Re-examining dashboard development: putting the horse back in front of the cart

Paper presented in track 3 at the

EAIR 37th Annual Forum in Krems, Austria

30 August till 2 September 2015

Name of Author(s)

Elizabeth Archer
Glen Barnes

Contact Details

Elizabeth Archer
University of South Africa
PO Box 392, UNISA
Tshwane, 0003
South Africa
E-mail: archee@unisa.ac.za

Key words

Management, Strategic planning

Abstract

Re-examining dashboard development: putting the horse back in front of the cart

During this age of data proliferation, heavy reliance is placed on data visualisation to support users in making sense of vast quantities of information: Finding the signal in the noise (Silver 2012). Informational Dashboards have become the must have accoutrement for Higher Education institutions with various stakeholders jostling for development priority. Due to the time pressure and user demands, the focus of development process is often on designing for each stakeholder and the visual and navigational aspects. Dashboards are designed to make data visually appealing and easy to relate and understand; unfortunately this may mask data issues and create an impression of rigour where it is not warranted. This paper explores an alternative approach to data visualisation and dashboard design. It suggests that the first step should be the development of an enriched database which integrates key indicators from various data sources and then acts as a database for various dashboards. The emphasis is thus on the underlying value-added database which can then be overlaid with any user interface for dashboard generation for a multitude of stakeholders. The enriched central database thus becomes a menu of available quality assured data which can easily be drawn into various dashboards at various levels of granularity ensuring improved data quality management, responsiveness, efficiency and flexibility. A case study is employed illustrating the benefits, showcasing various views developed for diverse stakeholders employing this approach.

Presentation

Re-examining dashboard development: putting the horse back in front of the cart

The aim of the paper is to explore sound approaches of meeting stakeholder data visualisation requirements in the era of over-whelming data demand and supply, while maintaining data integrity, consistency and flexibility. Traditionally higher education institutions always had access to relatively large data sets and tools for analysis. This is growing exponentially with the ever increasing amount of digital student data that can be harvested and analysed (Swain, 2013), as well as increased technological and analytical capabilities (Wishon & Rome, 2012). Learning analytics has been described as the "new black" (Booth, 2012), and student data as the "new oil" (Watters, 2013). The *2013 NMC New Horizon report: Higher Education Edition* (New Media Consortium, 2013) identifies learning analytics as one of the key emerging technologies to enter mainstream use from 2015-2016. The report (New Media Consortium, 2013) also identifies dashboards as a key technology in leveraging the power of data at all stakeholder levels.

Rationale

With the exponential growth of data, with multiple complex relationships, differing constraints, the ever changing environment, uncertainty and time pressure, problem-solving and decision-making can be overwhelming. Dashboards have emerged as a technology to enable stakeholders to engage with the data effectively through various data visualisations (Abd-elfattah, Alghamdi, & Amer, 2014). A dashboard can be defined as: "… a visual display of the most important information needed to achieve one or more objectives; consolidated and arranged on a single screen, so the information can be monitored at a glance" (Few, 2006, p. 26). Dashboards started to gain traction in the 1980's in the business world and have since gained popularity in Higher Education (Newman, Thomas, & Webber, 2014). Many tools have been developed for the purpose of designing dashboards and disseminating information in graphic formats, these include IBM Cognos, Oracle BI Foundation Suite, SAS Enterprise Intelligence Platform, SAP Business Object BI Platform, MicroStrategy, QlikView and Web Focus, amongst others.

The proliferation of tools and demands by various stakeholders for immediate development of their own personal dashboards, has often lead to a focus on the visualisation design aspects of the dashboards or the particular functionality of the tools at the expense of the message being communicated. Dashboards are by definition custom designed and as such the design process often starts with determining the stakeholder needs and elements to be present in the dashboard. Once this is determined the required data is accessed, transformed and integrated from various data sources in order to provide the data for the visualisation. This results in a multitude of smaller enriched data sources, each designed to service a particular dashboard. This proliferation makes it more complex to apply changes to business rules consistently, be responsive, manage data quality and remain flexible as user needs change. (Abd-elfattah et al., 2014; Few, 2006; Newman et al., 2014)

Another aspect of data dashboards which has received growing attention, is that of memorability (Borkin et al., 2013). These authors (Borkin et al., 2013) state that memorability is intertwined with cognition and understanding and an important step in designing visualisations such as dashboards.

In this paper an alternate approach is explored where the focus is on establishing a comprehensive, highly navigable data sphere, which allows users to navigate and interrogate data in real time to address their dynamic informational needs. The focus is on designing and integrating enriched databases with an intuitive navigation system allowing users to easily move between various data nodes and explore each node in detail to gain a comprehensive understanding of the real time data relating to their specific challenge. Data presented is as close to real time as possible, focusing rather on allowing problem identification and exploration at will as opposed to memorability and visualisation.

Contextualisation

This paper explores a case study at the University of South Africa (Unisa) where an alternative approach is employed to dashboard development. The University of South Africa is one of the mega Open Distance Learning (ODL) institutions in the world with more than 360,000 students and a range of courses and programmes. Unisa already has access to a staggering amount of student data, hosted in disparate sources, and governed by different processes. The data covers the entire student journey from admissions and registrations, to learning activities, course success, retention, graduation, employment and citizenship. The existing data sources lie in a number of disparate operational systems maintained by various functional units within the institution. The disparate systems typically have grown from the need of different functional areas to have customised functionality built into operational systems. These developments take place without a reporting or analytic objective in mind with the result that levering and integrating sensible information from these sources becomes a problem. (Prinsloo, Archer, Barnes, Chetty, & Zyl, 2015)

As the university moves to mainstreaming online learning, the amount of and need for analyses of data are increasing, raising important questions regarding our assumptions, understanding, data sources, systems and processes (Prinsloo et al., 2015). The case documents the development of an enriched central database, created to extract, combine, transform and house as many possible data elements which may have value to a diverse body of users. This database serves as the basis for many different dashboards, serving the needs of stakeholders ranging from students, staff, student support services to management. This approach allows for greater consistency of data presentation and rule application across various views and shifts the focus from the dashboard software capabilities to the core principles of relevance, clarity, data quality, consistency and ability to explore. It also allows for greater independence from any individual dashboard design tool.

Methodology

The authors engaged with literature and the current status of student data at Unisa in a systematic way over the course of six months. Several meetings were held with various stakeholders, and the analysis in this article resulted from the field notes from these meetings. Construct validity was ensured by engaging with multiple sources of evidence. The team confirmed internal validity by explanation building and addressing rival explanations as suggested by Yin (2009). The resulting theses were compared with the literature and available evidence in the field. This paper employs the insights, experience and thoughts of these researchers faced with the often undocumented realities of engaging with vast amounts of data and providing actionable information to stakeholders. As an interpretive case study we attempt to not only describe a specific case of the use of student data, but interpret the case to move to propositions or theses (Thomas, 2011). Our aim is therefore to advance phronesis or practical knowledge.

Exemplary case

An example of how the data sphere can be used to facilitate real time data exploration will be presented in order to illustrate how the central, enriched database can be leveraged to provide actionable, timely data to stakeholders with diverse needs. We will launch the discussion by discussing the design logic employed in the development of this data exploration environment, followed by an example of the application and illustration of the node access screens (dashboards) as well as a discussion of the various data exploration approaches.

Design logic

The data sphere development employed the following design principles:

1. The database should be aggregated and built up from the lowest level of granularity (in the case of our example, the student)
2. The database should be enriched with additional variables and calculations which are often required in exploration and reporting
3. Data will exist in layers as they relate to key nodes
4. Users must be able to enter the data sphere at any layer through one of the key nodes
5. Users must be able to explore each layer in depth and move between layers at will to extend their exploration.
6. It must be possible to aggregate and disaggregate data at various levels and for various purposes.
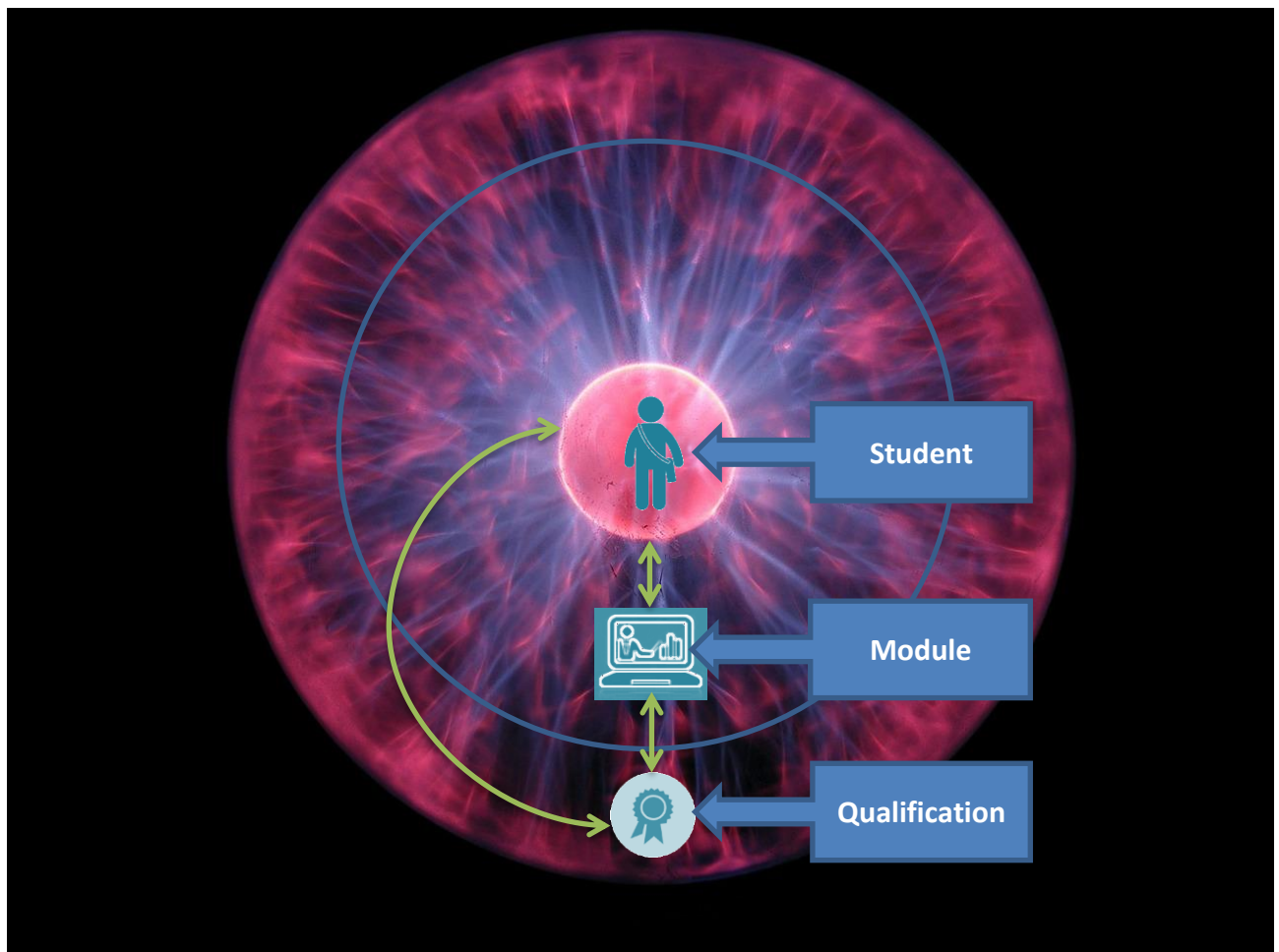
(Public Domain, 2015)

Figure 1: Data Sphere

In this illustrative case we will examine data with pertaining to teaching and learning. The data sphere in this case is layered around three entry nodes (See Figure 1):

- Student
- Module
- Qualification

The data sphere holds the real-time, enriched and quality assured data warehouse. Users can enter the data sphere to start their exploration at any one of these nodes using either a student identifier (e.g. name, date of birth, student number), module identifier (e.g. module code, module name, navigation through the curriculum structure) or qualification identifier (e.g. qualification code, qualification name, navigating

through the organisational architecture). Once a user has entered any of the layers, they are presented with a dashboard which provides key information for the particular student, module or qualification they are exploring. A signboard to orientate the user to what is available for exploration in that layer of data, based on a certain level of granularity. These dashboards act as the layout map for each level to show what aspects can be explored at the particular level of granularity. Once a user has engaged with the primary dashboard for the level they can click on various aspects that they wish to explore further either moving through data on the same level (drill through, 360°) or shifting their exploration of the data to another level (drill down or roll up). Moving through and between levels is facilitated through catalysts in the form of student, module and qualification identifiers.

The idea is similar to exploring a multi-storey department store. Each node (student, module or qualification) would represent a different department (data level). Each department is located on its own floor. You navigate between the floors (data levels) by using the elevator (catalysts - student, module and qualification identifiers). When you reach the specific store you need to select which department (data layer) to go to and use the lift to go to the relevant floor by using your entry key (node identifier). Once the elevator door opens you are confronted with a sign-board showcasing what is available on that particular floor and you can decide which areas to visit (floor map or node dashboard). When you have finished on a particular floor, you may want to move to another floor and you thus use the elevator (catalysts or node identifiers: student, module or qualification identifiers) again to navigate to a new floor, to explore what is available in more depth.

Illustrative case
We will now illustrate how the data sphere can be employed through an enrolment management example relevant to the higher education environment.

Scenario
A college (faculty) planning workshop is taking place. The purpose of the meeting is to establish qualification enrolment targets for the 2016 cohort.
The workshop is attended by the:
* Top Management: Institutional development and Academic Planning,
* College level: Executive Directors, Deputy Deans, Heads of Schools, Heads of Departments
* Support: Strategic Planning, Quality Assurance, Information Analysts Purpose – setting enrolment targets for each qualification for 2016 cohort

Initial discussions and engagement has resulted in the development of draft targets which have been captured on a planning spreadsheet to capture targets from 2016 to 2019. These targets now need to be interrogated to examine if they are feasible and where action is required to achieve the targets.

View point identification
A data exploration will now take place using the data sphere to determine which qualification targets may be problematic and to explore how to either achieve these targets or adjust the targets. An example of such a spreadsheet is provided in Table 1 below.

Table 1: Excerpt of qualification history and targets set for 2016

Qual Code	New Qualification Name	Strategy	2010	2011	2012	2013	2014	2015	2016
98615	HIGHER CERTIFICATE IN ADULT BASIC EDUCATION AND	Growth strategy	2709	3721	5735	11318	10644	9307	13084
98999	MASTER OF EDUCATION in Open and Distance Learning	Sustainable strategy				13	12	19	15
99001	DIPLOMA in Adult Basic Education and Training	Declining strategy	4460	4608	3314	2206	1725	1316	1006
0264X	ADVANCED CERTIFICATE IN EDUCATION (FOUNDATION	Phased out	106	13	4	1			0
0328X	ADVANCED CERTIFICATE IN EDUCATION: TOURISM EDU	Phased out	37						
0376X	ENDORSEMENT: SPECIALISATION IN INCLUSIVE EDUCAT	Phased out	14	25	21				0
2032X	DIPLOMA IN EDUCATION (SECONDARY PHASE)	Phased out	1						

The data exploration approach employed in this case emphasises identification of problem areas and then diagnostic exploration to have comprehensive data to facilitate planning and decision making. As such, the first step is to identify the targets that warrant further exploration in this case.

From Table 1 it can be seen that the targets envisaged for the Qualification 98615 (Higher Certificate in Adult and Basic Education and Training) may be problematic given the recent enrolment history. Whilst enrolment increased steadily from 2010 to 2012 and showed a jump in 2013, there has been a steady decrease in enrolments since 2014. The new enrolment targets require a more than 40% growth, not only altering the trend of decreased enrolment, but requiring higher enrolments than have been achieved in the last 6 years. This qualification should thus be explored in greater depth. As we are using the new data sphere exploration approach we will not request a full examination by our Institutional Research directorate, but request that the analyst uses the data sphere to explore the data for this qualification in real time so that we may engage with it.

Possible viewpoints
In exploring this qualification we want to examine a number of aspects. These aspects can be seen as possible viewpoints during our data exploration journey.

	Possible question	Viewpoints
1	How do previous versions of the qualification contribute to enrolments and graduates?	Equivalents view
2	What is the pool of interest and potential uptake?	Attrition analysis
3	What are the inflows and outflows –intake, first time intake, returning, dropouts and graduates?	Qualification Flow Planning View
3	How many provisional enrolments are required to achieve the official census date targets[1] and what is the required workload?	Qualification Flow Planning View
4	What is the spatial distribution of these students?	Cohort View
4	What is the race, gender, matric score, age distribution of the students in this qualification?	Cohort View
5	What risk aspects are defined for this qualification?	Risk Management view
5	What modules are included in this qualification and how many of these modules have been deemed as "at risk"?	Risk Management View

Our first step in this journey is to enter the data sphere at one of the nodes. In this case it is the qualifications node, however before we explore this fully, we need to determine if there are any equivalents to the current qualification naming and number allocation: ***How do previous versions of the qualification contribute to enrolments and graduates?*** This is explored through the *Equivalents view* which in this case shows us that while there is a previous equivalent qualification. Only one student from

---

[1] Officially audit data captured on a specific date for Higher Education Institution in South Africa is known as HEMIS (Higher Education Information Management System) data. This is the basis for subsidy allocation by government.

the previous qualification is still busy completing and students will no longer be able to enrol in the old qualification code. The old qualification enrolment numbers have already been incorporated into the new code. We are thus free to explore this qualification employing only the new qualification code.
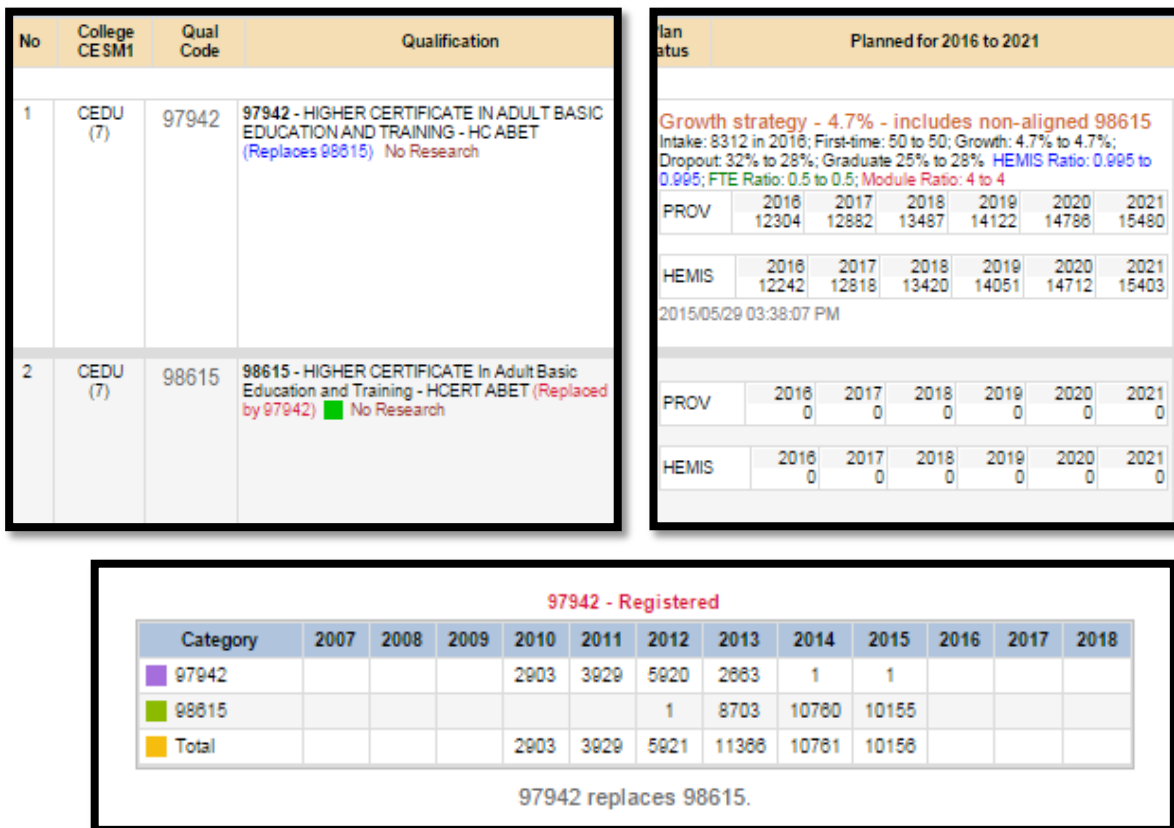

Figure 2: Excerpts Equivalence View

Now that we have identified the relevant qualification code, we can examine this data level further. Our first stop is the attrition view: ***What is the pool of interest and potential uptake?*** Here we have an overview of the various types of attrition taking place (see Figure 3).

Figure 3: Attrition View

The attrition view considers data from applications through to final registrations and active students submitted for statutory purposes. This view enables a rich understanding of the various points of attrition along that trajectory.

Our next point of interest is to see the inflow/outflow and planning over time, thus the Qualification Flow Planning View. We can explore current and historical enrolment data: ***What are the inflows and outflows – intake, first time intake, returning, dropouts and graduates?*** We can also explore scenarios to achieve growth: ***How many provisional enrolments are required to achieve the official census date targets and what is the required workload?*** The two excerpts from the Qualification Flow Planning view are provided below.

Actual Inflow & Outflow

Category	2011	2012	2013	2014	2015	2016
Intake	3585	4234	8696	6561	5190	
% of Total	91.2%	71.5%	76.5%	61%	51.1%	-
First-time	3026	3792	7291	5605	4610	
% of Intake	84.4%	89.6%	83.8%	85.4%	88.8%	-
Returning	335	1683	2645	4098	4760	
% of Total	8.5%	28.4%	23.3%	38.1%	46.9%	-
Dropouts	1633	1825	5042	3291	10144	
% of Total	41.6%	30.8%	44.4%	30.6%	99.9%	-
Graduated	613	1451	2226	2710	12	
% of Total	15.6%	24.5%	19.6%	25.2%	0.1%	-
Registered	3929	5921	11366	10761	10156	
% of Total	100%	100%	100%	100%	100%	-
Total Growth	-	50.7%	92%	-5.3%	-5.6%	-100%

97942 replaces 98615.

Figure 4: Excerpt Qualification Flow Planning View - Actual Inflow and Outflow
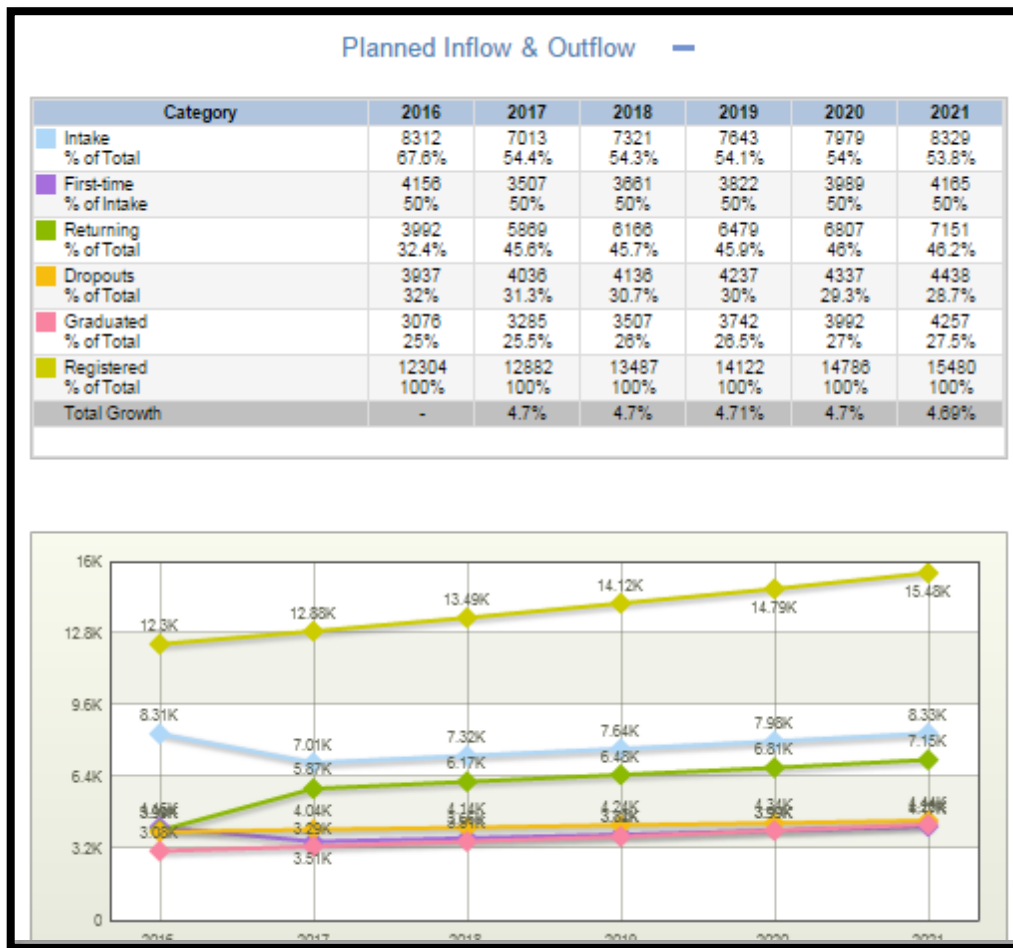
Figure 5: Excerpt Qualification Flow Planning View - Planned Inflow and Outflow

As a next step in our exploration we wish to examine (profile) the students within this qualification. As with all the other views, the cohort view is highly flexible and the user can select the information he or she wishes to engage with (see Figure 6)

Figure 6: Excerpt Cohort View – Profile Display Options

In this first selection, we are examining: *What is the race, gender, matric score, age distribution of the students in this qualification?*
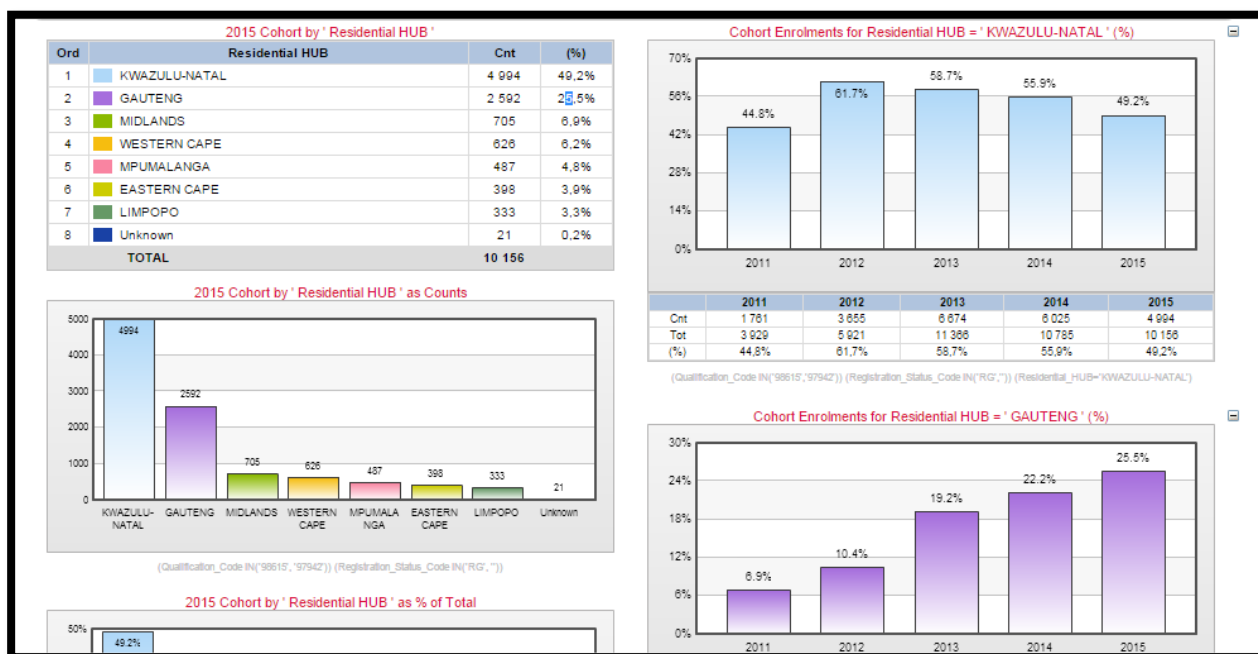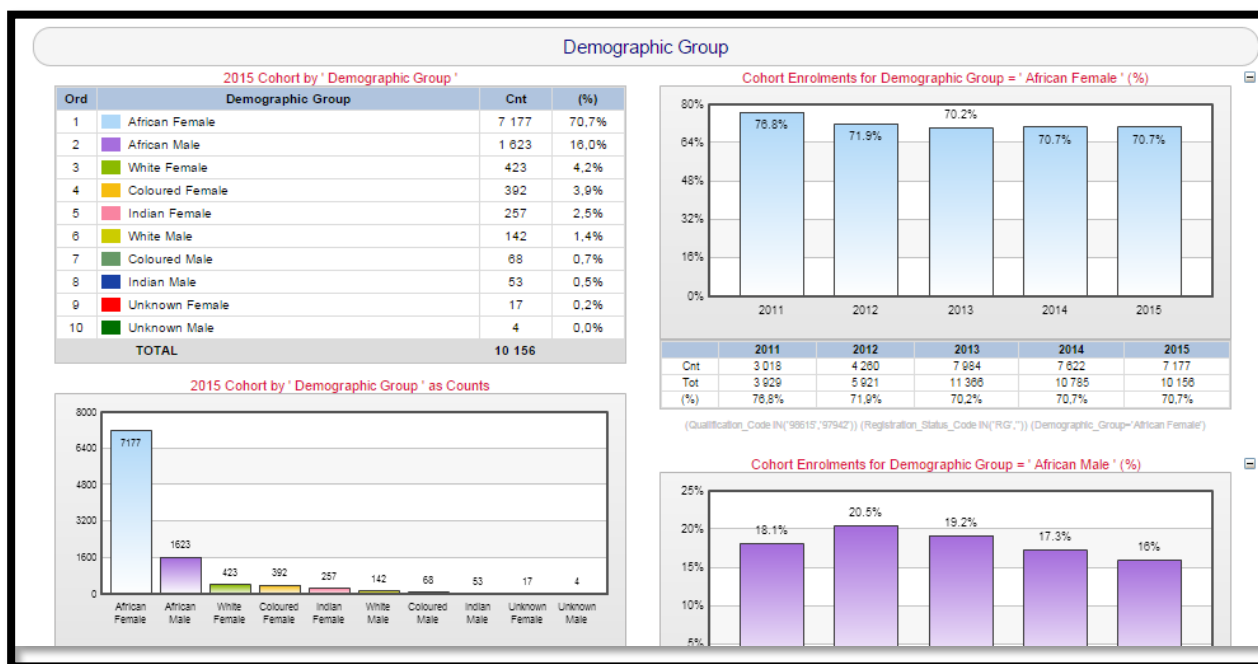
Figure 7: Excerpts Cohort View

Finally, we pay a visit to the Risk Management View: *What modules are included in this qualification and how many of these modules have been deemed as "at risk"? What risk aspects are defined for this qualification?* The main Risk Management View condenses multiple indicators into one succinct view. In our case we are interested in which modules have been deemed 'at risk' or 'high risk' (last 2 columns). We are also interested in any modules included in these qualifications which have seen a decrease in the Normal exam pass rate (NPR)[2]. These three modules can thus be identified for additional support to

---

[2] Normal exam pass rate – the number passed relative to those that wrote (no deferments) – the field

improve the pass rate. (see Figure 8)

Figure 8: Excerpts Risk Management View

Discussion

This data exploration process takes place in real-time during the college planning workshop, ensuring relevant, accurate and timely data is available to make the decisions. The data exploration is also logged and exported into a report to support and document the decisions taken at the workshop. Based on the comprehensive real time exploration the group could establish the pool of interest, pool of suitable candidates, points of possible attrition, geographical distribution of students, student profiles, success and barriers to graduation. With this information at hand it is possible to decide whether the envisaged enrolment target is feasible and if so what is required to achieve it.

In this example the participants in the workshop decide that although the enrolment in this qualification has decreased over the last few years, there is a high enough pool of interest that the 2016 target could be achieved. It is however clear that some interventions will be required to attain this target. The attrition view shows that while there is a high interest in the qualification (80 195 applications), this does not necessarily convert into enrolments (36 759 registrations); this may be because many students see this qualification as a second or third choice. The attrition rate for this qualification is very high (54%) and there are also a number of high risk modules in the qualification. Providing additional support for these modules may ameliorate some barriers to success for this qualification. The envisaged 2016 enrolment target is thus approved with the following interventions to be put in place:

- Marketing;
  - General marketing to increase interest and applications for the qualifications
  - Targeted marketing aimed at students who have applied for this qualification in order to increase the conversion of applications into registration
- Resource allocation;
  - Assigning additional online tutors in line with increased enrolments
  - Assigning face-to-face tutoring for students in 'at risk' modules
  - Scheduling additional face-to-face tuition sessions based on the geographical distribution data available

Catering for various users

The example provided above was for management planning purposes at a high level. The data sphere is however designed in such a way to allow for multiple levels and various users. Exploration must be made possible for users dealing with various entry nodes and users who have varying levels of data literacy and knowledge of the system. In addition, role based access is facilitated to ensure dissemination of appropriate data and information to users. As such three supporting features are essential: the entry node dashboard (data level floor plan), providing suggestions for exploration (suggested itineraries) and the ability to move between the various data levels at will (catalysts or elevators). These are discussed shortly.

shows the previous year pass rate, an indicator of change (up or down) and the current year pass rate

Node dashboards

As discussed earlier, users can access the data sphere at three nodule entry points: student, module and qualification. The node dashboard is the first data view that the user will be confronted with. This will provide a quick overview of the data available at that data level with click through and click through capabilities. This is a simple diagnostic view which allows the user to see which aspects they are concerned about for the student, module or qualification they are exploring and provide them with the opportunity to visit different aspects providing more information about these areas of concern. The student node dashboard is shown in Figure 9 below to illustrate this.
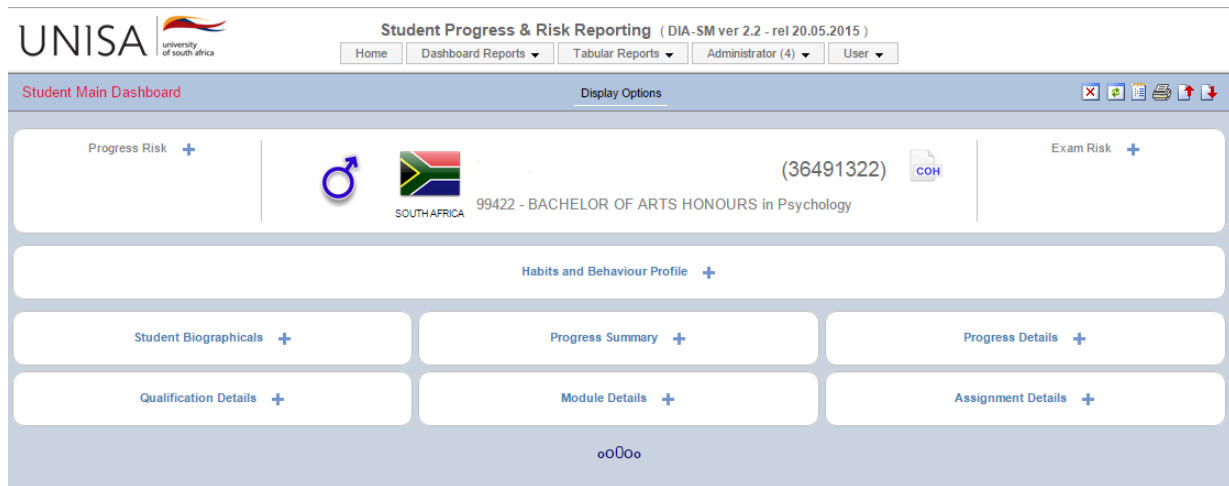
Figure 9: Excerpt Student Node Dashboard

Information is 'on demand'; the various sections open depending on the interest and requirements of the user. An example of an 'expanded' section is given in Figure 9.
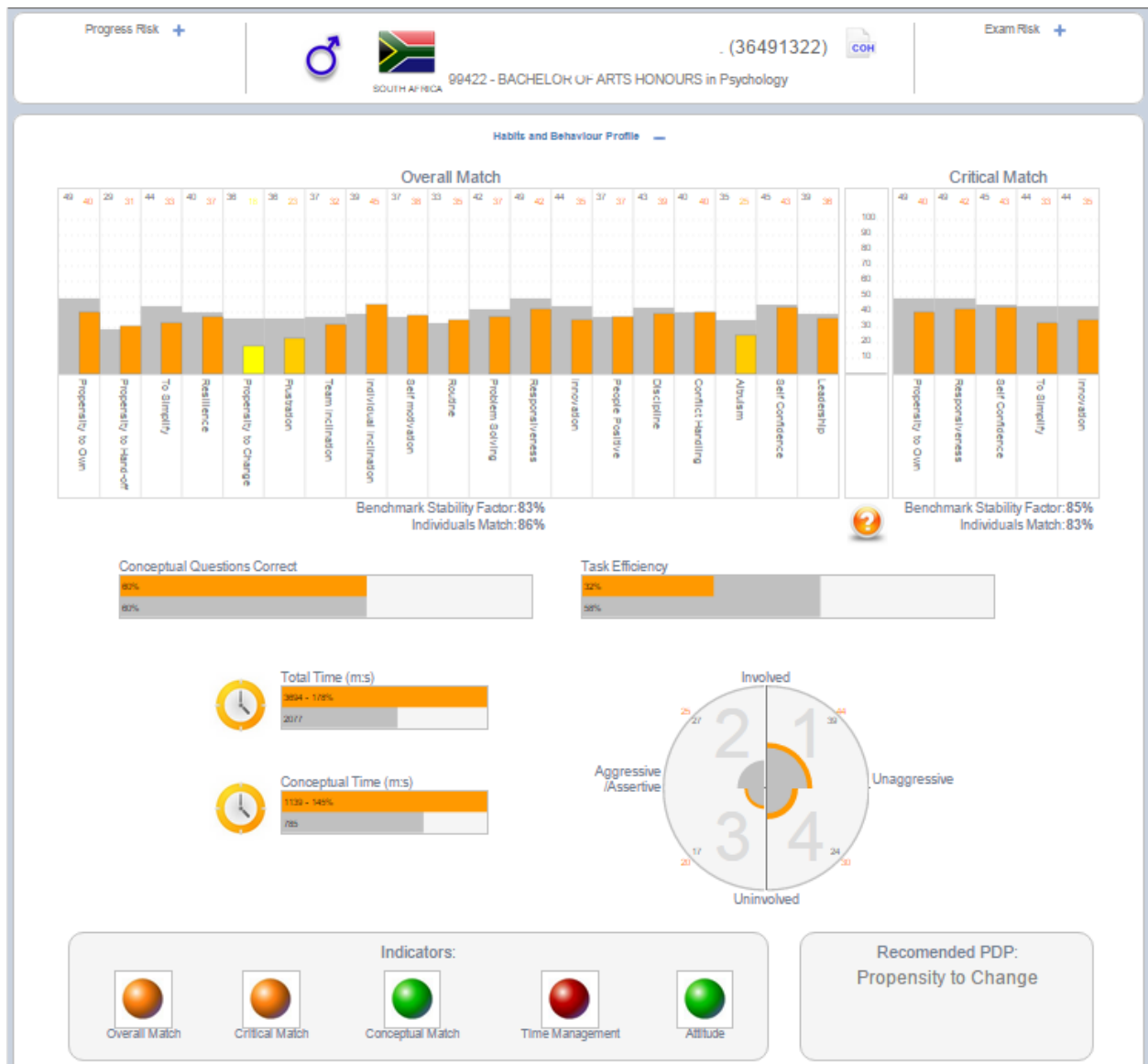
Figure 10: Excerpt Student Node Dashboard expanded for Habits and Behaviour

Suggested itineraries
The data exploration sphere can be quite overwhelming for new users, particularly as most users move from having difficulty to access any data to having a wealth of navigable data at multiple levels of granularity and aggregation at your fingertips. To support users who are making their first forays into exploring the sphere, certain suggested itineraries have been developed for common types of explorations.

Table 2: Example Itineraries

Reason for exploration	Suggested View points
Quality Assurance	Qualification or Module Quality Assurance Metrics or Aggregated Quality Assurance Metrics
Risk analysis	Qualification Risk Rankings; Module Risk Identification; Module 'At Risk' Report; Student Progression Risk
Profiling	Qualification Cohort Profile; Module Cohort Profile; Current and Planned Enrolment Profile; Current Student Profile; Student Habits and Behaviour Profile
Predictive analytics	Qualification Inflow/Outflow Modelling; Qualification Retention and Success Predictions; Module Attrition and Success Predictions
Success Analyses	Qualification Throughput and Success; Module Examination Success

Re-examining dashboard development: putting the horse back in front of the cart

Catalysts

The real power of the data sphere is the ease of navigation, not just within each data level but between the various levels. This is facilitated through a navigation bar consisting of a collection of icons that allows the user to jump from one data layer to another at will to further the data exploration. These bars are located in various places in the design of the dashboard (see Figure 11 below). It is also possible to click on any of the modules in the curriculum window to act as catalysts to explore the module level in depth.
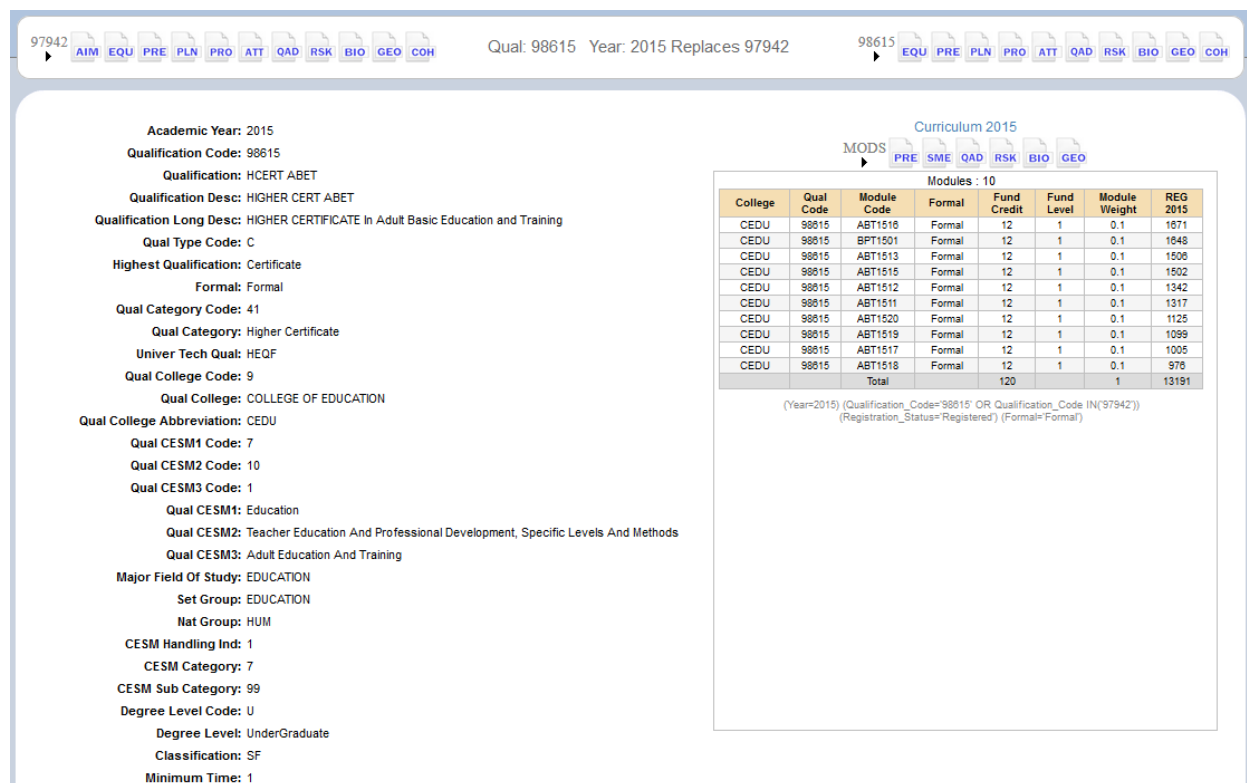
Figure 11: Example Report with various Navigation bars

The definitions of the individual icons on the navigation bar are:

AIM   Qualification or module structure or AIMS (Academic Information Management System) data

EQU   Qualification or module equivalents or 'alternate codes'

PLN   Inflow outflow modelling and planning

PRE   Qualification or module predictions and targets

RSK   Qualification, module or student risk identification and analyses

QAD   Quality assurance data reports

GEO   Geographical profiles

BIO   Biographical profiles

PRO   Progress against targets

COH   Cohort studies and analyses

EXA   Module Exam success model

SME   Student Module Evaluation reports

LAN   Module language targets

Conclusion

The data sphere approach to data exploration presented in this paper represents several shifts in the way that dashboard design and visualisation is usually approached. The primary design concern is data integrity and linkage with a central enriched data based. This means that exploration logic now becomes the focus as opposed to the needs of a particular user or the capabilities and interface of any particular dashboard development software. The focus is also not on the data visualisation and memorability, but rather on having real-time navigable data with various displays for users to choose from. User may thus identify and explore areas of concern of data surrounding key nodes, with the ability to extensively drill through, drill down and aggregate upwards. The shift is thus from canned (pre-packaged) reporting in the form of dashboards to the flexibility of exploration (on demand) of data at will with an export functionality to capture the real time exploration.

The data sphere approach caters for the highest level of user with the highest level of data literacy. It becomes a bank of all the possible data available. Any subset of this data can now be easily drawn into dashboards and score cards for users with lower levels of data literacy and more basic data needs. As all data is located in one warehouse with enrichment taking place prior to extraction, it becomes easier to apply business rules consistently and ensure data quality. This approach to data exploration has the added benefit of shifting from a situation where data is pushed onto the users to a pull approach where users can identify and explore only the data which is of real concern to them.

The next step in the research and development will focus on visualisation in order to make navigation and exploration more intuitive. There will also be a focus on how to help users to identify possible problem areas for further exploration more easily out of the node dashboards.

References

Abd-elfattah, M., Alghamdi, T., & Amer, E. (2014). Dashboard technology based solution to decision making. *International Journal of Computer Science Engineering*, *4*(2), 59–70.

Borkin, M. A., Vo, A. A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A., & Pfister, H. (2013). What makes a data visualisation memorable ? *IEEE Trans Vis Comput Graph*, *19*(12), 2306–2315. doi:10.1109/TVCG.2013.234.

Few, S. (2006). *Information Dashboard Design: The effective Visual communication of data*. O'Reilly.

New Media Consortium. (2013). *Horizon Report : 2013 Higher Education Edition*. Austin.

Newman, K., Thomas, E., & Webber, E. (2014). Dashboards: A window into igher Education risks. *URMIA Journal*, 49–58.

Prinsloo, P., Archer, E., Barnes, G., Chetty, Y., & Zyl, D. Van. (2015). Big (ger) Data as Better Data in Open Distance Learning. *The International Review of Research in Open and Distributed Learning*, *16*(1), 284–306.

Public Domain. (2015). Plasma Globe. Pixabay. Retrieved from https://pixabay.com/en/plasma-globe-electricity-plasma-566202/

Silver, N. (2012). *The signal and the noise: Why so many predictions fail — but some don't.* London, UK: Penguin Books.

Thomas, G. (2011). *How to do your case study. A guide for students and researchers*. London, UK: SAQGE.

Watters, A. (2013, October 13). Student data is the new oil: MOOCs, metaphor, and money. [Web log post]. Retrieved from http://www.hackeducation.com/2013/10/17/student-data-is-the-new-oil/

Wishon, G.D., & Rome, J. (2012, 13 August). Enabling a data-driven university. *EDUCAUSEreview* [online]. Retrieved from http://www.educause.edu/ero/article/enabling-data-driven-university

Yin, R.K. (2009). *Case study research. Design and methods,* 4th edition, Applied Social Research Methods Series, Volume 5. London, UK: Sage.