Education / Workforce Analytics *and* Big Data – A Joint AIR and Virginia Tech Workshop

SALLIE KELLER, DIRECTOR
SOCIAL AND DECISION ANALYTICS LABORATORY
VIRGINIA BIOINFORMATICS INSTITUTE AT VIRGINIA TECH

Outline

- Pressures of Today
- Big data
  - Why important?
  - What about privacy?
- Education/Workforce analytics
  - What makes it big data?
  - How does big data change current approaches?
- Methodology challenges




Education/Workforce Pressures of Today

- Management of costs and expenditures
  - Be less reliant on public funding
- Achieve higher standards and greater outcomes
  - Across populations of students with diverse backgrounds, abilities, and aspirations
  - Produce graduates with job-ready skills whose collective productivity can immediately impact the economy
- Training for higher cognitive capabilities
  - Utilize a greater diversity of learning ecosystems
    - Social networks, multidisciplinary, reverse classrooms, workforce training, life-long learning, etc.
  - Incorporate an ever changing flow of technological innovations

Education Analytics Opportunities

For example….

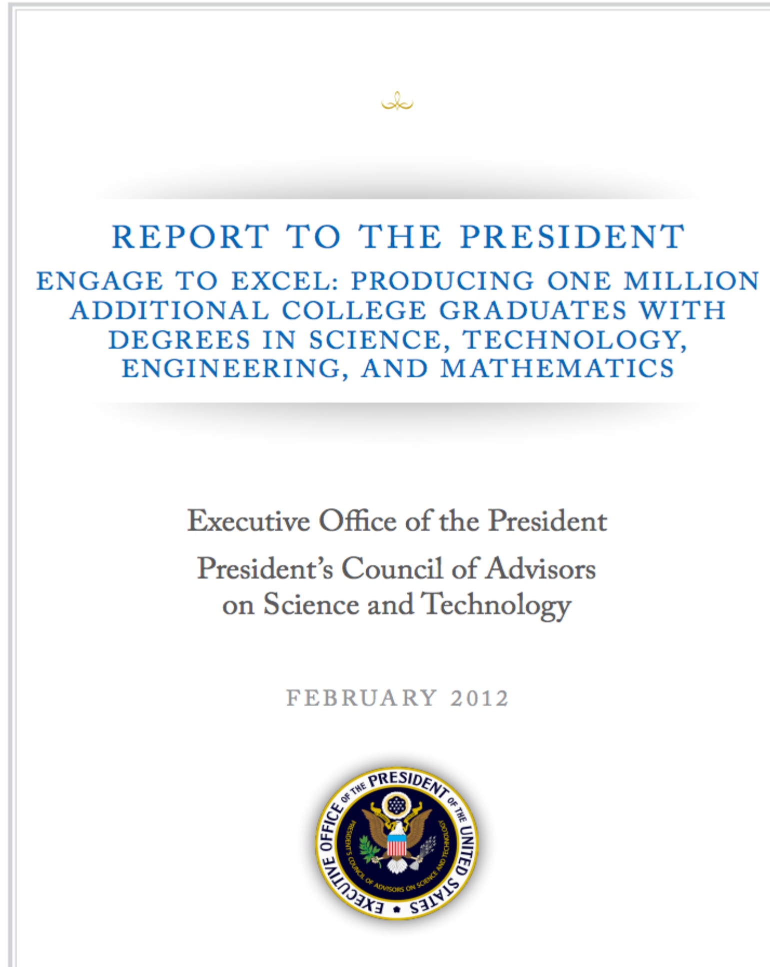
- Learning and demography?
- Learning and experiential characteristics of the learner?
- Learning and content presentation sequencing effectiveness?
- Learning and environments to stimulate curiosity?
- Learning and social interaction effects on learning progress?

Workforce Analytics Opportunities

Developing a STEM-educated and capable workforce

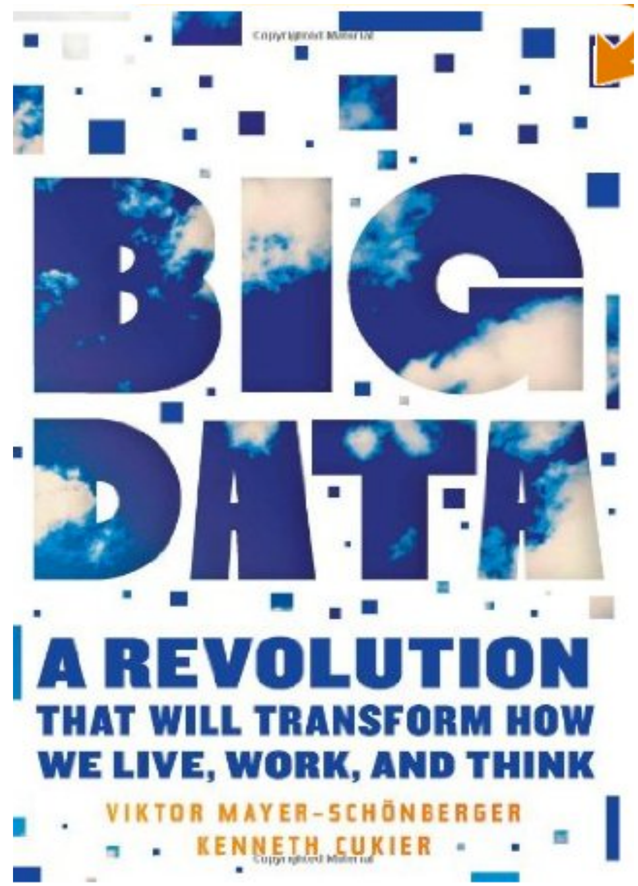
- Equilibrium of demand and supply?
- Measures of success?
- Improved data collection approaches?

Education and workforce development are difficult to decouple



Big Data - *Doesn't matter what its called, only matters what you do with it*

- Big data
  - Structured & unstructured
  - Collections
    - Designed
    - Observational/convenience
- Statistics / analytics
  - Replication, reproducibility, representativeness
  - Description, association, causation
    - prediction ≠ correlation
- Cost drivers
  - Analytics and informatics, NOT data collection

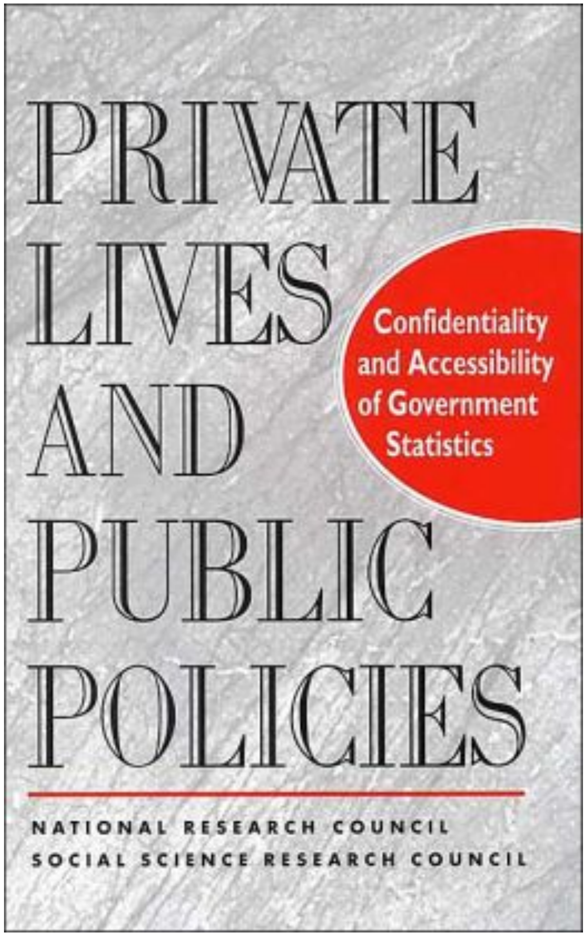


Now Big Data is Changing Social Sciences

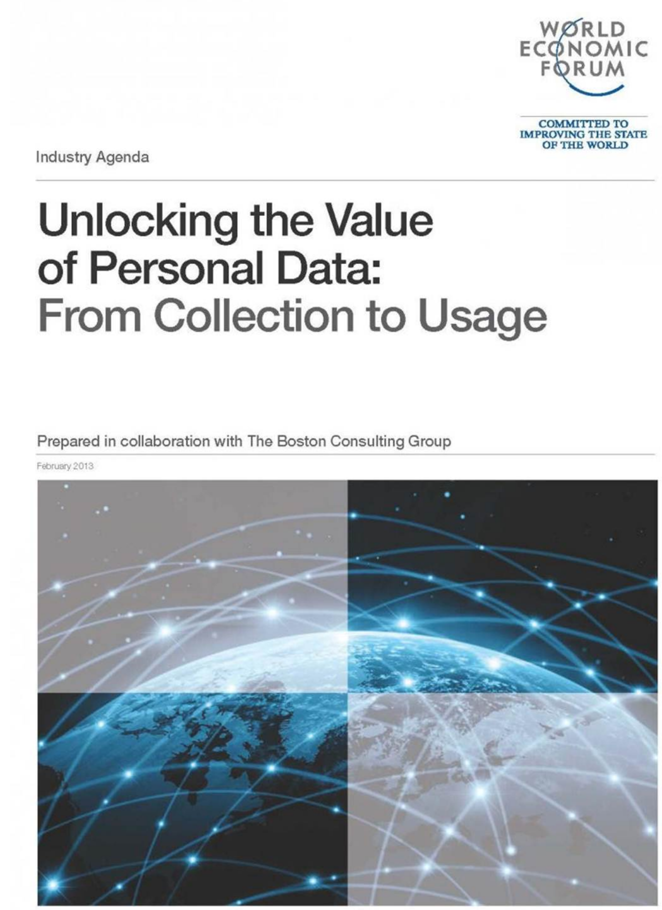
(C) 2012 THE HUMAN FACE OF BIG DATA

- Social science research has traditionally been informed by surveys and statistically designed experiments
  - Clean, well-controlled, limited in scale (~10^3) and/or resolution
- Bringing "Big data" to bear for social policy
  - Data informed computational social science models
  - Quantitative social science methods and practice at scale

What about Privacy?



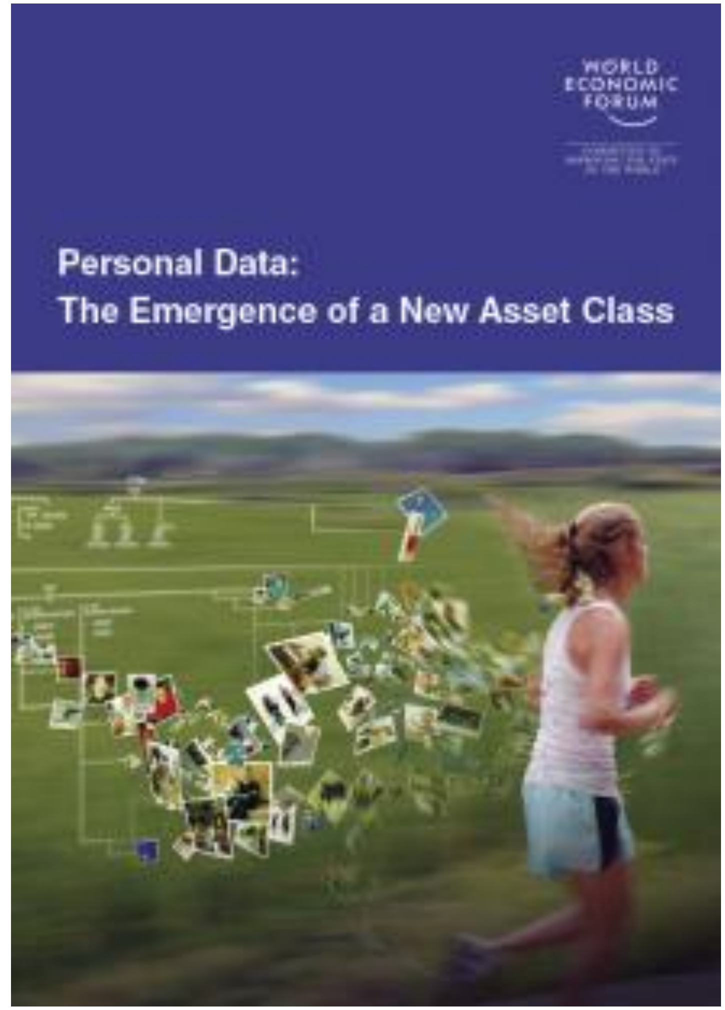
1993



2013

Personal Data - New Asset Class

- European Council 1995/1996:
  - *"… any information relating to an identified or identifiable natural person; an identifiable person is one who can be identified (data subject), directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity."*
- World Economic Forum 2011:
  - *"… digital data created by and about people."*



World Economic Forum 2013

Yesterday

- Definition of personal data is predetermined and binary
- Individual provides legal consent but not truly engaged
- Policy framework focuses on minimizing risk to individual

Today

- Definition of personal data is contextual and dependent on social norms
- Individual engaged and understands how data is used and value created
- Policy *needs to* focus on balancing protection with innovation and economic growth

Further Privacy Thoughts

- Will people voluntarily give up their data if they can see a personal or societal benefit?
- Are norms/expectations changing with generations?
- What are technical fixes for multi-level privacy/ classification?
- What is the optimal level of privacy for studies of interest?

Can we table privacy for the duration of the workshop?

- Deserves serious, devoted conversation
- We should be leaders in this conversation
- Will need to specifically address as projects develop

What Makes it Big Data?

Data Characteristics

- Multi-sourced
- Observational
- Noisy
- Multi-purposed

Multi-Sourced Data

Learning and development occurs within context

- Learner - including past experiences & mental processes
- Learning situation and content
- Department and institutional environment
- Local, state, and national education systems
- Political and economic factors

Information communication technology opens opportunity to capture meta data and provenance of the information

Challenge: integration and interpretation of data captured under such varied circumstances

DECEMBER 23, 2013, 11:19 AM | 17 Comments

A Start-Up Moves Teachers Past Data Entry

By QUENTIN HARDY

From left: Rafael Garcia, Dan Carroll and Tyler Bosmeny of Clever.

Apps Galore!


ALGEBRA NATION
Algebra Nation
ALL In Learning
A
Amplify Access
Benchmark EDUCATION
Benchmark Education Company
COPIA
Copia
DIBELS
DIBELSnet
Digital Passport
DreamBox
edmentum
Edmentum
eSpark
eSpark Learning
GoalBook
Gobstopper
Gobstopper

Observational Data / Convenience Sample

- Can come from every stakeholder, source, or technology that interacts with the learner
- Little discrimination on what is captured
  - Key strokes, eye movement, content accessed, test scores and testing attempts, etc.
- On-demand data from multiple systems
  - Social networks, school records, work history, medical records, extramural activities, etc.

Presents opportunity to study the learning processes as it naturally occurs

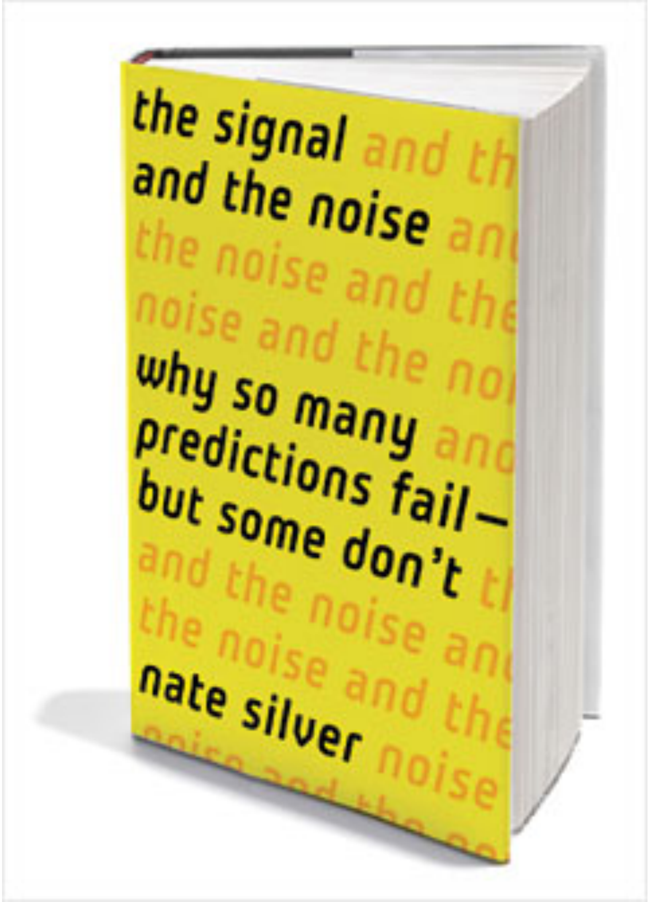
Challenge: manage biases and target group coverage

Noisy data

Meanwhile, if the quantity of information is increasing by 2.5 quintillion bytes per day, the amount of useful information almost certainly isn't. Most of it is just noise, and the noise is increasing faster than the signal.

Nate Silver, 2013

Challenge: uncertainty quantification



Multi-Purposed Data

Data reuse for multiple purposes

- Macro-level: regional, state, national, and international
- Meso-level: institution-wide
- Micro-level: individual learners, cohorts, and groups

An opportunity to more fully use data

Challenge: coherent aggregation across the levels

Source: Buckingham Shum, S. (2012)

Analytics Basics

Four key stages in application of analytics:

- Stage A – Extraction and reporting of inputs, outputs, responses, links, transfers, transactions, and other data
- Stage B – Analysis and monitoring of learner, teacher, operations, and other performance
- Stage C – What-if decision support
- Stage D – Modeling and simulation

Key Questions for Education Analytics

	Past	Present	Future
Descriptive	**What happened?**	**What is happening now?**	**What will happen?**
	Historical Reporting (Stage A)	*Assessment Reporting: Alerts* (Stage B)	*Extrapolations: Alerts* (Stage B)
Associative Inference	What variables best explain what happened?	What interventions seem reasonable?	What is the best/worst that can happen?
	Relationships and Modeling (Stage B)	*Options* (Stage C)	*Optimization, Simulation* (Stage D)

Learning Analytics Dashboards

1 of 1 Find | Next

Learn Course At-a-Glance

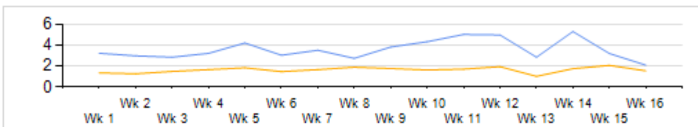
Report Help

CRS.225.M001.FALL12.Public Advocacy (M001)

Learn Course Information

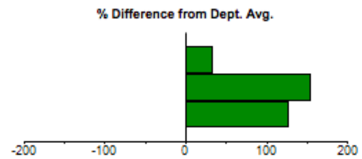
Instructor:	[redacted]	**Term:**	Fall 2012	**Delivery Method:**	In Person	**College:**	Visual & Performing Arts
Course ID:	28348.1131	**Status:**	Active	**Students Enrolled:**	136	**Department:**	Comm & Rhetorical Studies

Course Accesses Avg. vs. Department Avg. (same Instruction Method) — Section — Comm & Rhetorical Studies (53 Sections)

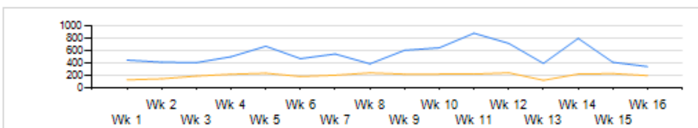


Course Item Count vs. Department Avg. (same Instruction Method)

Course Item Category	Course Total	Dept. Avg.
Assessment	25	18.1
Content	112	14.5
Tool	46	10.4

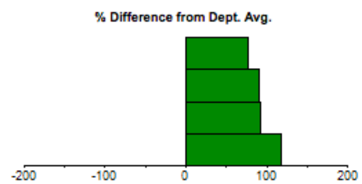


Minutes Avg. vs. Department Avg. (same Instruction Method)

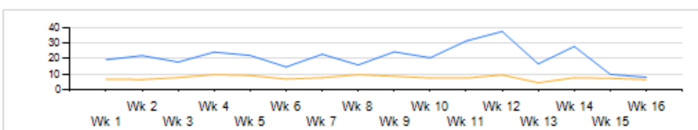


Avg. Activity Per Student vs. Department Avg. (same Instruction Method)

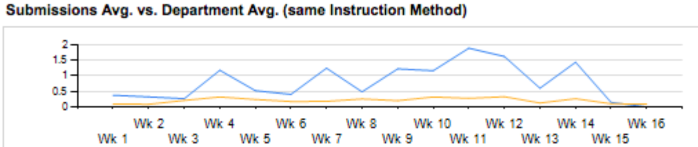
Activity Type	Course Avg. Per Student	Dept. Avg. Per Student
Accesses	57	25.6
Minutes	8,588	3,214.1
Interactions	335	123.7
Submissions	13	3.3



Interaction Avg. vs. Department Avg. (same Instruction Method)



Submissions Avg. vs. Department Avg. (same Instruction Method)



Legend

↑ > Avg + 10%	→ Within Avg +/- 10%
↓ < Avg - 10%	NA

Student Activity Summary (136 Students)

Student	Course Access Trend	Date of Last Access	Date of Last Submission		Course Accesses Student	Course Accesses Avg.		Total Minutes Student	Total Minutes Avg.		Total Interactions Student	Total Interactions Avg.		Total Submissions Student	Total Submissions Avg.	
[redacted]		12/20/2012	11/27/2012	↑	87	57.0	↑	12865	8587.7	↑	722	334.5	↑	19	12.9	→
[redacted]		12/19/2012	12/06/2012	↑	84	57.0	↑	13120	8587.7	↑	408	334.5	→	14	12.9	→
[redacted]		12/23/2012	11/27/2012	→	56	57.0	↓	7327	8587.7	↓	280	334.5	↓	9	12.9	→
[redacted]		[illegible]	[illegible]		[illegible]	[illegible]		[illegible]	[illegible]		[illegible]	[illegible]		[illegible]	[illegible]	

A Major Missing Component

Integration of theory into analytics

- Education, pedagogy, learning, and development

This eclectic approach is both a strength and a weakness: it facilitates rapid development and the ability to build on established practice and findings, but it – to date – lacks a coherent articulated epistemology of its own.

Clow, 2012

If we are helping people make decisions, then learning analytics is a moral and ethical endeavor. Being able to predict is not a high enough standard, we must understand why and how before we can ethically recommend.

Atkisson, 2011

Learning Disposition Theories

Dimensions	Positive characteristics of learners	Negative characteristics
Changing & Learning	Energy to learn and improve their minds	Static and satisfied with current knowledge
Critical Curiosity	Desire to learn and explore areas of knowledge	Passive and not curious
Meaning Making	Links current knowledge with material to be learned	Accumulating with minimal linking attempted
Dependence & Fragility	Resilient, persevering, and challenge-seeking	Avoids challenges/risks and is overwhelmed by mistakes
Creativity	Visualizes many perspectives and use imagination.	Limited in perspective and bound to rules
Learning Relationship	Balances social and private aspects of learning	Either too dependent on others or too isolated
Strategic Awareness	Self-aware and tries different learning strategies	Robotic

Five Stages in Application of Analytics

- Stage A – Extraction and reporting of inputs, outputs, responses, links, transfers, transactions, and other data
- Stage B – Analysis and monitoring of learner, teacher, operations, and other performance
- Stage C – What-if decision support
- Stage D – Modeling and simulation
- Stage E – Combined education, pedagogy, learning, development, statistical modeling, and decision support

Key questions for education analytics

	Past	Present	Future
Descriptive	**What happened?**	**What is happening now?**	**What will happen?**
	Historical Reporting (Stage A)	*Assessment Reporting: Alerts* (Stage B)	*Extrapolations: Alerts* (Stage B)
Associative Inference	What variables best explain what happened?	What interventions seem reasonable?	What is the best/worst that can happen?
	Relationships and Modeling (Stage B)	*Options* (Stage C)	*Optimization, Simulation* (Stage D)
Causative Inference	How and why did it happen?	What interventions are prescribed?	What does theory suggest can happen?
	Theory-based Relationships and Modeling (Stage E)	*Recommendations* (Stage E)	*Theory-based Predictions, Optimization, Simulation* (Stage E)

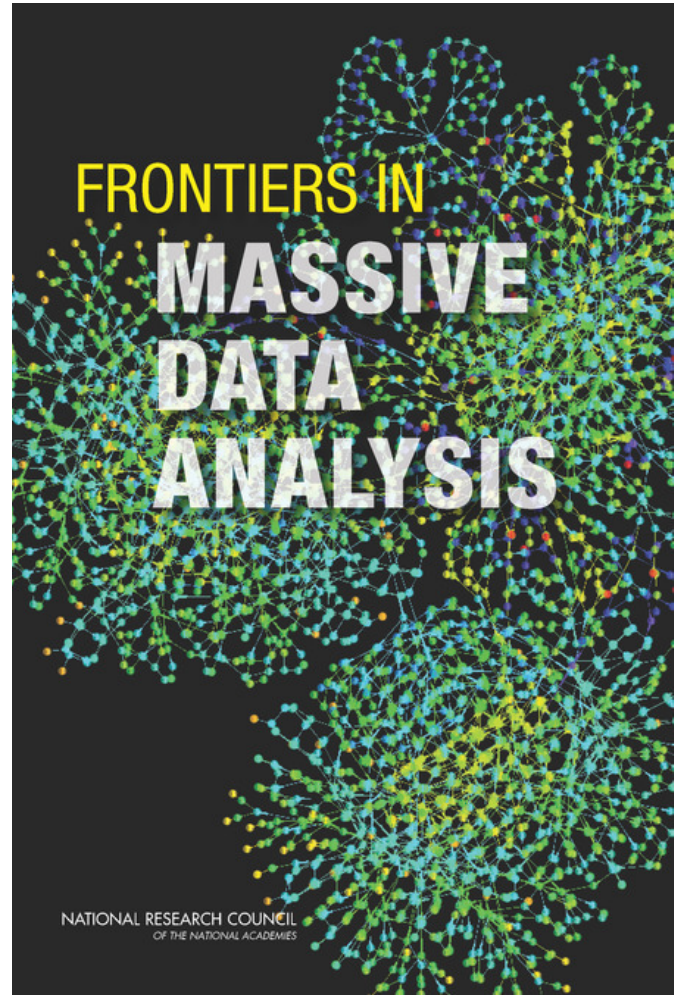
Big Data Analytics Framework – Focus on Purpose

- Learning Analytics
  - To manage the courses, learning experiences, and learner progress to maximize education effectiveness
- Academic Analytics
  - To manage institutions and maximize the institution's operational efficiency
- Evaluative Analytics
  - To assess institutional compliance with regulations and evaluate the outcomes of educational programs
- Comparative Analytics
  - To assess the relative standing and effectiveness of local, state, regional, national, and international educational systems

What are Methodological Issues?

New methods and tools are needed to ensure

- Data quality
- Representativeness
- Replication
- Reproducibility
- Characterization of noisy data
- Managing biases
  - Selection bias
  - Measurement bias



National Research Council 2013

Big Data Education/Workforce Analytics Conclusions

- Moving past tracking with large-scale data analytics

"What you have is a thermometer with no theory of action behind it. If I have a fever, nothing here is going to tell me how to deal with the fever. All it's going to do is tell me I have a fever."

- Mark Schneider, AIR, 2013

Goals for the Workshop

- Imagine a different world
- Look for synergistic capabilities to build partnerships
- Assess opportunities to integrate multiple sources of data and approaches to comprehensively understand education/workforce issues
- Incorporate theory into our thinking
- Propose prototype projects to work on together to set the stage for future projects

Virginia Tech's Social and Decision Analytics Lab

- SDAL joins Virginia Bioinformatics Institute
- Central to "Information Biology" theme
  - Study of massively interacting systems, from molecular to social phenomena
- Collaboration across VT and beyond
  - Embrace VBI mantra of transdisciplinary team science

Sallie Keller, Professor of Statistics and Director,
Social and Decision Analytics Laboratory,
Virginia Bioinformatics Institute at Virginia Tech
Sallie41@vbi.vt.edu