Measurement
- Physical/sensors (studied)
- Nonconventional (e.g. categorical)
  - E.g. relationship models
  - Some models exists
  - Extensive statistical theory
  - Computational theory needs development
- Data quality process (automation, computation)
  - Understand the pipeline
  - Provenance
  - Code that generates

Simulation data
- Indirect data
  - algorithms infer things
  - a model that transforms measurements
  - E.g. reflectance converted to cloud cover

Errors that you are not aware of
- Training bias
- Provenance of scientific data
  - Particularly shared data
  - Computationally derived error
  - Process or best practices

Systematic error

Representation
- Data products
  - Include "interpretation" error
  - E.g. interpolation of geospatial data (e.g. NASA)
  - Need models of error
  - Algorithms have errors
- Dislocation of data
  - Not produced where it is consumed
  - Purposes and uses change
  - Describing limitations becomes more challenges

Examples
- NASA
  - Atmospheric sciences (cloud cover, radiances, interpolation, CFD model)
  - Drift in a satellite (Charles)
- Medical records
  - New Item
- Annotations of kinase sub families (Vasant)
  - Downstream subfamilies
- Data deep dive
  - Automatically generated Paleo data

Documentation
- How to incentivize
- Citations
- Funding

Representations are important
- Producing realizations
  - E.g. from a generative model
  - E.g. clouds

Experimental work
- Formalizing the description experiments
- Provenance for experimental data

Infrastructure to support big data
- + uncertainty

Inherently stochastic models
E.g. stochastic PDEs

**Modeling**
- Uncertainty in parameter values
- Handles as distributions
- Error from processing
  - Numerics
  - Compression
  - Truncation

**Refinements of models**
- Uncertainty is a disincentive
- Models can make a difference
- Compounding errors vs compensation

**Interacting components**
- Some components we don't have data
- How to quantify confidence

**Perturbations in components**
- Change behavior of system

**Interactions are complex**
- Things are not additive
- Feedbacks may be disguised
- Emergent properties
- Multimodal distributions

**Reduced order models?**
- How do you gain confidence
  - Error bounds
  - Knowing that you have covered the set of possibilities

**Tools to support this**
- E.g. languages with built in distributions

**Uncertainty/statistical models**
- Complex
- Heterogeneous
- Rare events
  - Study the thresholds where things break

**Perception of rare events**
- communication
- understanding their frequency
- evolving boundary conditions change notion of rare events