

Welcome to the Biocomplexity Institute of Virginia Tech

Information Biology - from Molecules to Policy

Christopher Barrett, PhD
Executive Director & Professor

Biocomplexity Institute Organizations

Laboratories:

MBL: Mathematical Biocomplexity Laboratory

NDSSL: Network Dynamics and Simulation Science Laboratory

NIMML: Nutritional Immunology and Molecular Medicine Laboratory

SDAL: Social and Decision Analytics Laboratory

BI Fellows: Cellular Signaling
Cell Division Processes
Circadian Processes
Epigenetics

Technical Services Centers:

Network-Centric High Performance Computing Center

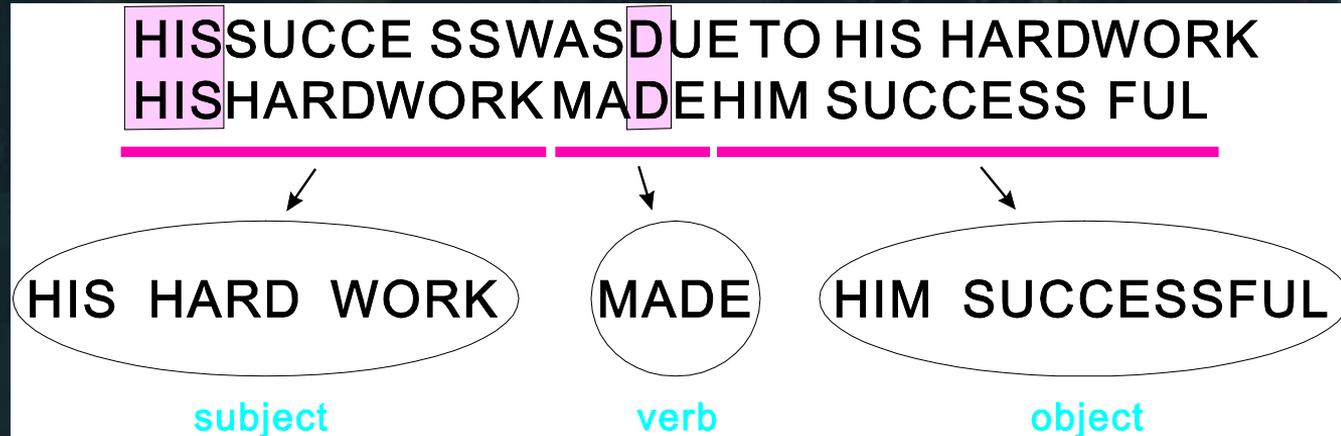
Genomics Sequencing Center



MBL: Mathematical Biocomplexity Laboratory

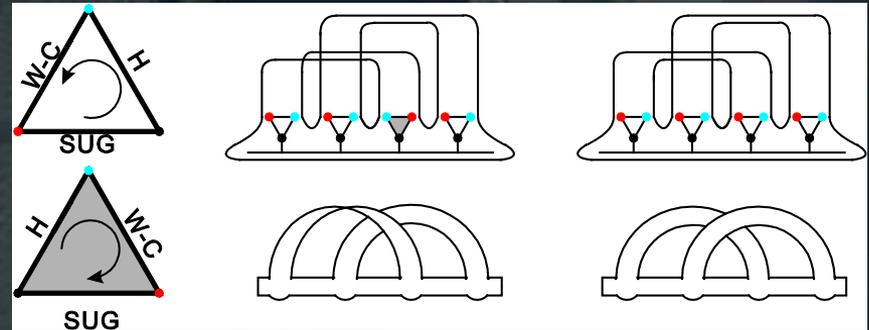
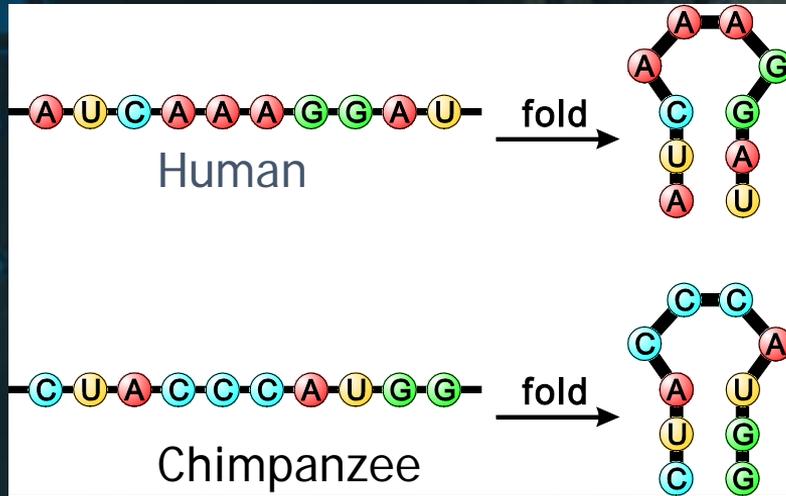
(re)Interpreting DNA data

According to research at Cambridge University, it doesn't matter what order the letters in a word are, the only important thing is that the first and last letters are at the right place. The rest can be a total mess and you can still read it without a problem. This is because we do not read every letter by itself but the word as a whole.



Toward a molecular grammar

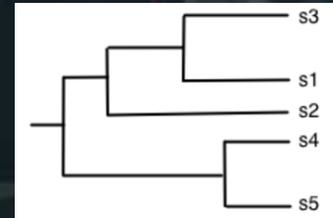
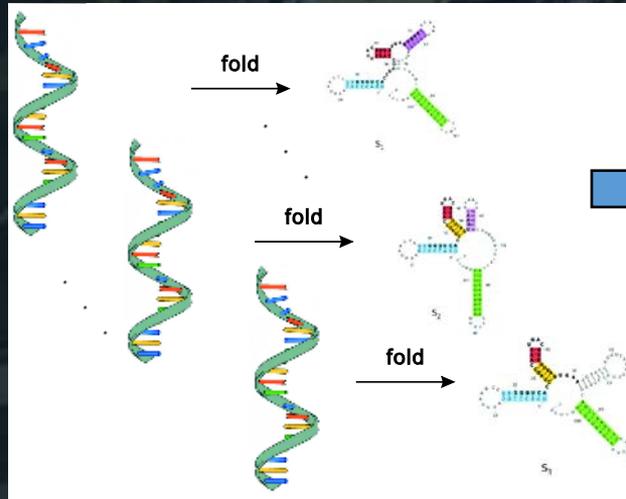
Moving beyond sequence alignment



Fat-graphs

Paradigm shift: information biology

s1: ACCGTGAAGCCAATAC
s2: ACGTGCAACCATTAC
s3: AGCGTGCAAGCCAATAC
s4: AGGGTGCCGCAATAC
s5: AGGGTGCCACAATAC

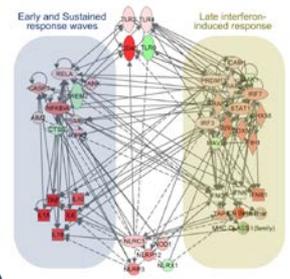
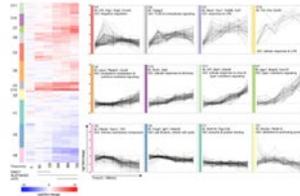


+ evolutionary
heat parameter

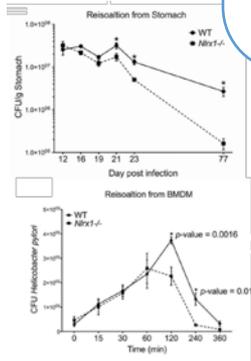
NIMML: Nutritional Immunology and Molecular Medicine Laboratory

Combining Expertise Across Disciplines

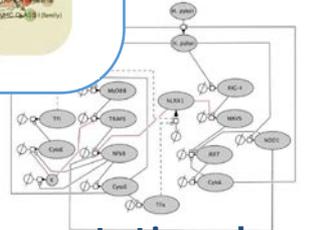
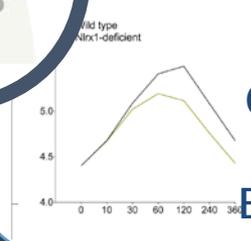
Bioinformatic Analysis



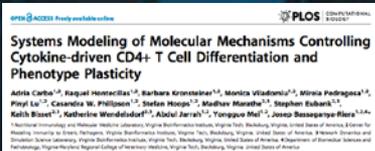
Laboratory or Clinical Validation



Computational and Mathematical Biosystem Modeling

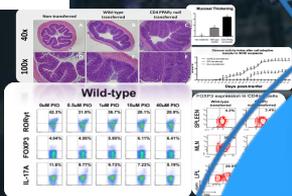


Advanced Computation for Bioscientists



Refinement of the model with new generated data

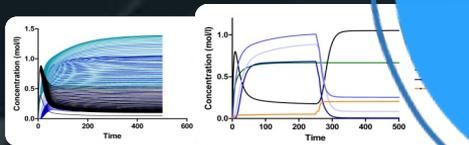
In vivo/vitro validation
Hypothesis testing



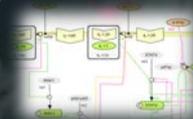
**SYSTEMS
IMMUNOLOGY
APPROACH**



Literature mining & in-house generated data



In silico experimentation
Prediction generation

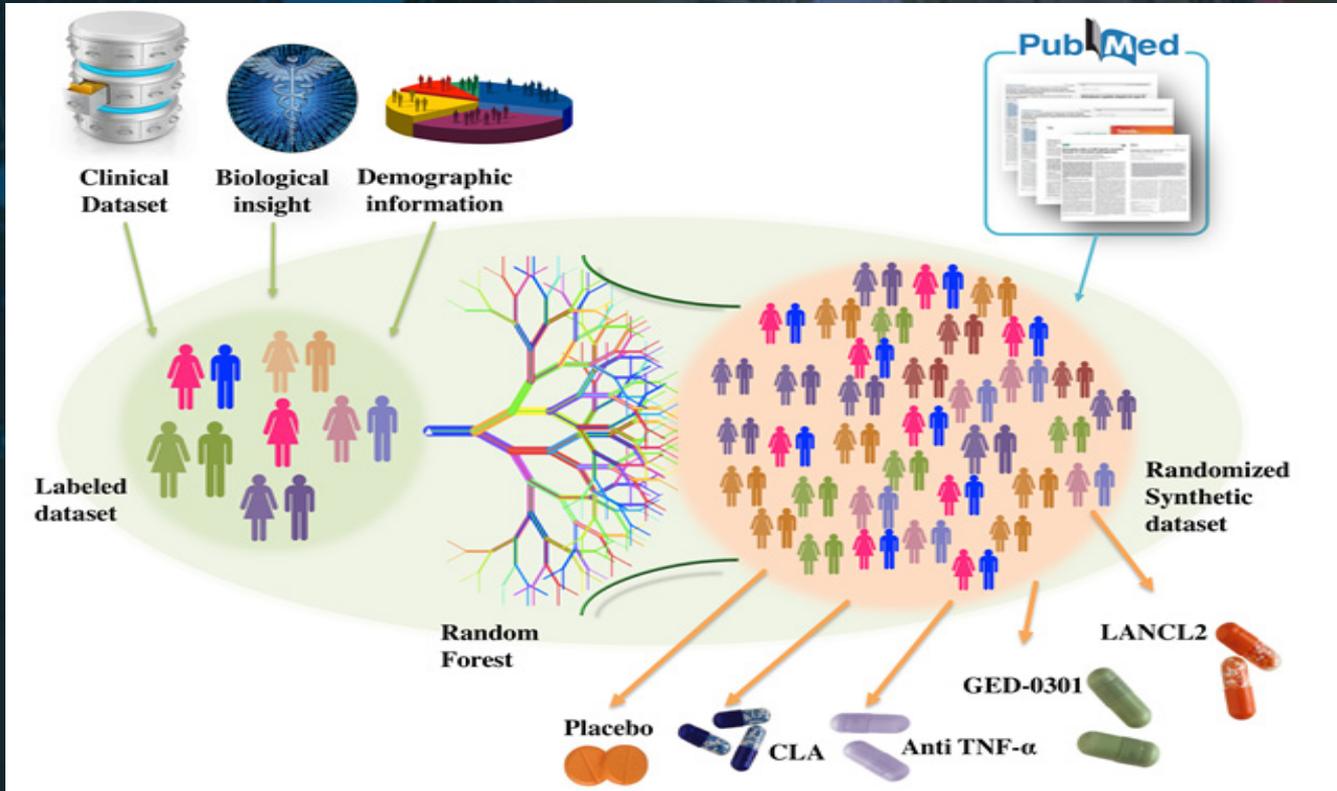


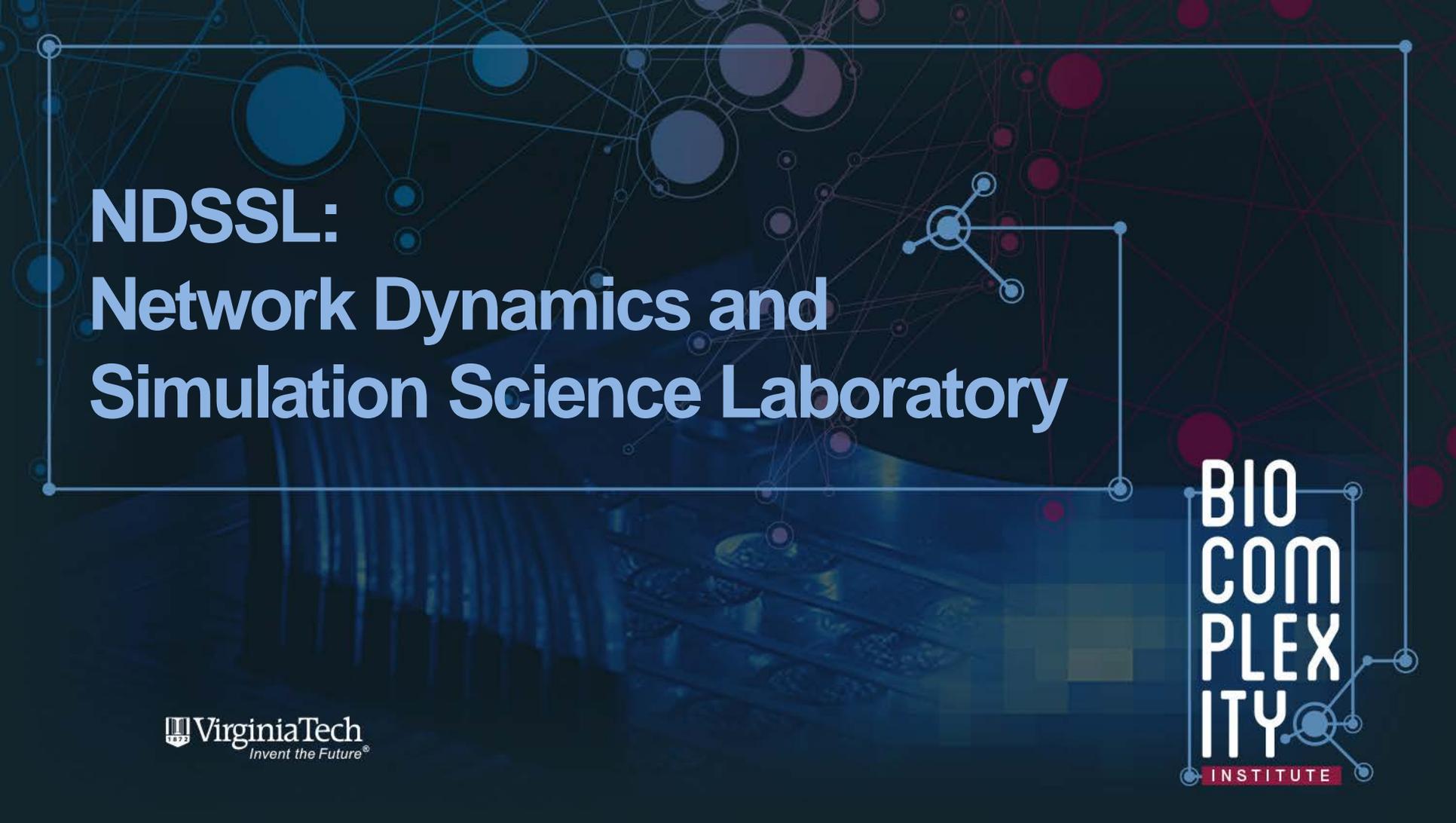
Creation of the network



Parameter estimation and ODEs adjustments

Predicting Drug Responses

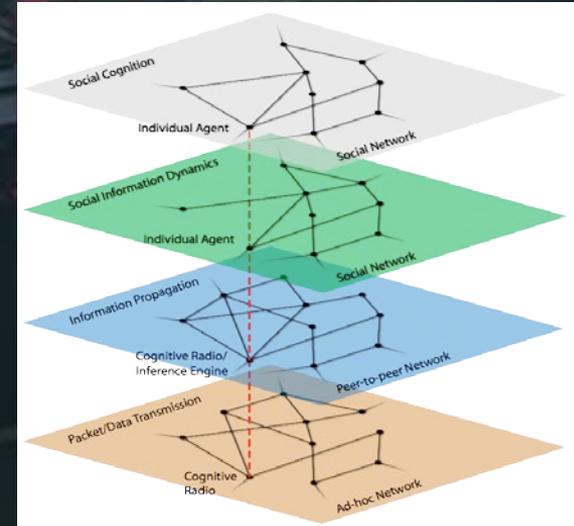


A complex network diagram with various nodes and connecting lines, overlaid on a blurred image of a computer keyboard. The nodes are represented by circles of different sizes and colors (blue, purple, red), and the lines are thin and light-colored.

NDSSL: Network Dynamics and Simulation Science Laboratory

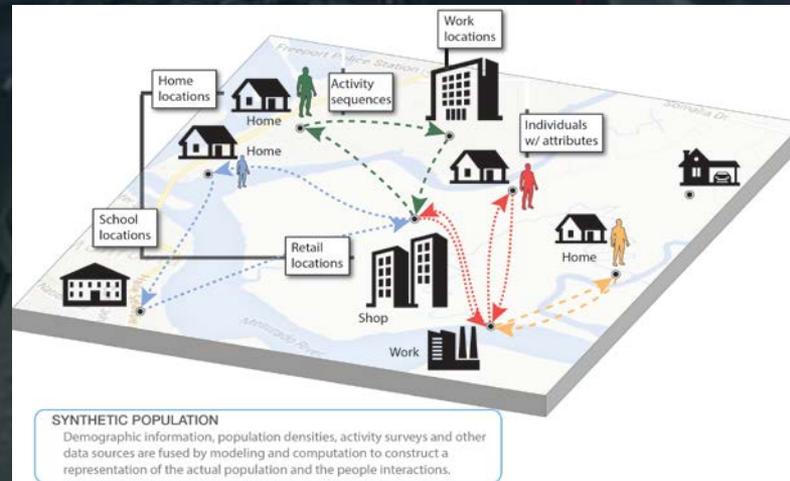
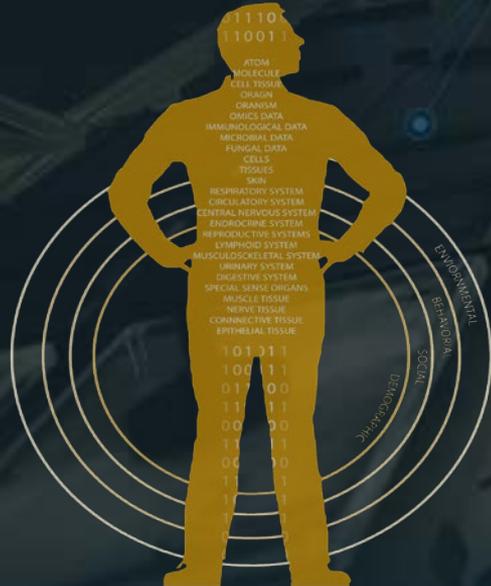
Synthetic Information Systems

- Flexible approach to organizing, updating, and using multiple sources of data
- Not “artificial” information
 - Synthesized data: multiple data sources form contextualized information resources
- Captures *interactions and relations* among entities
 - Within a certain context (information layer)
 - Between different contexts (through the layers)



Creating an *in silico* information resource

- Synthetic information library; un-encapsulated agency
- Represent individual, normative “avatars;” inherently anonymous
- Represent community dynamics incorporating detailed individuals
- Social and environmental determinants analytics



Benefits of Synthetic Information Libraries

- **Data Integration** - Synthesizes disparate data sources in a natural “coordinate system” (person, activity location, vehicle, etc.)
- **Privacy** - Maintains anonymity, but preserves important causal relationships
- Natural representation of target information sources
- Learning from others applications, etc.
- Web-delivered applications for physicians, administrators, insurers, regulators, & patients
- **Data & Service Organization** - distributed “hub and spoke”-style service architecture

Portals for “Big Data” Analytics

- Pathogen Portal:
 - 5-year DoD-funded biodefense-related research project
- Middle Atlantic Regional Center of Excellence Biodefense and Emerging Infectious Diseases:
 - 11-year NIH-funded vaccine and countermeasure development project
- Administrative Center of the Proteomics Research Centers:
 - 5-year NIH-funded support 7 Proteomics Research Centers project
- Pathosystems Resource Integration Center:
 - 15-year NIH-funded support infectious disease research project
- CIEpi:
 - 10-year infectious disease computational epidemiology project
- CINET:
 - NSF-funded Network Science project
- CNIMS:
 - 11-year DoD-funded integrated population-coupled analytics & global data platform



SDAL: Social and Decision Analytics Laboratory

The Social and Decision Analytics Laboratory

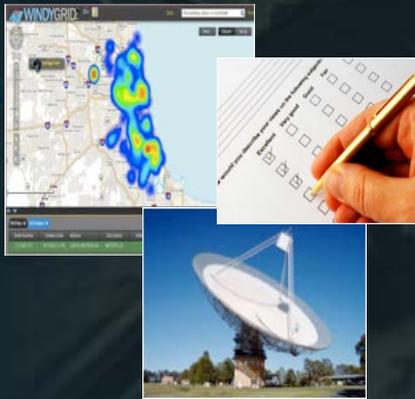
Interdisciplinary scientist teams engaged in today's ICT data revolution using evidence-based problem-driven research and developing advanced quantitative methods to inform policy decision-making.

- Science of *All Data*
- Community Learning Data-Driven Discovery
 - Education and Labor Force Analytics
 - Health and Well Being Analytics
 - Emergency Management Analytics
- Industrial Innovation Analytics
- Information Diffusion Analytics

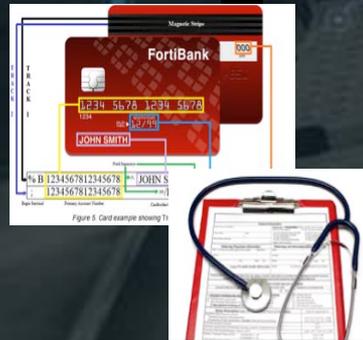
The All-Data revolution

Data integration, all types and sources, from molecules to policy

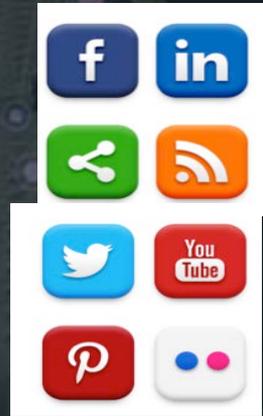
Designed Data
Collections



Administrative
Data



Opportunity
Data



Procedural
Data



Community Learning Data-Driven Discovery

Use data to serve communities,
enhancing resilience

Must be through the
lens of the community

Provide the technical pathway to
enable a community to become a
data-driven learning community



Technical Services Centers

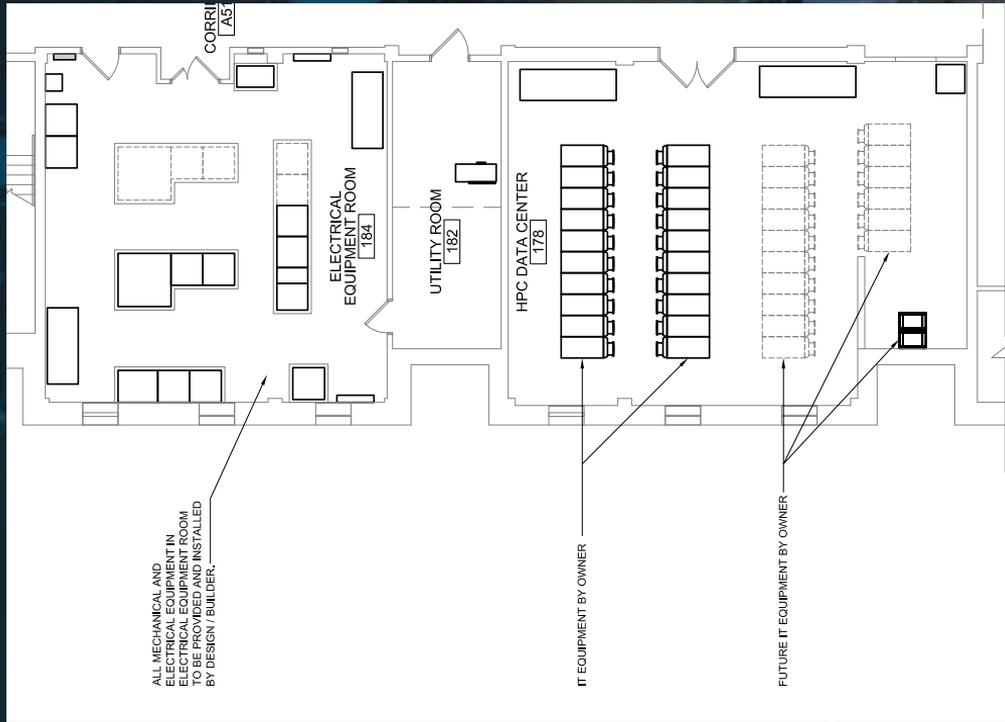
HPC: Shadowfax & Beyond



BI's current HPC capabilities:

- Five Scalable Compute Clusters
- Over 10,000 total cores: Dell, Cray, and IBM
- Over 2PB of GPFS Storage
- Will be doubled in April 2017

HPC: New Data Center



- 2 MW of new power
- Chilled water racks, support 50 kW per rack heat dissipation
- Initial 20 racks, future 36 racks
- Designed to meet our expansion needs into the future

Genomics Sequencing Center

- Illumina HiSeq, MiSeq, Thermo Fisher S5
- Automation Robots
- Quality Control Instrumentation
- Lab Info Management System
- Highly Trained and Dedicated Personnel



Current Services Offered by GSC

- **Laboratory Services Offered (Major)**
 - Next-Generation Sequencing
 - DNA whole-genome sequencing
 - RNA sequencing
 - Microbiome genome analysis
 - Sanger Sequencing
- **Training**
 - Individualized instruction and advice to faculty and graduate students
 - Consulting on grant proposals and publication submissions
 - Bioinformatics support
- **Innovation**
 - New process improvements
 - Automation for high throughput genomics
 - Adoption of new genomics technologies and methods to non-life science fields

GSC Services for FY16

