

RMDS

Global Association for Research Methods and Data Science

Research Methods in a Big Data and Cognitive Era

Dr. Alex Liu

RMDS

Pasadena, CA, USA

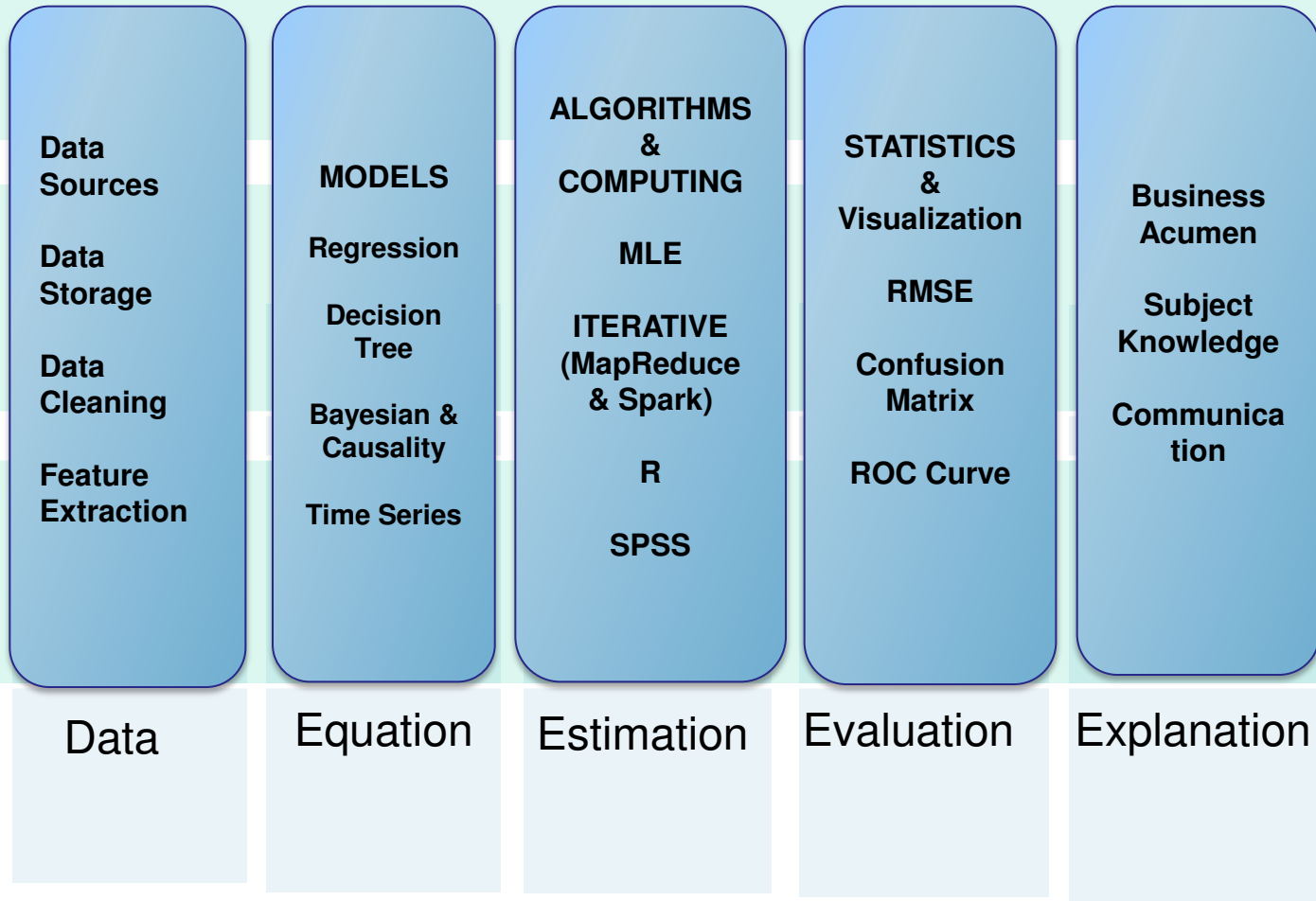
www.ResearchMethods.org

Updated October 8, 2015



Research Methods

are about optimal RM4Es workflows



Older Gen Research

- Literature Review in Library now Google
- Data in Excel Sheets
- Proprietary Computing with a Nicely Integrated Package
 - Stata, SPSS, Mathematica

New Gen of Research

- Open Source Computing Languages
 - R, Python, Scala, Julia
- Open Source Tools for Processing & Organizing Data and Analytics
 - Notebooks: Jupyter, Zeppelin
 - Visualization: D3.js, ggplot
 - IDE: R studio
 - Data Prep: Open Refine
- Open Source Execution Environments
 - Spark, Hadoop

We live in a moment of accelerated transformation

90%

of the world's data created in the last two years

75B

Devices connected to the internet by 2020

62%

of total workflows will be in the cloud by 2016

Big data in action



Financial services

- Risk and fraud management
- 360-degree view of customer



Retail

- Omni-channel marketing
- Click-stream analysis



Healthcare/Life sciences

- Medical record text analysis
- Genomic analytics



Law enforcement

- Real-time multimodal surveillance
- Cyber security detection



Telecommunications

- Call detail record processing
- Customer profile monetization



Energy and utilities

- Smart meter analytics
- Asset management



Digital media

- Real-time ad targeting
- Website analysis

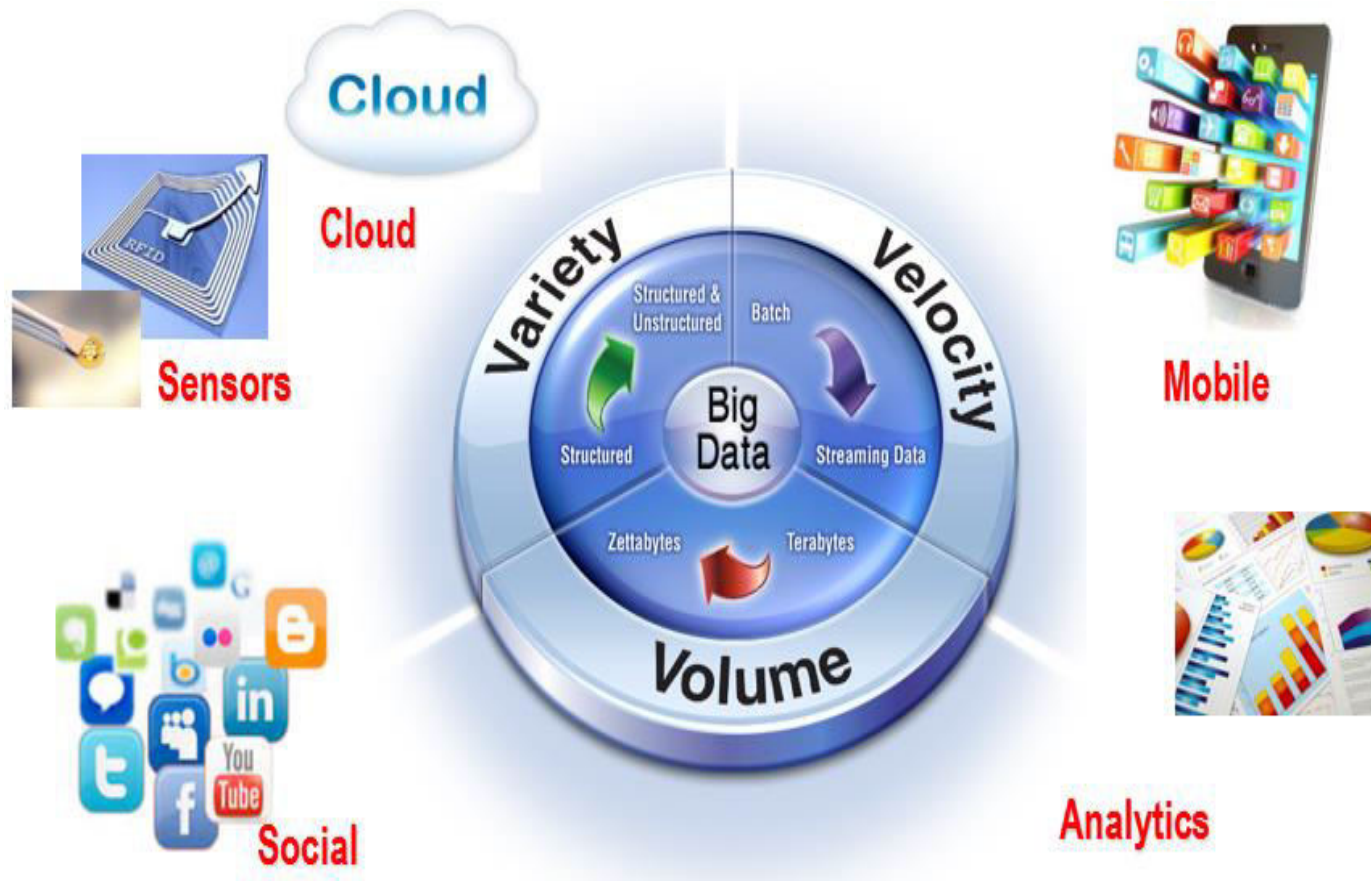


Transportation

- Logistics optimization
- Traffic congestion

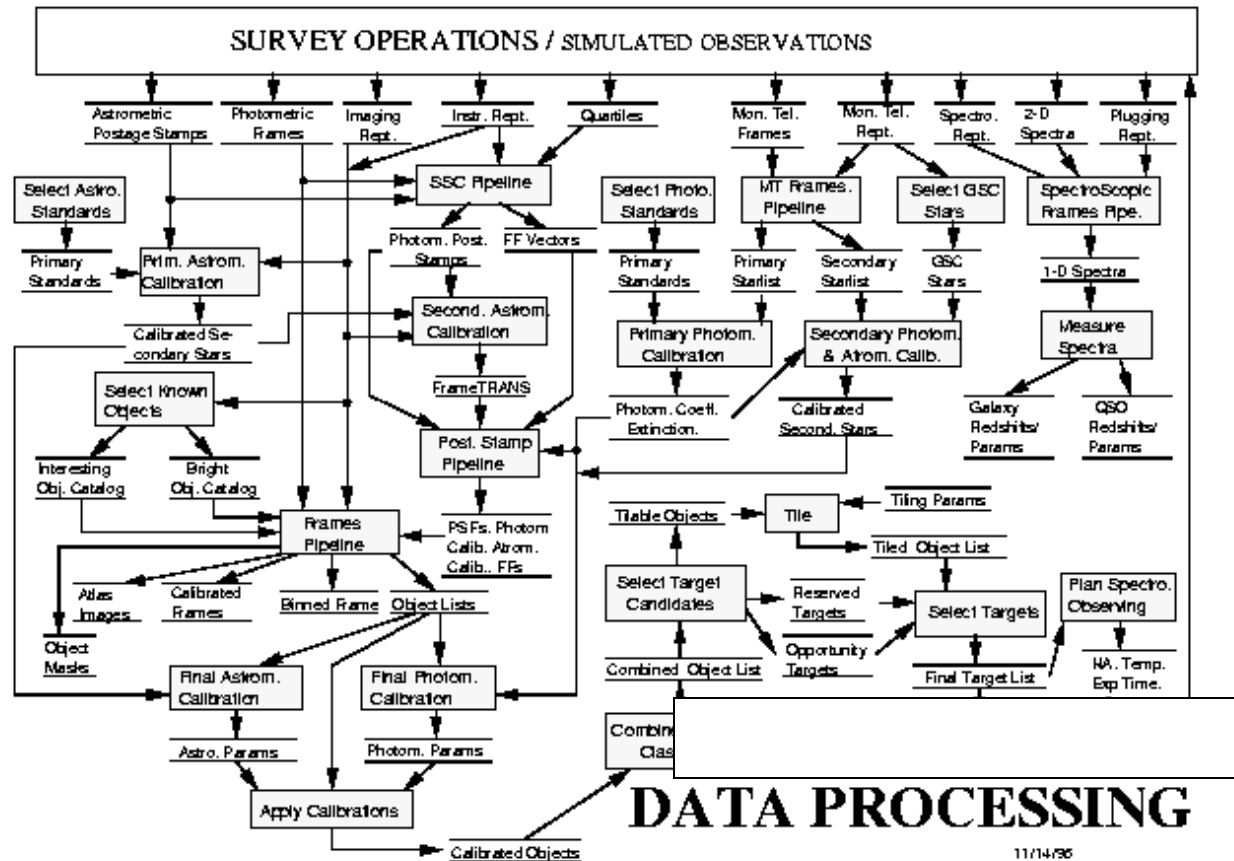
Big Data Era

– Too Much Data to Use

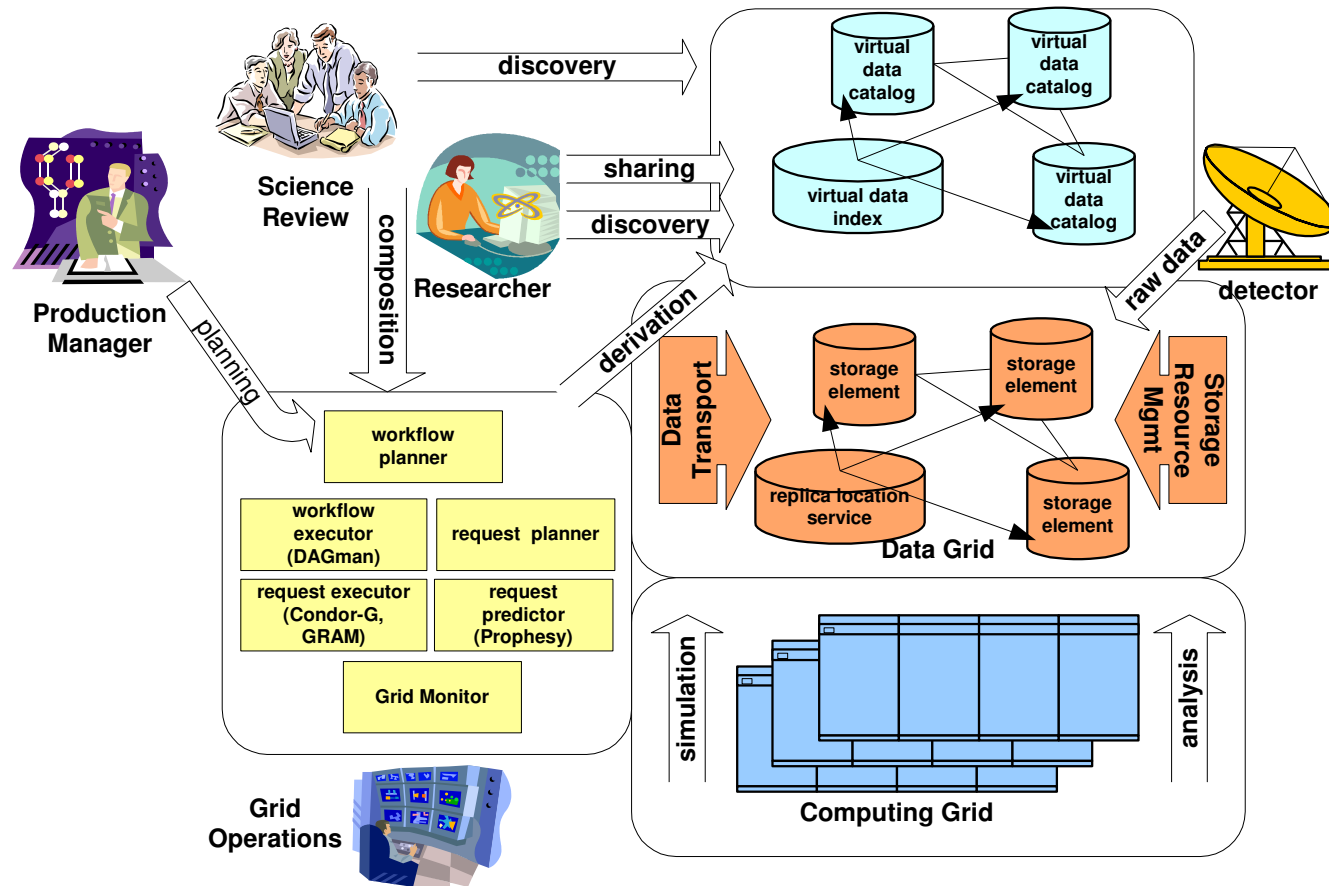


Too Many Analytical Steps

Research Flows Difficult to Manage



Too Much Resources to Coordinate



And a lot more to care ...

- **Safeguard Research Assets**
 - Control access
 - Timely tracking
 - Knowledge management
- **Regulatory compliance**
 - Book keeping
 - Versioning
 - Time recording

Challenges for Research Methods

- Too much data to import
- Too much data cleaning to complete
- Too many analytical methods to select
- Too many algorithms to select
- Too many computing tools to select
- Too many IT systems to select

Many new methods are coming

Traditional Method *Structured Analysis*

Researchers
Determine what
question to ask



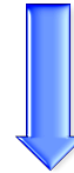
Research Support
Structures the data
to answer that
question



Monthly research reports
Profitability analysis
Customer surveys

Big Data Method *Iterative Analysis*

IT
Delivers a platform to
enable creative
discovery



Researchers
Explores what
questions could be
asked



Brand sentiment
Product strategy
Maximum asset utilization

Research Methods in a new era

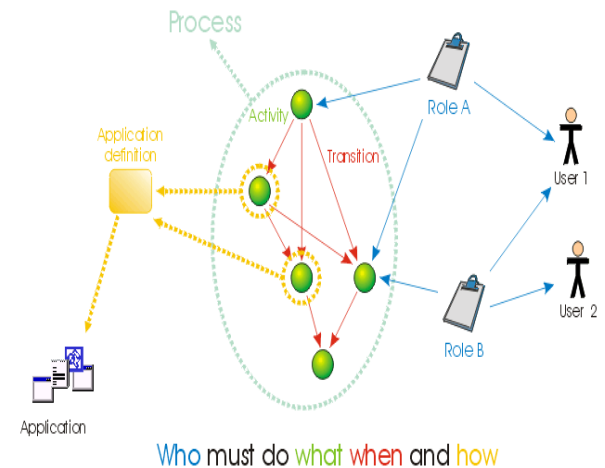
- Great opportunities for utilizing big data, high speed computing power and huge selection of analytical tools (models and algorithms)
- One researcher alone may not be able to solve all the problems faced
- Some intelligent assistance is needed to help every researcher

Intelligently Managing RFs Helps

Replicability is the foundation of scientific research.

RF management facilitates replicability.

- Replication (Provenance)
- Knowledge re-use & sharing
- Readiness for auditing
- Readiness for automation
- Removes much of the mundane data management burden, freeing scientists to do science



Research Flow (RF)

Need AI to automate and augment

- AI to automate some research flows
- AI to augment all researchers

