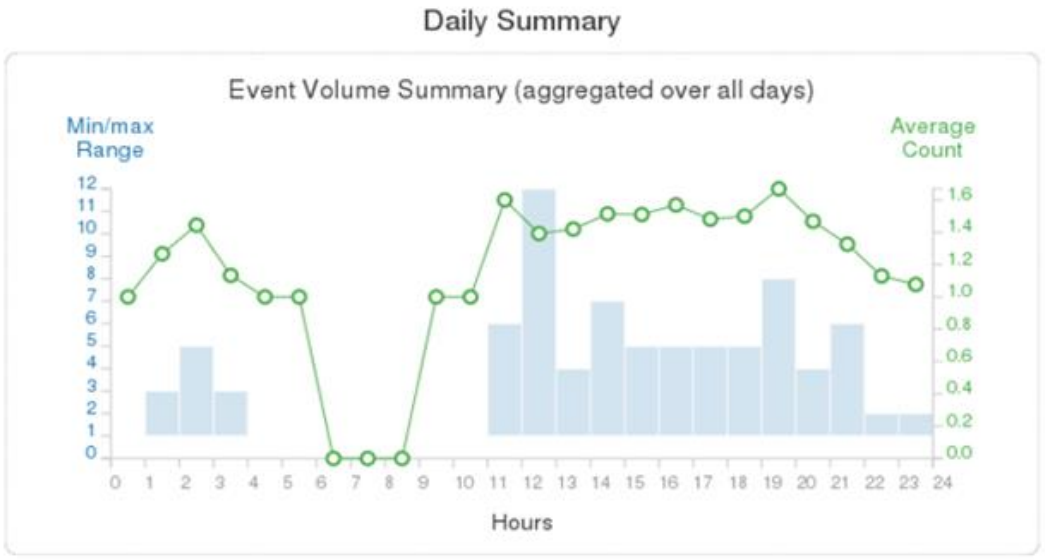
Visualization: Dr. Sid Thakur, RENCI

Time, People, Geography



Sid Thakur, RENCI



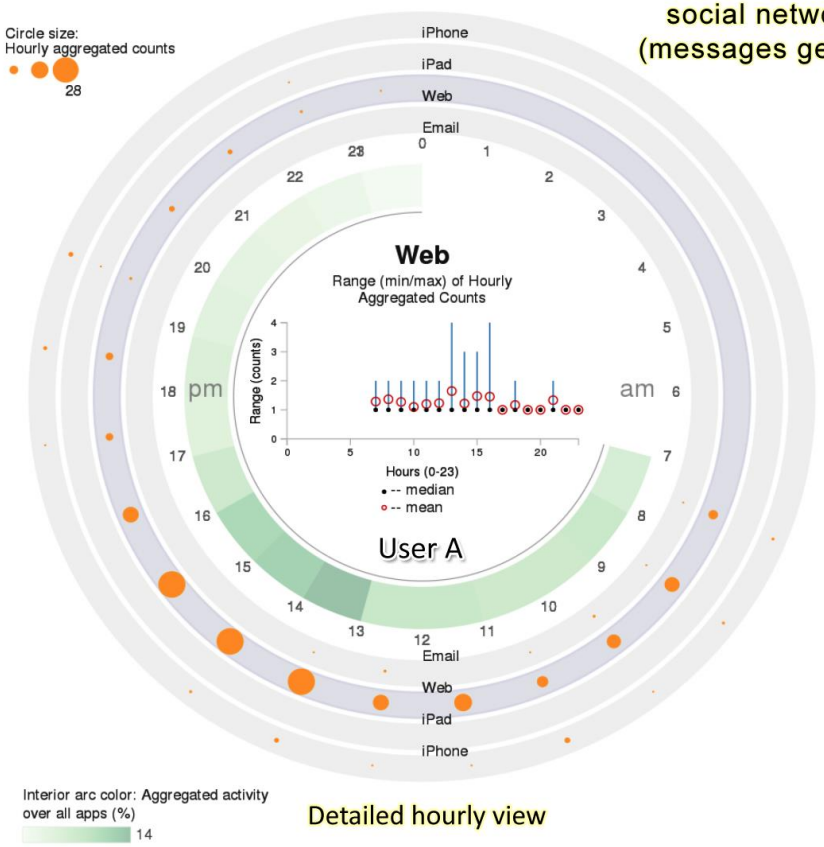


Visualization: Dr. Sid Thakur, RENCI

Plot #1

Activity of a user on a social networking site in 2015 (messages generated on Yammer)

Plot #2



Detailed hourly view

Maximum total events per time period:

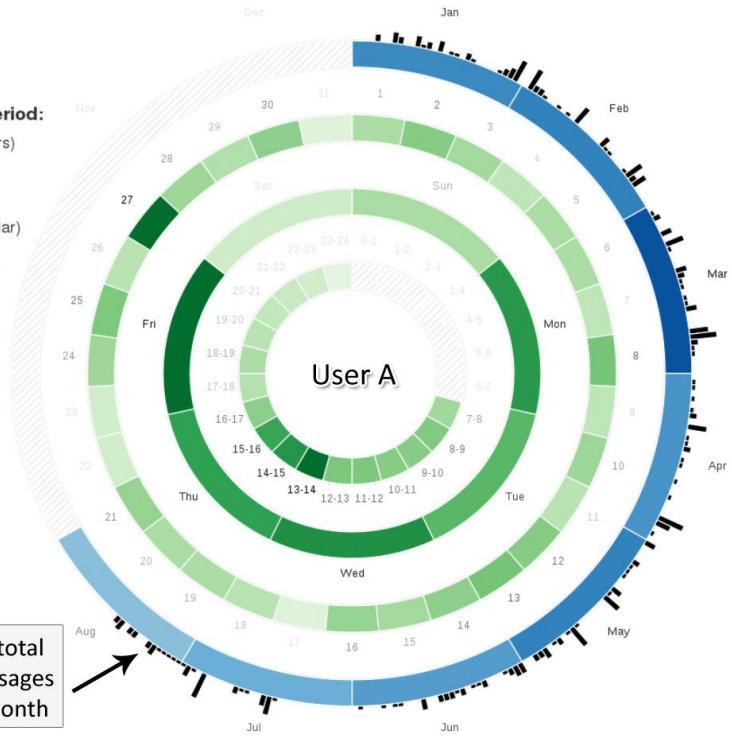
- Hour: 33 max events (during 13-12 hrs)
- Week day: 51 max events (on Fri)
- Day: 23 max events (on 27 day)
- Month: 40 max events (in month of Mar)

Maximum events per app:

- Web: 211
- iPhone: 26
- Email: 8
- iPad: 2

No data

Bars represent total number of messages each day in a month



Arc color: Aggregated number of events
Low (light) ---> High (dark)

Summary view (multiple time periods)

Visualization: Dr. Sid Thakur, RENCI