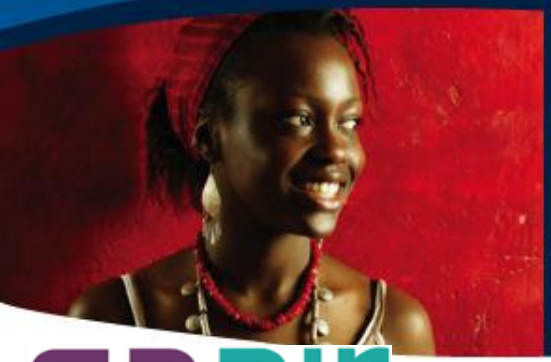
Can Big Data really be useful in Higher Education?

SAAIR Conference
18 October 2016

Herman Visser

Senior Specialist: Institutional Statistics & Analysis

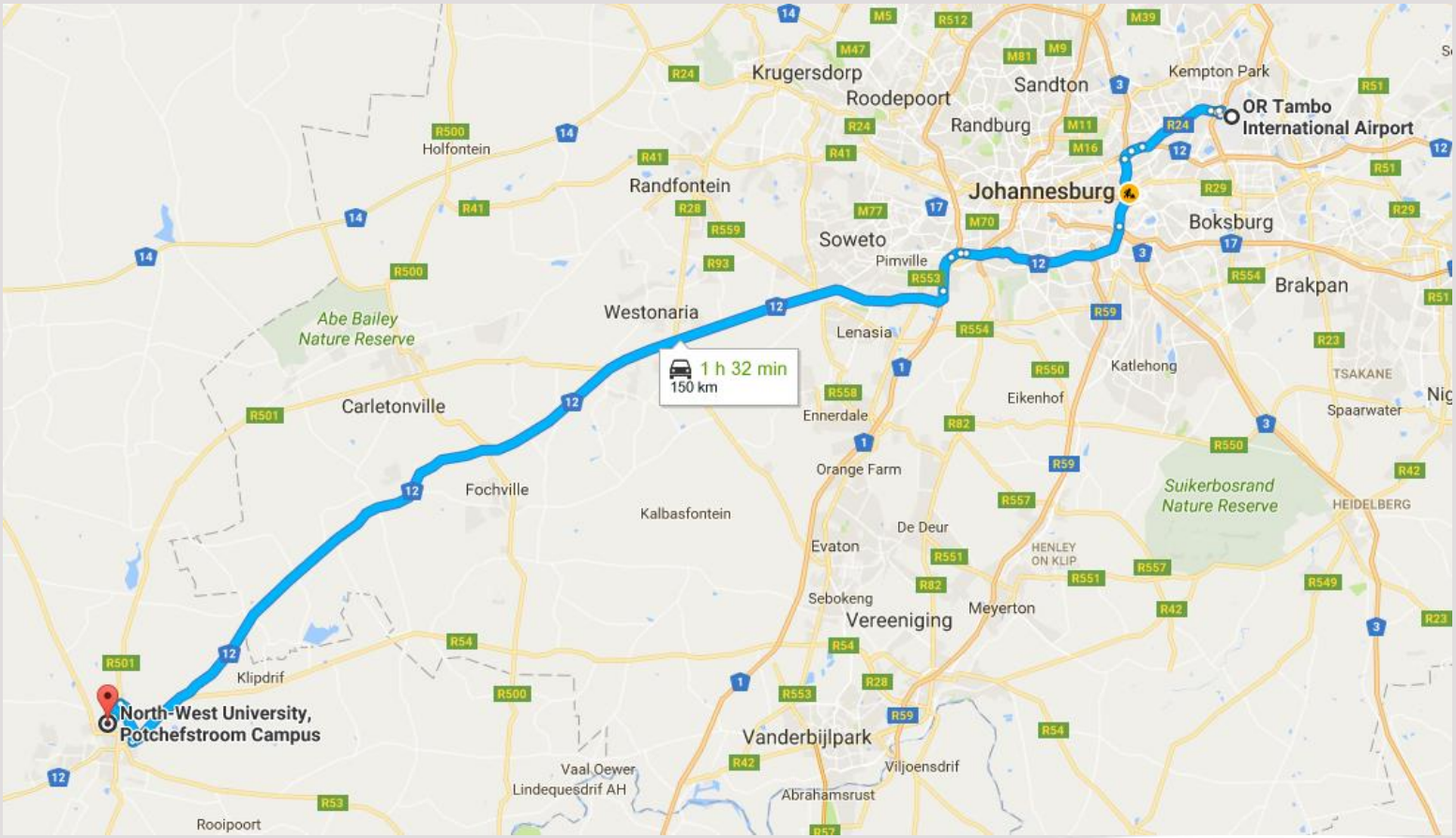
Directorate Information and Analysis




saair
Southern African Association for Institutional Research

UNISA university of south africa

Use of Geographical Positioning Systems (GPS) – Google maps: OR Tambo to Potchefstroom



UNISA university of south africa

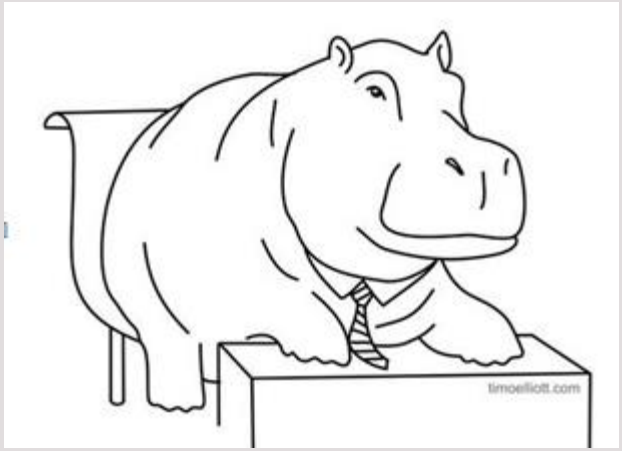
Big data is not a scary concept

- We are using it in our personal life regularly
- We are exposed to it regularly, but don't necessarily know what to do with it

UNISA university of south africa

Problem statement

- Some institutional researchers and campus leaders are not yet convinced that Big Data is important for Higher Education
- Many regard Big Data as a fad or more applicable to business than to Higher Education
- No Analytics?



Welcome to HIPPO – Highest Paid Person's Opinion

UNISA university of south africa

Purpose of this paper

The purpose of this paper i:

- To clarify some misconceptions about Big Data that may exist.
- To consider if it is a passing fad.
- To consider whether there are some potential uses of Big Data in Higher Education.

Megatrends identified by Gartner

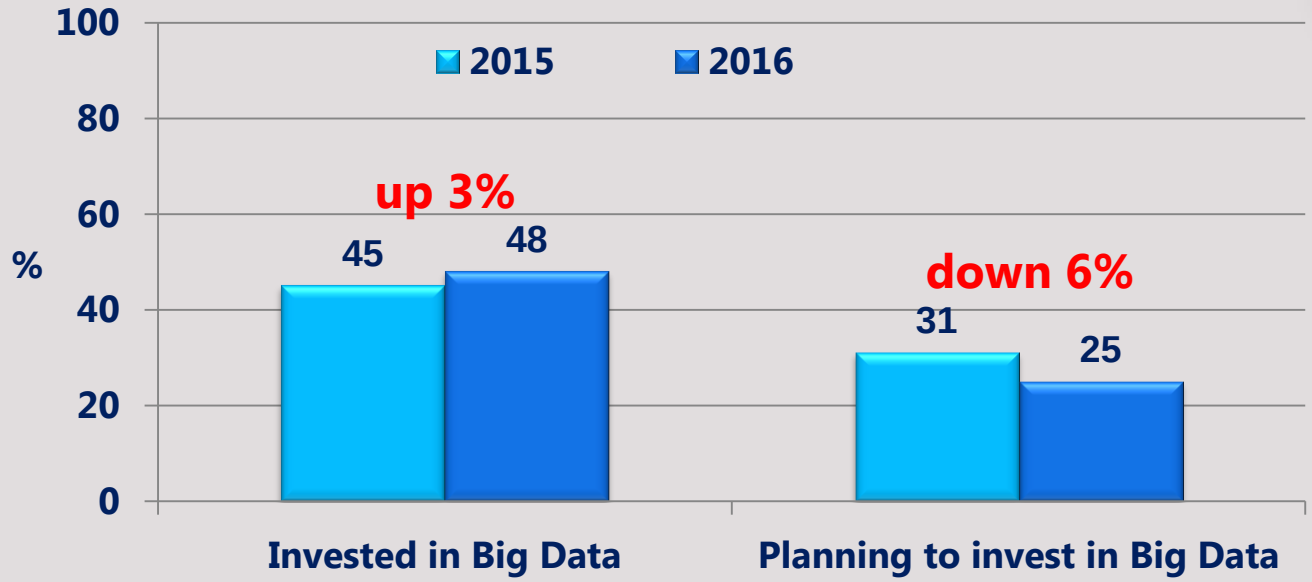
In their 2015 Hype Cycles, Garner identified five megatrends that shift the computing landscape, of which two are related to Big Data & Analytics:

- Analytics are at the *Peak of Inflated Expectations*

  Big data analytics for customer intelligence, predictive analytics, customer engagement and advanced analytics solutions are nearing the Peak.

- Big Data and Cloud moves toward the *Trough of Disillusionment*
  - Starting to see through the market hype and realities of strengths and challenges become clearer
  - In order to emerge toward the *Plateau of Productivity,* focus on application to address specific needs and opportunities and not on technology itself

Investment in Big Data



- Only 15% of projects in production, majority still at pilot stage
- Big data touted as a way to trawl through vast amounts of information to find trends that would otherwise remain hidden
- However, this proved a difficult goal to achieve – too many projects are poorly designed and therefore result in a lack of return on investment so organisations are spending on other priorities.

Gartner's research among Tech executives

UNISA university of south africa

Characteristics of Big Data
Laney's 3 Vs extended

- **Volume** – magnitude of the data (e.g. petabytes or zettabytes) generated by various sources, including sensors and the internet
- **Velocity** – rate at which data are generated and the speed at which it should be analysed and acted upon.
- **Variety** – diversity or structural heterogeneity of a dataset where the nature of the data varies from structured data to semi-structured data
- **Veracity** – refers to unreliability inherent in some data sources. There is a need to deal with imprecise and uncertain data because there is not sufficient time to clean the data before using it
- **Value** – Integrating different types of data and, putting them all together in order to extract hidden knowledge represents value

Some figures illustrating the characteristics of big data …

- 90% of the world's data has been generated in the last 2 years (SINTEF 2013)
- About 2,5 exabytes of data created each day, doubling every 40 months (*McAfee & Brynjolfsson 2012*)
- Every day Google alone processes about 24 000 terabytes of data, yet very little of the information is formatted in traditional rows and columns of conventional databases (*Davenport, Barth & Bean 2012*)
- Walmart's data warehouse includes some 2,5 petabytes of information (*Manyika et al 2011*)
- Tata Motors analyze 4 million text message every month from product complaints and reminders about service appointments to announcement about new models and customer satisfaction polling (*Agarwal & Weill 2*012)

BI & A evolution: Key characteristics and capabilities

BI&A phase	Key characteristics	Gartner BI Platform Core Capabilities	Gartner Hype Cycle
BI&A 1.0	Database management system-based (DBMS) structured content • Relational DBMS (RDBMS) & data warehousing • Extraction, transformation and loading (ETL) & online analytical processing (OLAP) • Dashboards and scorecards • Data mining and statistical analysis	• Ad hoc query & search-based BI • Reporting, dashboards & scorecards • Interactive visualisation • Predictive modelling & data mining	• Column-based DBMS • In-memory DBMS • Real-time decision • Data mining workbenches
BI&A 2.0	Web-based unstructured content • Opinion mining • Question answering • Web analytics and web intelligence • Social media analysis • Social network analysis • Spatial-temporal analysis		• Information semantic services • Natural language question answering • Content and text analysis
BI&A 3.0	Mobile and sensor-based content • Location-aware analysis • Person-centered analysis • Context-relevant analysis • Mobile visualisation & human-computer interaction (HCI)		• Mobile BI

Chen, Chiang & Storey 2012

UNISA university of south africa

Gartner Hype Cycle for Advanced Analytics and Data Science, 2015

Benefit	Years to mainstream adoption		
	less than 2 years	2 to 5 years	5 to 10 years
Transformational		Citizen data science Machine learning Smart data discovery	Analytics marketplace Crowdsourcing of microwork Deep learning Event stream processing
High		Ensamble learning Geospatial and location intelligence Hadoop-based data discovery Load forecasting Predictive analytics R Self-service data preparation Text analytics	Chief Analytics Officer (CAO) Graph analysis in-DBMS analytics Linked data Natural language generation Natural language question answering Optimization Presctiptive analytics Real-time analytics Spark Speech analytics
Moderate	Video analytics	Model factory Model management	Data lakes Emotion detection/recognition Predictive Model Markup Language Simulation Uplift modeling

Gartner, 2015

Why is Big Data important? How can Big Data add value?

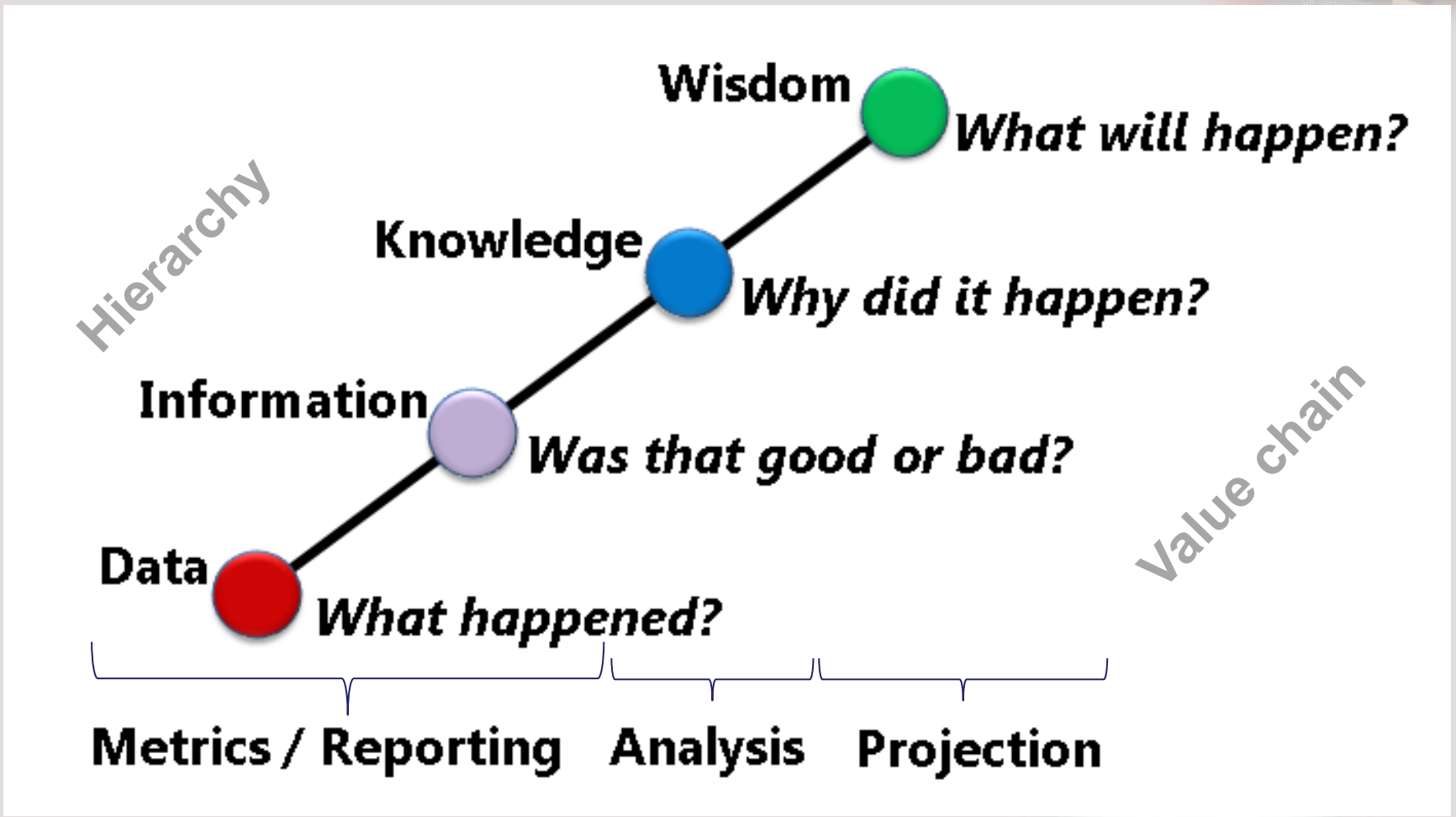
- "Business intelligence and analytics (BI&A) and the related field of big data analytics have become increasingly important in both the academic and the business communities over the past two decades."

  Chen, Chiang & Storey (2012:1165)

- They refer to two important studies.
  - **2011 *IBM Tech Trends Report*** → identified business analytics as one of four key technology trends in the 2010s
  - **McKinsey Global Institute study**, predicted that by 2018 the United States alone will face a shortage of between 140 000 and 190 000 people with deep analytical skills and a shortfall of 1,5 million data proficient managers with the capacity to analyse big data to make effective decisions.
  - Both of these findings are likely to have significant implications for institutional researchers and their role in higher education institutions of the future.

UNISA university of south africa

Data Science: From data to wisdom



Changing landscape of higher education

Globally, higher education institutions are under increasing pressure to change because of:

- Decline in government funding
- Reduction in endowments from alumni and other stakeholders
- Declining support from business
- Increasing regulatory requirements
- Increased performance monitoring, accountability and transparency
- Massification of higher education
- Pressures to improve retention, success and throughput
- Technology changes

UNISA university of south africa

Static data and fluid data

The ***UK Higher Education Commission*** on the potential of data and analytics in higher education (2016) distinguish between the following categories of data:

- **Static data**

  Data which is recorded and stored by institutions and traditionally includes student record data, staff data, research output data, financial data, alumni data, estates data

- **Fluid data**

  Data generated through increasingly digital interactions of students with their university such as interactions with the virtual learning environment/learning management system, library and swipe card data

Learning analytics

- Learning analytics is the measurement, collection, analysis and reporting of data about learners and their contexts, for the purposes of understanding and optimising learning and the environments in which it occurs

  Society for Learning Analytics Research (SoLAR)
- Personalising learning or adaptive learning

Major motivations for introducing learning analytics

- Improving retention
- Providing students with better feedback on their progress
- Enhancing teaching and learning
- Capturing attendance data

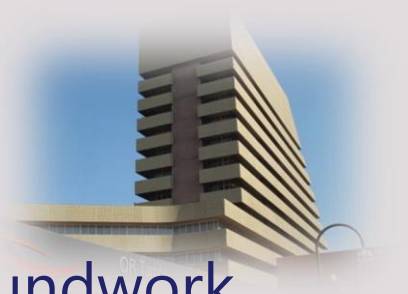
Applications of big data in Education
Data Mining and Learning Analytics

- Performance prediction
- Attrition risk detection
- Data visualization
- Intelligent feedback
- Qualification recommendation
- Student skill estimation
- Behavior detection
- Grouping and collaboration of students
- Social network analysis
- Developing concept maps
- Constructing coursework
- Planning and scheduling

Sin & Muthu (2015)

UNISA university of south africa

Student dropouts



- Tinto, Spady, Bean, to name a few, laid the groundwork for theories on student attrition in the 1970s – 1980s
- However, student dropouts still remain a major issue
- Roughly 30% of 1st year students at US baccalaureate institutions don't return for their 2nd year → over US$9 billion is spend education these students (*Aulck, Velagapudi, Blumenstock & West 2016:16*)
- *Brown (2016)* highlights the problem of non-returning students who are not identified timely as at risk students in order to receive targeted interventions
  - Historically the students were blamed by institutions for dropping out because of poverty, under-preparation and other extra burdens
  - Georgia State University as an exception

UNISA university of south africa

International comparison

- UK Higher Education Commission found that US and Australia are ahead of the UK in terms of adoption of data analytics across institutions
- Lot of research in the US into the potential of learning analytics and predictive analytics but there is a deficit of evidence of the impact of learning outcomes with some exceptions, e.g. University of Georgia
- The US and Australia expect to see an increase in the use of learning analytics over the next five years, in particular a move to whole system models, personalisation and predictive analysis
- Concerns about the ethics of data usage among students
- Learning analytics still in relative infancy in UK but it can be expected that the pace of adoption will increase and widespread adoption can be expected within the next 3-5 years
- To be most cost effective analytics systems must be fit-for-purpose and appropriate to each student

Examples of data and analytics applications used

University	Data and analytics application
Nottingham Trent University	Student dashboard Business intelligence dashboard Predictive analysis
Open University	Identifying 'at-risk' students Access and recrutiment Developing an 'analytics mind-set'
University of Leeds	Client Relationship Management system
Oxford Brookes University	Dashboards
Purdue University	Signals project
University of Maryland, Baltimore County	Analysing usage of the VLE
New York Institute of Technology	Identifying at-risk students
California State University	Fine-graned analysis of student data
Marist College	Predictive modeling
Edith Cowan University	Enhancing retention
University of New England	Student engagement
University of Wollongong	Analysing social networks
Open Universities Australia	Personal pathway planning Input for curriculum redesign

Some ethical concerns and impacts

- Automated data collection needs to be handled with care – because it is available and accessible does not make it ethical
- Consider moral, ethical and legal implications of surveillance
- Understand the origin, properties and limits of a data set
- Consider power relations carefully
- Proper informed consent crucial to comply with moral, ethical and legal requirements
- Flows and integration of data from multiple sources needs to be handled with care to ensure integrity of the data
- Data used out of context may lose its meaning and/or value

Some ethical concerns and impacts (continued)

- Use of data outside the institution may not take context into account
- Big data may not be complete
- Aggregation of data may lead to use of data in an inappropriate context
- Automated analysis and machine learning algorithms should be checked for validity
- Be careful to cluster data in a way that loses the context
- Classify access to different types of data
- Availability of results to subjects of the analysis
- Develop policies and use best practices

Analytics gap

- With more access to useful data, institutions are **increasingly using sophisticated analytical methods**
- As a result of this, there is often a **gap between** the institution's **capacity to produce analytical results** and its **ability** to **apply** them **effectively to business issues**
- *Ransbotham, Kiron & Prentice (2015)* identified five ways for managers to **improve** their **comfort in consuming analytics**:
  - Bolstering your knowledge base
  - Building off prior experience
  - Creating analytical options
  - Capitalizing on domain knowledge
  - Recognizing the limitations of models

UNISA university of south africa

Skills gap (continued)

- Growing usage of big data analytics, social media platforms and mobile devices require employees to acquire skills in these areas (Capgemini Consulting):
  - **Mobile skills** – over 50% of organisations identified this as one of the two most important skills but more than 80% of organisations face a talent shortage in this area
  - **Analytical capabilities** – over 85% of organisations indicated that they have a big data initiative in place or are planning one but only 21% regarded their analytical capabilities as more than adequate
  - **Social media skills** – only 13% of organisations described their social media skills as advanced
  - **Robust understanding of business**
- By 2018 the United States alone could face a shortage of between 140 000 and 190 000 people with deep analytical skills as well as 1,5 million managers and analysts with the know-how to use the analysis of big data to make effective decisions *(McKinsey Global Institute*)

Traditional Institutional Research and Modern analytics

	More traditional institutional research	Modern analytics
Data	• Scarcity of data • Focus on data generation • Generally historical, trend or snapshot orientation	• Big data • Multiple data sources • Generally trend and real-time data
Skillset emphasis	• Science of measurement including instrument design and construct measurement • Data collection to ensure representativeness and generalisations	• Finding meaning in big data and making relevant connections • Programming skills • Multi-source data mining • Statistical modelling • Development of algorithms
Approach	• Multitude possible questions narrowed down to more specific research questions	• Big data not tailored to any questions, narrowed down to information needs for specific questions
Main complexities	• Limited granularity and ability to segment of due to small numbers • Representativeness of sample and ability to make inferences about the larger population	• More data = more noise, difficulty to determine which data is meaningful and what the patterns are reflecting • Missing values • Data quality
Driver	• Understanding what drives student learning and success	

Prinsloo, Archer, Barnes, Chetty & Van Zyl (2015)

UNISA university of south africa

Importance of change management

- *McAfee & Brynjolfsson (2012)* identified **change management** as crucial for success of big data initiatives.
- They identified five important **management challenges**:
  - Leadership
  - Talent management
  - Technology
  - Decision-making
  - Organisational culture

Should you do it?

- If you have an appropriate use for Big Data, go for it.
- However, remember it is not a silver bullet and will require new and innovative thinking to ensure a successful implementation.
- Make sure your ETL process and data cleansing is sufficiently applicable.
- Add analytic components as required.
- Remember the skills required will be important for success.

Thank you!

UNISA university of south africa