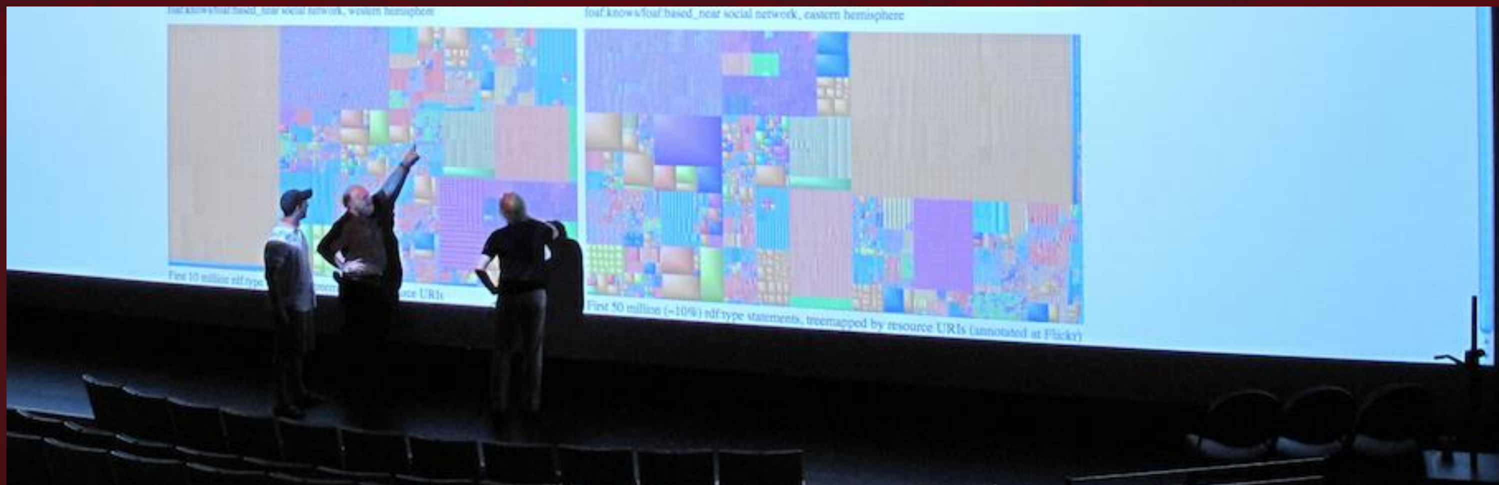
Data Stuff



The Rensselaer
Institute for Data Exploration and Applications

Prof. Jim Hendler
Tetherless World Chair of Computer, Web and Computer Sciences

Rensselaer

What I was asked to talk about

- Thoughts on ontologies, Semantic Web and data integration
  - Wither OWL: ***http://www.slideshare.net/jahendler/wither-owl***
- Modern AI meets GOFAI
  - KR in the age of Deep Learning, Watson and the Semantic Web
    - ***http://www.slideshare.net/jahendler/knowledge-representation-in-the-age-of-deep-learning-watson-and-the-semantic-we***

What I was going to talk about (Plug my book)

Tetherless World Constellation, RPI

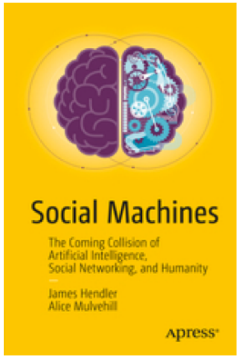
Springer

Search

Home | Subjects | Services | Products | Springer Shop | About us

» Computer Science » General Issues

© 2016



Social Machines

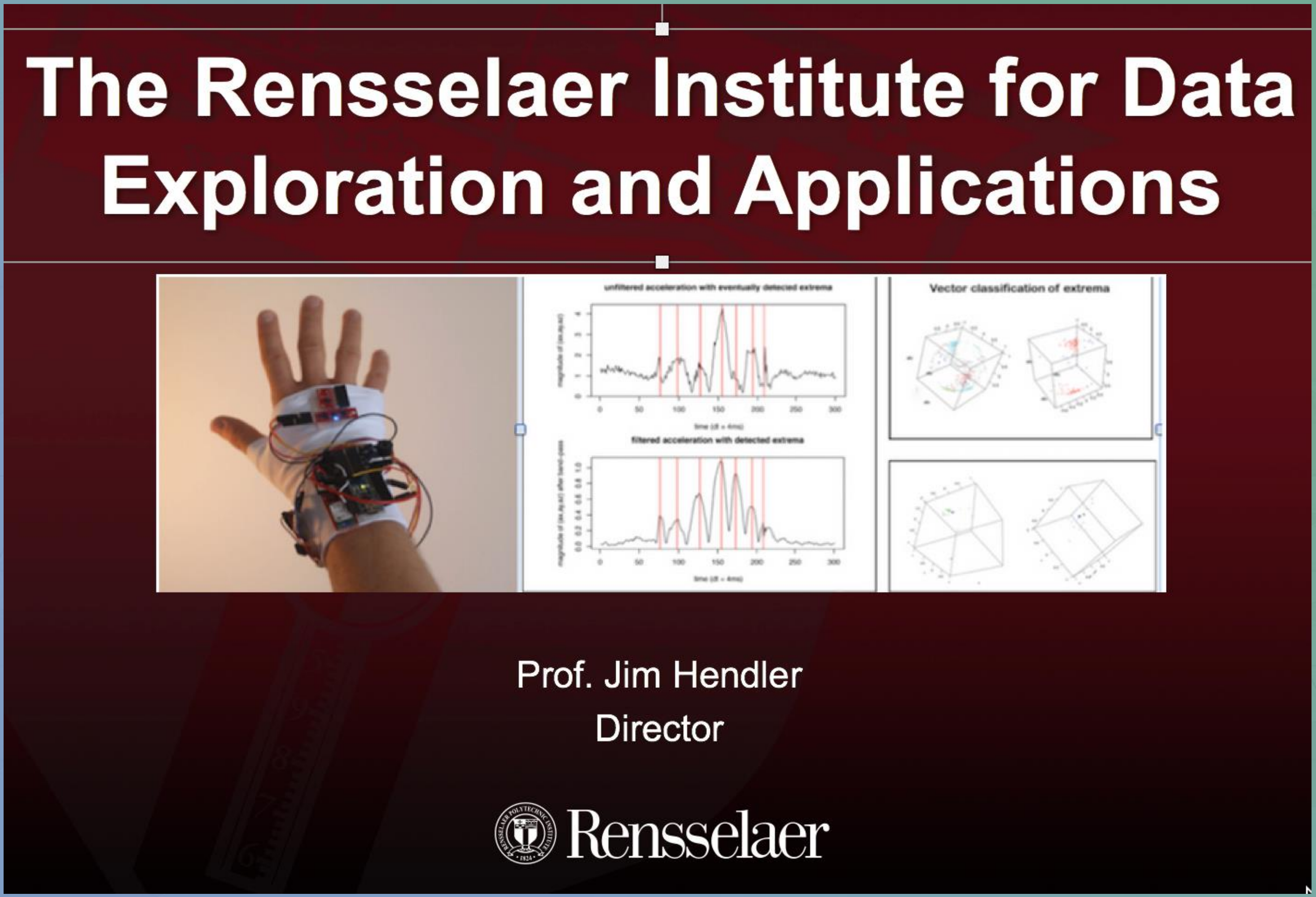
The Coming Collision of Artificial Intelligence, Social Networking, and Humanity

Authors: **Hendler**, James, **Mulvehill**, Alice

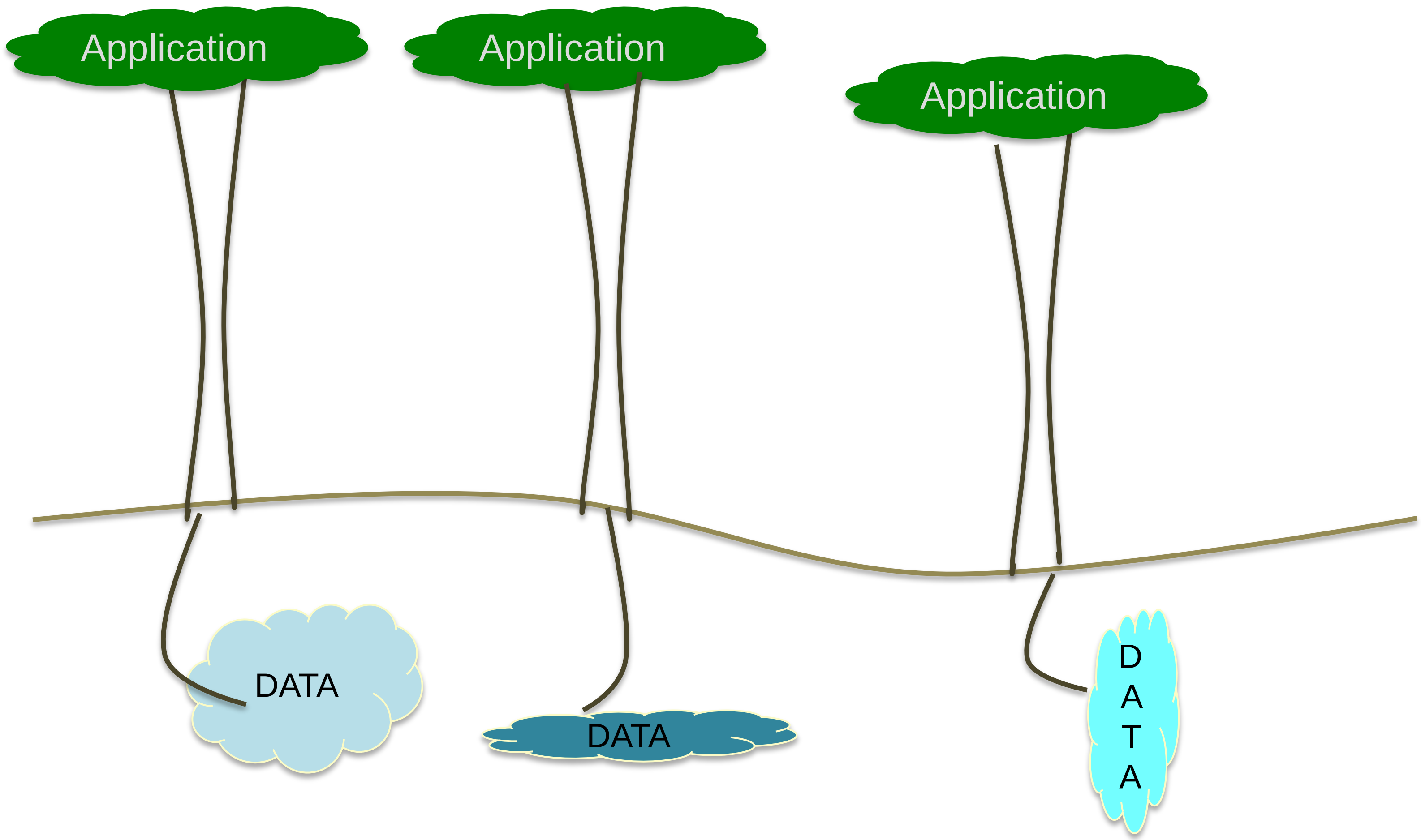
Rensselaer

What I'm going to talk about

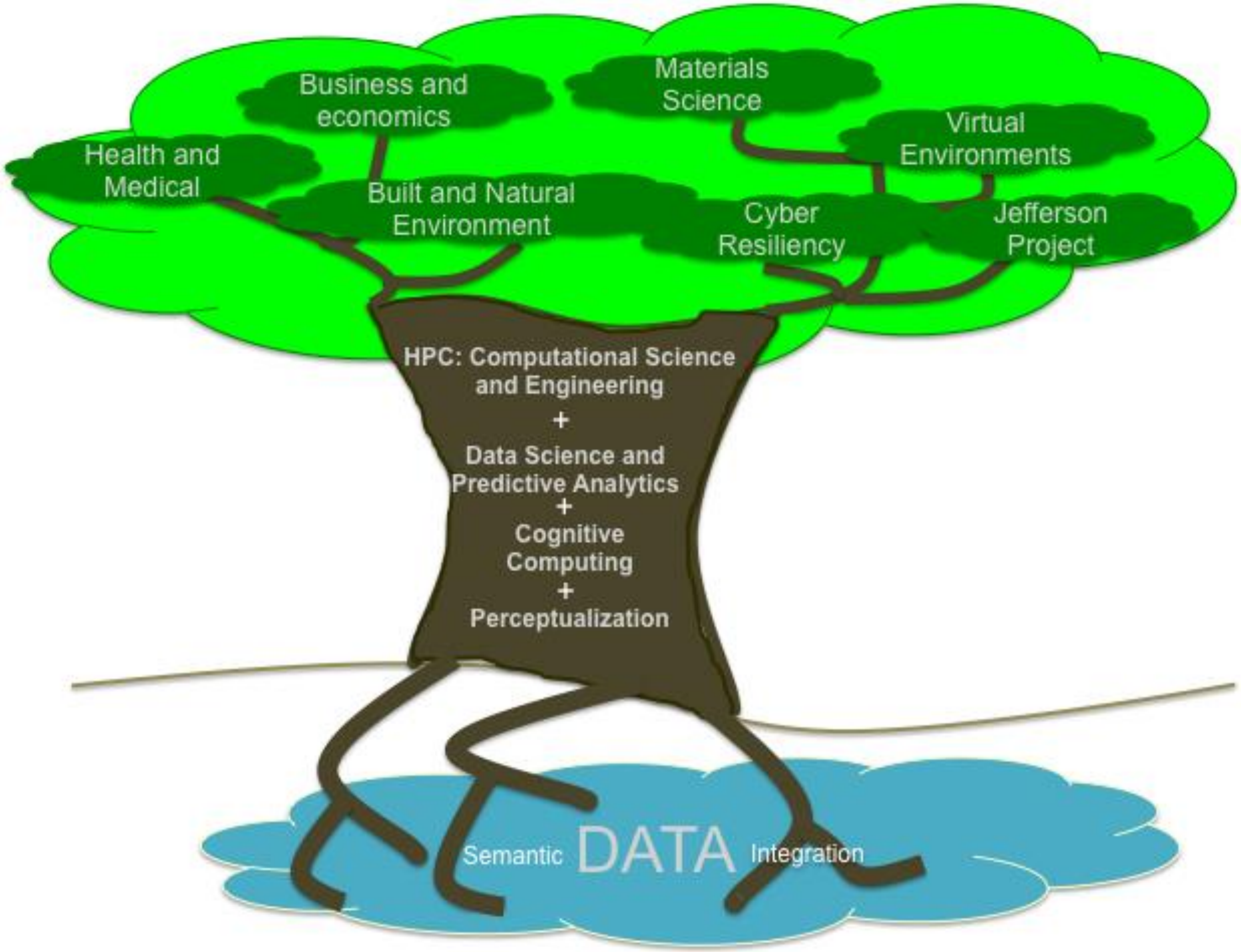
Tetherless World Constellation, RPI



What we have



What we need


Business and economics
Materials Science
Health and Medical
Built and Natural Environment
Virtual Environments
Cyber Resiliency
Jefferson Project
HPC: Computational Science and Engineering
+
Data Science and Predictive Analytics
+
Cognitive Computing
+
Perceptualization
Semantic
DATA
Integration

Discover

Use analytics to find relationships inherent in the data

Integrate

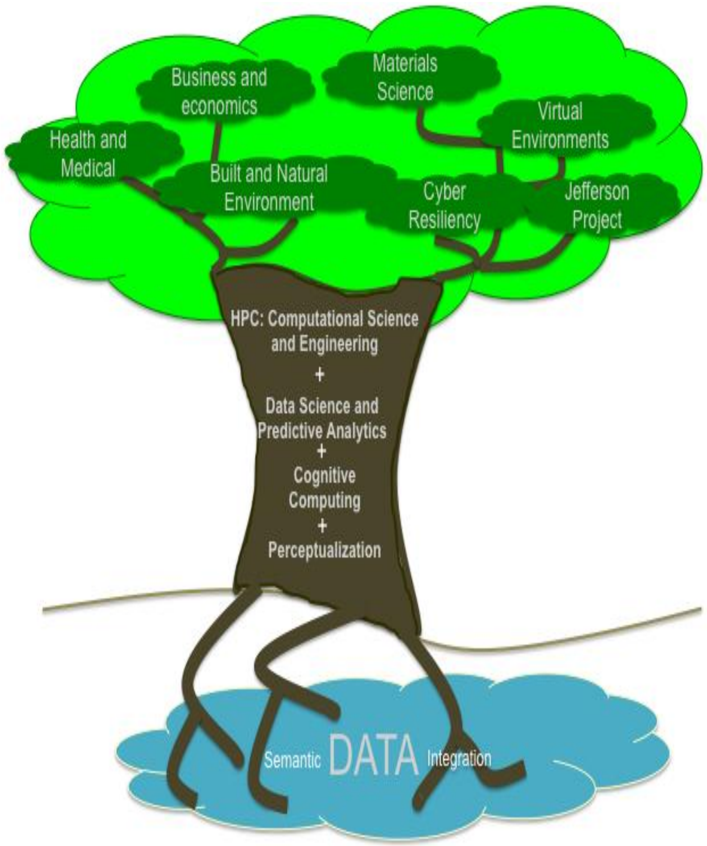
Link the relations using meaningful labels

Validate

Provide inputs to modeling and simulation systems

Explore

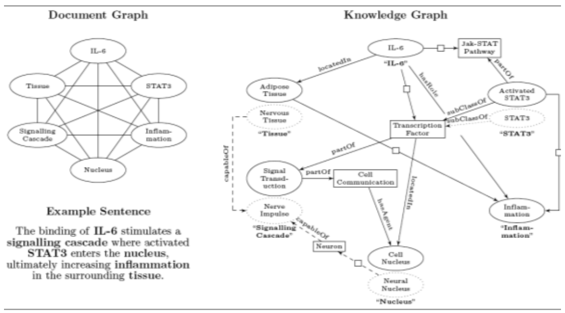
Develop multimodal approaches to turn data into actionable knowledge



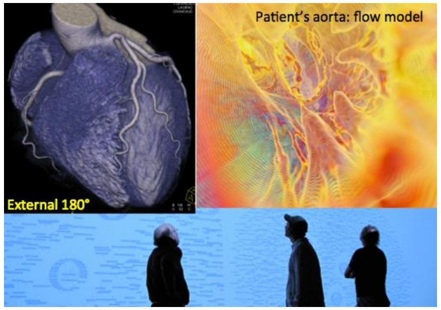
IDEA is not (just) about Big Data

We are also about the data science areas

Next-Gen Analytics & ML



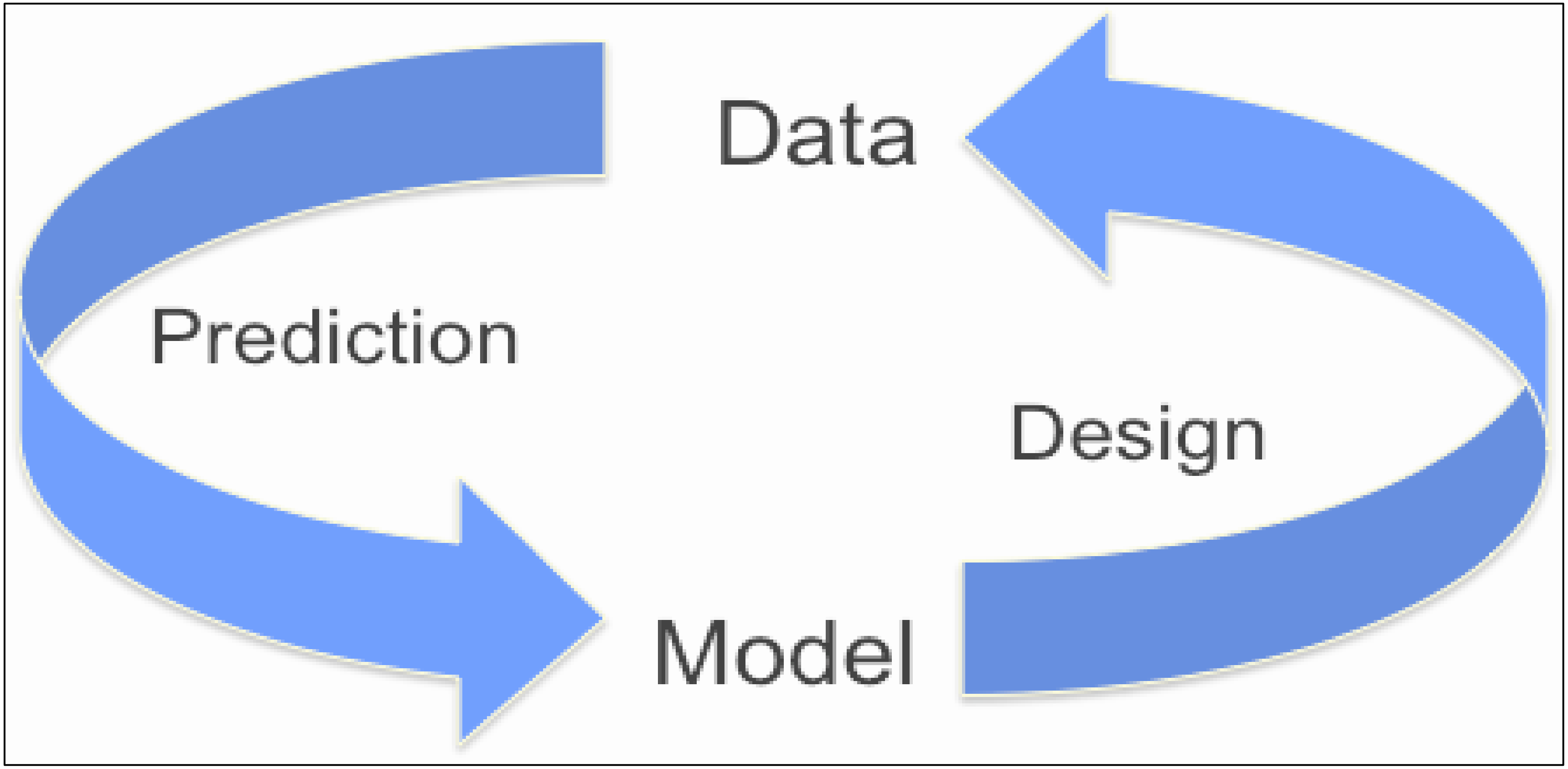
Discovery Informatics



Data Exploration

which are revolutionizing engineering, science and business with significant social impact

Data Science needs to combine correlative and causal



These capabilities are critical in "closing the loop" between data, simulation and modeling in scientific discovery, engineering design, and business innovation.

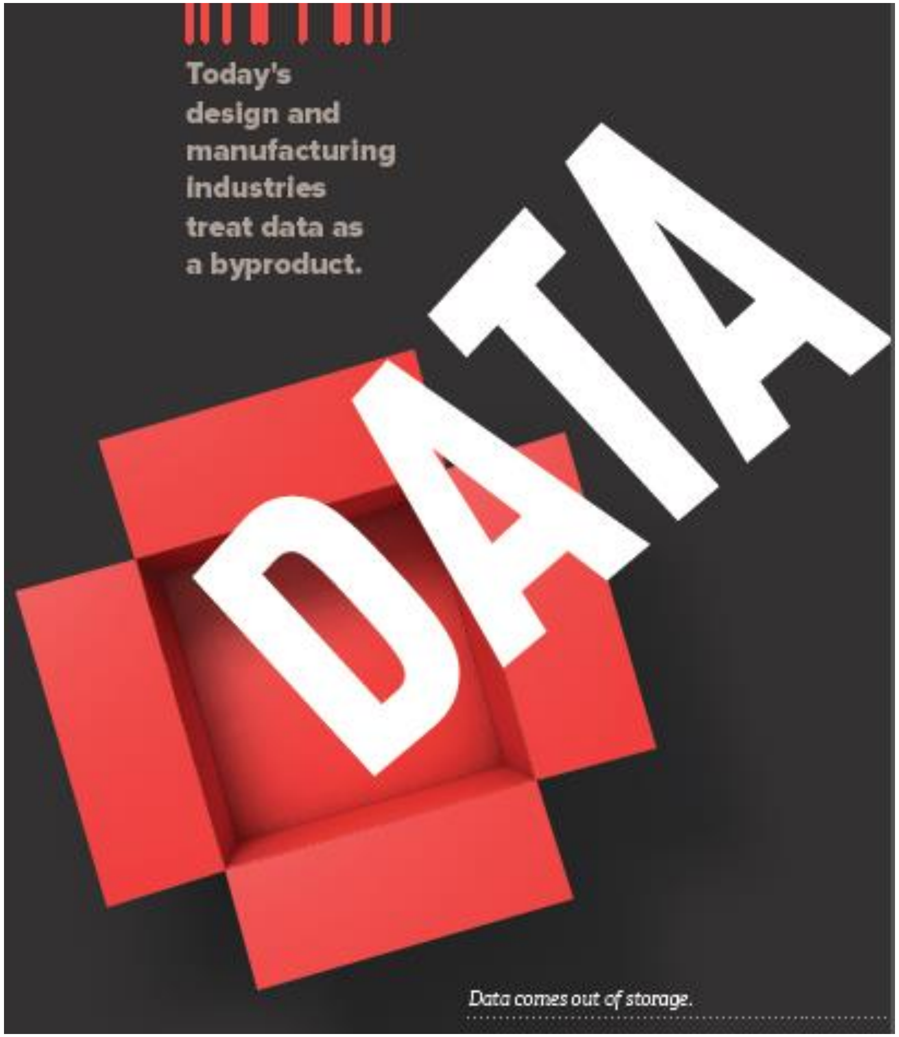
Data Analytics
Applied to
Advanced Manufacturing

By

Johnson Samuel

Assistant Professor, Mechanical Aerospace & Nuclear Engineering

Rensselaer Polytechnic Institute, Troy NY 12180



Data comes out of storage.

AND MANUFACTURING INNOVATION

Many of today's designers and manufacturers view data that's generated during the development of a new product or manufacturing technique as a mere byproduct of those processes.

As a result, only the most rarified of the data produced during design and manufacturing processes is curated in digital formats that make it accessible and meaningful. Too often, we leave potentially valuable data in a state that realizes no current or future value. Enterprises of all sizes orphan important data on the shop floor.

This is a lost opportunity. A sufficiently rich data set that is fully accessible enables designers to discover previous processes and leads— including false starts and dead ends— that could develop into new solutions. Rather than throwing out this valuable data or leaving it in inaccessible forms, industry, researchers, and others may soon be able to use tools to explore this information and amplify their intelligence and experiences.

Before we can get to that point, though, we have to rethink the relationship between data and manufacturing innovation. We will have to understand that data is the central and most essential product of engineering design activity.

By William Regli

"DATA IS NOT A MERE PRODUCT OF PRODUCT LIFECYCLE ACTIVITIES— IT IS DATA THAT GIVES RISE TO THESE ACTIVITIES IN THE FIRST PLACE."

Keeping data isn't enough.

Data can be meticulously archived but also rendered utterly useless. For instance, it could be kept on paper or in an analog data format such as old Applicon files printed to aperture-style punch cards. Digital data stored in unsupported storage technology, such as tapes or floppies, is just as inaccessible. Digital data could be in a lossy or derivative format, such as a 3-D CAD drawing archived as a 2-D PDF, or it could lack the context or metadata to make it discoverable.

Another element usually missing from stored data is the thought process behind its creation. Design produces many branches that—as a collection—can be valuable; yet those design decisions, explorations, R&D tests, and alternative analyses are typically discarded.

While industry decision makers recognize that product and manufacturing data is important, they often lack an understanding of what constitutes product-related data and the actual value of that data.

For instance, industry today is rapidly adopting something called the

Mechanical Engineering magazine, Vol 138, No. 9, September 2016, ASME.

Dr. William Regli
Defense Sciences Office (DSO)
Deputy Director

Rensselaer

Rethinking data vs. innovation

Dr. William Regli
Defense Sciences Office (DSO)
Deputy Director

" Many of today's designers and manufacturers ***view data that's generated during the development of a new product or manufacturing technique as a mere byproduct of those processes.***

*.....**we have to rethink the relationship between data and manufacturing innovation**.*

*We will have to understand that **data is the central and most essential product of engineering design activity**."*

"Transformation of design & manufacturing into information-centric disciplines."

Metal-based AM: State-of-the-field

Rastered laser beam sinters/consolidates metal powder to create high resolution structural parts

Source: DARPA: Open Manufacturing

- **(Metal) AM systems** are typically "closed" – limited control
- Expensive systems (min $750 k), no modularity, **lack of open knowledge base.**
- The technology is at a nascent stage with few "turn-key" systems.

Metal Additive Manufacturing Process

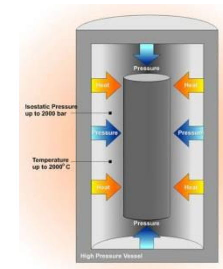
Material Properties

Titanium Alloy

Desired Geometry

CAD Model

Process Parameters



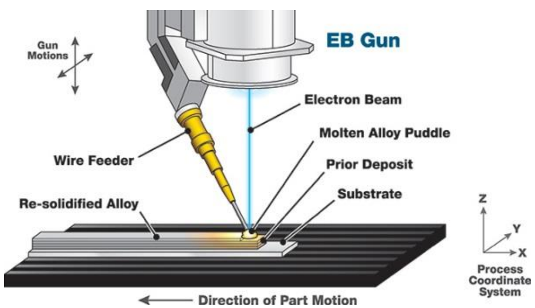
Process Parameters

E-Beam Deposition

Intermediate Product

Hot Isostatic Pressing (HIP)

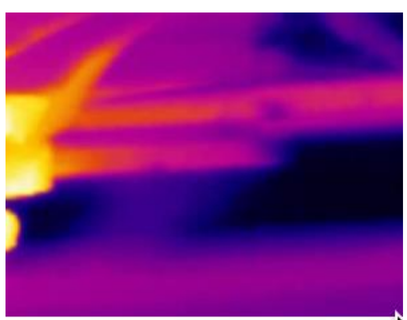
Final Product



CT Scans
Images
Chemistry
Mechanical Props.

.stl/.prt files

CT Scans
Images
Chemistry
Mechanical props



Equipment Specifications

Processing Route

Extrinsic Data

Intrinsic Data

Manufacturing Data Problem

– DARPA Open Manufacturing Performers **(Honeywell, Lockheed Martin, Boeing etc.)** generated TBs of metal AM **process, testing and characterization** data.

– Data management requirements (Materials Genome Initiative)

– Over a period of time…..DARPA's data server looks like this

www.existentialennui.com

"Good data"
but
Little use in its current form !

Rensselaer

Relevant Questions

1. How do we create **meaningful visualizations** of this data ?
2. Can we find **meaningful interrelations between the data sets**?
3. If so - can we do machine learning and **make prediction in domains where the tests have not been conducted ?**

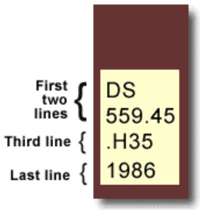
More Fundamentally

Can data-driven analytics

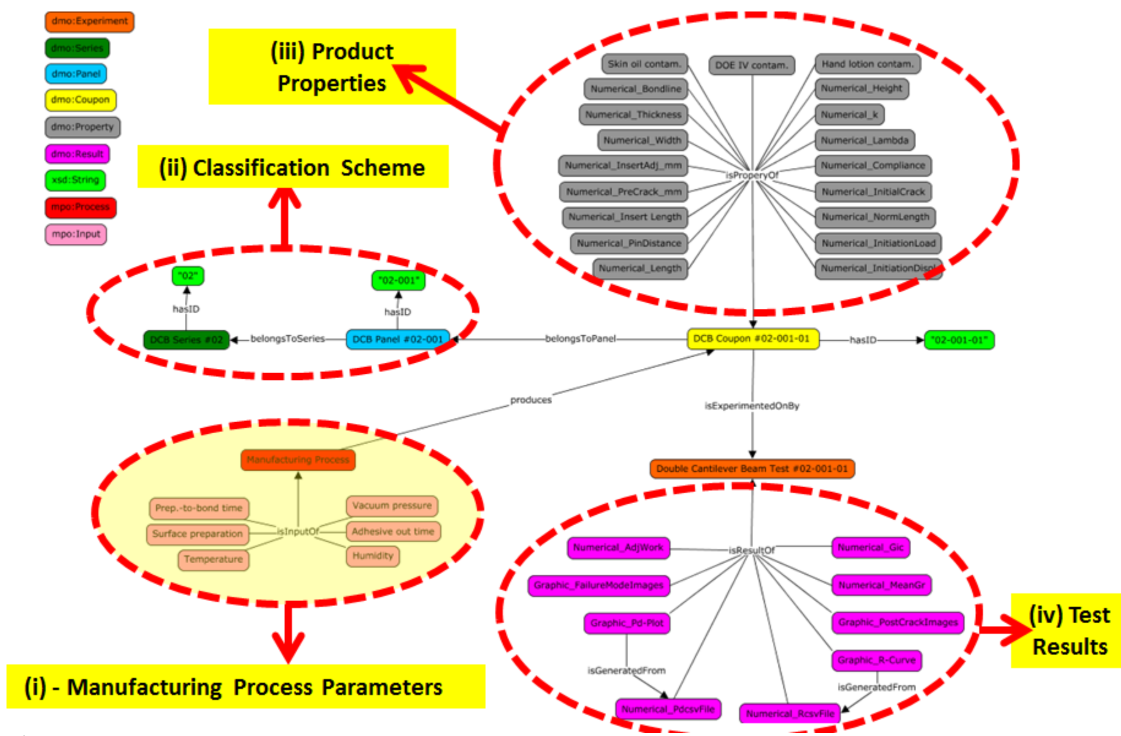
1. **Enable process planning and part qualification for metals ?**
   - **Biggest bottle neck in the "democratization" of AM**
2. **Enable the creation of processing recipes for functionally-graded AM**
   - **"Programmed" metal microstructures**

Our Approach

Step 1: "Pick up the books" → Drill into the data files



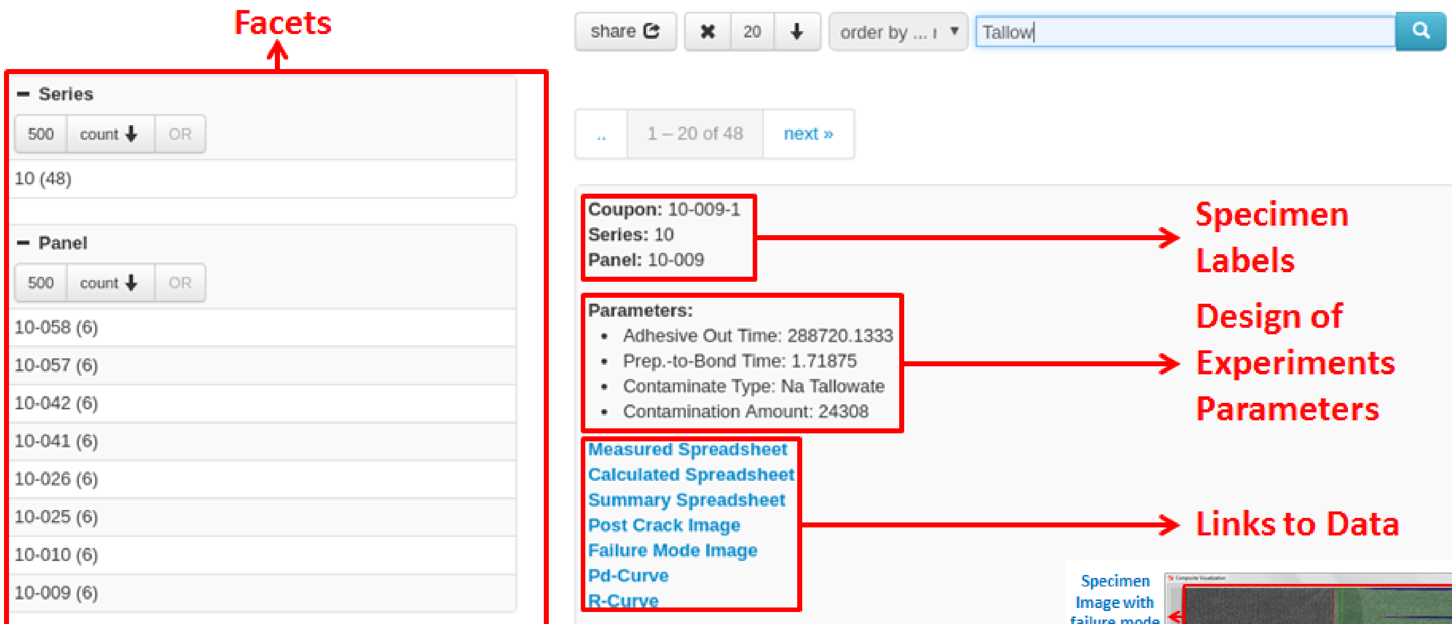
Step 2: "Develop basic Dewey decimal system" → Use domain expertise to realize *"functional ontologies"* to anchor the data sets.

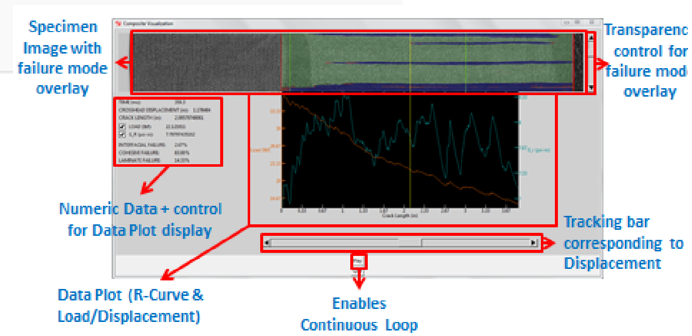


Our Approach

Step 3: “What Type of Display Case ? ”

- Faceted search-based visualization of data
- Meaningful interaction with data





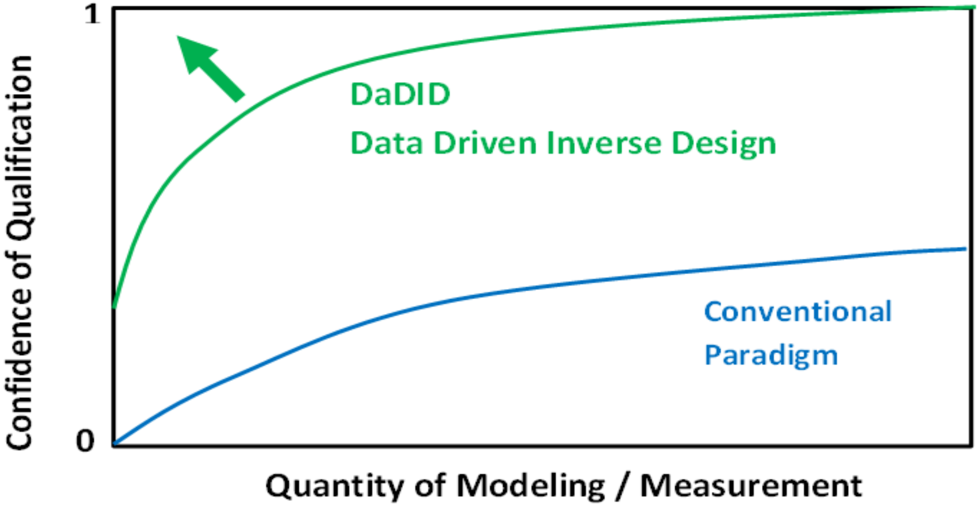
Rensselaer

Our Approach

Step 4: "Read & Discover New Knowledge"

- Apply machine learning on the data sets.
- Train & then **Predict for untested conditions.**

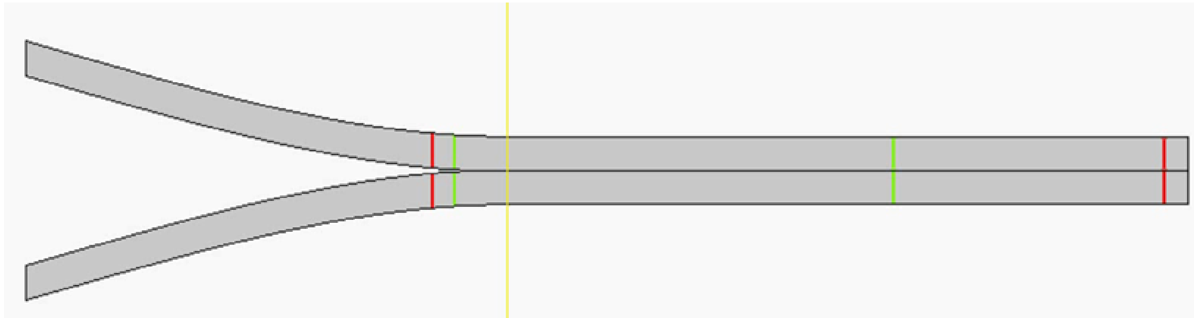
Grand Vision: Data-driven Inverse Design for AM Part Qualification Paradigm



Rensselaer

Machine Learning Example

(Composites Testing Data)



Objective: Classify **majority failure modes** ***(interfacial/cohesive)*** based **on input parameters** ***(Surface Preparation, Contaminate Type, Contaminate Amount)***

- Data set (n=562) **randomly partitioned into training set (n=395)** and **test set (n=167).** Each trial partitions the data differently.

Typical validation output (confusion matrix) from a single trial. Green cells are correct predictions. Gray cells are incorrect predictions

		Predicted	
		Interfacial	Cohesive
Actual	Interfacial	82	8
	Cohesive	15	62
		Correct	0.86

Machine Learning Predictions: Untested Parameter Combinations

n=28 combinations of parameters for which there was no data were chosen and run through Bootstrap Aggregating model

n	Surface Preparation	Contaminate Type	Contaminate Amount	Failure Mode Predictions
1	XX	XX	XX	**Cohesive**
2	XX	XX	XX	**Cohesive**
3	XX	XX	XX	**Cohesive**
4	XX	XX	XX	**Interfacial**
5	XX	XX	XX	**Interfacial**
6	XX	XX	XX	**Interfacial**
7	XX	XX	XX	**Cohesive**
8	XX	XX	XX	**Cohesive**

Predictions can be verified through future experimentation

Rensselaer

Interdisciplinary Team

Center for Materials, Devices, and Integrated Systems

Institute for Data Exploration and Applications

Dr. Robert Hull

Expertise:
Materials Genomics, Materials Processing, Measurement & Control

Dr. Johnson Samuel

Expertise:
Additive manufacturing: Process development, Process planning

Dr. Peter Fox

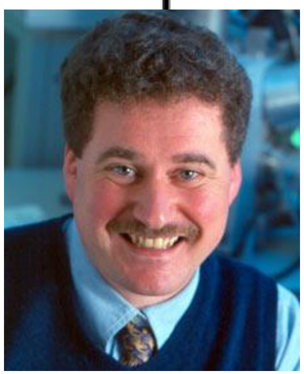
Expertise:
Data Science, Materials Informatics, Semantic eScience

Dr. James Hendler

Expertise:
Semantic Web, RDF triple stores, Supercomputing

Dr. Deborah McGuinness

Expertise:
Ontologies, Semantic Data Integration, Linked Data, Semantic eScience, Explanation

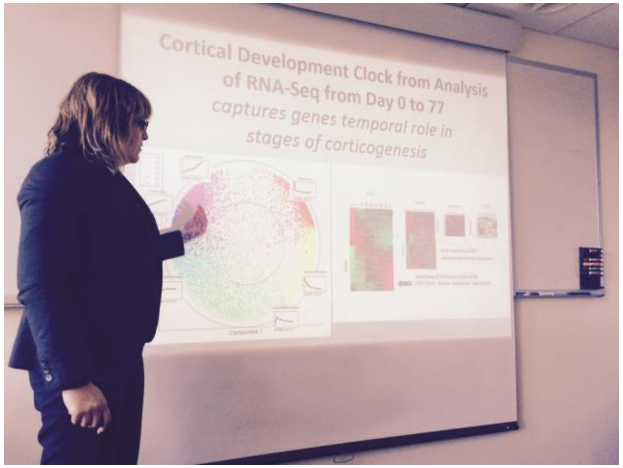


- **Dr. Bryan Chu (Post-doc)**
- **Graduate students: Congrui Li, Greg Echeverria , Charles Parslow**

Using Human Perception to deduce patterns in data

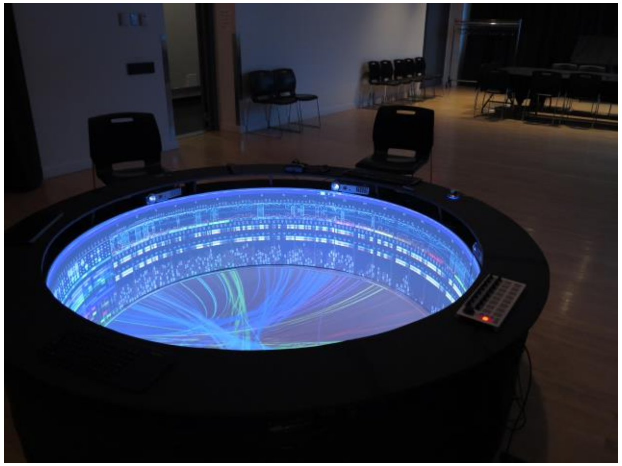
Data Exploration is an important direction

- Visualization techniques coupled with data analytics has major potential
  - Especially for collaborative exploration of complex data
- For example, “Campfire” gives IDEA a unique platform well-suited to “radial” visualizations used heavily in analytics



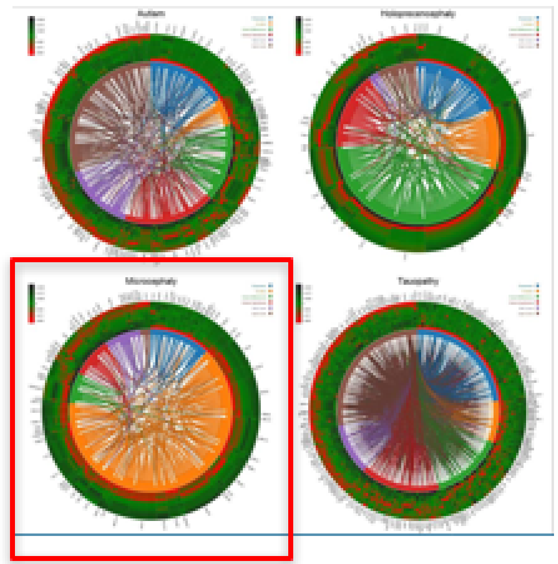
From this

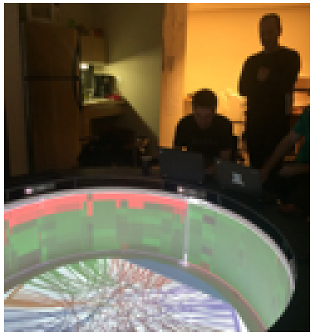
To this



More significant results require mutliple datasets (remember DIVE)

- **IDEA is developing data technologies for revolutionary improvement in global children's health**
  - Rensselaer tool for analyzing Corticogenesis (brain development) identifies windows of susceptibility
    - Team Led by Deborah McGuinness and Kristin Bennett
      - Freshman Hannah de Los Santos developed clustering technique in summer program
      - Junior Matt Pogel produced campfire version
  - Zika virus causing "microcephaly" in newborns has unusual window of susceptibility
    - Microcephaly gene expression pattern recognized in Campfire session
  - Result verified w/colleague at Neural Stem Cell Institute (2/16)





Transformative Educational Impact

Develop Data Dexterity in *Every* Rensselaer Student

- Data Dexterity: Institute Wide Initiative (Lead: Prof. K. Bennett, Assoc. Dir. IDEA)
    - Data Awareness core curriculum for *all* undergraduates
        - Require data-intensive courses for all students
        - Add concentrations, certificates, minors to many of our majors
    - Building interdisciplinary courses and programs
        - *eg.* courses launched in: data ethics, cognitive computing, Big Data projects
        - *eg.* digital ethnography project, data analytics masters, Increased campus participation in Production/Installation/Presentation (PIP) program
    - Data Interdisciplinary Challenge Intelligent Technology Exploration (Data-INCITE) Laboratory
        - Work directly with established and emerging companies
        - Students do real projects on real data (outcomes unknown)
    - Create data-related coop/internship opportunities
        - Benefit to corporate partners and to our students

The Rensselaer Institute for Data Exploration and Applications

* Developing and expanding Rensselaer's research strength in data science
* Exploring new directions in pedagogical innovation
* Creating new opportunities for cross-disciplinary research
* Building new partnerships for internships and off-campus cooperative learning