



**JISC** cetis

centre for educational technology  
& interoperability standards

**Analytics Series**

**Vol.1, No.11.**

**Analytics  
Tools and  
Infrastructure**

By Wilbert Kraan and David Sherlock  
(CETIS)

CETIS Analytics Series: ISSN 2051-9214

Produced by CETIS for Jisc

## Analytics Tools and Infrastructure

Wilbert Kraan and David Sherlock (CETIS)

### Table of Contents

1. Introduction .....	4
2. Background .....	4
2.1 The Analytic Communities.....	4
2.2 Analytics Workflow.....	5
2.3 Analytics Area .....	6
2.4 Data Sources .....	7
3. Communities and Their Tools.....	7
3.1 Business Intelligence Community.....	7
3.2 Business IntelligenceTools.....	8
3.3 Web Analytics .....	9
3.4 Web Analytics Tools.....	10
3.5 Information Visualisation .....	11
3.6 Information Visualisation Tools .....	12
3.7 Operational Research .....	14
3.8 Operational Research Tools.....	14
3.9 Data Mining.....	15
3.10 Data Mining Tools .....	16
3.11 Social Network Analysis .....	17

3.12	Social Network Analysis Tools .....	18
3.13	Learning Analytics.....	19
3.14	Learning Analytics Tools .....	20
4.	Summary.....	21
5.	References.....	22

## 1. Introduction

Analytics is notable in that it is a headline grabbing trend in many domains, but has also been around for a long time under various other labels. One consequence of that longevity is that there is a bewildering array of tools available that can support an analytics process in some way.

An exhaustive overview of all such tools is near impossible, and probably out of date the moment it's finished. What is possible, however, is to provide a map of the major categories of tools, and highlight some landmark tools that are available now.

Because of the diverse history and practice of analytics, many different categorisations are possible, but we choose to group them by tradition, or established approach. One reason is that such an approach makes tools more easily comparable, because they have been developed to meet the needs and expectations of their communities over time. The other reason is that it tallies closely with other papers in the CETIS Analytics Series<sup>1</sup> of which this briefing is a part.

## 2. Background

The tools we consider here have been organised into communities based upon tradition or established approach. This briefing paper also explores how they fit into an analytics workflow, the data sources they are designed to deal with and any alternatives that are available. All website addresses referred to can be found online at [https://delicious.com/atai\\_cetis](https://delicious.com/atai_cetis) for convenience.

### 2.1 THE ANALYTIC COMMUNITIES

The tools and infrastructure discussed here are split into broad groups based upon the communities identified in 'A Brief History of Analytics', which is part of the CETIS Analytics Series (Cooper, 2012b). A Brief History's "Analytics at Work" matrix categorises analytic approaches by the time frame a community deals with and whether a community is interested in fact based questions (information) or questions regarding understanding (insight). It can also be read as going from the relatively simple in the 'reports & description' cell to the most sophisticated techniques in the 'prediction' cell.

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

Fig 1: The "Analytics at Work" Matrix

<sup>1</sup><http://publications.cetis.ac.uk/c/analytics>

## 2.2 ANALYTICS WORKFLOW

Even though there are a number of different approaches to analytics, they all share a recognisably similar set of steps. The steps themselves are relatively simple, but the success of the analysis does hinge on having a very clear idea of what the goal of the exercise is.

Identifying the goal of an analytics workflow is never just a matter of technology. Selecting a goal such as increasing student success, and then finding a quantifiable indicator for that goal such as graduation rates or marks involves more than tool configuration. An organisation-wide consensus around the goal is needed, as is a deep, non-reductive understanding of the process of achieving it. Only then does a consideration of quantifiable variables come into the picture.

Nonetheless, once a goal and indicative variables have been decided on, a logical analytics tool workflow can be identified. For the purposes of comparing analytics tools, as well as relating particular tools to different conceptions of analytics, it is useful to plot a variety of analytics process conceptions against each other and distill a generic analytic workflow from the m. As such, the analytics tool workflow can be thought of as a column with each step as a row that links to other conceptions of analytics flows such as those identified by Elias (2011):

Knowledge Continuum	Five steps of analytics	Web analytics objectives	Collective applications model	Processes of Learning Analytics	Analytics tool workflow
Data	Capture	Define goals	Select	Select	
		Measure	Capture	Capture	Collection and acquisition
Aggregate	Aggregate & report		Storage		
			Cleaning		
Information	Report			Integration	
Knowledge	Predict		Process	Predict	Analysis
Wisdom	Act	Use	Display	Use	Representation & visualisation
	Refine			Refine	
		Share		Share	Alerting

Table 1 Various analytics approaches mapped to an analytics tool workflow

For the purposes of tool comparison, it is worth going into some further detail about each step of the analytics tool workflow:

**Collection & Acquisition:**

The extraction of source data

**Storage:**

Storing source data in a data warehouse, where relevant

**Cleaning:**

Rectifying anomalies and inconsistencies, and normalising the syntax of the data

**Integration:**

Aligning the data to either existing datasets, or a common vocabulary

**Analysis:**

Analyse the data, in order to build descriptive or predictive models

**Representation and Visualisation:**

Creating reports and diagrams that illustrate the models for a wider audience

**Alerting:**

Operationalising the models to (near) real time to enable alerting of relevant stakeholders



Fig 2 Example of tool used for analysis, representation and visualisation flow

In the remainder of the briefing, we shall use a shaded area on this flow as an indicator of which part of the analytics process is covered by a particular tool. Figure 2 shows an example flow describing a tool where analysis, representation and visualisation processes are covered, this is shown by the area covered by red.

### 2.3 ANALYTICS AREA

Not all analytics traditions and tools are relevant to all areas where analytics might be applied. A business intelligence tool that is fine tuned for financial market predictive insights won't be much use when analysing student engagement in VLE forums, for example. For that reason, we'll also try to position each tool with regard to the major analytics areas relevant to higher and further education.

In the 'What is Analytics?' instalment of the CETIS Analytics series (Cooper, 2012a), three broad domains of analytics in higher and further education are identified, that we'll re-use here:

1. Domain of Learning Analytics: insights to support education aims and objectives
2. Domain of Academic Analytics: Insights to support operational and financial activity

- 3. Domain of Research Management Analytics: insights to support the management and funding of research

## 2.4 DATA SOURCES

A key point of differentiation between analytics tools is the range of data sources they are designed to deal with. Some tools deal with the data of only one particular system, others with a small, predetermined set, and yet more deal with virtually any sort of data source, provided there is some custom integration. The variance is partially inherent in the technical difficulties of integrating different data sets, and partially to do with how many different systems a vendor sells, and also any collaboration agreements between different vendors.

It is, for example, relatively easy for a VLE vendor to offer an analytics capability on top of the data the VLE itself collects or generates. Doing the same for a student record system that may not provide access to the data, and the shape of whose data is unknown, is another matter.

How serious a limited set of data sources is, depends on the goal of the analytics exercise: if the goal is to optimise student engagement with online resources, a small number of data sources may be fine. If the goal is to optimise the way the institution manages relations with students over the whole lifecycle, a small set of sources is a problem.

Custom integrations can be a solution to increase the number of data sources in the mix, but tools can also differ in their ability to integrate multiple, possibly unknown data sources. This will be noted, where relevant.

## 3. Communities and Their Tools

### 3.1 BUSINESS INTELLIGENCE COMMUNITY

Business intelligence (BI) can be considered ‘analytics before it was called analytics’. Though recent advances in size and scope of data analysis capabilities (“big data”), and the rise of real-time analysis means that there is a difference in emphasis, analytics and BI still make use of similar tools and techniques.

With regard to tools, the important aspect of BI is that the tools are designed to be general purpose, and are sold by major vendors in all the vertical markets in which they participate. This means that a degree of customisation is likely to be necessary, even if some education specific features are built in, but the trade-off is that the number of data sources that can be included is very wide.

Key Questions Positioning			
	Past	Present	Future
Information	■	■	
Insight			

Figure 3: Key Questions positioning for Business Intelligence, Related: Data mining, information visualisation

### 3.2 BUSINESS INTELLIGENCE TOOLS

#### Microsoft BI suite

Though Microsoft currently targets its Business Intelligence solutions more directly at schools than higher education, the basic application stack that make up Microsoft's offering is fairly widely used in universities and colleges too:

- Excel and the PowerPivot analysis tools
- SQL server and its analysis and Power View services
- SharePoint

The whole stack covers the full analysis flow, which is one of the attractions of solutions of this type.

One particularity of the Microsoft offering is that one tool – Excel – can play a role in nearly all steps of the process. While that's great for self-service analysis and easy data gathering, some specialised skills in SQL are still required in the storage, cleaning and integration stages to address the full range of data sources. Custom coding and advanced data wrangling is likely to be increasingly necessary; the more one uses the set for more predictive insights rather than reports of past events.

<b>Name:</b> Microsoft BI stack	<b>Area:</b> All areas in principle, but most readily used in Academic Analytics
<b>URL:</b> <a href="http://www.microsoft.com/en-us/bi/default.aspx">http://www.microsoft.com/en-us/bi/default.aspx</a>	<b>Data sources:</b> Any conventionally structured data (i.e. spreadsheets and database tables)
<b>Workflow:</b>	
<pre> graph LR     A[Collection &amp; Acquisition] --&gt; B[Storage]     B --&gt; C[Cleaning]     C --&gt; D[Integration]     D --&gt; E[Analysis]     E --&gt; F[Representation &amp; visualisation]     F --&gt; G[Alerting]             </pre>	
<b>Alternatives:</b> Oracle BI suite <a href="http://www.oracle.com/us/solutions/business-analytics/business-intelligence/overview/index.html">http://www.oracle.com/us/solutions/business-analytics/business-intelligence/overview/index.html</a> SAP Business Objects <a href="http://www54.sap.com/solutions/analytics/business-intelligence.html">http://www54.sap.com/solutions/analytics/business-intelligence.html</a> IBM Cognos <a href="http://www-142.ibm.com/software/products/us/en/category/SWQ20">http://www-142.ibm.com/software/products/us/en/category/SWQ20</a> SAS Business Intelligence <a href="http://www.sas.com/technologies/bi/">http://www.sas.com/technologies/bi/</a>	

Table 2: Summary of Microsoft BI tools and the analytic workflow they cover

#### Pentaho

Pentaho is an emerging offering that is based on an open source platform. The basic data integration, reporting, connectivity and analytics packages come with the open source package, but the more advanced functionality and support come via annual subscription packages. A variety of open source add-ons to Pentaho from other vendors exist, however, which also work on the supported versions of the platform.



Functionally, the suite is claimed to be more lightweight and simpler than the large vendor offerings in the BI area. Customers include UK public sector organisations such as the NHS, but no educational establishments.

<b>Name:</b> Pentaho	<b>Area:</b> All areas in principle, but most readily used in Academic Analytics
<b>URL:</b> <a href="http://www.pentaho.com/">http://www.pentaho.com/</a>	<b>Data sources:</b> Any structured data (including NoSQL)
<b>Workflow:</b> 	
<b>Alternatives:</b> Palo <a href="http://www.palo.net/">http://www.palo.net/</a> SpagoBI <a href="http://www.spagoworld.org/">http://www.spagoworld.org/</a> Fluidops Workbench <a href="http://www.fluidops.com/information-workbench/">http://www.fluidops.com/information-workbench/</a>	

Table 3 Summary of Pentaho tools and the analytic workflow they cover

### 3.3 WEB ANALYTICS

The ‘web analytics’ term usually refers to an ‘on-site’ flavour of analytics regarding an organisation's web presence and concerning fact based reporting on items such as number of page visits, search terms, visitor geographical location and product downloads. However, it can also refer to mining and analysing data on the web that somehow involves the organisation; this second ‘off-site’ style of analytics is often used to gauge public sentiment on an organisation, its products or services. The off-site style of analytics crosses over with tools and techniques found in both Social Network Analysis and Data Mining communities.

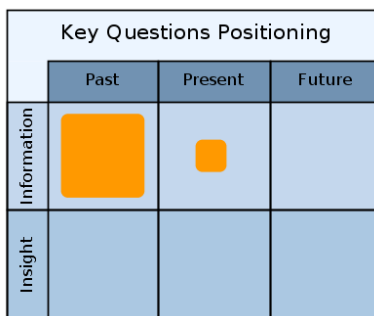


Fig 4: Key Questions Positioning for Web Analytics Tools

### 3.4 WEB ANALYTICS TOOLS

#### AWStats

AWStats takes server logs, such as those created by a web or mail server and analyses them to generate web pages with graphical usage statistics. In general, the intelligence in such log analysis tools is somewhat limited and may do little to distinguish between ‘real’ traffic and automated traffic, such as web indexing crawlers. As such, log analysing tools such as AWStats are aimed at a more technical user and while it may not be giving a clear picture on audience outreach it does accurately answer technical questions such as resource usage. AWStats is installed on the same server that stores the web logs, the generated web pages can be viewed by any operating system with a web browser.

<b>Name:</b> AWStats	<b>Area:</b> Academic Analytics
<b>URL:</b> <a href="http://awstats.sourceforge.net">http://awstats.sourceforge.net</a>	<b>Data sources:</b> Web server logs
<b>Workflow:</b>	
<pre> graph LR     A[Collection &amp; Acquisition] --&gt; B[Storage]     B --&gt; C[Cleaning]     C --&gt; D[Integration]     D --&gt; E[Analysis]     E --&gt; F[Representation &amp; visualisation]     F --&gt; G[Alerting]             </pre>	
<b>Alternatives:</b> Webalizer <a href="http://www.webalizer.org/">http://www.webalizer.org/</a> W3Perl <a href="http://www.w3perl.com/">http://www.w3perl.com/</a>	

Table 4 Summary of AWStats tools and the analytic workflow they cover

#### Google Analytics

An alternative and more modern approach to on-site web analytics over server log analysis is to add small pieces of Javascript code to web pages. By adding code snippets Google Analytics is able to track visitors and various statistics regarding their site usage. While this approach gives a much clearer picture of the usage of online resources by users its reliance of Javascript rather than actual web server logs means that it does not catch all transactions with the server and does not measure resource usage accurately.

A small piece of Javascript needs to be installed on any webpages that are to be tracked. Stats can be viewed through a web browser.

<b>Name:</b> Google Analytics	<b>Area:</b> Academic Analytics
<b>URL:</b> <a href="http://www.google.com/analytics/">http://www.google.com/analytics/</a>	<b>Data sources:</b> Small piece of Javascript inserted into website HTML
<b>Workflow:</b>	
<pre> graph LR     A[Collection &amp; Acquisition] --&gt; B[Storage]     B --&gt; C[Cleaning]     C --&gt; D[Integration]     D --&gt; E[Analysis]     E --&gt; F[Representation &amp; visualisation]     F --&gt; G[Alerting]             </pre>	
<b>Alternatives:</b> Open Web Analytics <a href="http://www.openwebanalytics.com/">http://www.openwebanalytics.com/</a>	

Table 5 Summary of Google Analytics tools and the analytic workflow they cover

### Bluefin Signals

Bluefin Signals is the flagship service offered by analytics company Blue Fin Labs. The service analyses online user comments and discussion related to television programming. The service offers a dashboard of social analytics that shows the amount and style of attention that television programme are receiving, and how this compares to other shows at any specific time of the day.

The service also offers an insight into audience distributions between difference types of shows, giving advertisers and broadcasters an insight into how to create and lucrative advertising slots.

<b>Name:</b> Blue Fin Labs: Social TV Analytics	<b>Area:</b> Academic Analytics
<b>URL:</b> <a href="http://bluefinlabs.com/">http://bluefinlabs.com/</a>	<b>Data sources:</b> Social Media
<b>Workflow:</b>	
<pre> graph LR     A[Collection &amp; Acquisition] --&gt; B[Storage]     B --&gt; C[Cleaning]     C --&gt; D[Integration]     D --&gt; E[Analysis]     E --&gt; F[Representation &amp; visualisation]     F --&gt; G[Alerting]             </pre>	
<b>Alternatives:</b> Nielsen Media Research <a href="http://nielsen.com/us/en.html">http://nielsen.com/us/en.html</a> AT&T SocialTV <a href="http://www.research.att.com/projects/Video/SocialTV/">http://www.research.att.com/projects/Video/SocialTV/</a>	

Table 6 Summary of BlueFin tools and the analytic workflow they cover

### 3.5 INFORMATION VISUALISATION

Information Visualisation is concerned with taking information and representing it in a visual form so that it is easily digestible by the human eye. There is a crossover with many of the other communities we have described as data will need to be collected and prepared first. Tools within the Information Visualisation community come with a wide array of required skill

sets. Typically the more customization offered by a tool the higher the technical demands on the user, which usually means proficiency in a programming language.

Key Questions Positioning			
	Past	Present	Future
Information	■		
Insight	●		

Fig 5 Key Question Positioning for Information Visualisation

### 3.6 INFORMATION VISUALISATION TOOLS

#### Many Eyes

Many Eyes is an online service giving users a selection of common graphical representations for their data. The service is hosted by IBM and allows users to upload datasets in CSV for visualisation and sharing. While users have a selection of visualisation types to choose from, there is little scope to modify or tweak them. As such Many Eyes gives users the ability to quickly try different graphical representations of their data quickly with little technical input, but does not give much in the way of customization.

Data has to be prepared in CSV format, visualisations can be shared and viewed through a web browser.

<b>Name:</b> Many Eyes	<b>Area:</b> All
<b>URL:</b> <a href="http://www-958.ibm.com/software/data/cognos/manyeyes/">http://www-958.ibm.com/software/data/cognos/manyeyes/</a>	<b>Data sources:</b> Dataset in CSV
<b>Workflow:</b> 	
<b>Alternatives:</b> Google Charts: <a href="https://developers.google.com/chart/">https://developers.google.com/chart/</a> Gapminder: <a href="http://www.gapminder.org/">http://www.gapminder.org/</a>	

Table 7 Summary of Many Eyes tools and the analytic workflow they cover

### Tableau Software

Tableau Software is a commercial analytics visualisation company offering services and tools for organisations to analyse and visualise existing data sets. Tableau offers a mix of both desktop and server side applications so that visualisation projects can be shared easily throughout the organisation.

A basic 'Tableau Public' tool is offered for free, and it offers functionality for datasets of up to 100,000 rows of data under the caveat that the resulting analytics is viewable on publicly accessible servers. Data input and manipulation is done via a Windows only desktop tool which connects to an online offering where the data is stored and a selection of visual views on this data are available.

Paid editions allow for larger datasets and for visualisations to be worked on privately. There are many similar commercial companies in the visual analytics space offering different levels of their products and services for free.

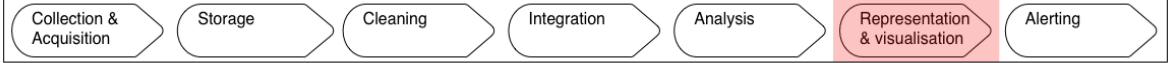
<b>Name:</b> Tableau Software	<b>Area:</b> Any
<b>URL:</b> <a href="http://www.tableausoftware.com/">http://www.tableausoftware.com/</a>	<b>Data sources:</b> CSV
<b>Workflow:</b>  <pre> graph LR     A[Collection &amp; Acquisition] --&gt; B[Storage]     B --&gt; C[Cleaning]     C --&gt; D[Integration]     D --&gt; E[Analysis]     E --&gt; F[Representation &amp; visualisation]     F --&gt; G[Alerting]             </pre>	
<b>Alternatives:</b> Spotfire <a href="http://spotfire.tibco.com/">http://spotfire.tibco.com/</a>	

Table 8 Summary of Tableau tools and the analytic workflow they cover

### Processing

Processing is both a language and its integrated development environment, Processing was created around the idea of creating electronic sketchbooks to teach programming concepts through visual design. The sketchbooks are an environment for users to organise and experiment with different visual designs and though it was originally intended as a teaching tool, a large visualisation community has gathered around it.

Processing is available on Linux, Mac OS X and Windows. Dynamic visualisations can be exported as Java applets.

<b>Name:</b> Processing	<b>Area:</b> Any
<b>URL:</b> <a href="http://processing.org/">http://processing.org/</a> (Sharing community at <a href="http://www.openprocessing.org/">http://www.openprocessing.org/</a> )	<b>Data sources:</b> Any structured data (spreadsheets, database tables)
<b>Workflow:</b>	
<b>Alternatives:</b> Prefuse <a href="http://prefuse.org/">http://prefuse.org/</a> Adobe Flash <a href="http://www.adobe.com/uk/products/flash.html">http://www.adobe.com/uk/products/flash.html</a>	

Table 9 Summary of Processing tools and the analytic workflow they cover

### 3.7 OPERATIONAL RESEARCH

Operational research's roots are in optimisation and decision support. Those roots go quite deep, as its statistical modelling and data analysis techniques have been in use since World War II. Tools used in this approach vary from spreadsheets to custom programming, with specialist mathematical modelling tools and industry specific BI type solutions in between. Tool choice is fairly critical since the creation of sound models is both central to the approach and potentially resource intensive.

Operational research is still particularly popular with the UK Civil Service and other organisations with large, capital intensive, but predictable processes such as logistics companies, airlines, and oil and gas explorers.

Key Questions Positioning			
	Past	Present	Future
Information			
Insight			

Fig 6 Key Questions Positioning for Operational Research

### 3.8 OPERATIONAL RESEARCH TOOLS

#### IBM ILOG

More a suite than a single application, ILOG was offered until 2006 by an independent company that specialised in process optimisation and decision automation with mathematical modelling. The heart of the ILOG set is formed by the CPLEX Optimization Studio; a toolkit for mathematical programming that allows formally stated problems to be solved. Models for specific kinds of problems can be built from scratch in ILOG's own language or out of pre-existing modules. The suite also comprises tools that allow more self-service solutions for automated decision making (e.g. for credit card applications or

course choices), visualisation toolkits and a variety of pre-built applications for industries such as logistics and supply chain management.

ILOG is claimed to be in wide use in universities, but that's probably mostly down to research use of the constraint programming solver within CPLEX studio.

<b>Name:</b> IBM ILOG	<b>Area:</b> Academic Analytics
<b>URL:</b> <a href="http://www-01.ibm.com/software/websphere/ilog/">http://www-01.ibm.com/software/websphere/ilog/</a>	<b>Data sources:</b> relational databases, spreadsheets, and .Net, Java, and C++ APIs
<b>Workflow:</b>	
<pre> graph LR     A[Collection &amp; Acquisition] --&gt; B[Storage]     B --&gt; C[Cleaning]     C --&gt; D[Integration]     D --&gt; E[Analysis]     E --&gt; F[Representation &amp; visualisation]     F --&gt; G[Alerting]     style E fill:#f8d7da     style F fill:#f8d7da             </pre>	
<b>Alternatives:</b> AIMMS <a href="http://www.aimms.com/">http://www.aimms.com/</a> Vanguard Software Business Analytics Suite <a href="http://www.vanguardsw.com/products/business-analytics-suite/">http://www.vanguardsw.com/products/business-analytics-suite/</a>	

Table 10 Summary of IBM ILOG tools and the analytic workflow they cover

### 3.9 DATA MINING

Although the Data Mining term has been used loosely for any type of data analysis, as well as data preparation prior to analysis, the more narrow meaning of the term relates to just the discovery of new information in data sets. For that reason, the terms Knowledge Discovery in Databases (KDD) or ‘machine learning’ are also used to emphasise both the discovery of new information and the automation aspects of data mining. In practice, data mining means using a variety of algorithms, including:

- regressive analysis - classic statistic model validation
- classification - putting entities in known classes
- association - finding out which attributes are often associated
- anomaly detection - detecting outliers
- clustering - grouping entities into a new and unknown structure
- factor analysis - describing the dataset with the fewest attributes

Software components that implement just these algorithms exist, but they are usually found embedded in tool sets that facilitate data clean-up, visualisation and other stages of the data analysis process.

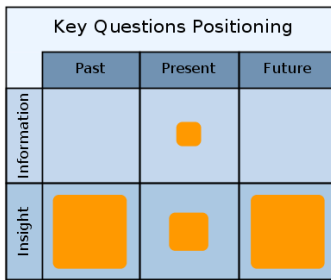


Fig.7 Key Questions Positioning for Data Mining,

### 3.10 DATA MINING TOOLS

#### Oracle Data Miner

Oracle Data Miner is a separate download that extends the Oracle SQL developer tool. Together with an integrated version of R (see below), it is marketed as an Advanced Analytics Option on the standard Oracle database server product.

The main thing that Data Miner does is provide a graphical environment where developers can compose and represent a data analysis workflow. The data mining stages of it are represented as steps after data import, preparation and so on. The limitation as well as the claimed advantage of the tool is that it runs entirely inside the Oracle database. This presumes that data either is already kept in there, or is imported first. The idea is that the process after that is shortened because there doesn't need to be any data duplication or other import and export routines.

Because it is part of the widely used Oracle database server offering it is likely that the tool is accessible to many in the UK education sector.

<b>Name:</b> Oracle Data Miner	<b>Area:</b> All
<b>URL:</b> <a href="http://www.oracle.com/technetwork/database/options/advanced-analytics/odm/index.html">http://www.oracle.com/technetwork/database/options/advanced-analytics/odm/index.html</a>	<b>Data sources:</b> Oracle Relational Database
<b>Workflow:</b> 	
<b>Alternatives:</b> SAS Enterprise Miner: <a href="http://www.sas.com/technologies/analytics/datamining/miner/">http://www.sas.com/technologies/analytics/datamining/miner/</a> Microsoft Analysis Services: <a href="http://www.microsoft.com/sqlserver/en/us/solutions-technologies/business-intelligence/analysis.aspx">http://www.microsoft.com/sqlserver/en/us/solutions-technologies/business-intelligence/analysis.aspx</a>	

Table 11 Summary of Oracle Data Miner tools and the analytic workflow they cover



**R**

R' is both a statistical programming language and a set of open source libraries that implement that language. Many of the common data mining algorithms are included in the standard distribution of the software, but hundreds of additional specialised R libraries have been developed for all sorts of functions. Much the same goes for a number of graphical user interfaces that facilitate data manipulation with R. As noted above, R has also been integrated in commercial offerings such as Oracle's, as well as in comprehensive open source suites such as Rapidminer, or representation tools such as shiny and rapache.

R is widely used for research in UK higher education, and probably increasingly for operational purposes as well. Its popularity in research means that courses in its use should be fairly readily available in many institutions.

<b>Name:</b> R	<b>Area:</b> All
<b>URL:</b> <a href="http://www.r-project.org/">http://www.r-project.org/</a>	<b>Data sources:</b> Any data that is amenable to statistic modelling
<b>Workflow:</b>	
<pre> graph LR     A[Collection &amp; Acquisition] --&gt; B[Storage]     B --&gt; C[Cleaning]     C --&gt; D[Integration]     D --&gt; E[Analysis]     E --&gt; F[Representation &amp; visualisation]     F --&gt; G[Alerting]             </pre>	
<b>Alternatives:</b> IBM SPSS <a href="http://www.ibm.com/software/uk/analytics/spss/">http://www.ibm.com/software/uk/analytics/spss/</a> SAS analytics <a href="http://www.sas.com/technologies/analytics/">http://www.sas.com/technologies/analytics/</a> Stata <a href="http://stata.com/">http://stata.com/</a>	

Table 12 Summary of R tools and the analytic workflow they cover

**3.11 SOCIAL NETWORK ANALYSIS**

Social Network Analysis (SNA) is the analysis of data on social networks to ask questions regarding individuals and the relationships between them. Recently SNA has had a surge of popularity due to the rise of online social networks, that is not to say however that SNA is limited to the online space as the term has been used since the 1950s.

As much of SNA is related to Network Theory and while the tools discussed here are often associated with SNA, many of the tools can be applied to data from any discipline that is concerns with the study of graphs .

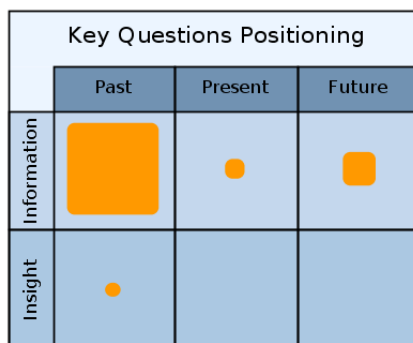


Fig 8 Key Questions Positioning for Social Network Analysis,

### 3.12 SOCIAL NETWORK ANALYSIS TOOLS

#### Gephi

Gephi is a tool for exploring complex networks through analysis and visualisation of graph data. Exploring these networks often involves detecting the presence of communities or understanding communication patterns. This can be particularly relevant for the analysis of teaching and learning in online communities such as forums. Gephi's 'on the fly' data exploration and layout engines has made it popular with the SNA and data journalism communities.

Gephi is available on Windows, OS X and Linux platforms.

<p><b>Name:</b> Gephi</p>	<p><b>Area:</b> Academic and Research Analytics Learning Analytics</p>
<p><b>URL:</b> <a href="http://gephi.org">http://gephi.org</a></p>	<p><b>Data sources:</b> Graph Data</p>
<p><b>Workflow:</b></p> <div style="display: flex; justify-content: space-around; align-items: center;"> <div style="border: 1px solid black; border-radius: 15px; padding: 5px; text-align: center;">Collection &amp; Acquisition</div> <div style="border: 1px solid black; border-radius: 15px; padding: 5px; text-align: center;">Storage</div> <div style="border: 1px solid black; border-radius: 15px; padding: 5px; text-align: center;">Cleaning</div> <div style="border: 1px solid black; border-radius: 15px; padding: 5px; text-align: center;">Integration</div> <div style="border: 1px solid black; border-radius: 15px; padding: 5px; text-align: center; background-color: #f8d7da;">Analysis</div> <div style="border: 1px solid black; border-radius: 15px; padding: 5px; text-align: center; background-color: #f8d7da;">Representation &amp; visualisation</div> <div style="border: 1px solid black; border-radius: 15px; padding: 5px; text-align: center;">Alerting</div> </div>	
<p><b>Alternatives:</b> NodeXL <a href="http://nodexl.codeplex.com/">http://nodexl.codeplex.com/</a> Tulip <a href="http://tulip.labri.fr/TulipDrupal/">http://tulip.labri.fr/TulipDrupal/</a></p>	

Table 13 Summary of Gephi tools and the analytic workflow they cover

#### Network Workbench

Network Workbench is a multipurpose toolkit aimed at research topics based around Network Theory. It takes users through a set workflow of data management, analysis and visualisation.

Network Workbench’s visual tools are not as immediate as Gephi’s, making it less of a playground to try different techniques and as such users tend to know beforehand what operations they wish to perform upon the data.

Network Workbench’s workflow and ability to work with large datasets has made it popular with the academic community. It is available on Windows, Linux and OS X.

<b>Name:</b> Network Workbench	<b>Area:</b> All
<b>URL:</b> <a href="http://nwb.cns.iu.edu/">http://nwb.cns.iu.edu/</a>	<b>Data sources:</b> All structured data (spreadsheets, database tables)
<b>Workflow:</b>	
<b>Alternatives:</b> Pajek: <a href="http://vlado.fmf.uni-lj.si/pub/networks/pajek/">http://vlado.fmf.uni-lj.si/pub/networks/pajek/</a>	

Table 14 Summary of Network Workbench tools and the analytic workflow they cover

### 3.13 LEARNING ANALYTICS

If ‘learning analytics’ is considered as an application of analytics to gain insights to support educational aims and objectives (Cooper, 2012a), it quickly becomes clear that most tools in this area centre around the use of Virtual Learning Environments by learners. The typical application of analytics in VLEs is performance management: monitoring which activities by students and teachers correlate with attainment, and warning which students are at risk. Since VLEs typically have fairly detailed assessment data in a grade book, and very detailed data about online engagement with resources and in forums, the only data used ‘from the outside’ tends to be group and personal attributes from a student record system.

This is not the only type of learning analytics, however. Other applications in this area include things like group work analysis, traditional assessment psychometrics and course choice planning

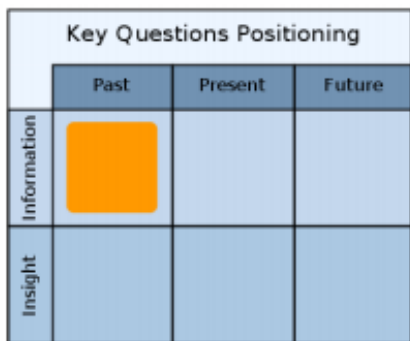


Fig 9 Key Questions Positioning for Learning Analytics

### 3.14 LEARNING ANALYTICS TOOLS

#### Blackboard Analytics for Learn

Analytics for Learn is one of a suite of Blackboard tools that are ‘Analytics’ badged. Other applications deal with alumni and donor management, finance, enrollment and human resources. Analytics for Learn is focussed on the Blackboard Learn VLE, and is about the classic learning analytics questions: what learner activities correlate with high grades?, which students are at risk?, who teaches most effectively?

Dashboards and reports are available for students and staff, and administrators have access to a self-service analysis tool that allows them a degree of report customisation. Architecturally, it is a data warehouse built on Microsoft SQL server, with the main differentiator being the range of pre-built, HE specific models and metrics, as well as connectors to popular student records that are included. In most cases, further customisation is still likely to be required, but might well be less time consuming compared to a from-scratch custom build.

<b>Name:</b> Blackboard Analytics for Learn	<b>Area:</b> Academic Analytics
<b>URL:</b> <a href="http://www.blackboard.com/Platforms/Analytics/Products/Blackboard-Analytics-for-Learn.aspx">http://www.blackboard.com/Platforms/Analytics/Products/Blackboard-Analytics-for-Learn.aspx</a>	<b>Data sources:</b> VLE database (logs, gradebooksetc), Student Record System
<b>Workflow:</b>	
<pre> graph LR     A[Collection &amp; Acquisition] --&gt; B[Storage]     B --&gt; C[Cleaning]     C --&gt; D[Integration]     D --&gt; E[Analysis]     E --&gt; F[Representation &amp; visualisation]     F --&gt; G[Alerting]             </pre>	
<b>Alternatives:</b> Desire2Learn analytics <a href="http://www.desire2learn.com/products/analytics/">http://www.desire2learn.com/products/analytics/</a> Instructure Canvas (API included with VLE) <a href="http://www.instructure.com/features-higher-education">http://www.instructure.com/features-higher-education</a>	

Table 14 Summary of Blackboard Analytics for Learn tools and the analytic workflow they cover

#### Austin Peay Degree Compass

Not available as a separate product yet, Degree Compass is an application that guides a student through the course/module selection process. Taking into account the constraints of a particular degree programme, and the attainment of a particular student so far, it suggests a module selection with which a student is most likely to succeed. The application works by comparing as many data points about a student as possible, and comparing them to what similar students did with modules in the past. Early results suggest the recommendations do lead to higher success rates.

The application has received some grant funding, and it is now being rolled out in other institutions in Tennessee.

<b>Name:</b> Austin Peay State University Degree Compass	<b>Area:</b> Academic Analytics
<b>URL:</b> <a href="http://www.apsu.edu/information-technology/degree-compass-what">http://www.apsu.edu/information-technology/degree-compass-what</a>	<b>Data sources:</b> student record system, grade books
<b>Workflow:</b>	
<pre> graph LR     A[Collection &amp; Acquisition] --&gt; B[Storage]     B --&gt; C[Cleaning]     C --&gt; D[Integration]     D --&gt; E[Analysis]     E --&gt; F[Representation &amp; visualisation]     F --&gt; G[Alerting]             </pre>	
<b>Alternatives:</b> Custom build	

Table 15 Summary of Austin Peay Degree Compass tools and the analytic workflow they cover

## 4. Summary

In general, the tool choices facing many colleges and universities can be summarised as either easy to install but limited in scope and inflexible, or flexible and comprehensive, but expensive and time consuming. In some cases, it might, therefore, be attractive to experiment with ready-made tools such as those from the learning analytics or Business Intelligence traditions, not least because they frequently include 'self-help' user interfaces. More bespoke tools or custom development can be added later when there is a proven need. Unfortunately, some of these ready-made solutions may be too expensive for experimentation. Also, analytics initiatives depend heavily on identifying the right variables, if a ready-made solution doesn't cover it, it may be of little use.

In those cases, investing in staff experimentation with low cost components from a range of traditions may be a more prudent initial move, even if the most effective tool subsequently turns out to be a ready-made suite.

## 5. References

- Baker, R. S. J. d., S. M. Gowda, and A. T. Corbett. 2011. "Automatically Detecting a Student's Preparation for Future Learning: Help Use Is Key." Proceedings of the 4th International Conference on Educational Data Mining 179–188. <http://users.wpi.edu/~rsbaker/PFL-EDM-2011-v19.pdf>
- Baker, R. S. J. d. 2011. "Data Mining for Education." In *International Encyclopedia of Education*, 3rd Ed., edited by B. McGaw, P. Peterson, and E. Baker. Oxford, UK: Elsevier.
- Cooper, A. 2012a. *What is Analytics? Definition and Essential Characteristics* Vol 1, No 5. CETIS Analytics series. Bolton: The University of Bolton. <http://publications.cetis.ac.uk/c/analytics>
- Cooper, A. 2012b. *Brief History of Analytics*. Vol 1. No. 9 CETIS Analytics Series. Bolton: The University of Bolton. <http://publications.cetis.ac.uk/c/analytics>
- Elias, T., 2011. Learning Analytics: Definitions, Processes and Potential. In *LAK '11 1st conference on learning analytics*. LAK '11 1st conference on learning analytics. Banff, Alberta, Canada: Athabasca University.
- Pham, M. C., Cao, Y., Petrushyna, Z., & Klamka, R. (2012). Learning Analytics in a Teachers' Social Network. In V. Hodgson, C. Jones, M. de Laat, D. McConnell, & T. Ryberg (Eds.), *Proceedings of the Eighth International Conference on Networked Learning 2012* (pp. 414-421). Maastricht.
- Siemens, G. (2012, April 11). *Eli 2012 Sensemaking Analytics*. Technology. Retrieved from <http://www.slideshare.net/gsiem>

## About the Authors

Wilbert Kraan joined CETIS in 2002 as the CETIS Journalist. From there, he entered specification development work, starting with IMS Content Packaging 1.1.4, moving on to become the chair of the Content Packaging 1.2 group. That specification has recently been turned into the first ISO standard for educational content. Wilbert has also worked on software architectures, first in a partnership between JISC, the Australian Department of Education Science and Training, Surf foundation of the Netherlands and the New Zealand Ministry of Education. Later, he moved into Enterprise Architecture modelling, and was instrumental in introducing the ArchiMate Modelling language to the UK higher education sector. More recently, Wilbert has championed the IMS Question and Test Interoperability 2.1 specification. As a co-chair of the IMS working group, he has worked with content publishers and software developers on maturing and profiling the specification. Wilbert has also worked as the technical lead on the development of the Resource Aggregation Model for Learning, Education and Training (RAMLET) ontology which has recently been published as an IEEE standard. He is currently an Assistant Director at CETIS.

David Sherlock joined CETIS in 2007 after completing his MSc in Systems development at the University of Bolton. He is responsible for the development and maintenance of CETIS's web presence as well as aspects of communication policy. Recently David has designed and developed more specialised software for a number of projects, including the addition of linked data capabilities to the CETIS project directory (<http://prod.cetis.ac.uk>), widgets for the Wookie educational widget server (<http://getwookie.org/>) and a Firefox extension for sharing resources. David was also involved in developing a validator and renderer for the eXchanging Course Related Information (XCRI) specification. He is currently also working on his part-time PhD on agent based networks.

## CETIS Analytics Series

- Vol.1 No.1. Analytics, What is Changing and Why does it Matter?
- Vol.1 No.2. Analytics for the Whole Institution; Balancing Strategy and Tactics
- Vol.1 No.3. Analytics for Learning and Teaching
- Vol.1 No.4. Analytics for Understanding Research
- Vol.1 No.5. What is Analytics? Definition and Essential Characteristics
- Vol.1 No.6. Legal, Risk and Ethical Aspects of Analytics in Higher Education
- Vol.1 No.7. A Framework of Characteristics for Analytics
- Vol.1 No.8. Institutional Readiness for Analytics
- Vol.1 No.9. A Brief History of Analytics
- Vol.1 No.10. The Implications of Analytics for Teaching Practice in Higher Education
- Vol.1 No.11. Infrastructure and Tools for Analytics

<http://publications.cetis.ac.uk/c/analytics>

## Acknowledgements

The CETIS Analytics Series was commissioned by Myles Danson (JISC programme manager) to give an overview of current thinking around analytics in post-16 education in the UK. In addition to the authors the following people have contributed to the production of the CETIS Analytics Series; Lorna Campbell (CETIS), Adam Cooper (CETIS), Rob Englebright (JISC), Neil Jacobs (JISC), Sheila MacNeill (CETIS) and Christina Smart (CETIS). Design and Artwork by:<http://www.consul4design.com>

## About this White Paper

Title: CETIS Analytics Series Volume 1, No. 11 Analytics tools and Infrastructure

Authors: Wilbert Kraan and David Sherlock (CETIS)

Date: January 2013

URI:<http://publications.cetis.ac.uk/2013/535>

ISSN 2051-9214

Text Copyright © 2013 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 paper has been produced by CETIS for JISC: <http://www.jisc.ac.uk>