Forecasting Innovation Pathways: Using Data on 'Big Data'

Alan Porter
Georgia Tech
alan.porter@isye.gatech.edu
And
Search Technology, Inc.

Thanks to **Ying Huang, Jannik Schuehle, Jan Youtie, Jianhua Liu, Ying Quo & Doug Robinson** for key ideas and research.
We acknowledge support from the US National Science Foundation (Award #1527370 – "Forecasting Innovation Pathways of Big Data & Analytics"). The findings and observations contained in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Agenda

1. My slanted view on Management of Technology **(MOT)** "intel"
2. Tech Mining: A classic case – advanced ceramics
3. Forecasting Innovation Pathways **(FIP)**: The case of data on "Big Data" (**BD**)
4. Message: Think empirically – use those 'big' data (+ more)

PICMET Mission

- Portland International Conference on Management of Engineering & Technology
- **Advance analyses of changing technologies** to
- **Inform Management of Technology (MOT)**

Managing (other than 'technology'): On the Shop Floor

1. The font of knowledge: The machine operator's intuition (experience)
2. Role of data: Deming's Quality Control? [nonsense; see #1]
3. Action: Japan – empirical data provides enhanced knowledge to manage production processes far better
4. Result: **Six Sigma** manufacturing world (contrast to hospital world)

Managing (other than 'technology'): Athletics

1. The font of knowledge: The old veteran coach/manager ~ intuition (experience)
2. Role of data: [nonsense; see #1]
3. Action: **Moneyball** – empirical data provide enhanced knowledge to get superior major league baseball players on a tight budget [Oakland A's quite successful]
4. One result -- NBA-2015/16 – Golden State Warriors
   - Stats inform personnel decisions
   - Stats inform player combinations & matchups [~Big Data = "Advanced Scout reveals hidden patterns from play-by-play data]
   - Andre Iguodala – first start after 82 regular season + 3 playoff series – in the 2015 finals
   - **NBA champs** (2015)

     [but, hey, you can't win 'em all!]

Managing: Science, Technology & Innovation (ST&I)

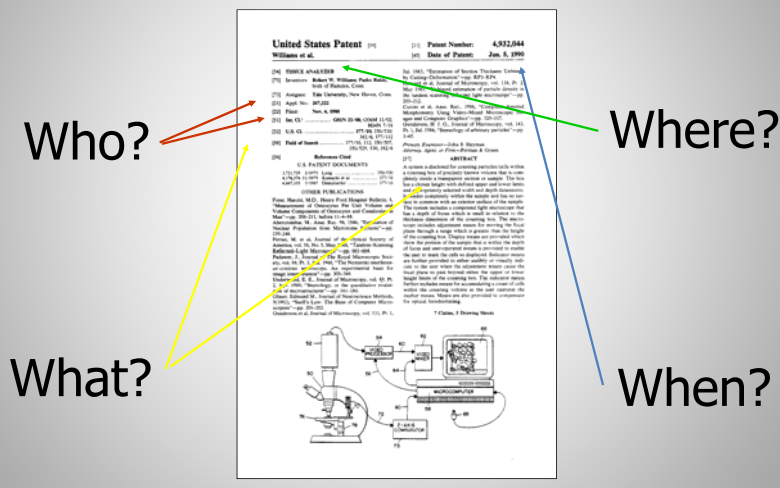
1. The font of knowledge in the US = personal judgment
   - scientists (peer review), or
   - engineers ~ collective experience, or
   - MOT managers ~ tacit knowledge
2. Role of data: [nonsense; see #1]
   - Scientometrics to inform R&D funding decisions? (in the US, request ear plugs)
3. Action: "Tech Mining" for **empirical intelligence** to complement expertise
4. Result → Better decisions!?

Tech Mining

- **To generate effective intelligence**
- **From ST&I information resources**
- **on topics of concern**

Alan L. Porter and Scott W. Cunningham
John Wiley & Sons Inc., 2005

Questions to Answer from field-structured data: e.g., R&D publication or patent abstract records gathered via global database search



How? & Why? – Need human analyst to interpret the data

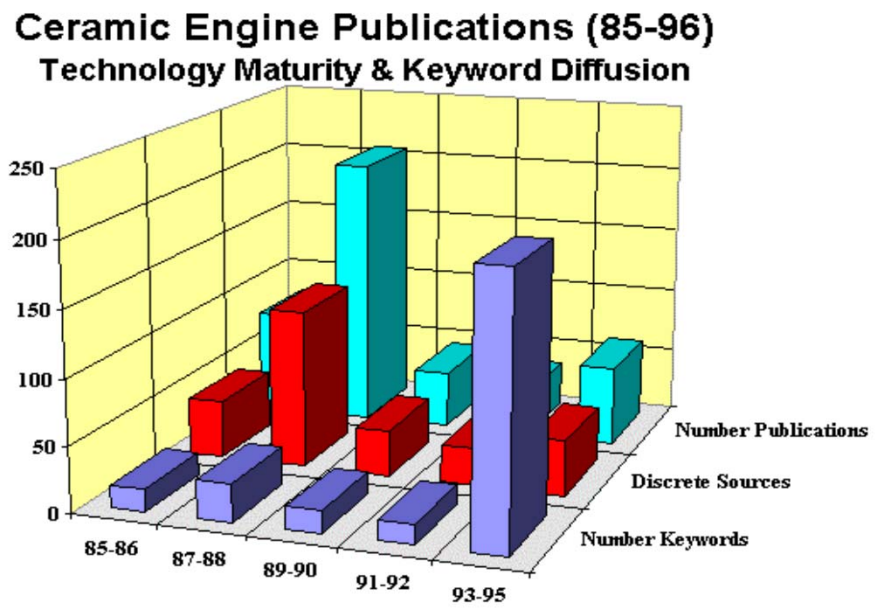
TechMining success story: Ceramics in Engines
by Bob Watts (US Army)
****PICMET Best Student Paper**

- Overcoming Management Resistance
- Jumping Domains
- "Discovering" new technology

Informing a tough decision

- **US Army Tank-Automotive Research, Development & Engineering Center**
- **Task in 1996: Reassess a "loser technology" – could thin-film ceramics be used in tank engines?**
- **TechMining: R&D Profile -- Amount of activity up a little -- but clues of significant maturation (next slide)**

The data speak:



The rest of the story

- Experts support the empirical findings
- Management buys in – search out potential in "coating engine parts"
- Who to go to? search finds ~95% of the research is NOT in their mechanical engineering domain -- so who would lead in ceramic coatings R&D?
- Identify R&D leaders – in semiconductor ceramics!
- $million projects funded with Sandia National Lab and a company to adapt "vapor deposition" to turbine blades
- Production plant coats used (Gulf War) Abrams tank turbine blades back to spec begins successful operation (2004)

Forecasting Innovation Pathways (FIP)

10 Steps (non-linear!) to Forecast Newly Emerging Science/Technology (NEST) Innovation Pathways

1. Understand the NEST and its TDS (Technology Delivery System)	Step A: Characterize the technology's nature
	Step B: Model the TDS
2. Tech Mine	Step C: Profile R&D
	Step D: Profile innovation actors & activities
	Step E: Determine potential applications
	Step J: Engage experts
3. Forecast likely innovation paths	Step F: Lay out alternative innovation pathways
	Step G: Explore innovation components
	Step H: Perform Technology Assessment
4. Synthesize & report	Step I: Synthesize and Report

Tech Mining for "Big Data" (BD)

Why study BD? To inform a potent Social Innovation

- US Government Accountability Office (GAO) doing a Technology Assessment of 21st Century Data
- To inform the US Congress and stakeholders re:
  - Issues and opportunities
  - Potential legislation – e.g., to protect privacy & security
  - Potential incentives to promote commercial development
- NSF support for our Georgia Tech group to improve our "Forecasting Innovation Pathways" methodology, using BD case
  - Interacted informally with GAO

Various Data on Big Data

Our initial focus is on R&D data:

- *INSPEC, EI Compendex* – treat the strong computer science core
- *Derwent Innovation Index* – patents
- *Research awards* – NSF, NSFC
- **Web of Science** – fundamental research; include conference proceedings [focus here]

+ Social Innovation facets:

- Commercial activity (databases like ABI/Inform)
  - Databases like Lexis Nexis ~10,000
- Popular attention
  - Google hits ~> 274 million (as of 7/31/2016)

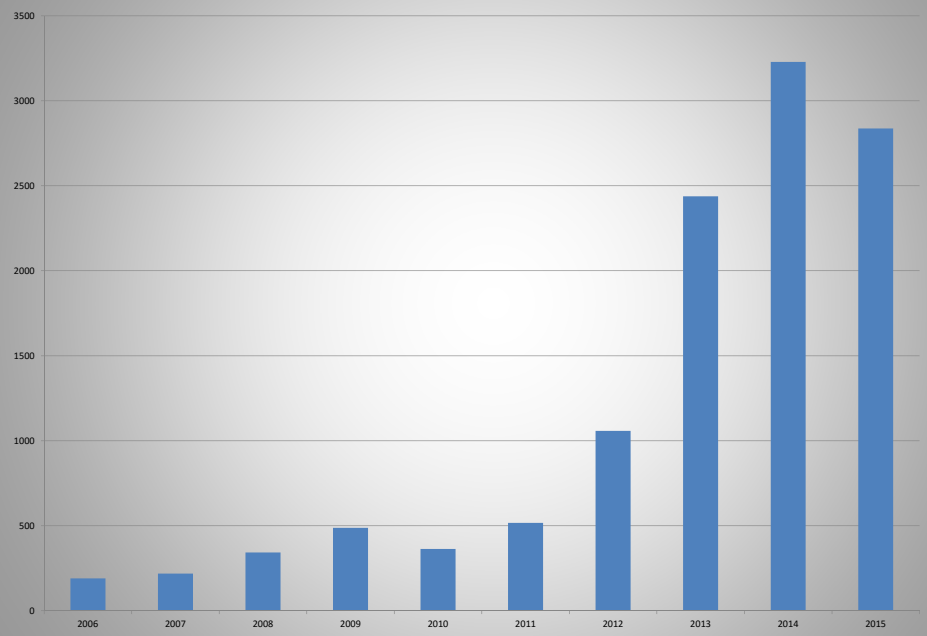
Locating & Retrieving Data

No.	Search Strategy	Search Terms
1	Core Lexical Query	TS= ("Big Data" or Bigdata or "Map Reduce" or MapReduce or Hadoop or Hbase or Nosql or Newsql)
2	Expanded Lexical Query	TS=((Big Near/1 Data or Huge Near/1 Data) or "Massive Data" or "Data Lake" or "Massive Information" or "Huge Information" or "Big Information" or "Large-scale Data" or Petabyte or Exabyte or Zettabyte or "Semi-Structured Data" or "Semistructured Data" or "Unstructured Data") AND TS=("Cloud Comput*" or "Data Min*" or "Analytic*" or "Privacy" or "Data Manag*" or "Social Media*" or "Machine Learning" or "Social Network*" or "Security" or "Twitter*" or "Predict*" or "Stream*" or "Architect*" or "Distributed Comput*" or "Business Intelligence" or "GPU" or "Innovat*" or "GIS" or "Real-Time" or "Sensor Network*" or "Smart Grid*" or "Complex Network*" or "Genomics" or "Parallel Comput*" or "Support Vector Machine" or "SVM" or "Distributed" or "Scalab*" or "Time Serie*" or "Data Science" or "Informatics*" or "OLAP")
3	Specialized Journals	The papers published in these specialized journals are not indexed by WOS
4	Cited References	The publications, which were cited more than 20 times did not fulfill the criteria for inclusion (see paragraph "Cited Reference Analysis")

Big Data Research Publication Profile

- **Web of Science**
- **2008 thru 2015 (as of early 2016)**
- ***11684 records***, of which 53% are conference papers, with 39% journal articles
- Leading sources are ***IEEE International Conference on Big Data*** and ***IEEE International Congress on Big Data***
- Lots of computer science attention
- Notable health Big Data research activity

Research Trend: "Big Data" [Web of Science]{"when??}



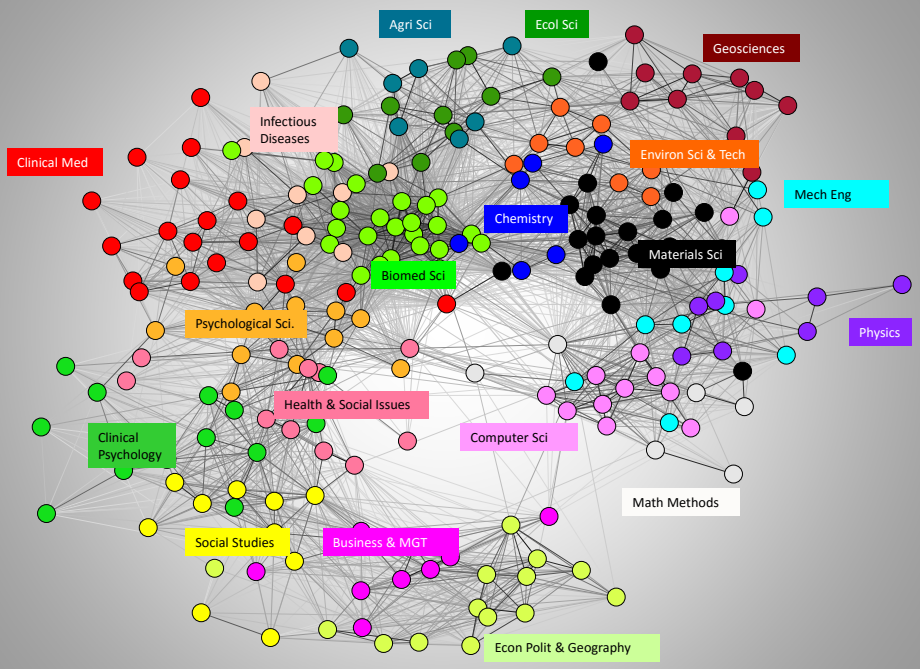
Similar strong growth in business & public attention

"Who Is doing Big Data Research? Top 10 Organizations publishing

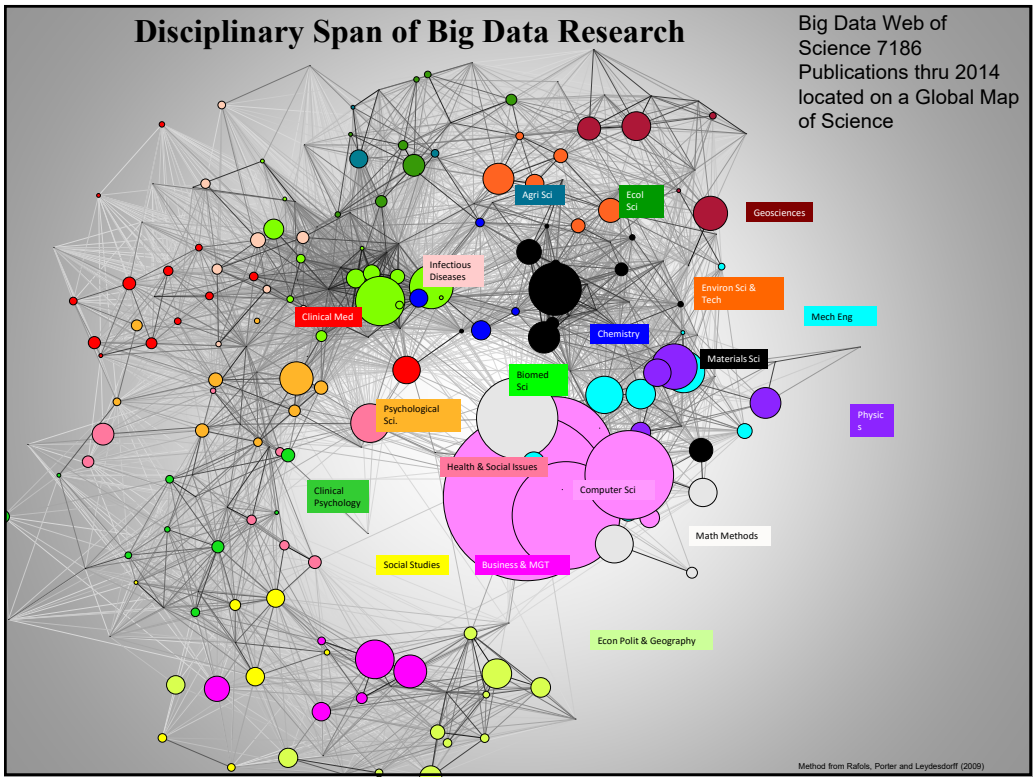
Author Organization	Records
Chinese Acad Sci	293
Tsinghua Univ	151
IBM	101
Harvard Univ	95
MIT	93
Beijing Univ Posts & Telecommun	90
Univ Calif Berkeley	87
Stanford Univ	86
Univ Illinois	86
Huazhong Univ Sci & Technol	85

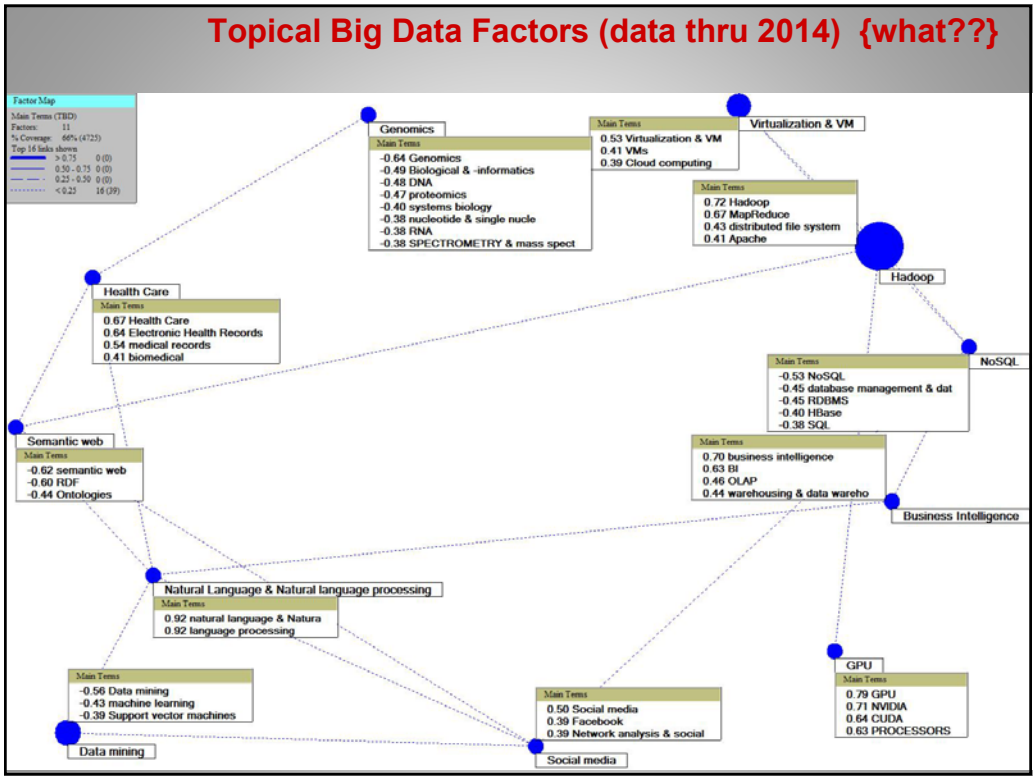
Of Top 30 -- **17 American, 13 Chinese**;
Of 11684 WoS papers, 3656 with US author; 3022 with Chinese author
[trailed by UK (672) & Germany (594)

Global Map of Science, Nodes = 224 Web of Science Categories; in 19 macro-disciplines {what??}

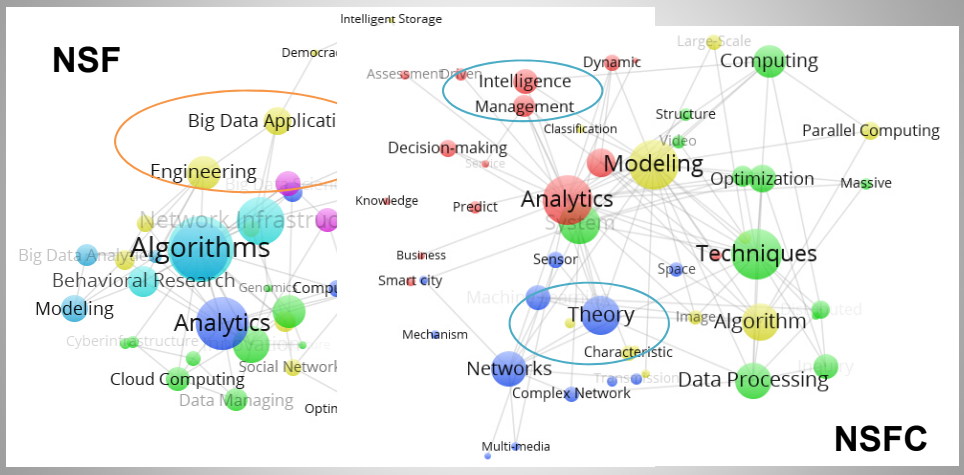


Method from Rafols, Porter and Leydesdorff (2009)


Disciplinary Span of Big Data Research
Big Data Web of Science 7186 Publications thru 2014 located on a Global Map of Science
Agri Sci
Ecol Sci
Geosciences
Infectious Diseases
Environ Sci & Tech
Mech Eng
Clinical Med
Chemistry
Materials Sci
Biomed Sci
Psychological Sci.
Physics
Health & Social Issues
Computer Sci
Clinical Psychology
Math Methods
Social Studies
Business & MGT
Econ Polit & Geography
Method from Rafols, Porter and Leydesdorff (2009)


Topical Big Data Factors (data thru 2014) {what??}
Factor Map
Main Terms (TBD)
Factors: 11
% Coverage: 66% (4725)
Top 16 links shown
> 0.75 0 (0)
0.50 - 0.75 0 (0)
0.25 - 0.50 0 (0)
< 0.25 16 (39)
Genomics
Main Terms
-0.64 Genomics
-0.49 Biological & -informatics
-0.48 DNA
-0.47 proteomics
-0.40 systems biology
-0.38 nucleotide & single nucle
-0.38 RNA
-0.38 SPECTROMETRY & mass spect
Virtualization & VM
Main Terms
0.53 Virtualization & VM
0.41 VMs
0.39 Cloud computing
Main Terms
0.72 Hadoop
0.67 MapReduce
0.43 distributed file system
0.41 Apache
Hadoop
Health Care
Main Terms
0.67 Health Care
0.64 Electronic Health Records
0.54 medical records
0.41 biomedical
NoSQL
Main Terms
-0.53 NoSQL
-0.45 database management & dat
-0.45 RDBMS
-0.40 HBase
-0.38 SQL
Semantic web
Main Terms
-0.62 semantic web
-0.60 RDF
-0.44 Ontologies
Main Terms
0.70 business intelligence
0.63 BI
0.46 OLAP
0.44 warehousing & data wareho
Business Intelligence
Natural Language & Natural language processing
Main Terms
0.92 natural language & Natura
0.92 language processing
GPU
Main Terms
0.79 GPU
0.71 NVIDIA
0.64 CUDA
0.63 PROCESSORS
Main Terms
-0.56 Data mining
-0.43 machine learning
-0.39 Support vector machines
Data mining
Main Terms
0.50 Social media
0.39 Facebook
0.39 Network analysis & social
Social media

BD R&D Policy Analyses {what??}



Tech Mining of Big Data Research: Who? What? Where? When? → what to do??

- Amazing increase in research on BD 2011-2014
- BD research is dominated by two countries – implications?
- Multidisciplinary attention, centered on computer science, but involving most R&D areas
- Social Innovation is a priority: sharp increases in social science analyses, patenting, commercial interest, and popular coverage
  [4,300,000 hits in Google Scholar; 274,000,000 in Google]
- We are pursuing Social Innovation aspects:
  - technology delivery system (players & influences)
  - technology assessment

Exploring Big Data Social Innovation

- **Technology Assessment**
  - Likelihood X Magnitude impact screening
  - Mitigation options
  - Survey + open web opinion seeking
    [Jianhua Liu, Ying Guo, et al. PICMET paper]
- **GAO "21st Century Data" TA study ongoing**

Big Data Social Innovation Scenario Analyses

Key on application variations

- Morphological strategy:
  - List factors
  - Identify alternatives for each
  - Package – driven by applications (real + potential)
- Construct ~ 15-20 diverse application scenarios
  - Present via blog or such to elicit feedback (elaboration, correction, addition)
  - Of various stakeholders
- Scan for
  - Sensitivities – factors (states) affecting multiple development paths
  - Opportunities
  - Emergent issues

Morphological Structuring for Scenario Analyses

Squeezed Version

Factor	States			
Product segments [letters key to 'market segments']	A1. E-med records	A2. Activity/fitness tracking	A3. Medical diagnostics	A4. Smart prostheses
Market segments	A. Health	B. Gov't oversight (detect misdeeds)	C. Taxing	D. transport
Notable Apps	F. Google Flu	S1. OK Google	S1. Dr. Watson	
Architecture	Distributed data	Distributed processing	Remote interfaces to central processing (Vs. desktop comp)	Multi-channel processing
Regions	US	China	W Europe	Other
Standards	Bodies	International		
Regulations	International collaboration	Proactive regs	Soft regs	Covering multi-actors handling data streams
BD Development guiding policies	Open source emphases	platforms	Scale up	
Boosting policies	R&D funding	Free vs. Fee		
Industry Structures	Start-ups	Consolidation (M&A)	Multi-actors	Alt business models
External forces	Economic health	Interest rates	war	terrorism
Infrastructures	Knowledge reservoirs	Cross-data type combinations	Datafication (e.g., collect & provide all Fed data)	Linkulation
Key Actors	Google	IBM	Hackers	
Societal concerns	Privacy	IP	New governance structures	Digital divide
Privacy Issues	Awareness (of threats to us)	Privacy policies	Data protection actions	Compliance checking
- Security	Hacking – private	Hacking – gov't	Data protection actions	
Subsystems	Distributed storage	Parallel processing	Remote Apps	
Synergistic Technologies (combo's with BD)	AI			
Key Tech Capabilities	Sensors development			
Underlying IT	Memory	Processing power	Quantum computing?	
R&D Thrusts	Data science	BD analytics	Viz (Tableau)	
(Data resources)	R&D funding	Res pubs	Patents	Newspaper coverage
(comparables – to compare trends)	Electrification	Other IT capabilities	Digitalization of pictures (cameras to smart phones)	TV – generations

Our Big Data Papers

1. Zhang,Y., Chen, H., Zhang, G., Porter, A.L., Zhu, D., & Lu, J. (to appear), Topic analysis and forecasting for science, technology and innovation: Methodology with a case study focusing on big data research, *Technological Forecasting and Social Change.* DOI: 10.1016/j.techfore.2016.01.015.
2. Zhang, Y., Robinson, D.K.R., Porter, A.L., Zhu, D., Zhang, G., & Lu, J. (to appear), Technology roadmapping for competitive technical intelligence, *Technological Forecasting and Social Change.* DOI: 10.1016/j.techfore.2015.11.029.
3. Huang, Y., Schuehle, J., Porter, A. L., & Youtie, J. (2015). A systematic method to create search strategies for emerging technologies based on the web of science: illustrated for 'Big Data'. *Scientometrics*, 105(3), 1-18. DOI 10.1007/s11192-015-1638-y.
4. Youtie, J., Porter, A.L., and Huang, Y. (2016), Early social science research about big data, *Science and Public Policy*, http://spp.oxfordjournals.org/cgi/reprint/scw021?ijkey=kuwzJ4wJmF7R4dz&keytype=ref
5. Huang, Y., Zhang, Y., Youtie, J., Porter A.L., and Wang, X. (2016), How does national scientific funding support emerging interdisciplinary research: A comparison study of Big Data research in the US and China, *PLoS One* 11 (5): e0154509. doi:10.1371/journal.pone.0154509.
6. Porter, A.L., Huang, Y., Schuehle, J., and Youtie, J. (in press), MetaData: BigData research evolving across disciplines, players, and topics, *IEEE BigData Congress 2015*.

Selected FIP References

- Daim, T. Porter, A.L., Chiavetta, D., and Saritas, O. (Eds.) (to appear 2016), **Anticipating Future Innovation Pathways through Large Data Analytics**, Springer, New York.
- Watts, R.J., and Porter, A.L. (1997), Innovation Forecasting, *Technological Forecasting and Social Change,* 56, 25-47.
- Porter, A.L., and Huang, Y. (2016), Forecasting future innovation pathways with big data analytics, *CIMS Innovation Management Report*, 8-13 (July/August), Poole College of Management, NC State University, Raleigh.
- Robinson, D.K.R., Huang, L., Guo, Y., and Porter, A.L. (2013), Forecasting Innovation Pathways for New and Emerging Science & Technologies, *Technological Forecasting & Social Change*, 80 (2), 267-285.
- Global Tech Mining Conference **Special Issues**: *Technology Analysis & Strategic Management; Technological Forecasting & Social Change; Scientometrics; International Journal of Technology Management.*

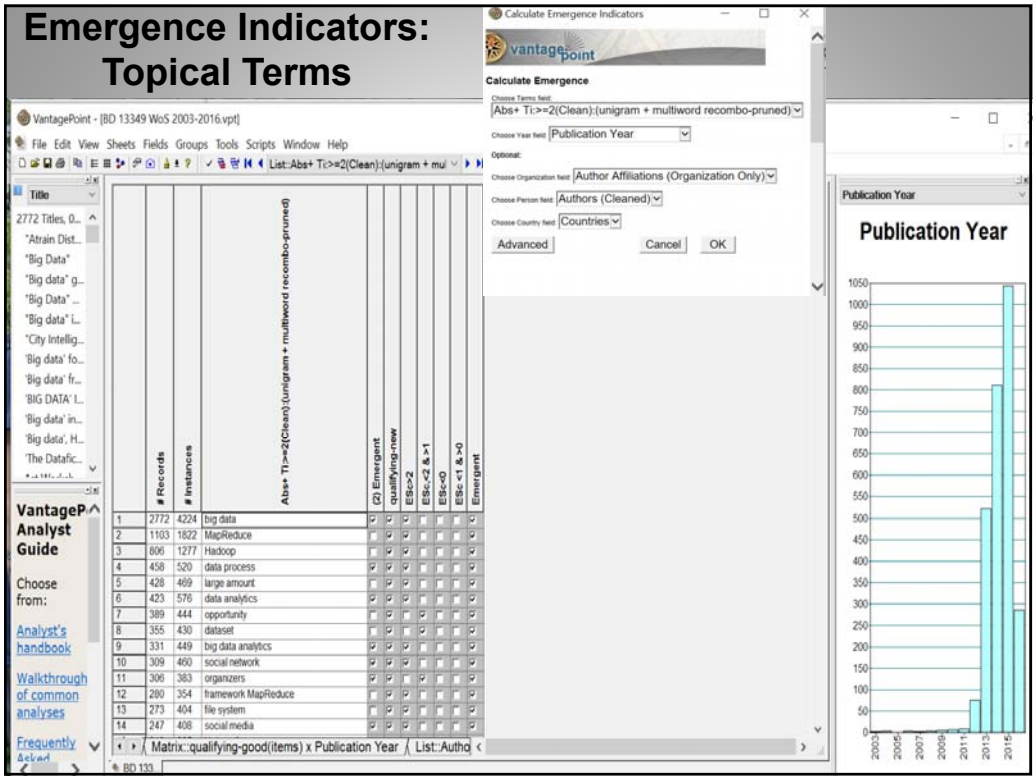
Resources

- The text mining software used: *VantagePoint* www.theVantagePoint.com [or *Thomson Data Analyzer*]
- Global Tech Mining Conference, in conjunction with Atlanta Conference on Science & Innovation Policy, September, 2017

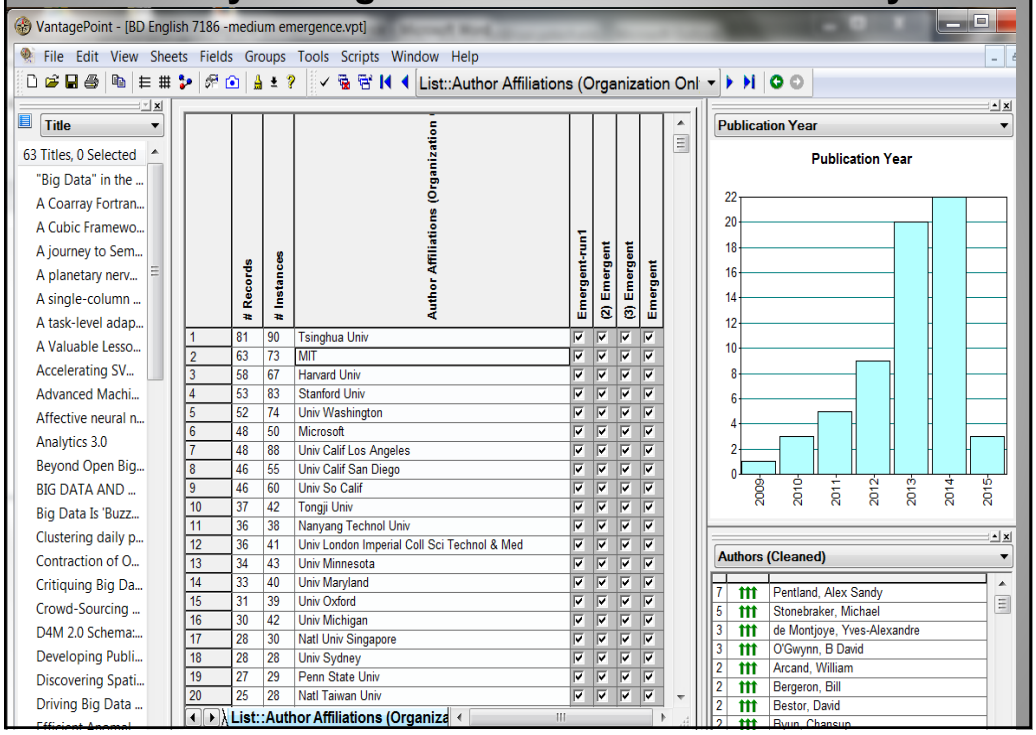
Managing Science, Technology & Innovation (ST&I)

One more data-based example:
**** Indicators of technical "emergence"**

Emergence Indicators: Topical Terms



Secondary Emergence Indicators: Frontier Players



Managing ST&I

Combine multiple knowledge sources
(Hal Linstone's multiple perspectives):

**** My charge: ST&I management is too intuitive**
[exceptions]
⇒ **Competitive Technical Intelligence (CTI)**

**** Direct experience + expert opinion + empirical intelligence**
⇒**Better decisions!**