



JISC cetis

centre for educational technology
& interoperability standards

Analytics Series

Vol. 1, No. 2. Analytics for the Whole Institution: Balancing Strategy and Tactics

By David Kay (Sero Consulting) and
Mark van Harmelen (Hedtek)

CETIS Analytics Series ISSN 2051-9214

Produced by CETIS for JISC

Analytics for the Whole Institution: Balancing Strategy and Tactics

David Kay (Sero Consulting) with Dr. Mark van Harmelen (Hedtek)

Table of Contents

1. Executive Summary	4
1.1 Scope.....	4
1.2 Context	5
1.3 Analytics Potential.....	5
1.4 Business Benefits	5
1.5 Headline Challenges	5
1.6 The Data Ecosystem	6
1.7 Data Management.....	6
1.8 Organisational Readiness.....	6
1.9 Strategy for the Journey	7
1.10 Starting Points for the Journey.....	7
1.11 Measuring Progress.....	8
2. PART ONE – Landscape	9
2.1 Sector Context.....	9
2.2 Analytics Potential.....	10
2.3 Business Benefits	12
2.4 Headline Challenges	14
3. PART Two – Challenges.....	16
3.1 Variety: The Data Ecosystem.....	16
3.2 Volume: Data Management	18

3.3	Velocity: Organisational Readiness	19
4.	PART Three – Response Strategies	23
4.1	Mapping the Journey: Strategy	23
4.2	Mapping the Journey: Starting Points	23
4.3	Measuring Progress	25
5.	References	29
	About the Authors	30
	CETIS Analytics Series.....	30
	Acknowledgements	30
	About this White Paper.....	31

1. Executive Summary

1.1 SCOPE

The benefits afforded by the longitudinal collection and analysis of key institutional data are not new to enterprise IT managers nor to senior management more generally. Data warehousing and Business Intelligence (BI) dashboards are integral to the modern management mindset and part of the 'enterprise IT' architecture of many Higher Education Institutions and Further Education Colleges.

However some things are changing that pose questions about how business intelligence and the science of analytics should be put to use in customer facing enterprises:

- The demonstration by online services ranging from commodity sales to social networks of what can be done in **near-real time** with well-connected data..
- The emphasis brought by the web to the detection, collection and analysis of **user activity data** as part of the BI mix, ranging from clicks to transactions..
- The consequent **changes in expectations** among web users of how services should work, what businesses could do for them, accompanied by shifts in legal and ethical assumptions..
- The availability of **new types of tools** for managing, retrieving and visualizing very large data that are cheap, powerful and (not insignificantly) accessible to grass roots IT users..

Set against that backdrop, this paper aims to;

- Characterise the educational data ecosystem, taking account of both institutional and individual needs.
- Recognise the range of stakeholders and actors – institutions, services (including shared above-campus and contracted out), agencies and vendors.
- Balance strategic policy approaches with tactical advances.
- Highlight data that may or may not be collected.
- Identify opportunities, issues and concerns arising.

Our focus is therefore not on technology but rather on high value gains in terms of business objectives, the potential for analytics and opportunities for new thinking across the organisation.

This paper has three parts as follows:

PART One – Landscape: Covering the Sector Context, Analytics Potential, Business Benefits and Headline Challenges.

PART Two – Challenges: Covering the Data Ecosystem, Data Management and Organisational Readiness.

PART Three – Response: Covering Mapping the Journey, Measuring Progress and concluding with some Lessons To Consider.

PART ONE – Landscape: explores the landscape relating to the adoption and advancement of analytics in UK Further and Higher Education.

1.2 CONTEXT

A number of factors in the current operational landscape are influencing the approach Further and Higher Education organisations are taking to analytics:

- The experience and satisfaction of the service user is central.
- The economic challenges and new models facing institutions and their fee-payers demand increased levels of control and agility.
- Student retention and achievement represent increasingly high priorities.
- Clients are increasingly 'born digital', bringing new expectations.

Consequently there are growing institutional motivations for analytics to support a wide range of functions, for general and individual good. However the sector faces deeply rooted system-wide challenges:

- Cultural – the sector takes seriously the legal and ethical considerations of exploiting user data.
- Organisational – institutions are not always organised around the premises of central controls and evidence-based decision-making that necessarily underpin analytics.
- Systems disconnect – there may be a disconnect between systems managed at centre and those serving teaching and learning and research.
- Data ecosystem – institutions operate within a complex data ecosystem, which involves reliance on data from shared services, partnerships and agencies.

1.3 ANALYTICS POTENTIAL

For Further and Higher Education, the analytics 'silver bullet' lies in the potential to derive and to act upon pre-emptive indicators, on 'actionable insights', stepping beyond the long haul reactive measures typical of early generation BI systems applied to traditionally cyclical systems such as financials. Analytics can help most where:

- The data is collected through **standard business processes**.
- Data in multiple systems can be joined together based on **commonly agreed coding frames** for key elements.
- Collection extends **over time**.
- Analysis can serve **multiple areas of the business**.

1.4 BUSINESS BENEFITS

Having established that analytics should matter to any customer facing business, it is evident that the range of opportunities is potentially large. It is therefore essential to identify what strategic problems or issues analytics can help address and what the institution's initial focus could be. For example:

- **Assessing performance** in terms of efficiencies, effectiveness, economies and value. E.g. student success, research productivity, etc.
- **Informing segmentation and targeting**. E.g. tuning course focus and curriculum design, distinguishing learning modes and styles.
- **Identifying business models and predicting trends**. E.g. the student marketplace, the research landscape, cash / resource utilisation.

1.5 HEADLINE CHALLENGES

The scale, frequency and range of potentially relevant data sources exceed human analytical capacity (even with desktop productivity tools) and therefore demand a step change in terms of automation backed up by corporate commitment and staff skills.

Whilst the challenges of scale are serious at the global level, they may be less threatening in education where there is a valuable range of near real-time indicators available from established systems and where the data is relatively small scale. It is therefore suggested that the first order challenges for analytics in education are likely to relate to priority, implementation and sustainability within the institution, rather than to underlying technological barriers.

PART TWO – Challenges: *explores the challenges relating to the data ecosystem, data management and organisational readiness.*

1.6 THE DATA ECOSYSTEM

There are three core challenges common to institutional and external data:

- **Diversity** – diversity of source systems typifies the analytics mission.
- **Consistency** – as a consequence of diversity, institutions need to identify and consolidate systems of record, to standardise key coding frames across applications and to establish shared definitions of key terms.
- **Data quality** - data quality issues range from cleanliness to provenance, and notably involve trust, completeness and currency.

1.7 DATA MANAGEMENT

There is a balance to strike between answering known questions and discovering unknown narratives. Experience commends atomic approaches linked to the raw potential for data capture as well as requirements driven down from KPIs. Data managers are therefore faced with key decisions about what to keep and for how long, what to aggregate and what to delete.

The tendency is to preserve wherever feasible, which puts significant pressure on defining collection and maintenance strategies including the level of the data, the provision of the hardware and the responsibility for the process. There are key choices between highly centralised ‘enterprise’ approaches and platforms that enable more distributed strategies.

Dialogue within the community and the experience of JISC projects in Activity Data, Business Intelligence and CRM suggest that institutions will benefit most from a combination of these approaches, promoting practitioner innovation whilst securing data for institution wide and as yet unknown long term use.

1.8 ORGANISATIONAL READINESS

First order technical issues are not preventing the development of institution wide analytics. The core challenges exist rather in the areas of culture, policy and skills, in institutional confidence, control and capability.

Culture: the pressure for near to real time access to consistent data to drive evidence-based decisions may threaten a collision of cultures and management approaches.

Policy: effective governance requires both appropriate authority and clear policies. UK and US universities contributing to recent JISC case studies emphasised the role of directive governance and policy to ensure accountability to data subjects and owners, to analytical partners, and to stakeholders such as parents who need to trust the data. This requires clear policy to ensure quality and responsibility.

Roles: analytics places dependencies on tight interaction between institutional and external actors that differ from the demands of transactional IT and 'reporting'. It is therefore likely that those responsible for central IT, for institutional data research and for a range of autonomous systems will need to review and strengthen their relationships.

Capability: skills and tools cannot be separated, though the right tools can lessen the skills entry level. In the current phase of 'bootstrapping' of analytics capability, the issues of data skills, tools and IT roles to underpin analytics are critical. Institutions may in the immediate term benefit from support facilitated by JISC working with UCISA and Open Source communities.

PART Three – Response: *suggests strategies of value to institutions and above-campus services seeking to develop a comprehensive approach to analytics based on a dynamic combination of local innovation and corporate direction.*

1.9 STRATEGY FOR THE JOURNEY

An institutional analytics road map will benefit from iterative consideration of strategy, taking account of three high level approaches:

- **Local initiatives** - let a thousand flowers bloom.
- **Single enterprise solution** – channel all effort through the IT service.
- **Hybrid** – encourage and exploit local innovation by embedding key parameters to channel data from local to central.

This paper argues that the hybrid model can maximise benefits in terms of building momentum, developing practice, enabling reuse and securing long term data value. One way or another, it is essential to assess the implications of such a choice in terms of governance, processes and skills.

1.10 STARTING POINTS FOR THE JOURNEY

Given increasing awareness of how a wide range of operational IT systems and other data sources might play into the development of analytics, there are challenges in deciding where to start. An audit and readiness evaluation of existing systems and operational processes is therefore recommended to identify systems collecting data that could address core business challenges. Immediately fruitful areas benefitting from analytics might include:

- **Effectiveness:** student lifecycle, research processes.
- **Experience:** personalisation and recommendation services for students and researchers, IT availability and performance, client surveys.
- **Efficiency and economy:** optimising resource investments, tuning campus services and waste, budget trend analysis.

Project work and individual institutional explorations also suggests a wide range of potential targets that are less well explored but may suit local priorities and preparedness. These may include a variety of personal interventions, relationship management, staff performance management, portfolio and course design and research success factors.

Regardless of the focus of enquiry, in order to maximise the narratives and analyses that can be derived from locally collected data, it should be a priority to secure access to data from key external sources and to make it accessible for consumption across the institution. Some of this will be 'open data', some provisioned by a sector service.

1.11 MEASURING PROGRESS

A range of IT developmental initiatives have found a maturity model to be a good way to gauge progress at each stage. For example, in 2012, EDUCAUSE developed a maturity model to enable institutions to self-assess their progress towards embedding analytics in everyday business practice. The key criteria in the model are culture, governance, process, data, infrastructure and tools, expertise and investment.

It is also important to consider whether the focus of any self-assessment model should be on the whole institution, on subsidiary project developments or on both. Experience in projects has suggested a number of 'before' and 'after' success factors. Whilst some are general to strategic ICT projects, they typically require specific consideration in the case of business intelligence and analytics on account of the exploratory intent, the variety of approaches and the potentially cross-institution nature of the undertaking.

In these respects there is a strong tradition in the education community of developing shared models, often internationally applicable, to enable institutions to self-assess and derive benchmarking intelligence in support of continuous improvement. The institution wide challenges and opportunities raised here suggest that an institutional toolkit could be developed for analytics, encompassing the EDUCAUSE Maturity Model, thus providing opportunity for international benchmarking and professional dialogue.

2. PART ONE – Landscape

The section explores key features of the landscape relating to the adoption and advancement of analytics in UK Further and Higher Education.

2.1 SECTOR CONTEXT

Whilst there is much to be learned from developments in other sectors, not least from consumer facing online businesses, this exploration is placed in the context of post-compulsory education, specifically the UK Further and Higher Education sectors. This is characterised as follows:

- The mission of Further and Higher Education is focused on the advancement of teaching, learning and research, and the associated reputation of the institution and the UK sector at large.
- The experience and satisfaction of the service user, whether a student, researcher or lecturer, is central to that mission.
- The economic challenges and new models facing institutions and their fee-payers demand increased levels of control and agility regarding resource utilisation and funds management.
- Student retention and achievement represent increasingly high priorities in the fee-paying environment.
- The sector's clients are increasingly 'born digital', bringing new expectations of services, and of what constitutes good business and good practice¹.
- Consequently there are growing institutional motivations for the use of analytics to support a wide range of functions and decision making in business and research, for general and individual good.

Given this context, it would be logical to recommend to the sector the immediate and on-going business benefits of analytics, and especially near to real-time analysis of user activity data. The implications of doing nothing in this space appear self-evident. However we should recognise that the sector faces deeply rooted, perhaps systemic (or at least system-wide), challenges:

- Cultural – it is fundamental that the sector takes very seriously the ethical considerations of exploiting user data and the research challenges of identifying significant indicators; there may also be a cultural mistrust of the analytical role of data.
- Organisational – HEIs are not always organised around the premises of central controls and evidence-based decision making that necessarily underpin effective use of analytics.

¹ In his presentation 'Making data a way of life for public servants' to the 2012 Eduserv Symposium, Max Wind-Cowie (Head of progressive conservatism, DEMOS) observed that Google and Amazon have embedded their referral services as an expectation and suggested that public services are well behind and could be using the available Big Data to adapt <http://www.eduserv.org.uk/newsandevents/events/2012/symposium/making-data-a-way-of-life-for-public-servants>

- Systems disconnect – there may exist a disconnect between the enterprise business systems managed centrally (e.g. financials, registration, estates) and those newly serving the purposes of teaching and learning and research; whilst the back story can be readily explained in terms of organisational culture and systems maturity, this is unusual in business terms and exacerbates the challenges of analytics.
- Data ecosystem – HEIs are not unusual in that they operate within a complex data ecosystem, which extends beyond and crosses institutional boundaries; for example, reliance on external data (e.g. UCAS), shared services (e.g. JUSP e-journal statistics) and partnerships (e.g. international research groups) builds significant dependencies in to the analytics mission.

It is therefore unsurprising that commentators are moving from bullish cheerleading to a more reflective recognition of the problem space and its undoubted opportunities. This is reflected in the position adopted by EDUCAUSE, the US-based Higher Education technology partnership group. Chief Executive Diana Oblinger reflected in October 2011 that, “Analytics holds transformative promise for education, but the field is still in the developmental stages.”²

2.2 ANALYTICS POTENTIAL

Wherein lies the ‘transformative promise’ of analytics? Amazon and Google have led the way for a wide range of businesses to take advantage of user behaviours and ‘activity data’ to improve their businesses at a variety of levels. The use of analytics ranges from supermarkets optimising their stocking and shelving to comparison websites tracking decision paths, to everyday consumer and public services cutting out the middleman. This is not just about wholly online operations. We should recognise that the likes of Amazon, eBay and Ticketmaster, like much of Further and Higher Education, are strongly focused on ‘physical’ fulfilment – stuff in the post and face-to-face experiences – mediated by intelligence gathered online.

For education, in the current climate of customer expectation and economic austerity, the potential ‘silver bullet’ lies in the potential to derive and to act upon pre-emptive indicators, on ‘actionable insights’. This offers benefits above and beyond the long haul reactive measures typical of early generation BI systems applied to traditionally cyclical systems such as financials and performance management. Adam Cooper of CETIS provides a working definition for Analytics as;

The process of developing actionable insights through problem definition and the application of statistical models and analysis against existing and/or simulated future data.

[Cooper, 2012] [1]

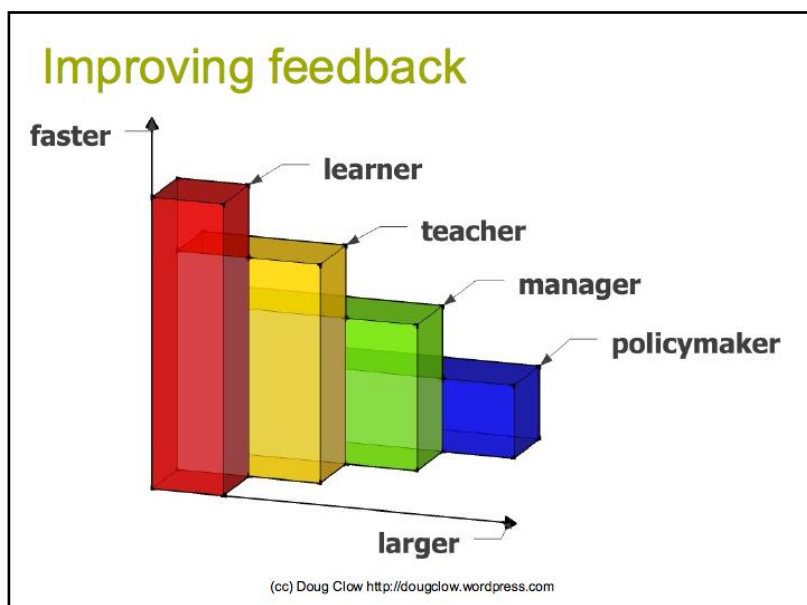
Myles Danson, JISC Programme Manager for Business Intelligence has described this as “Spotting the actionable and moving from knowledge to intervention.”

² Compare, for example, the position taken in the 2010 white paper ‘What’s New in Analytics in Higher Education?’ (<http://www.edu1world.org/Home/35362>) with the 2011 EDUCAUSE analytics initiative press release (<http://www.educause.edu/About+EDUCAUSE/PressReleases/EDUCAUSEAnnouncesNewInitiative/239558>)

Analytics can help most where:

- The data is collected through **standard business processes**, such as the selection of a purchase or wish list item in e-commerce or interaction with course management, library or attendance systems in education.
- Data resides in multiple systems (such as library, registry and VLE) but can be joined together based on **commonly agreed coding frames** for key elements; in the case of learner support these would be about the learner (notably registration ID) and their studies (e.g. course, unit).
- Collection extends **over time** as part of core business requirements..
- Analysis can be seen to serve **multiple areas of the business** – for example, benefitting both student experience and resource management.

In his presentation to the LAK12 conference [2], Doug Clow of the Open University illustrated differences in how analytics might bring differing demands and benefits in different areas of the education business.



The diagram indicates that learners will typically be looking for feedback in as close to real time as possible, whilst requiring relatively little data to be processed (even in a comparative 'others like me' recommendation scenario). Meanwhile teachers will want to see exception and comparative analyses based on similar data to identify 'early warning' signals. At the other end of the spectrum, institutional managers and policymakers will be looking on a cyclical basis (e.g. monthly, annually) for analysis of potentially broad swathes of data that will need to be connected and analysed and then ideally visualized to expose both preconceived and novel patterns.

Analytics can support each of these use cases, though the demands on the systems originating and processing the data will be very different. It is however important to recognise that education presents no challenges to the available technology, it

needs nothing more immediate than the use of a consumer recommender or comparison service, nor is it as large scale as businesses like Facebook.

2.3 BUSINESS BENEFITS

We have established why analytics should matter to any customer-facing business that is already geared to collect data. Furthermore it is evident that the range of opportunities is potentially large. It is therefore essential to identify what strategic problems or issues analytics can help the institution address and what the institution's initial focus will be for analytics.

It may be useful to distinguish between three levels of practical impact, which can be illustrated from projects that have achieved traction within JISC programmes:

Assessing performance in terms of efficiencies, effectiveness, economies and value, for example:

- Student success and satisfaction.
- Research productivity.
- Staff performance management.
- Other indicators such as brand recognition and reputation.

University of Huddersfield – The Library Impact Data Project (LIDP) tested the hypothesis that ‘there is a statistically significant correlation across a number of universities between library activity data and student attainment.’ The project used the student’s final grade, course title and variables relating to library usage, including books borrowed, library e-resources access and entry to the library. At the end of the project the hypothesis was successfully supported for student attainment in relation to use of e-resources and books borrowed for all eight institutional partners.

<http://library.hud.ac.uk/blogs/projects/lidp/>

Roehampton University – This CRM project used a variety of activity data to support student progression and retention, especially addressing causes of non-completion. An ‘Early Warning System’ approach to collate key activity data and flag up students who were at risk of failing to progress was trialled in the Department of Psychology and is now being implemented more widely. The indicators ranged from poor lecture attendance to receiving a fail grade or a warning for plagiarism. In its first year of operation, the system saw student progression rates improve by 14% amongst first year Psychology undergraduates.

<http://www.roehampton.ac.uk/fulCRM/Project-documents/>

Informing segmentation and targeting, for example:

- Differentiating pricing and offers.
- Tuning course focus and curriculum design.
- Generating appropriately targeted recommendations.
- Distinguishing learning modes and styles.
- Designing personal interventions to assist students, researchers and staff.

University of East London – UEL has used UCAS, HESA, local economic and institutional data to predict likely demand based on demographic trends, traditional catchment and competition before deciding what new courses develop. This involved collaboration with the London Borough of Newham, illustrating how two institutions have worked together with HESA data for operational and policy-making purposes. The university has also drawn on a broad portfolio of data to provide actionable student lifecycle intelligence with support from the JISC BI programme.

<http://www.uel.ac.uk/it/news/JISCBusinessIntelligence-QlikView.htm>

The Open University – The RISE project investigated whether activity data generated by users of online resources could provide useful recommendations along the lines of ‘students on my course are looking at these resources’, thus enhancing learner experience by providing real time personalised feedback. The vital catalyst for analytical thinking was to evaluate the readily available activity data as the key to being able to draw in contextual data from other local and sector sources.

www.open.ac.uk/blogs/RISE/

Identifying business models and predicting trends, for example:

- Forecasting service demand.
- Predicting the student marketplace.
- Mapping the research landscape.
- Mitigating unsteady flows such as cash and resource utilisation.
- Identifying trending from social media.

University of Glasgow – The Engage business intelligence project investigated data about research clusters with a view to identifying indicators of grant applications success in terms of team make up based on track record and the potential for automated identification of new research opportunities tailored to project teams approaching the end of funding.

The project integrated and visualised information about research clusters, including external RCUK classifications, in order to support strategic decision-making and to enhance interdisciplinary research and research data reporting.

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

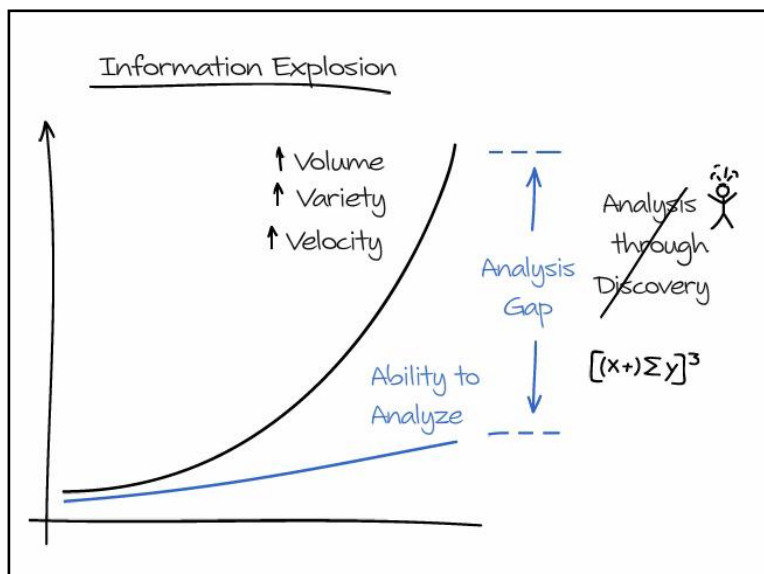
Use of analytics can apply in many areas of the institution's business, ranging from catering to IT support or the library. The challenge is where to start and what to prioritise.

For example, some observers regard the area of learning analytics, where a variety of data already exists to support the student journey, to be most ripe for immediate development. As observed by EDUCAUSE,

The strategic value of analytics includes improved institutional productivity, decision-making, and learner success. While better decision-making about campus operations is an important aspect of analytics, its potential to support students is of paramount interest. [3]

2.4 HEADLINE CHALLENGES

The scale, frequency and range of potentially relevant data sources represent challenges that exceed human analytical capacity and therefore demands a step change in terms of automation, backed up by corporate commitment and staff skills.



Prof. Anthony D. Joseph of the University of California set out a 'Berkeley view of Big Data' at the 2012 Eduserv Symposium [4]. He headlined three challenging trends:

- **Algorithms** – timely analytics will involve analysing user behaviour ('now casting') rather than relying on explicit, often post hoc, input. Illustrations are found in the Twitter-based Earthquake Detector, which monitors earthquake related tweets and at Google Flu Trends.
- **Machines** – to feed analytics, data collection and processing needs to be timely and longitudinal. On the global stage, data is potentially growing faster than available storage and processing capacity, raising questions of what data to keep and what to delete, bearing in mind that future narratives and lines of enquiry are unpredictable.
- **People** - it is hard to extract value from data in ways that make analysis accessible to all, not just data scientists.

These challenges are very considerable at the macro or global level. However, they may be less threatening in the domain of Further and Higher Education where:

- There exists a valuable range of near real-time indicators available from established systems such as VLEs.
- The data is relatively small scale compared with applications such as Facebook or scientific research data.
- The audience is potentially well placed in terms of digital and analytical literacy.

We do need to beware of generic 'big data' scaremongering, readily amplified by the IT supply-side, taking up the baton from the heavyweight enterprise architecture and data warehouse lobby. Rather, it is suggested here that the headline first order challenges for analytics in Further and Higher Education more likely relate to priority, implementation and sustainability within the institution rather than to underlying technological barriers.

This perspective is evident in the basket of indicators proposed by EDUCAUSE for its forthcoming analytics maturity model. The organisational challenges of culture, governance, process and investment stand side by side with the more technical capacity relating to data, reporting, infrastructure and expertise.

In Part Two (Challenges) and Part Three (Response) we consider these factors in more detail.

3. PART Two – Challenges

This section explores the challenges presented by analytics in relation to the data ecosystem, data management and organisational readiness. These challenges should be regarded as pervasive and systemic, they potentially exist irrespective of the focus of the analytics effort and especially in cases where the target problem cannot be addressed through data from a single system.

3.1 VARIETY: THE DATA ECOSYSTEM

There are three core challenges that impact the whole institutional and external data landscape:

- Diversity of systems.
- Consistency of sources, coding and terminology.
- Quality of data.

Diversity

Diversity³ exacerbated by a multiplicity of operational systems is generally characteristic of the raw materials contributing to institutional analytics, that is the nature of the game.

- Systems implemented over long periods to satisfy essential but specialised requirements (for example VLEs and Library Management Systems) have coding frames and data in common.
- Analytics could significantly benefit from data in above-campus applications operated by bodies such as funding and research councils, JISC services, HESA and UCAS.
- Analytics needs to draw on complementary data from such as ONS, local government, sector agencies, market research and publishers.

University of Pennsylvania – the Metridoc systems addresses this problem space by providing an extensible framework that supports library assessment and analytics, using a wide variety of activity data from heterogeneous sources. It currently involves fund accounting systems, discovery tool logs, publisher reports and authentication logs. These are only the beginning of potential targets, which Pennsylvania is gradually expanding, guided by management and planning needs. This

³ Data sources could involve any or all of login/access, clicks/keystrokes and transactions:

Systems Interactions – e.g. Enquiries, Service Desk, Lecture Recordings

Human recorded activity – e.g. Student academic and financial records

Electronic transactions – e.g. Student / Staff Card, Library Card

ID related actions – e.g. System log-ins, e-resource access, RFID tagging

Back channels - Social Networks such as Facebook, Twitter, Foursquare

approach eases the burden of generating business intelligence by providing scalable and sustainable aggregation processes and a normalized data repository. The approach isolates extraction, transformation and repository services from the functions of specialist IT systems, enabling the creation of new collectors with manageable technical effort.

<http://metridoc.library.upenn.edu/>

Consistency

As well as providing mechanisms for coping with diverse data sources, it is essential that the institution should take steps to address the inconsistencies that frequently arise:

- Identification of and consolidation of systems of record where key data may be created and updated in one place only, even though it may be duplicated elsewhere on a read only basis, e.g. course codes or students' personal details.
- Standardisation of key coding frames to be used in all applications, notably for organisational units, courses and budget headings.
- Shared definitions of key terms, e.g. staff FTE or student modes of study.

Cornell University – Cornell reflects that determining the place to start with embedding analytics in corporate practice is a key issue for any institution, raising challenges relating to governance, coordination and culture. In order to catalyse our institutional requirement to develop analytics capabilities, we set ourselves the target for 2012-13 of refining existing administrative and financial dashboards to better meet the needs of our Trustees. This has allowed us to take some mission critical steps that will benefit our subsequent development of analytics in other areas - to identify systems and data elements of record, to engage data owners as stewards and co-contributors and to agree data governance principles.

We are only at the beginning of a journey towards mutual understanding of our data and embedding analytics in institutional practice. However we're confident that our starting point will create a solid foundation. We're also clear that the journey is essential.

From an interview with Ted Dodds, CIO - <http://www.cornell.edu>

Data Quality

Data quality is surfaced in issues ranging from cleanliness to provenance; for example:

- Can the data source be trusted?
- Is the data complete?
- Is it up to date?
- Is the data validated against standard coding frames?

Michigan State University – MSU emphasises that the key to the institutional analytics mission is quality of data. You cannot derive valid actionable insights if the data are unreliable. Our emphasis is therefore on establishing 'systems of record' that ensure that key shared data items used by many different systems, such as names and addresses, are entered and managed in one place. To date, we have been principally focused on developing sound analytics capabilities around

finance and human resources systems. However, we are attentive to a new wave of developments in the area of learning support, focusing 'in time' personalized support at the point of need.

From an interview with David Gift, CIO – <http://www.msu.edu>

3.2 VOLUME: DATA MANAGEMENT

The Requirement

There is a balance between answering known questions versus discovering unknown narratives. Experience commends atomic approaches linked to the potential for data capture as well as those dictated by established KPIs; 'Deductive' plus 'Inductive' data analysis. Adam Cooper of CETIS sums up this position:

My take home, then, is ... if we are serious about analytics we need to be thinking about exploratory data analysis and confirmatory data analysis and the label "analytics" is certainly inappropriate if neither is occurring. [Cooper, 2012][5]

To build those inductive narratives, the inclusion of generic 'connectors' will be essential, including institutional IDs, IP addresses and time, in addition to operational coding frames, such as course and budget codes.

Given these imperatives, data management is faced with key decisions about what to keep and for how long, what to aggregate and what to delete. Such decisions are susceptible to being made on the fly, influenced by immediate availability and affordability of storage, perceived processing demands and, perhaps most significantly, the current focus of analysis. However, to collect and manage analytics data on a 'just in case' basis is both challenging and essential.

University of Huddersfield The library has accumulated almost 4 million circulation transaction records over 15 years to 2012. These have been preserved in their atomic form. However, early analytical interests in resource utilisation could have led to space saving aggregation of that data, for example, by title and by period. With analytical focus now on the student experience, it is essential that the data preserved exist at the level of the individual transaction in order to make personalised recommendations and also to provide student progress indicators.

<http://eprints.hud.ac.uk/8453/>

The Tightrope

The lesson is to preserve wherever feasible, which puts significant pressure on defining collection and maintenance strategies in terms of the level of the data, the provision of the hardware and the responsibility for the process.

Institutions are therefore faced with the question of how that should be achieved in terms of infrastructure and management. As indicated, the potential scale of analytical data in educational institutions does not present a challenge to the available and proven technologies for storage, retrieval and analysis of analytical data. There are however choices between highly centralised 'enterprise' approaches and platforms that enable more distributed strategies. In extreme terms these may be characterised as:

- **Controlled / centralised** – deployment of data warehousing (e.g. from Oracle, Cognos, etc.) situated within a single homogeneous enterprise architecture.
- **Devolved / distributed** – local or problem specific approaches based on a range of locally useful stores and tools (e.g. NoSQL databases, Lucene / Solr indexing, text engines – the Learning Registry being an example).

Tightrope Walkers

Conversations within the community and the experience of JISC projects, synthesised at <http://www.activity.org> and in the JISC advisory paper on activity data [6] suggest that institutions will benefit most from a combination of these technical approaches, simply because they promote local innovation whilst securing data for institution wide, and as yet unknown, long term use. The following exemplars are detailed in the JISC advisory paper.

The Open University RISE development illustrates that model, delivering near-real time operational analytics serving users of e-resources, whilst at the same time making the data available to the institutional warehousing function. The Metridoc platform developed by University of Pennsylvania represents a scalable and reusable approach to the technical and operational challenges of diversity, providing a consistent method for local data acquisition and processing whilst supporting downstream integration in to corporate data warehousing.

As emphasised by Cornell, Michigan State and Roehampton exemplars, this model requires strategic direction at institutional level covering the data standardisation issues outlined above, notably identification of essential connector attributes and coding frames, plus a drive for data consistency by establishing systems of record. It also requires one or more methods of amalgamation and therefore some sort of data upload (push) or harvesting (pull) strategy, which might use a widely understood standard such as OAI-PMH.

However, we should conclude that there are no first order technical platform capability or capacity issues preventing the development of institution wide analytics. As set out in the next section, the core challenges exist in the areas of culture, policy and skills, in institutional confidence, control and capability.

3.3 VELOCITY: ORGANISATIONAL READINESS

We have suggested that the core first order challenges to the establishment of institution wide analytics of long term value exist in the areas of culture, policy and skills – in institutional confidence, control and capability. These challenges are set against global predictions of increasing pace of data growth (which may not apply as dramatically in education), escalation of operational and customer demand (which certainly applies as described above) for evidence-based near to real-time data, and also a market shortage of the necessary technical and business skills.

This position is clearly reflected in EDUCAUSE analysis in North America [7] as well as in the experiences of JISC funded projects in programmes addressing Activity Data, Business Intelligence and CRM.

Culture

EDUCAUSE questions about culture:

- What are the institution's cultural issues around buy-in and adoption of analytics?
- How does analytics relate to accountability?

The pressures arising from analytics for near to real time access to consistent data to drive evidence-based decision making may threaten a collision of cultures and management approaches.

Commentators highlight the importance of:

- Fostering a reflective and responsive culture.
- Rethinking attitudes to risk, to local innovation, and to the role of the IT service.
- Recognising that cultural change must be aligned with appropriate central and local (e.g. directorate or faculty) governance, covering both the data itself and data based decisions.

Policy

EDUCAUSE questions about policy:

- What institutional policies need to be updated or put in place in response to analytics initiatives?

Effective governance requires appropriate authority and clear policies. UK and US universities contributing to recent JISC activity data case studies emphasised the role of directive governance and policy to ensure accountability to data subjects and owners (including staff, students and researchers), analytical partners (such as funders), and others who need to trust HE data (such as parents and employers). This requires clear and committed policy in the areas of quality and responsibility.

Quality

- Identifying and positioning systems of record.
- Adoption of consistent coding frames for key elements across systems.
- Positioning of analytics in relation to personal and transactional data.

Responsibility

- Understanding and mitigation of risk.
- Appropriate interpretation of the legal position under UK law. [8]
- Considerations of ethical implications.

Roles

EDUCAUSE questions about roles:

- What is the role of the IT organization in analytics initiatives?
- Will it lead the provision of analytics services, or will it simply be a link in the overall supply chain?

Analytics implies dependencies on tight interaction between both institutional and external actors that differ from the demands of transactional IT and traditional 'reporting'. It is likely that those responsible for central IT, for institutional data research and for autonomous systems (e.g. the faculties, the library, outsourced suppliers and such as UCAS) will need to review and strengthen their relationships to recognise critical, complementary (even symbiotic) capabilities.

This may raise tensions in terms of:

- Local versus central imperatives.
- Service level agreements at all levels.
- Historic overestimations of capabilities and of quality of data.

Capability

EDUCAUSE questions about capability:

- What are the institution's existing capabilities and requirements in terms of data, infrastructure, processing, governance, and staffing for analytics?
- How will the institution address the shortage of professional analysts across all industries that is expected to continue for a number of years?

Skills and tools cannot be separated, though the right tools can lessen the skills entry level. This has been illustrated in the development of programming and report writing that have over twenty years taken on 'user interfaces', metaphors and layers of abstraction in order to hide the machine and the database. Currently analytics suffers from gaps both in skills and tools, a flashback to the position of relational database technology in 1982 when Oracle entered the market.

In the current phase of 'bootstrapping' of analytics capability, the following issues are critical and may in the short term benefit from direction and facilitation above the level of the institution, from bodies such as JISC working with UCISA and open source communities.

Availability of big data skills and tools:

- Global talent gap in terms of analytics expertise and experience.
- Development (including Digital Literacies) of the existing workforce.
- Longer term provision for the future workforce.
- Gaps in terms of user accessible and affordable (ideally open) tools, which may be filled by apps or other generic services for data transformation, indexing, processing and presentation.

Key technical / IT roles required to underpin analytics:

- Safeguarding the data within a diverse ecosystem.
- Supporting the tools in terms of infrastructure ('Can I set up a Non-SQL database?'), premium skills, reusability, advice and training.
- Partnering with internal stakeholders in exploratory analytical projects.

4. PART Three – Response Strategies

This section suggests response strategies of value to institutions and above-campus services seeking to develop a comprehensive approach to analytics that derives synergy, momentum and practice from a dynamic combination of local innovation and corporate direction and ensures that analysis leads to actions and benefits.

4.1 MAPPING THE JOURNEY: STRATEGY

An institutional analytics road map will benefit from iterative consideration of strategy. This paper has presented three high level approaches to the organisation of analytics, options that can inform the road map:

Choice	Key Pros	Key Cons
1) Local initiatives - let a thousand flowers bloom.	Targets real problems and short-term benefits.	Likely to reinvent the wheel in terms of data challenges, tools and skills.
2) Single enterprise solution – channel all effort through the IT team.	Maximises data consistency and cleanliness, potential for reuse and exploration.	Limited in terms of resources and agility; inability to respond to local service interests.
3) Hybrid – embed key controls and channel the data from local to central.	Enables a navigable data ecosystem and encourages local drive and innovation whilst safeguarding long term value.	Depends on organisational maturity and sensitive governance; overheads escalate if poorly conceived.

Institutions will, explicitly or implicitly, be favouring one or other of these choices. In Part Two of this paper, we have argued that the hybrid model offers the best opportunities to maximise benefits in terms of building momentum, developing practice, enabling reuse and securing long term data value. One way or another, it is essential to assess the implications of such a choice in terms of governance, processes and skills.

4.2 MAPPING THE JOURNEY: STARTING POINTS

Given increasing awareness of how a range of operational IT systems and other data sources might play in to the development of analytics, plus the response of business systems vendors to the opportunity (e.g. Blackboard, ExLibris), there is some danger of paralysis in deciding where to start.

Low Hanging Fruit

Immediately fruitful areas benefitting from analytics might include:

- Effectiveness
 - Student lifecycle – choices, wellbeing, success and retention.

- Research processes – grant success factors, research output measures.
- Student and researcher experience
 - Effective personalisation of learning.
 - Resource recommenders.
 - IT availability and performance.
 - Client surveys.
- Efficiency and economy
 - Optimising the library collection and learning resource investments.
 - Tuning estates and catering services and waste.
 - Finance – budget trend analysis.

Based on such areas, a relatively simple audit and readiness evaluation of existing systems and operational processes is recommended to identify whether systems are already collecting data that could address core business challenges.

Fundamental feasibility questions include:

- Is relevant data already being collected? e.g. timed stamped activity data.
- Does the collecting system enforce standard coding? e.g. for courses, individuals, organisational units.
- Can the data be regularly extracted in machine-readable format?
- Is the analytical value of this data dependent on data from elsewhere inside or outside the institution? E.g. Student registry, UCAS.

Bubbling Under

Project work and individual institutional explorations suggests a wide range of potential targets that are less well explored but may suit the local priorities and preparedness. These may include:

- Personal interventions - drawing on help desk, call centre and social network data to enhance the trends and interventions analysis derived from enterprise applications managing enquiry, admissions, student welfare, advice and guidance.
- Relationship management – Use of analytical techniques to enhance value of business, community and alumni networks, perhaps involving external business sector or destinations data.
- Performance management – staff performance management is a strong target for analytics in some business sectors where strongly role related data are readily available (e.g. call centres, point of sale); personal performance management in education needs has particular sensitivities and therefore requires careful consideration.
- Portfolio coverage and course design – mapping local demographic and public examinations data to identify pockets of interest and specialism for potential course or unit development.
- Research success - mapping research profiles, grant opportunities and outcomes to facilitate fruitful partnerships and successful applications.

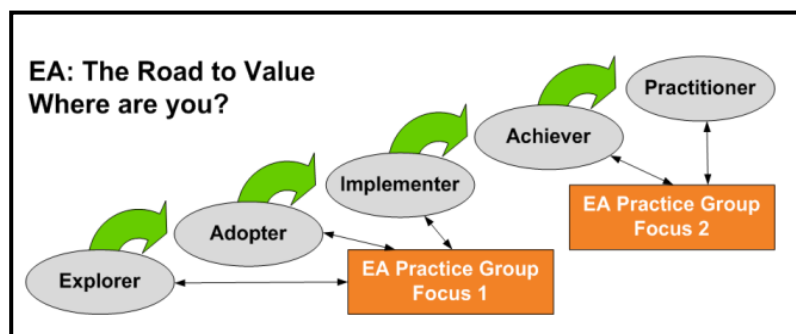
External Data Sources

In order to maximise the narratives and analyses that can be derived from locally collected institutional data, it should be a priority to secure access to data from trusted external sources and to make it accessible for consumption across the institution. Some of this will be 'open data' (e.g. openly licensed under an Open Government Licence or a Creative Commons CC0 licence for unrestricted reuse) or provisioned by a sector service.

- Above-campus / shared service sources, e.g. JANET (UK), JISC Collections.
- Sector-related agencies and national services, e.g. HESA, LSIS, National Student Survey, REF, The Information Authority, UCAS, funding bodies, research councils.
- Central and local government datasets from bodies such as ONS, e.g. target demographics, economic and employment statistics by sector, destinations.
- Independent third party collectors, e.g. social networks, publishers.

4.3 MEASURING PROGRESS

There are several frameworks for professional and organisational readiness and maturity in specialised ICT areas such as enterprise architecture and e-learning. The idea of the road to value was central to the JISC Flexible Service Delivery (FSD) programme, linking investment in new approaches with tangible institutional value in the eyes of senior management [9]. A range of IT developmental initiatives have found a simple maturity model to be a good way to gauge progress and the activities that help at each stage.



Drawing on the JISC Enterprise Architecture model, illustrated above, this might be presented for analytics development as follows:

- **Explorer:** researching, investigating, identifying potential change projects, developing a case.
- **Adopter:** planning, orienting, engaging with colleagues, designing a live project.
- **Implementer:** initial project under way, with training and support.
- **Achiever:** first results, where impact and value are evident – though that may be hard to quantify at this stage.
- **Practitioner:** analytics has been established as a professional approach for operational and strategic decision support.

Maturity Checklist

In 2012, EDUCAUSE consulted with its community to develop a maturity model to enable institutions to self-assess their progress towards embedding analytics in everyday business practice. The online tool, which will be piloted later in 2012, will present around thirty criterion statements that fall into readily recognisable categories, broadly as follows. Responses will be on a standardised scale, e.g. from 'strongly agree' through to 'strongly disagree'.

The categories illustrate the value of examining on an ongoing basis the development and use of analytics from a range of perspectives and therefore the importance of an organisational commitment to respond to the opportunities and the challenges identified in the process.

Category	Expectations to be covered by criterion statements
Culture	Buy-in and understanding is required across the range of faculty, service and management levels.
Governance	Clear policies and responsibilities will ensure legal and ethical obligations are fulfilled.
Process	Process must support identification of target outcomes, and moving from what the data says to making decisions and enacting change.
Data	The outputs can only be as good as the data so the right data should exist in a consistent and clean form.
Infrastructure and Tools	Capability and capacity is required for storage, processing, analysis, reporting and visualisation.
Expertise	Training must compensate for the shortages of professional analysts across all industries expected to continue for a number of years.
Investment	The critical enabling investment will be in practical business-facing initiatives and in skills.

As an illustration, draft criterion statements under 'Process' include:

- There is effective communication between our IT and IR (Information Research) departments.
- Our senior institutional researcher is involved in the planning process for addressing high-level strategic initiatives or questions.
- We have identified the key institutional outcomes we are trying to improve with better use of the data.
- Use of data is part of our strategic plan.
- We have a process for moving from what the data says to making changes and decisions.

Learning From Local

It is important to consider whether the focus of any self-assessment model should be on the whole institution, on subsidiary project developments or on both. For example, the following table shows how a basket of indicators or measures could usefully be applied to a local project and seen as a stepping-stone to the wider adoption and embedding of analytics practice and policy across the institution.

Experience in the MOSAIC activity data project [10] suggested a number of critical success factors. Whilst some factors generally apply to strategic ICT projects, they typically require specific consideration in the case of business intelligence and analytics on account of the exploratory intent, the variety of approaches and the potentially institution wide nature of the undertaking.

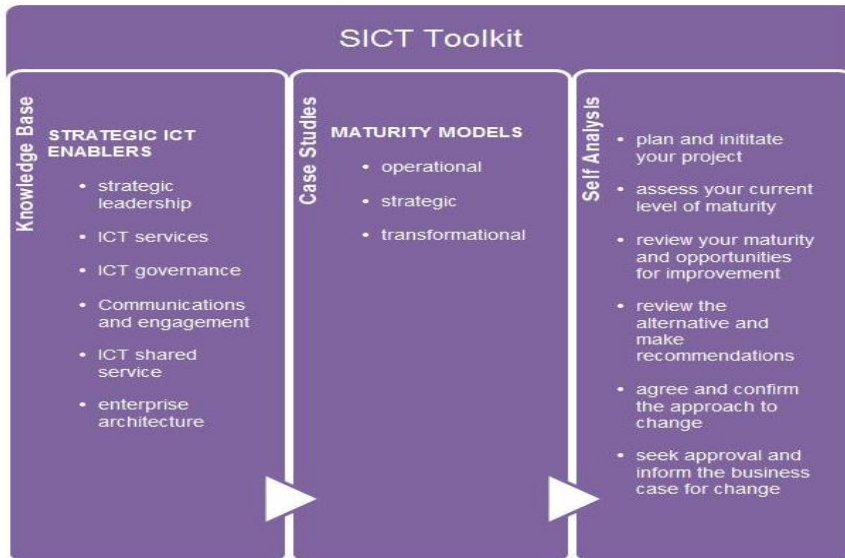
In the following table the project is scored (in this case a self-assessment by the team members) before, during and after using a four point scale from **Weak** > **Adequate** > **Good** > **Strong** plus **'Not Considered'** (a very significant indicator in exploratory developments).

Success factors	Up front	During project	Future prognosis
1 Problem definition	Adequate	Strong	Strong
2 Risk management	Weak	Adequate	Good
3 Data mapping	Adequate	Good	Good
4 Data quality	Weak	Adequate	Good
5 Technical platform	Adequate	Good	Strong
6 Implementation	Adequate	Adequate	Adequate
7 Operational enactment	Not considered	Good	Adequate
8 Strategic impact	Adequate	Good	Good
9 Future potential	Good	Good	Strong
10 Sustainability	Weak	Weak	Adequate
11 Skills	Not considered	Adequate	Adequate
12 Advocacy	Weak	Good	Good
13 Governance	Not considered	Adequate	Good

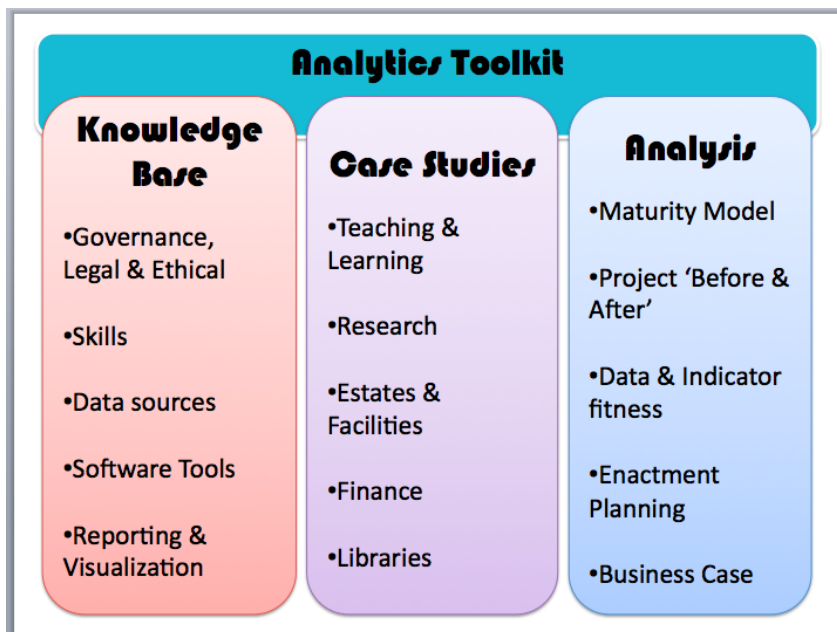
It should be noted that the thirteen factors used here, which are derived from reflection on the JISC Activity Data programme projects (2011), are closely aligned to the categories under consideration for the EDUCAUSE institutional model.

Shared Models

There is a strong tradition in the education community of developing shared models, often internationally applicable, to enable institutions to self-assess and derive benchmarking intelligence in support of continuous improvement. Recent benchmarking work by EDUCAUSE, the HE Academy and JISC is notable in the general IT and e-learning spaces, e.g. the JISC InfoNet Strategic ICT Toolkit [11] developed with the University of Nottingham to help higher education institutions (HEIs) to analyse, assess and develop their strategic use of ICT.



As imagined below, the institution wide challenges and opportunities raised here suggest that a similar toolkit could be developed to encompass the range of analytics themes covered in this CETIS Analytics Series.



In the 'whole institution' context, the observations in this paper indicate scope for a range of supporting tools, potentially including the EDUCAUSE Maturity Model, thus providing opportunity for international benchmarking and dialogue.

5. References

1. A. Cooper, "What is Analytics? Definition and Essential Characteristics", *CETIS Analytics Series*, Vol.1 No.5. 2012. <http://publications.cetis.ac.uk/2012/521> (to be published in December).
2. D. Clow, (2012), "The learning analytics cycle: closing the loop effectively", in *LAK12: 2nd International Conference on Learning Analytics & Knowledge*, 29 April - 2 May 2012, Vancouver, BC. <http://oro.open.ac.uk/34330/1/LAK12-DougClow-personalcopy.pdf>
3. <http://www.educause.edu/About+EDUCAUSE/PressReleases/EDUCAUSEAnnouncesNewInitiative/239558> - October 17, 2011 — EDUCAUSE will work with the higher education community to provide thought leadership and education to develop institutional capability for analytics.
4. A. D. Joseph, (2012), "A Berkeley View of Big Data", EDUSERV Symposium, 10 May 2012. http://eduserv2012.s3.amazonaws.com/Anthony_Joseph.pdf
5. A. Cooper (2012), "Exploratory Data Analysis" <http://blogs.cetis.ac.uk/adam/2012/05/18/exploratory-data-analysis/>
6. JISC Activity Data: Delivering benefits from the data deluge (to be published in December) <http://www.jisc.ac.uk/publications/reports/2012/activity-data-delivering-benefits.aspx>
7. J. Bichsel, (2012), "Analytics in Higher Education: Benefits, Barriers, Progress, and Recommendations", EDUCAUSE Center for Applied Research. <http://net.educause.edu/ir/library/pdf/ERS1207/ers1207.pdf>
8. D. Kay, N. Korn and C. Oppenheim, "Legal, Risk and Ethical Aspects of Analytics in Higher Education", *CETIS Analytics Series*, Vol.1 No.6, 2012. <http://publications.cetis.ac.uk/2012/500> (to be published in December)
9. The Road to Value, JISCInfonet, <http://www.jiscinfonet.ac.uk/flexible-service-delivery/ea/road-to-value>
10. MOSAIC Activity Data Project <http://www.sero.co.uk/jisc-mosaic.html>
11. Delivering the Technology Enhanced Institution, JISCInfonet, <http://www.jiscinfonet.ac.uk/strategy/ict/maturity>

About the Authors

David Kay is a senior consultant at Sero Consulting Ltd (<http://www.sero.co.uk>). He joined Sero in 2004 after over 20 years involvement in the design and development of IT systems for education and library management and for learner support. David began investigating the potential of activity data and resulting analytics in connection with Higher Education libraries in the 2009 JISC TILE project, working with Mark van Harmelen (Hedtek) and Ken Chad. He has subsequently been involved in JISC's further examination of those concepts in the MOSAIC demonstrator project and in the University of Manchester synthesis of the 2011 Activity Data programme (<http://activitydata.org>).

Dr Mark van Harmelen is the Director of Hedtek Ltd, both a strategy consultancy and a software house that develops highly usable web-based services, including services for libraries. Mark is also an Honorary Research Fellow in the University of Manchester's School of Computer Science. Previously Mark has lectured and consulted broadly in industry, been a Senior Researcher for Matsushita Electric in Japan, and a South African Cabinet appointee tasked with formulating the establishment of the Meraka Institute in South Africa

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 Vol.1, No.2, Analytics for the Whole Institution, Balancing Strategy and Tactics

Authors: David Kay (Sero) and Mark van Harmelen (Hedtek)

Date: November 2012

URI: <http://publications.cetis.ac.uk/2012/513>

ISSN 2051-9214



Text Copyright © 2012 David Kay & Dr. Mark van Harmelen ; 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.

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/>

The Analytics Series has been produced by CETIS for JISC: <http://www.jisc.ac.uk/>