



**JISC** cetis

centre for educational technology  
& interoperability standards

**Analytics Series**

**Vol. 1, No 9.**

**A Brief History  
of Analytics**

By Adam Cooper (CETIS)

CETIS Analytics Series ISSN 2051-9214

Produced by CETIS for JISC

## A Brief History of Analytics

Adam Cooper (CETIS)

### Table of Contents

1. Introduction.....	3
2. The Key Questions of Analytics.....	3
3. Vignettes – Analytics Communities.....	5
3.1 Statistics.....	5
3.2 Business Intelligence (BI) .....	5
3.3 Web Analytics.....	6
3.4 Operational Research (OR) .....	7
3.5 Artificial Intelligence (AI) and Data Mining.....	8
3.6 Social Network Analysis (SNA) .....	10
3.7 Information Visualisation.....	11
4. Analytics Communities Rooted in the Education Sector .....	12
4.1 Bibliometrics & Scientometrics .....	12
4.2 Learning Analytics.....	13
5. Analytics Timeline .....	15
6. Bringing it All Together .....	15
7. Further Reading .....	16
8. References .....	18
About this White Paper.....	21
About CETIS.....	21

## 1. Introduction

The potential of analytics according to this definition is to help us to evaluate past actions and to estimate the potential of future actions, so to make better decisions and adopt more effective strategies as organisations or individuals. Analytics allows us to increase the degree to which our choices are based on evidence rather than myth, prejudice or anecdote.

Several factors are coming together at the moment to stimulate interest in making more use of analytics. One of these is the increased availability, detail, volume and variety of data from the near-ubiquitous use of ICT (Information and Communication Technology) throughout almost all facets of our lives. This aspect tends to be the focus of the news media but data alone is not enough to realise benefits from analytics. A less popularised factor driving effective exploitation of analytics is the rich array and maturity of techniques for data analysis; a skilled analyst now has many disciplines to draw inspiration from and many tools in their toolbox. Finally, the increased pressure on business and educational organisations to be more efficient and better at what they do adds the third leg to the stool: data, techniques, need.

This paper, one of the CETIS Analytics Series, is **aimed at readers who wish to be introduced to the range of techniques** that are being pieced together and labelled as Analytics. It does this by outlining some of the most important communities - each with their own origins, techniques, areas of limitation and typical question types - and suggests how they are contributing to the future, with special reference to the context of post-compulsory education.

The diversity and flexibility of some of the techniques lined up under the analytics flag is evidenced by the numerous different applications of analytics: financial markets, sports analytics, econometrics, product pricing and yield maximisation, fraud, crime detection, spam email filters, marketing, customer segmentation, organisational efficiency and even tracking the spread of infectious disease from web searches<sup>1</sup>. Behind these applications we can find the roots of analytics in the birth of statistics in the eighteenth century but since then different applications of statistics and IT have led to different communities of practice that now seem to be merging together. We see that Web Analytics pioneers are now exploiting data from the “social web” by using Social Network Analysis and that the techniques of Information Visualisation are supporting interactive and exploratory forms of analysis rather than just the graphs in management reports. Subjects that some see as old-hat such as Operational Research and others that are often perceived as futuristic such as Artificial Intelligence are each making contributions in surprising ways. Meanwhile, the education community has made its own contributions; Social Network Analysis and Artificial Intelligence have both emerged from academic research and we are now beginning to see sector-specific variants of analytics being put to work in the form of Educational Data Mining, Learning Analytics and Bibliometrics.

## 2. The Key Questions of Analytics

The specific questions that analytics can address are innumerable but it is helpful to organise them according to both the time dimension of the questions and their level of insight. These aspects, the Key Questions, are summarised on the matrix

in figure 1, which is adapted from the book “Analytics at Work” by Davenport et al [1].

Key Questions Positioning			
	Past	Present	Future
Information	Reports & Description	Alerting	Extrapolation
Insight	Models & Explanation	Recommendations	Prediction

Figure 1 - Key Questions Matrix from Davenport et al

The matrix distinguishes between more fact-based approaches (the “information row”) and approaches that are more targeted towards deeper understanding (the “insight” row) and segments these according to a temporal frame.

The questions are summarised as follows.

Questions of information and fact:

- What happened? Analytics produces reports and summarised descriptions of data, (the past).
- What is happening now? Analytics provides alerts in near-real time, (the present).
- Where are trends leading? Past data is extrapolated, (the future).

Questions of understanding and insight:

- How and why did something happen? Analytics builds models and explanation, (the past).
- What is the best next action? Analytics provides one or more recommendations, (the present).
- What is likely to happen? Analytics provides prediction, simulates the effect of alternative courses of action, or identifies an optimal course of action, (the future).

An understanding of the information / insight distinction is a critical one in working out how to maximise benefits from analytics. To quote from Davenport et al:

“You may find, for example, that many of your 'business intelligence' activities are in the top row. Moving from purely information-oriented questions to those involving insights is likely to give you a much better understanding of the dynamics of your business operations.”

Notice that this quotation refers to “better understanding”; we must in practice settle for incomplete or qualified insight since data is incomplete and contains errors and most analysed systems contain people and complexity that must be simplified in practical analytics.

The positioning matrix is used in the vignettes that follow to provide a qualitative indication of the balance of the key questions commonly addressed by the users of the technique described. This is done by showing a larger orange blob in those matrix cells that match the questions typically addressed by the technique being described. There will always be exceptions to this broad-brush characterisation.

### 3. Analytics Communities

This section outlines some of the most important communities and suggests how each is contributing to the future. The boundaries between these communities are fuzzy at best and sections are best understood as different perspectives rather than different kinds of analytics.

#### 3.1 STATISTICS

Statistical methods are a rather obvious underpinning for analytics but this fact has often been neglected, at least to judge by the marketing and publicity material, vendor analysis, etc. surrounding Business Analytics and Business Intelligence products and services. This neglect is now being challenged<sup>ii</sup>.

The mathematical foundations of statistics were laid in the 18th Century at which time the focus of interest was on using demographic and economic data to inform policy decisions (the Swedish Tabellverket is thought to be the first such data collection exercise in 1749<sup>iii</sup>). Statistics has come to be much more than tabulations and calculated means and medians; statistical methods for testing hypotheses, establishing confidence levels or the likelihood of a correlation are essential components for moving beyond fact-based reports to actionable insights.

Statistical methods are the fundamental basis for all of the “Key Questions” and many analytics problems can only be addressed with statistical methods. Fortunately, these statistical methods are well-established and they can sometimes be hidden inside software aimed at analytics novices, although any organisation that is serious about analytics should really have someone who understands the strengths and weaknesses of the statistical methods to help in proper interpretation; the tool that is so dull that you cannot cut yourself on it is not likely to be sharp enough to be either useful or helpful<sup>iv</sup>.

#### 3.2 BUSINESS INTELLIGENCE (BI)

As far as definitions go, “Business Intelligence” is essentially the same as “Analytics” but BI has its own character if we consider the typical applications and capabilities of the product category that is generally labelled “Business Intelligence”. One of the purposes of this “Brief History of Analytics” is to differentiate between the BI product category and other related fields that are converging as Analytics in 2012 and which the BI products are beginning to embrace or may do so in the future.

<p><b>Origins</b></p>	<p>H.P. Luhn is credited with inventing the term “business intelligence” in 1958<sup>v</sup> in an IBM Journal article called “A Business Intelligence System”, although its popularity stems from use by Gartner from 1989. The tradition that became BI is one of tools to make the production of management information reports easier and faster.</p>	<p><b>Aims and themes:</b> Empower users to access, combine and visualise data from a range of sources. Summarise key performance indicators in a form suited to continuous monitoring.</p>																
<p><b>Techniques &amp; tools</b></p>	<p>BI is typically concerned with data selection and presentation. Data is often extracted from “live” systems and organised into more user-friendly structures to allow “drill-down”, “roll-up” and “slice and dice” selection and filtering.</p> <p>The BI product category is characterised by quite large integrated software suites that work with structured data (such as from a mainstream database). Modern BI systems emphasise the online dashboard, a visual summary of key metrics with some personalisation and interaction possible.</p>	<table border="1" style="width: 100%; text-align: center;"> <thead> <tr> <th colspan="4">Key Questions Positioning</th> </tr> <tr> <th></th> <th>Past</th> <th>Present</th> <th>Future</th> </tr> </thead> <tbody> <tr> <th>Information</th> <td style="background-color: orange; width: 40px; height: 40px;"></td> <td style="background-color: orange; width: 20px; height: 20px;"></td> <td></td> </tr> <tr> <th>Insight</th> <td></td> <td></td> <td></td> </tr> </tbody> </table> <p>(shows mainstream)</p>	Key Questions Positioning					Past	Present	Future	Information				Insight			
Key Questions Positioning																		
	Past	Present	Future															
Information																		
Insight																		
<p><b>Trajectory</b></p>	<p>The major BI vendors are extending their offering in response to a widening of perspective on analytics, for example to offer support for analysis of unstructured data and for more sophisticated statistical and data-mining analysis.</p>	<p><b>Issues and Limitations:</b> BI products are often large and complex and can demand a large investment in people and IT before returns are realised.</p>																
<p><b>Examples</b></p>	<p>Mainstream: Microsoft BI is a good stereotype<sup>vi</sup>.</p> <p>Progressive: IBM's acquisition of SPSS (formerly the Statistical Package for the Social Sciences) and incorporation with its Cognos BI suite is a good example of where BI is going<sup>vii</sup>.</p> <p>Oracle and SAP form the remainder of the “big 4” BI vendors and SAS is particularly strong on advanced methods and statistics.</p>	<p>Basic BI offerings are strongly rooted in the tradition of reporting and dashboards and do not effectively address questions of insight.</p>																

**Relevance to the education sector:** the recently completed JISC Business Intelligence Programme<sup>viii</sup> (end August 2012) followed 11 Further and Higher Education Institutions as they explored more effective use of BI and the “Environment Scanning & Business Intelligence Infokit”<sup>ix</sup> produced to accompany it paint a good picture of early adoption of BI in the UK post-compulsory education sector. Acquiring a heavyweight BI suite is probably not an appropriate entry-point into analytics; there are likely to be cheaper and simpler routes to automated reporting and more incremental routes to increasing insight over information. Institutions with established BI may be able to follow the trajectory of their existing platform as the basis of a move towards greater insight.

### 3.3 WEB ANALYTICS

For many people, analytics *is* web analytics, thanks to the widespread use of tools such as Google Analytics that analyse and report on web page visits. Web analytics is divided into “on-site” and “off-site”, depending on whether the data is about activity on your own website or about activity occurring elsewhere on the web that is about your products and services.

<b>Origins</b>	Web analytics got started as soon as the public web began in the early 1990s, by using the automatic logs of page requests on web servers.	<p><b>Aims and themes:</b></p> <p>On site: Which pages do people visit? How does this change with date and time?</p> <p>Where do visitors come from (geographical)? Which site linked to us? What were the search terms that led people in? Is my site user-friendly?</p> <p>Off site: What is being said about us, or our products? What effect did our advertising have?</p> <table border="1" data-bbox="1011 745 1406 1068"> <thead> <tr> <th colspan="4">Key Questions Positioning</th> </tr> <tr> <th></th> <th>Past</th> <th>Present</th> <th>Future</th> </tr> </thead> <tbody> <tr> <th>Information</th> <td style="background-color: #f4a460; width: 40px; height: 40px;"></td> <td style="background-color: #f4a460; width: 20px; height: 20px;"></td> <td></td> </tr> <tr> <th>Insight</th> <td></td> <td></td> <td></td> </tr> </tbody> </table> <p><b>Issues and Limitations:</b></p> <p>Web analytics has to make assumptions about what a page visit actually means.</p> <p>BI suites with web analytics capabilities can be complex and there is a lack of mid-range tools available for users who grow out of Google Analytics.</p>	Key Questions Positioning					Past	Present	Future	Information				Insight			
Key Questions Positioning																		
	Past		Present	Future														
Information																		
Insight																		
<b>Techniques &amp; tools</b>	Web analytics began by using logs of each page visited but modern e-commerce sites often make use of very fine-grained “click-stream” data that records sub-page-level activity. Tracking cookies may be used to monitor multiple visits to the same site but can also be used to track activity across multiple websites, although this is widely considered to be unethical.																	
<b>Trajectory</b>	<p>Increasingly effective methods are being developed for “opinion mining” to identify negative online reviews for so-called reputation management.</p> <p>Web Analytics is being increasingly used alongside Social Network Analysis and Data Mining for Customer Analytics to provide integrated cross-channel market intelligence, campaign management, etc. (see following vignettes)</p>																	
<b>Examples</b>	<p>Mainstream: Google Analytics<sup>x</sup> is free, easy to start using and is beginning to include tools that have previously been the preserve of experts<sup>xi</sup>.</p> <p>Pioneering: Blue Fin Labs<sup>xiii</sup> are among the leaders in multi-channel web analytics.</p>																	

**Relevance to the education sector:** anyone with a web site should be making use of the mainstream analytics opportunities offered by Google Analytics etc., while recognising that the reports may need a skilled interpreter or supplementary analysis to reveal an actionable insight. Surveys by “higherexperts.com” indicate an increasing use of web analytics in the HE institutions they surveyed but also a significant under-exploitation of potential by those making some use of web analytics<sup>xiii</sup>. This under-exploitation can be largely remedied by staff development and a culture where insight leads to change. Off-site web analytics and opinion mining is worth investigating to better understand awareness and perception of your institution among potential students.

### 3.4 OPERATIONAL RESEARCH (OR)

Operational Research (or “Operations Research” in North America) is not so much referred to in the current “buzz” about analytics but it has some claim to be its progenitor; according to Wikipedia<sup>xiv</sup> it is “a discipline that deals with the application of advanced analytical methods to help make better decisions”. Operational Research may be tarnished in the minds of many due to its association with manufacturing but this should not detract from its intention to deal with complex and messy

real-world operational scenarios, especially but not exclusively in logistics, supply chain and process design and management.

<p><b>Origins</b></p>	<p>Statistical process control, pioneered by Walter Shewhart in the early 1920's, can be seen as the earliest attempt to understand how the components of a process contributed to product quality but it was not until 1937/8 when OR can be considered to have been born in the efforts of the Royal Air Force to integrate radar and ground observations into its warning and control system<sup>xv</sup>.</p> <p>After the war OR became applied to diverse industries throughout the 1950's and 1960's.</p>	<p><b>Aims and themes:</b></p> <p>How do operational factors influence real-world objectives and how can we design an operational environment to maximise (or minimise) critical objectives?</p> <table border="1" data-bbox="1029 633 1428 958"> <thead> <tr> <th colspan="4">Key Questions Positioning</th> </tr> <tr> <th></th> <th>Past</th> <th>Present</th> <th>Future</th> </tr> </thead> <tbody> <tr> <th>Information</th> <td></td> <td></td> <td></td> </tr> <tr> <th>Insight</th> <td style="background-color: #f4a460;"></td> <td></td> <td style="background-color: #f4a460;"></td> </tr> </tbody> </table> <p><b>Issues and Limitations:</b></p> <p>OR traditions come from a time when business culture and environment was different to what it is today. OR can easily become complex and costly.</p>	Key Questions Positioning					Past	Present	Future	Information				Insight			
Key Questions Positioning																		
	Past		Present	Future														
Information																		
Insight																		
<p><b>Techniques &amp; tools</b></p>	<p>OR relies heavily on both mathematical models and statistical methods but may have to resort to simulation methods to explain how several probabilistic sub-processes combine to create an overall effect that may not simply be due to a sum of the parts.</p> <p>It is an overtly practical discipline hence explicit account is taken of the human factors, such as how people actually use technology or interact in practice.</p>																	
<p><b>Trajectory</b></p>	<p>Operational Research as a profession entered the doldrums<sup>xvi</sup> during the latter quarter of the 20th Century but now organisations such as the Institute for Operations Research and Management Sciences are actively promoting the profession and it has published a magazine called "Analytics"<sup>xvii</sup> since 2008.</p>																	
<p><b>Examples</b></p>	<p>United Parcel Service (UPS) is a classic example of optimising parcel delivery logistics<sup>xviii</sup>.</p>																	

**Relevance to the education sector:** the OR community of practice provides inspiration for the application of analytics taking into account the practical, on-the-ground, reality of both social and technical aspects of operations. It is the practices rather than particular technologies that have most relevance and the US-based Association for Institutional Research<sup>xix</sup> is an example of an HE-specific community for Operational Research.

### 3.5 ARTIFICIAL INTELLIGENCE (AI) AND DATA MINING

Artificial Intelligence has something of a sci-fi reputation but AI research has produced some practical methods to allow machines (computers) to learn, in a limited sense of the word. Some of these methods became the basis for data mining, also known as "knowledge discovery in databases". Through these so-called "machine learning algorithms", computers are able to detect patterns in streams of complex and potentially incomplete input data. Many of the astonishing applications of predictive analytics in online and print literature rely on data mining.















<p><b>Origins</b></p>	<p>The term “artificial intelligence” was coined by John McCarthy in 1956 and began to be applied to “expert systems” in the early 1980’s. Data mining, a branch of AI, began to find use as a business tool from the early 1990’s when it was applied to retail basket analysis<sup>xx</sup>. Since then, an increasing range of new business opportunities have been discovered in the very large data-sets that are being accumulated through day-to-day business and consumer and personal activity.</p> <p>Educational applications of AI were explored by researchers through a series of conferences on AI in Education and then a dedicated conference on Educational Data Mining from 2007, now run by the International Educational Data Mining Society<sup>xxi</sup>.</p>	<p><b>Aims and themes:</b></p> <p>Data mining is generally concerned with patterns, e.g.: given a history of events and final outcomes, can we predict the outcome for an incomplete set of events? Are there clusters of some-how similar people or things? (The similarities might be outside human imagination or perception.)</p> <p>Educational data mining is principally concerned with Intelligent Tutoring Systems.</p>																
<p><b>Techniques &amp; tools</b></p>	<p>Data mining and AI involves heavily computational techniques using elaborate computer algorithms. Although these usually rely on statistical tests, data mining is a computer science rather than a mathematical science. Numerous data mining methods are well established in practical algorithms but research continues to find better methods and to improve their efficiency and reliability.</p>	<table border="1" data-bbox="1027 869 1422 1196"> <thead> <tr> <th colspan="4">Key Questions Positioning</th> </tr> <tr> <th></th> <th>Past</th> <th>Present</th> <th>Future</th> </tr> </thead> <tbody> <tr> <th>Information</th> <td></td> <td style="text-align: center;">■</td> <td></td> </tr> <tr> <th>Insight</th> <td style="text-align: center;">■</td> <td style="text-align: center;">■</td> <td style="text-align: center;">■</td> </tr> </tbody> </table>	Key Questions Positioning					Past	Present	Future	Information		■		Insight	■	■	■
Key Questions Positioning																		
	Past	Present	Future															
Information		■																
Insight	■	■	■															
<p><b>Trajectory</b></p>	<p>Data Mining has begun, but not completed, a transition from the research lab to practical use. There is an increasing list of standard methods that match the techniques to real-world problems.</p>																	
<p><b>Examples</b></p>	<p>Applications:</p> <ul style="list-style-type: none"> <li>• Market Basket Analysis<sup>xxii</sup> (e.g. Amazon “people who bought this item also bought...”)</li> <li>• Customer Relationship Management<sup>xxiii</sup>.</li> <li>• Insurance fraud (non-technical)<sup>xxiv</sup>.</li> </ul> <p>Software:</p> <ul style="list-style-type: none"> <li>• SAS Predictive Analytics and Data Mining<sup>xxv</sup>.</li> <li>• RapidMiner (open source)<sup>xxvi</sup>.</li> </ul>	<p><b>Issues and Limitations:</b></p> <p>Although easy-to-use software exists for data mining, correct use requires extreme care and expert knowledge.</p> <p>Many methods require large datasets.</p>																

**Relevance to the education sector:** It would probably be unwise for any organisation to attempt to adopt data mining techniques for practical use without already having established an analytics capability, no matter how appealing “predictive analytics” sounds. The relative immaturity of the education sector in this respect and the absence of data at the scale typical of much-publicised commercial-sector applications of data mining (e.g. ebay, Amazon) suggests that a period of experimentation within the sector would be more appropriate than adoption as a core business tool. Most educational establishments are more likely to benefit from data mining when it is used by suppliers of large scale, and probably cloud-computing based, applications or content services.

### 3.6 SOCIAL NETWORK ANALYSIS (SNA)

Social Network Analysis is concerned with the relationships between people, often captured in their communication networks but also in other forms of interaction such as trading. These relationships may be direct or technology-mediated and might be an active expression of some kind of “tie” between two people, something more than an interaction. Social Network Analysis may be applied to activity in social networking websites and web-based services but there are many other applications and it may even be applied to relationships between people and things.

The term “Organisational Network Analysis” is sometimes used to apply to SNA within a formal organisation, a setting where the technique can expose striking differences between how the organisation is intended to work and how it actually works.

<p><b>Origins</b></p>	<p>Social Network Analysis grew out of the discipline of sociology, with the modern conception emerging from the 1950's. Since then these methods have been adopted outside academic sociology, for example business studies, although often without great attention to theories such as “social capital”<sup>xxvii</sup>.</p>	<p><b>Aims and themes:</b> SNA typically asks: Who is influential? Who is powerful (who is controlling the flow of information)? What sub-groups, or cliques, exist? Who is engaged or disengaged?</p>																
<p><b>Techniques &amp; tools</b></p>	<p>SNA typically involves the calculation of various metrics at both individual and group level, for example to indicate how many of the shortest communication paths between two people pass through a third, and the visualisation of the social network as a “sociogram” (see below and right). The data processing is relatively simple and there are numerous computer programs available.</p>	<table border="1" data-bbox="1027 945 1423 1267"> <thead> <tr> <th colspan="4">Key Questions Positioning</th> </tr> <tr> <th></th> <th>Past</th> <th>Present</th> <th>Future</th> </tr> </thead> <tbody> <tr> <th>Information</th> <td style="text-align: center;"></td> <td style="text-align: center;"></td> <td style="text-align: center;"></td> </tr> <tr> <th>Insight</th> <td style="text-align: center;"></td> <td></td> <td></td> </tr> </tbody> </table>	Key Questions Positioning					Past	Present	Future	Information				Insight			
Key Questions Positioning																		
	Past	Present	Future															
Information																		
Insight																		
<p><b>Trajectory</b></p>	<p>Social network analysis has become increasingly popular due to the rise of the “social web”, notably twitter™ and is becoming embedded in, for example, web analytics targeted at marketing and advertising.</p> <p>Sociology academics are beginning to establish techniques to numerically model the effect of different social theories on the form of social network that emerges; this may allow us to gain deeper explanatory insight if applied to operational SNA.</p> <p>The concept of “organisational network analysis” as a practical business tool is gaining traction, as evidenced by the arrival of companies such as Optimice (see examples) and conferences such as “Circuits of Profit”<sup>xxviii</sup>.</p>	<p><b>Issues and Limitations:</b> SNA is usually a descriptive form of analysis; semi-quantitative answers to the above questions are possible. Actionable insight is constrained by limited explanatory power, the complexity of real social systems, and because many social theories are contested.</p>																
<p><b>Examples</b></p>	<p>In-Maps<sup>xxix</sup> allows you to create a sociogram of your Linked-In contacts.</p> <p>Optimice<sup>xxx</sup> is a commercial platform of tools and techniques for organisational network analysis.</p> <p>SNAPP<sup>xxxi</sup> (Social Networks Adapting Pedagogical Practice)</p>	<p>A complete analysis usually requires data about interactions that are not automatically-tracked.</p>																

analyses communication patterns in virtual learning environments.

KXEN InfiniteInsight™<sup>xxxii</sup> is SNA combined with predictive analytics and targeted at customer analytics.

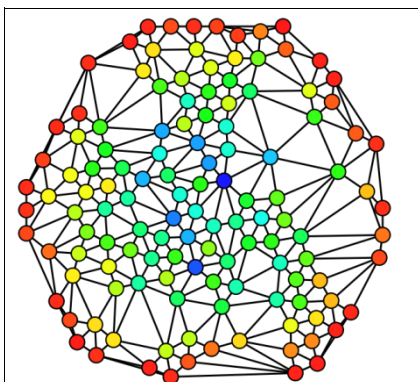


Figure 2: A Sociogram (by Claudio Rocchini)

**Relevance to the education sector:** SNA has potential to be applied wherever people interact, but particularly where there are traces left by their communications. It could, for example, help to uncover the potential significance of twitter in relation to marketing or student satisfaction, or to monitor collaborative online learning activities. It could also be used as the basis for a grounded re-structuring of departments/schools/faculties according to actual patterns of collaboration among staff.

### 3.7 INFORMATION VISUALISATION

Information Visualisation naturally occurs across all of the previous sections and yet it has been, and continues to be, a particular expertise and a specific target of innovation.

Information Visualisation is specifically defined in Wikipedia<sup>xxxiii</sup> as being concerned with visual representation of *non-numerical information*. Most times, however, the term is used to include non-numerical and numerical information. Popular appreciation of information visualisation is being shaped by the work of “data journalists” such as David McCandless<sup>xxxiv</sup> and by common use of “info-graphics” in the news media.

<p><b>Origins</b></p>	<p>The earliest use of statistical graphics is generally attributed to William Playfair (1786) but it was probably the influential statistician John Tukey who first guided us to a more powerful use of visualisation in his book “Exploratory Data Analysis” in 1977<sup>xxxv</sup>. Tukey emphasised the value of using visualisations to help in hypothesis forming prior to formal hypothesis testing.</p> <p>The most influential figure behind current practice is probably Edward Tufte<sup>xxxvi</sup>, author of the classic book “The Visual Display of Quantitative Information” in 1983<sup>xxxvii</sup>.</p>	<p><b>Aims and themes:</b></p> <p>Information visualisation seeks to make the interpretation of abstract or inter-related facts more intuitive. Effective information visualisation exploits innate aspects of visually-stimulated cognition.</p> <p><b>Issues and Limitations:</b></p> <p>Common pitfalls are to mis-apply visualisation techniques by focussing on visual attractiveness rather than supporting understanding through “visual reasoning”.</p>
<p><b>Techniques &amp; tools</b></p>	<p>The techniques and tools for information visualisation are diverse, stretching from hand-crafted work by graphic designers through to fully-automatic production.</p>	
<p><b>Trajectory</b></p>	<p>Modern analytics software is beginning to make effective use of information visualisation within the analytical/discovery process</p>	

	rather than purely as a means to report findings. In this respect, the analytics community is effectively re-discovering Tukey's exploratory data analysis.	
<b>Examples</b>	<p>Tableau<sup>xxxviii</sup> is a good indicator of commercial visualisation software targeted specifically at the analytics market.</p> <p>Google charts<sup>xxxix</sup> is a free-to-use toolkit for web-based presentation and has begun to support dashboard interaction.</p> <p>Gephi<sup>xl</sup> is a sophisticated open source tool for visualisation of a common kind of non-numerical information: relationships (see also Social Network Analysis).</p>	

**Relevance to the education sector:** focus on visualisation that is effective in supporting actionable insight through visual reasoning and making use of visual tools as part of the analytical process.

## 4. Analytics Communities Rooted in the Education Sector

There are some home-grown communities in the education sector that should be mentioned in this Brief History since they are likely to shape the way post-compulsory education adopts analytics.

The Educational Data Mining community has already been referred to in the vignette on Artificial Intelligence and Data Mining since it grew out of that wider research community but there are two further home-grown cases, which are summarised below.

### 4.1 BIBLIOMETRICS & SCIENTOMETRICS

Scientometrics is literally the measurement of science, with “science” to be understood in a very general sense to include diverse forms of scholarship and technology development, and is frequently undertaken through analysis of citations and co-authorship in scholarly works, i.e. “bibliometrics”.

<b>Origins</b>	Bibliometrics may be considered to be rooted in the creation of the Science Citation Index in 1961, which is currently owned by the business information giant Thomson Reuters.	<b>Aims and themes:</b> What impact or influence does a research output have? How are ideas developing in a discipline?  How are national research and innovation strategies performing? In which fields do different nations have strategic advantage?																
<b>Techniques &amp; tools</b>	The techniques of Social Network Analysis are particularly relevant but scientometrics also makes use of temporal, geospatial and topical (i.e. subject-matter) analytical methods. It makes strong use of information visualisation techniques to permit non-quantitative aspects to be interpreted.																	
<b>Trajectory</b>	Techniques are being devised to go beyond the traditional journal citation approach, to allow for richer kinds of relationships to be accounted for. Data-sets are an increasingly important public output from research which, along with software, may be widely used but are not tracked in the same way.  The bibliometrics/scientometrics community is increasingly addressing questions about the themes of research and of the development of ideas.	<table border="1" data-bbox="1027 678 1420 1003"> <thead> <tr> <th colspan="4">Key Questions Positioning</th> </tr> <tr> <th></th> <th>Past</th> <th>Present</th> <th>Future</th> </tr> </thead> <tbody> <tr> <th>Information</th> <td style="background-color: orange;"></td> <td></td> <td></td> </tr> <tr> <th>Insight</th> <td></td> <td></td> <td></td> </tr> </tbody> </table>	Key Questions Positioning					Past	Present	Future	Information				Insight			
Key Questions Positioning																		
	Past	Present	Future															
Information																		
Insight																		
<b>Examples</b>	The Sci <sup>2</sup> Tool <sup>xii</sup> is a tool-set for “temporal, geospatial, topical, and network analysis and visualization of scholarly datasets at the micro (individual), meso (local), and macro (global) levels.”  Scholarometer <sup>xiii</sup> “Academic Impact Analysis” sidebar extension for Firefox and Chrome.	<b>Issues and Limitations:</b> Bibliometrics/scientometrics analysis typically lacks an explanatory or predictive component; post-hoc evaluation dominates. Data sources other than citation indexes are not easily acquired.																

To date scientometrics has been little used for practical, as opposed to research, purposes with the notable exceptions of journal impact factor and research evaluation which are of disputable validity[2]. The methods of scientometrics offer some potential as a starting point to answer finer-grained questions at researcher or research-group level or to understand the effect of using new forms and newer modalities of scholarly communication such as open journals, data publication, podcasting for a general audience, etc. Scientometrics/bibliometrics falls under the more general heading of research management analytics, which is explored in more detail in another paper in the CETIS Analytics Series[3].

#### 4.2 LEARNING ANALYTICS

Learning Analytics as a term has been significantly popularised by the International Conferences on Learning Analytics and Knowledge (LAK), the first of which described learning analytics as “the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs.” Whereas we do not adopt this definition in the CETIS Analytics Series, there is certainly a community that identifies itself with the term Learning Analytics.

<b>Origins</b>	Although the first LAK Conference was in 2011 <sup>xliii</sup> , the emergence of Learning Analytics can be seen to be foreshadowed by papers from the Educause Learning Initiative in 2005 and it has aspects in common with Educational Data Mining[4].	<b>Aims and themes:</b> Within the field of learning analytics, attention has so far focussed particularly on matters of students at risk of dropping-out or under-performing and on learning resource recommender systems.
<b>Techniques &amp; tools</b>	A wide range of tools are being co-opted, from across many of the other communities referred to in this “Brief History”.	
<b>Trajectory</b>	<p>Learning Analytics is currently in its infancy and is attracting a range of creative new ideas as well as debate about ethical and other “soft” issues.</p> <p>It can be seen as an intra-institutional focus on data analysis of a kind that has previously been more coarse-grained and externally-processed, for example the Fischer Family Trust<sup>xliiv</sup> data analysis project for schools.</p>	<p><b>Issues and Limitations:</b> Learning Analytics has yet to become established in practice.</p> <p>There is a danger of disillusionment since some software that is being badged as “learning analytics” is little more than a make-over of basic tracking and statistical tools.</p>
<b>Examples</b>	Learning Analytics has been thoroughly reviewed in “The State of Learning Analytics in 2012: A Review and Future Challenges”[6] and is considered in another paper in the CETIS Analytics Series [5].	

## 5. Analytics Timeline

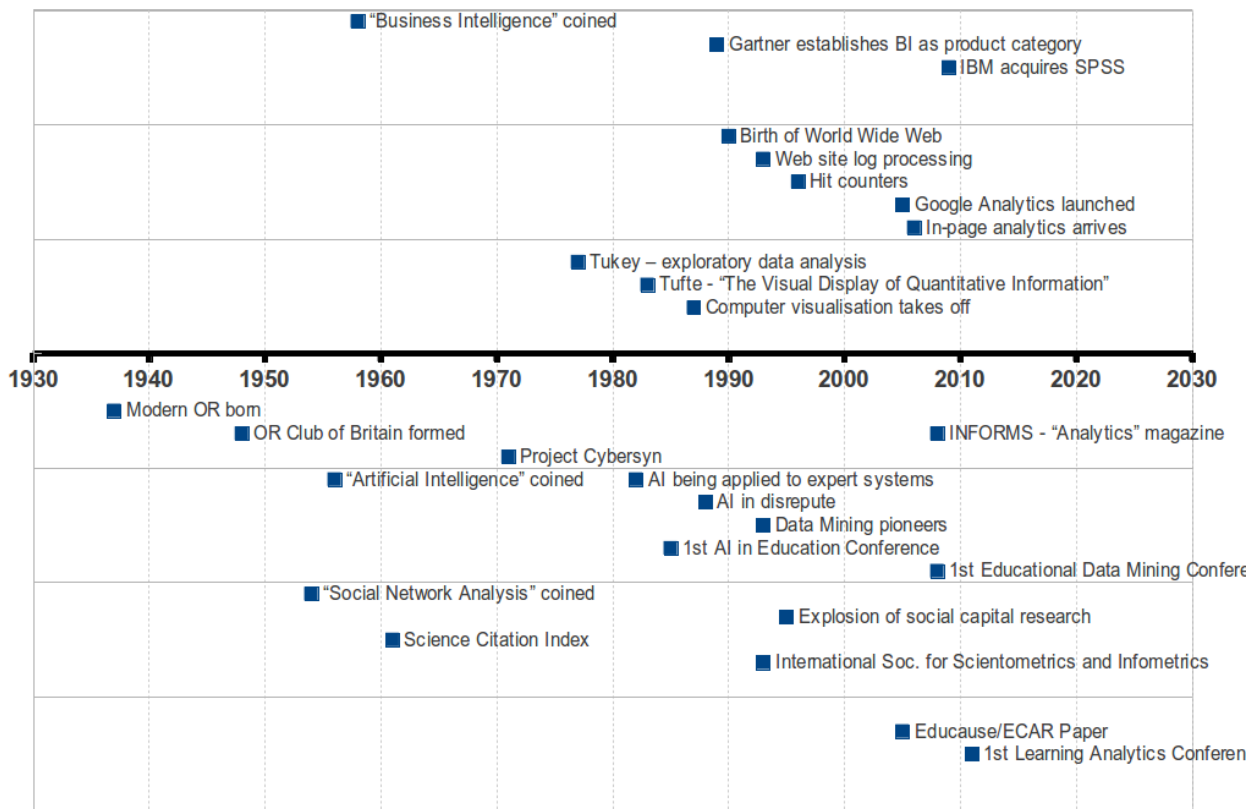


Figure 3: An Approximate Timeline of the Brief History of Analytics

What this timeline misses out is the long history of analytics that stretches way before the computer age. We should not forget that we are following in the footsteps of pioneers such as Florence Nightingale<sup>xlv</sup> and John Snow<sup>xlvi</sup> who both broke new ground in making the case for sanitary reform on the basis of statistical evidence in the middle of the 19th Century. Their actionable insights were based on much more simple analysis and much more limited data than we have available to us now and we would be wise to avoid assuming that analytics implies big data and the latest predictive data mining algorithms.

## 6. Bringing it All Together

This brief history has attempted to show the range of approaches that have contributed to the state of the art of analytics but even the largest and most progressive organisations cannot do it all. The most effective users of analytics select from this repertoire those techniques that best match their operational or strategic targets and what they have, or can build-up or acquire, in terms of data, technology, organisational culture, and skills, including leadership.

The message of this paper is not that doing analytics means embracing all of these techniques but that the next step from using management information systems as we do today, involves a range of options, some of which are lateral rather than evolutionary developments. The Further Reading section should help in exploring these in some more detail.

## 7. Further Reading

All of the website addresses in the Further Reading section and the footnotes are provided online at <http://delicious.com/bhofa> for convenient access.

### **Statistics:**

The Gapminder Foundation has some excellent videos about statistics, in particular the BBC documentary “The Joy of Stats” in which Hans Rosling covers the origin and applications of statistics to “replace devastating myths with a fact-based world view”. <http://www.gapminder.org/videos/the-joy-of-stats/>.

### **Business Intelligence:**

Gartner has an extensive selection of resources dealing with Business Intelligence - <http://www.gartner.com/technology/core/products/research/topics/businessIntelligence.jsp>.

Their “magic quadrant” for Business Intelligence was published in February 2012 - <http://www.gartner.com/technology/reprints.do?id=1-196WFFX&ct=120207&st=sb>.

### **Web Analytics:**

SEO Mix Tour newsletter: How to use web analytics to improve your web marketing strategy - <http://www.webceo.com/newsletter/issue-40/> (introductory).

Wikipedia - [http://en.wikipedia.org/wiki/Web\\_analytics](http://en.wikipedia.org/wiki/Web_analytics) (comprehensive).

### **Operations Research:**

The Wikipedia entry for Operations Research provides a good overview of the applications and development of OR – [http://en.wikipedia.org/wiki/Operations\\_research](http://en.wikipedia.org/wiki/Operations_research).

Project Cybersyn - [http://en.wikipedia.org/wiki/Project\\_Cybersyn](http://en.wikipedia.org/wiki/Project_Cybersyn) – is a historical footnote, an attempt at real-time analytics for the Chilean state-run sector from the early 1970's.

### **Artificial Intelligence and Data Mining:**

The American Association for Artificial Intelligence website has some clear descriptions and general readings. For example: data mining - <http://aaai.org/AITopics/DataMining>; intelligent tutoring systems – <http://aaai.org/AITopics/Education>.

Jean-Francois Belisle, “An Introduction to Data Mining for Marketing and Business Intelligence” - <http://jfbelisle.com/2011/03/an-introduction-to-data-mining-for-marketing-and-business-intelligence/>.

### **Social Network Analysis:**

Marin and Wellman, “Social Network Analysis: An Introduction” - <http://www.chass.utoronto.ca/~wellman/publications/newbies/newbies.pdf> – is a thorough exposition on the concepts and techniques of SNA.

Optimice white paper on SNA as an alternative to business process analysis - <http://www.optimice.com.au/documents/BusinessValueAnalysis.pdf>.

### **Information Visualisation:**



Stephen Few has written a number of books as well as published articles and examples on his website - <http://www.perceptualedge.com/>, that emphasise visual design to take advantage of human perception and embrace the principle of effective simplicity.

### **Bibliometrics/Scientometrics:**

JISC-funded TIDSR project (2008-9), “Toolkit for the Impact of Digitised Scholarly Resources” - <http://microsites.oii.ox.ac.uk/tidsr/welcome>.

Indiana University Cyberinfrastructure for Network Science Center “Science of Science” portal provides a range of resources on scientometrics – <http://sci.slis.indiana.edu/>.

The Wikipedia article on bibliometrics indicates history and current uses - <http://en.wikipedia.org/wiki/Bibliometrics>.

### **Learning Analytics:**

EDUCAUSE has published a number of papers since 2005 dealing with what they call “academic analytics” - and a more recent collection on Learning Analytics – <http://www.educause.edu/Resources/Browse/Learning%20Analytics/39193>.

The EDUCAUSE Review article “Penetrating the Fog: Analytics in Learning and Education” by George Siemens and Phil Long - <http://www.educause.edu/EDUCAUSE+Review/EDUCAUSEReviewMagazineVolume46/PenetratingtheFogAnalyticsinLe/235017>, gives a good overview of drivers and issues for Learning Analytics.

### **Other:**

A humorous take on the history of analytics can be found in the SAS (analytics software vendor) magazine - [http://www.sas.com/news/sascom/2008q4/column\\_bestblogs.html](http://www.sas.com/news/sascom/2008q4/column_bestblogs.html).

### **CETIS Analytics Series**

Other publications in the CETIS Analytics Series that pick up where this Brief History ends are:

- A more technical look at the tools of analytics [7].
- The capabilities – skills and literacies rather than IT - required within an educational organisation to apply analytics are explored through case studies [8].

Three papers in the CETIS Analytics Series explore the state of the art in each of three domains of interest within an educational institution:

- The teaching and learning domain [5].
- Research management [3].
- Whole-institutional concerns [9].

## 8. References

- [1] T. H. Davenport, J. G. Harris, and R. Morison, *Analytics at Work: Smarter Decisions, Better Results*. Harvard Business Press, 2010.
- [2] L. Leydesdorf, "The Evaluation of Research and the Scientometric Research Program: Historical Evolution and Redefinitions of the Relationship," *Studies in Science of Science*, vol. 22, no. 3, 2004.
- [3] M. van Harmelen, and D. Workman, "Analytics for Understanding Research", *CETIS Analytics Series*, vol. 1, no. 4, 2012.
- [4] G. Siemens and R. S. J. Baker, "Learning Analytics and Educational Data Mining : Towards Communication and Collaboration," in *LAK12*, 2012.
- [5] M. van Harmelen, "Analytics for Learning and Teaching", *CETIS Analytics Series*, vol. 1, no. 3, 2012.
- [6] R. Ferguson, "The State of Learning Analytics in 2012: A Review and Future Challenges (Techreport ID: kmi-12-01)," 2012.
- [7] W. Kraan and D. Sherlock, "Infrastructure and Tools for Analytics," *CETIS Analytics Series*, vol. 1, no. 11, 2012.
- [8] S. MacNeill and S. Powell, "Institutional Readiness for Analytics," *CETIS Analytics Series*, vol. 1, no. 8, 2012.
- [9] D. Kay, "Analytics for the Whole Institution; Balancing Strategy and Tactics," *CETIS Analytics Series*, vol. 1, no. 2, 2012

---

<sup>i</sup> <http://www.plosntds.org/article/info:doi/10.1371/journal.pntd.0001206>

<sup>ii</sup> See, for example IBM's suggestion that statistics should be incorporated into core business reporting: <http://www-01.ibm.com/software/analytics/business-intelligence/statistics/> .

<sup>iii</sup> [http://en.wikipedia.org/wiki/Statistics\\_Sweden](http://en.wikipedia.org/wiki/Statistics_Sweden)

<sup>iv</sup> To quote John Tukey, one of the prominent statisticians of the 20th century. See <http://cm.bell-labs.com/cm/ms/departments/sia/tukey/memo/techtools.html>

<sup>v</sup> [http://en.wikipedia.org/wiki/Hans\\_Peter\\_Luhn](http://en.wikipedia.org/wiki/Hans_Peter_Luhn)

<sup>vi</sup> <http://www.microsoft.com/en-us/bi/Capabilities.aspx>

<sup>vii</sup> <http://www-01.ibm.com/software/analytics/spss/>

<sup>viii</sup> <http://www.jisc.ac.uk/whatwedo/programmes/businessintelligence.aspx>

<sup>ix</sup> <http://www.jiscinfonet.ac.uk/infokits/business-intelligence/>

<sup>x</sup> <http://www.google.com/analytics/>

---

<sup>xi</sup>For example, cross-channel analysis, <http://www.google.com/analytics/features/multichannel-funnels.html>

<sup>xii</sup><http://bluefinlabs.com/>

<sup>xiii</sup>[http://media.higheredexperts.com.s3.amazonaws.com/state\\_analytics\\_2012.pdf](http://media.higheredexperts.com.s3.amazonaws.com/state_analytics_2012.pdf)

<sup>xiv</sup>[http://en.wikipedia.org/wiki/Operations\\_research](http://en.wikipedia.org/wiki/Operations_research)

<sup>xv</sup><http://people.brunel.ac.uk/~mastjib/jeb/or/intro.html>

<sup>xvi</sup><http://connection.ebscohost.com/c/articles/6690491/or-ms-dead-dying-rx-survival>

<sup>xvii</sup><http://www.analytics-magazine.org/>

<sup>xviii</sup><http://www.analytics-magazine.org/march-april-2010/154-corporate-profile-analytics-at-ups>

<sup>xix</sup><http://www.airweb.org>

<sup>xx</sup>The 1993 paper by Rakesh Agrawal - <http://rakesh.agrawal-family.com/papers/sigmod93assoc.pdf> – was stimulated by appreciation that the retailer Marks and Spencer was sitting on copious latent intelligence.

<sup>xxi</sup><http://www.educationaldatamining.org/>

<sup>xxii</sup><http://www.information-management.com/specialreports/20061031/1067598-1.html>

<sup>xxiii</sup><http://www.spss.ch/eupload/File/PDF/The%20Top%2010%20Secrets%20to%20Using%20Data%20Mining%20to%20Succeed%20at%20CRM.pdf>

<sup>xxiv</sup><http://www.iso.com/Research-and-Analyses/ISO-Review/Data-Mining-Essential-for-Exposing-Fraud.html>

<sup>xxv</sup><http://www.sas.com/technologies/analytics/datamining/index.html>

<sup>xxvi</sup><http://rapid-i.com/content/view/181/190/lang,en/>

<sup>xxvii</sup><http://www.socialcapitalresearch.com/>

<sup>xxviii</sup><http://circuitsoprofit.com/>

<sup>xxix</sup><http://inmaps.linkedinlabs.com/>

<sup>xxx</sup><http://www.optimice.com.au/>

<sup>xxxi</sup><http://research.uow.edu.au/learningnetworks/seeing/snapp/index.html>

<sup>xxxii</sup><http://www.kxen.com/Products/Social+Network+Analysis>

<sup>xxxiii</sup>[http://en.wikipedia.org/wiki/Information\\_visualization](http://en.wikipedia.org/wiki/Information_visualization)



## About the Author

Adam joined CETIS in 2006 and is currently one of the co-directors, with responsibility for CETIS's interoperability standards and technology strategy work. He has a long-standing interest both in quantitative modelling and in the appropriation of technology to support educational aims. He has worked on educational technology and innovation in both Further and Higher Education Institutions in the UK and in the private sector during which time he has taught, managed learning technology development, designed and built software, conducted R&D and analysed technology trends and their implications.

## About this White Paper

Title: CETIS Analytics Series Volume 1, No 9: A Brief History of Analytics

Author: Adam Cooper

Date: November 2012

URI: <http://publications.cetis.ac.uk/2012/529>

ISSN 2051-9214

Text Copyright © 2012 The University of Bolton; cover image courtesy of JISC



This work is licensed under the Creative Commons Attribution 3.0 UK Licence. For more information on the JISC CETIS publication policy see [http://wiki.cetis.ac.uk/JISC\\_CETIS\\_Publication\\_Policy](http://wiki.cetis.ac.uk/JISC_CETIS_Publication_Policy)

Published by The University of Bolton

## About CETIS

CETIS are globally recognised as leading experts on interoperability and technology standards in learning, education and training. We work with our clients and partners to develop policy and strategy, providing impartial and independent advice on technology and standards. CETIS are active in the development and implementation of open standards and represent our clients in national, European and global standards bodies and industry consortia, and have been instrumental in developing and promoting the adoption of technology and standards for course advertising, open education resources, assessment, and student data management, opening new markets and creating opportunities for innovation.

For more information visit our website: <http://jisc.cetis.ac.uk/>

This white paper has been produced by CETIS for JISC: <http://www.jisc.ac.uk>