

Big Data @ UC Davis

A Data Science Initiative



Patrice KOEHL, UC Davis

*Department of Computer Science
Genome Center*



The three I's of Big Data

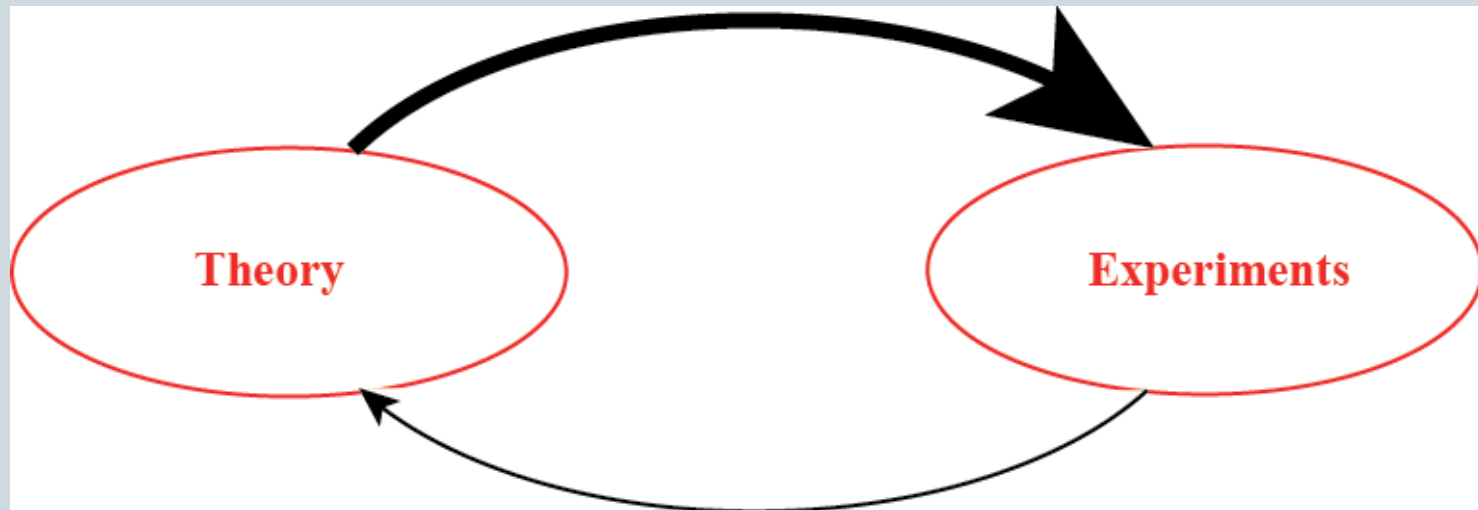


Big Data:

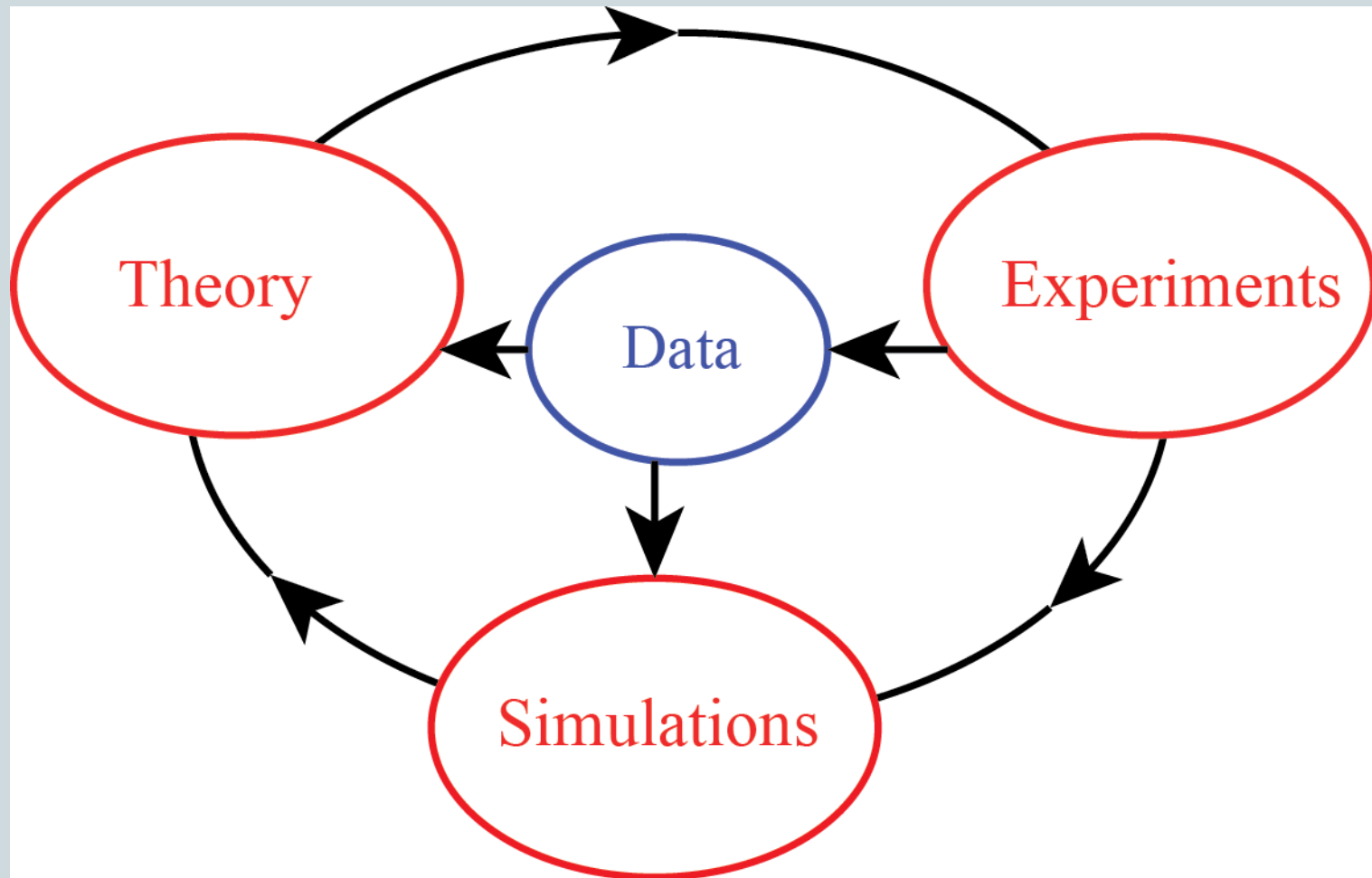
- **Immediate** (we need to do something about it now)
- **Intimidating** (what if we don't)
- **Ill-defined** (what is it?)

(loosely adapted from Forbes)

Biology Today



Biology Today



Big Data: Challenges



- Volume and Velocity
- Diversity
 - Structured, Unstructured...
 - Images, Sound, Numbers, Tables,...
- Security
- Reliability, Integrity, Validity

Big Data: Opportunities



- Fourth Paradigm: data driven science



- Holistic approaches to major research efforts
- From genotype to phenotype
- Personalized medicine

Handling Big Data



Collecting

*Retrieval
And
Management*

Learning

*Mining
For
Knowledge
Extraction*

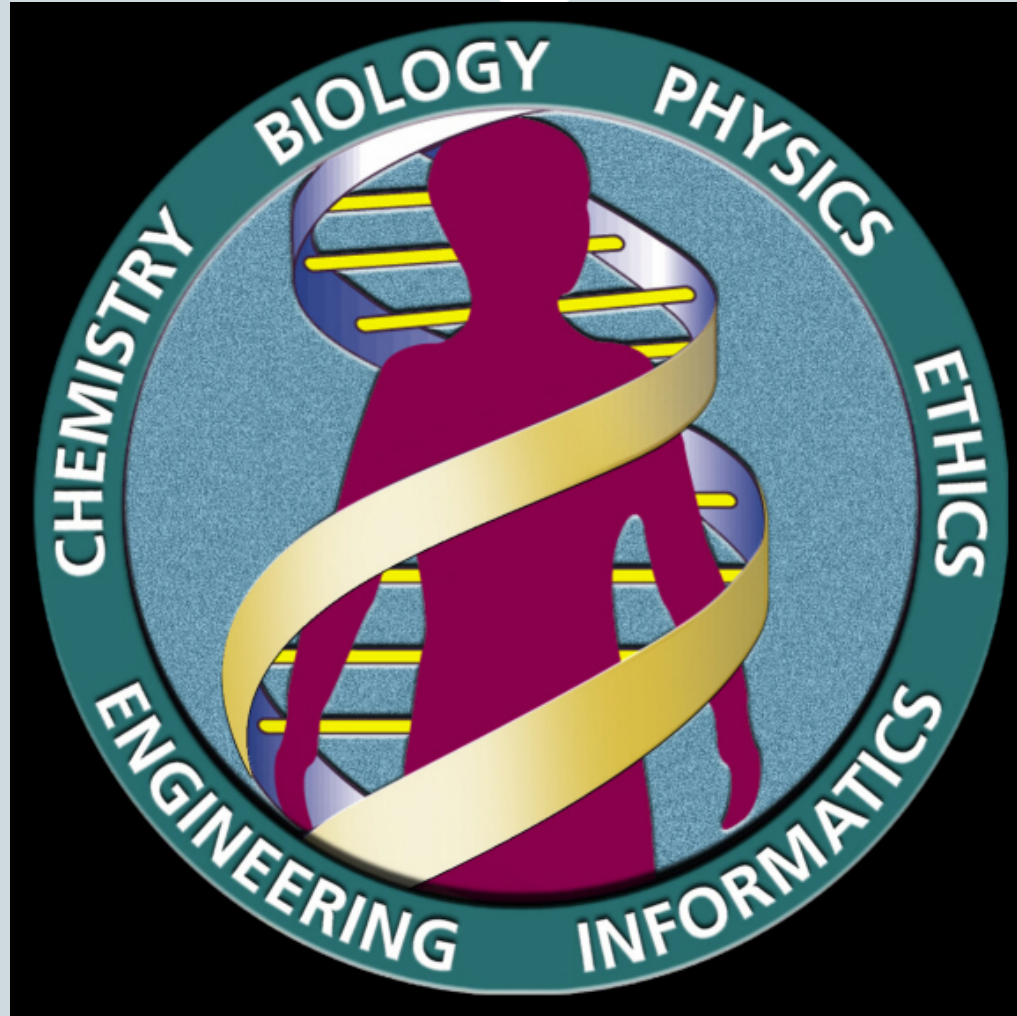
Summarizing

*Connecting
Presenting
Visualizing*

Predicting

*Inference
And
Prediction*

How to Approach Big Data



A model for DSI



Life Sciences

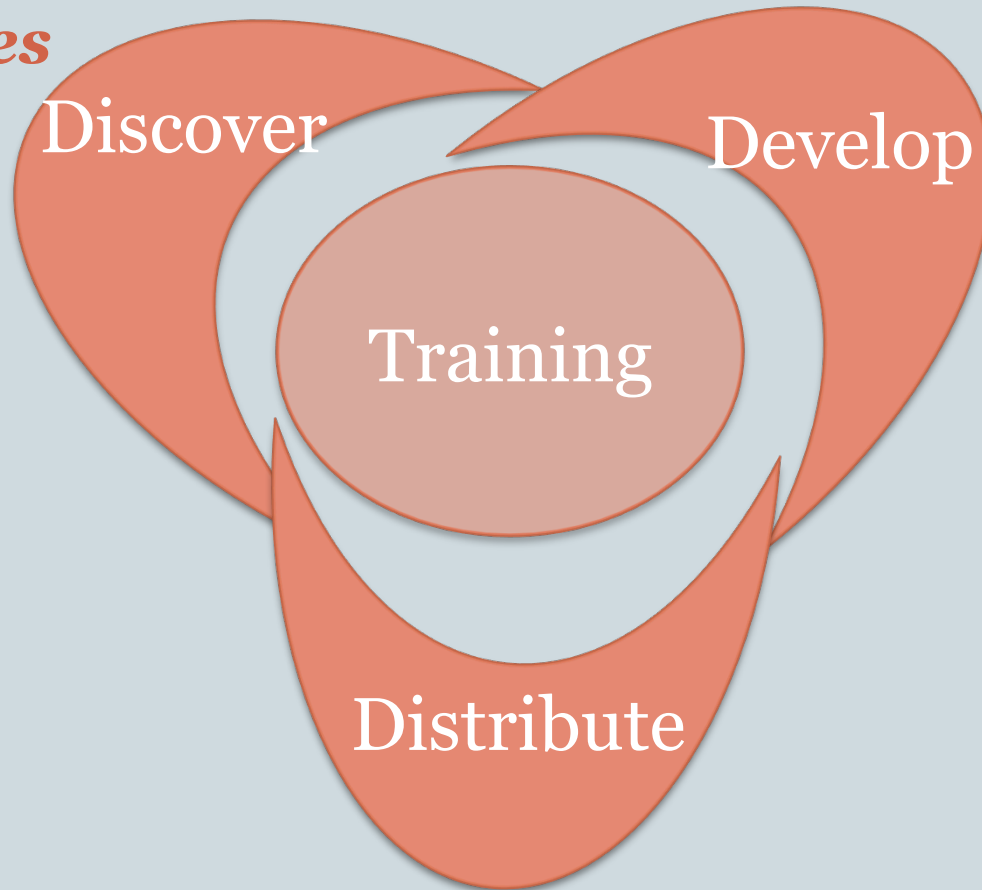
Discover

Develop

STEM

Training

Distribute



DSI: Research



- While data matter, the focus should be the scientific questions
- Need a physical space for collaboration
- DSI should serve as a support for synergy within the domain sciences
- Integrate with existing services, such as bioinformatics
- Good practice in Data Science
- Ethics in Data Science

Social Consequences of Commodity Sequencing



Based on your genes, what is your
Sensitivity to Warfarin?
(or Coumadin®, a common blood thinning drug)



may require
typical dose



may require
decreased dose



23andMe will tell you:
Your drug sensitivity
What to tell your doctor

Social Consequences of Commodity Sequencing



DSI: Education



- Students need to remain the priority
- Education at all levels: undergraduate students to faculty
- Integrated programs: embedding Data Science education in existing curricula + develop data science training for experts
- Motivate faculty to educate in Data Sciences

A model for DSI



Life Sciences

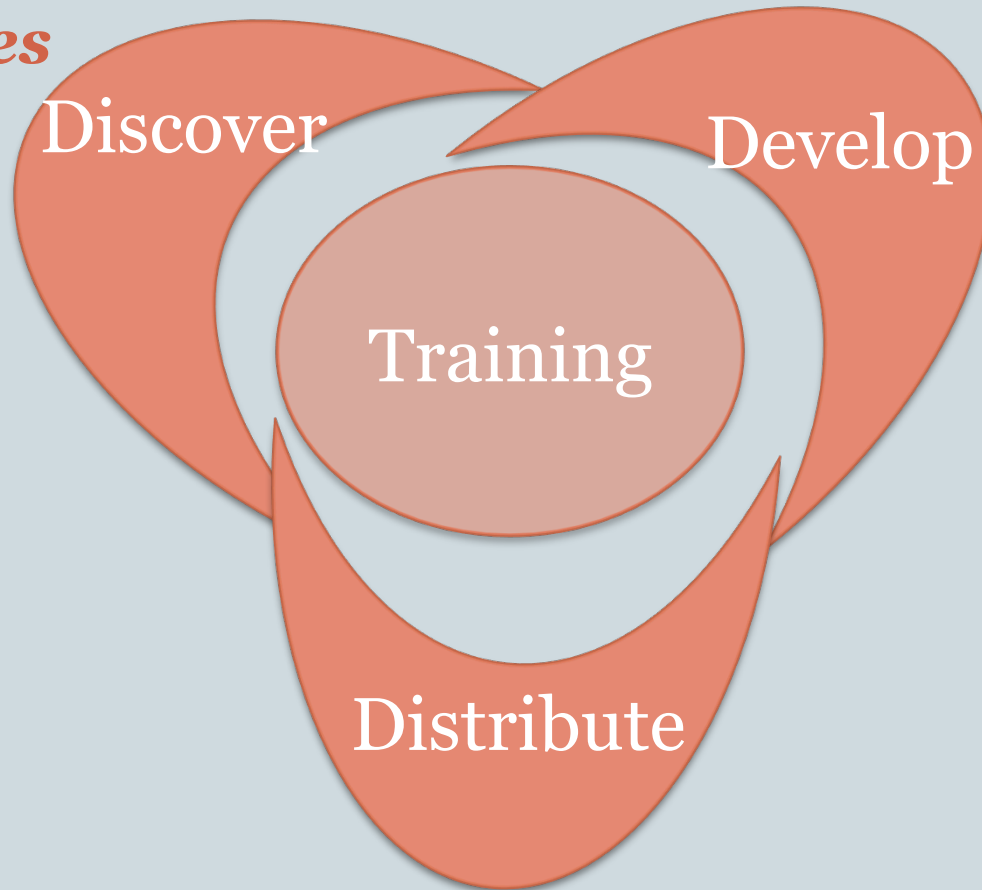
Discover

Develop

STEM

Training

Distribute



DSI



Thank You!