TECHNOLOGY: shifts in the foundational paradigms of Big Data to create next generation capabilities

October 2014

Howard Wactlar

Information and Intelligent System Division
CISE Directorate
National Science Foundation

Shifts in Paradigms for Data Analysis

- Dynamic and adaptive design of data analysis methods based on data characteristics
- Heuristics in statistics
  - AI paradigms applied in computational processing
- Systems that “reason”: analyzing data in context of existing knowledge
  - Information extraction and understanding of text
  - Semantic analysis and encoding of published works
- Model synthesis from first principles, hypotheses and data analysis/mining

Shifts in Paradigms for Data Analysis - 2

- Model guided data collection
  ←→
  Data guided model revision
- Collaborative synthesis of new knowledge
  - "Human Computation" applied to model formation and discovery
- *Discovery Informatics*
  - Automating methods for understanding causality and casual cascades
  - Representing and capturing scientific process (e.g., scientist activities)
  - Building more effective many-human-computer team interaction

Big Data Capabilities

	Small era	Big era	Next generation
Goals	Answer a specific question, establish correlations	Flexible goals, possibly ill-posed questions, probabilistic prediction	Knowledge assimilation and reasoning, understanding causality
Location	One place	Highly distributed	Amorphous
Data structure & Content	Highly structured	Absorbs unstructured data from many sources	Differing in uncertainty and quality; combined with certified knowledge
Data preparation	By user or small group	Many sources, many people, possibly unconnected to users	Captured raw, ad hoc; combined w/ certified, standardized data
Longevity	Limited	Perpetual	Perpetual and reuseable
Reproducibility	Repeatable	Not necessarily repeatable	New data, information, knowledge continuously alters results
Analysis	All data analyzed together, all at once	Analyzed in incremental steps, distributed	Continuous processing within a context

Adapted from Berman, J.K.(2013*) Principles of Big Data,* New York; Elsevier

Still in progress …

- Foundational shifts supporting:
  - real-time processing of continuous data / continuous analysis
  - customized hardware and networking architectures for infrastructure

- Sensing ➔ Data ➔ Knowledge ➔ **Practice**

Thanks!