

Leveraging on Data Visualisation & Analytics for Assessment & Innovation

ABSTRACT

Singapore Management University (SMU) Libraries embarked on its 'Culture of Assessment' journey in 2013, geared towards demonstrating fiscal accountability and driving a higher degree of stakeholder engagement. With more than 80% of library staff trained in the use of Lean Six Sigma for business process improvements, it was only natural that the library started putting in place procedures and systems to continuously assess and improve services. For effective decision making, the challenge lies in the normalization, consolidation and visualization of data from the varied library services into a single coherent platform that could be used for making decisions that are timely.

This paper will share SMU Libraries' experience in implementing a QlikView based dashboard for the visualization of operational data. It will discuss some of the challenges encountered in the following areas:

- *Data Scope – Identifying appropriate data sources and designing metrics*
- *Data Collection – Operationalizing timely information*
- *Data Quality – Ensuring accurate facts and figures*

INTRODUCTION

Singapore Management University (SMU) Libraries embarked on its 'Culture of Assessment' journey in 2013, geared towards demonstrating fiscal accountability and driving a higher degree of stakeholder engagement (Hanken, 2014). With more than 80% of library staff trained in the use of Lean Six Sigma for business process improvements, it was only natural that the library started putting in place procedures and systems to continuously assess and improve services (Singapore Management University, 2015). For effective decision making, the challenge lies in the normalization, consolidation and visualization of data from the varied library services into a single coherent platform that could be used for making decisions that are timely.

Like any academic library, SMU Libraries offers a variety of services to support teaching, learning and research to students and faculty. These services fall broadly under the following categories:

- Teaching & Learning
- Research Support
- Resource Management
- Learning Spaces
- Community Service

Within the Library, various teams take ownership of the services and manage the day-to-day operations as well as periodically assess, recommend and implement improvements to their work processes. For effective assessment to take place, teams have to analyze data from their work processes and systems. This proved challenging as the data required may reside in multiple systems and access to data may not necessarily be granted especially when systems, along with work processes and its service, belongs to another team.

In Oct'15, a project team was setup with representatives from the various Library teams to implement a platform that collates and visualizes data to support of teams in their assessments needs.

THE LIBRARY ANALYTICS PROJECT

With the support of the University's IT department, the project team started gathering project requirements for the Library Analytics Project in Oct'15 and through iterative prototyping, testing and deployment, completed the project within 6 months.



Figure 1: Project Milestones

There were several factors that enabled this agile development:

- The university IT department had a dedicated Business Intelligence & Analytics (BI) team specialising in QlikView.
- There was existing and readily available technical infrastructure and in-house development resources that supported QlikView development.
- There was strong sponsorship from the University Librarian and the CIO that formed the project steering committee.

Below are some screenshots of the outcome of the project:

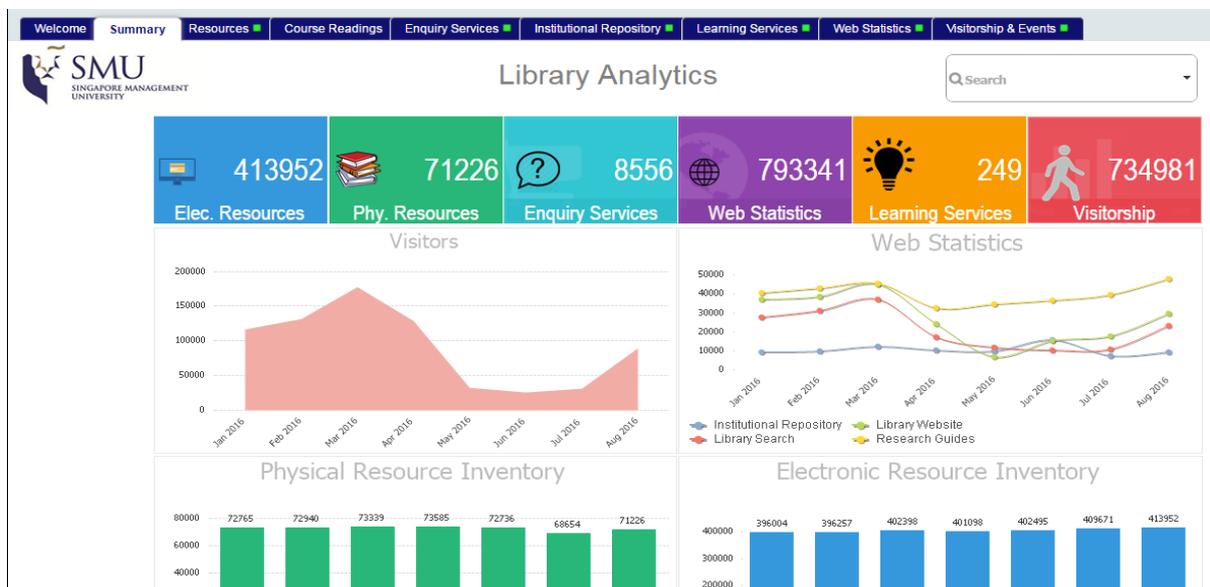


Figure 2: Summary Level Qlikview Dashboard

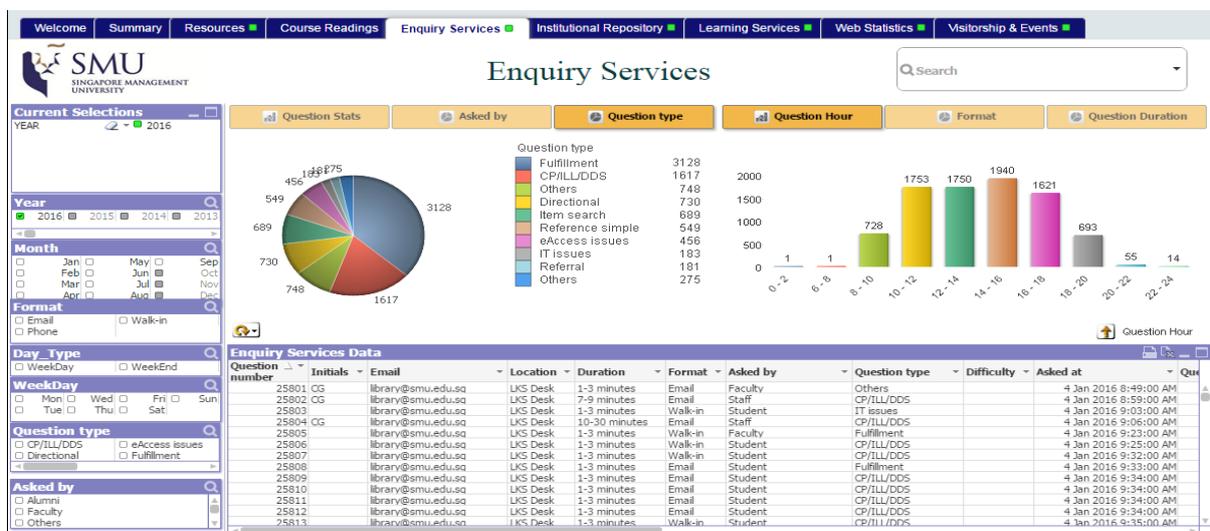


Figure 3: Enquiry Services Dashboard

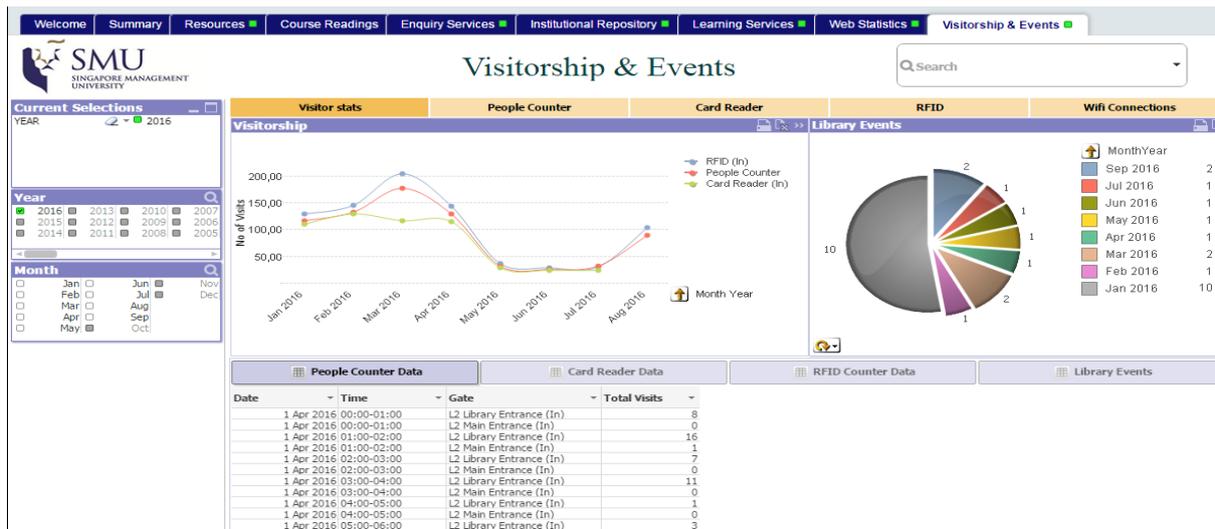


Figure 4: Visitorship Dashboard

The QlikView-based dashboards allow Library staff to visualize the high-level data trends of their service areas. They allow staff to filter or “click” into specific fields; for example, in the year view, clicking a month will filter to that specific month. They allow combination of fields; for example, selection of queries by walk-in and phone calls will automatically filter and displaying the graphs and pie-charts accordingly. Importantly, they allow analysis of datasets across different services; for example, we could find out if there was correlation between physical library visits and virtual online visits to the websites given a time period.

The dashboards would not have been realised without overcoming the challenges in identifying the appropriate data sources, collecting up-to-date datasets and ensuring data quality is not compromised. The rest of the paper will discuss broadly on these challenges and approaches to overcoming them.

DATA SCIENCE PROCESS

SMU Libraries is not new to the process of collecting data, producing reports and using them as narratives to highlight the value the Library bring to the higher education ecosystem. Our ‘Culture of Assessment’ mission underscores the need for an evidence-based approach when assessing our library services and how well they serve our campus community.

The Data Science Process is a framework for approaching data science tasks – collecting, cleaning, exploring, modelling, communicating (O’Neil & Schutt, 2013). The process work flow is a non-linear and iterative. For example, you can collect and process data, and find out during exploratory data analysis that the data collected is not suitable or not properly cleaned. It could be after a few iterations that the data is ready and usable for communication and visualisations.

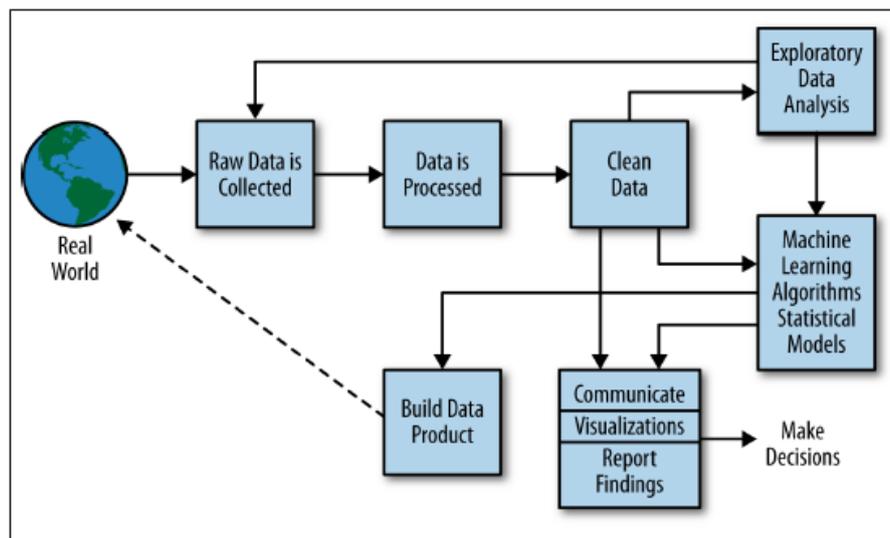


Figure 5: The Data Science Process

We loosely adopted this framework as a guideline for us to select, collect and clean our data.

DATA SCOPE

What data goes into the dashboard? How much? At which level - the summary or individual details? These are the questions that we encountered at the start of the project. The “natural” data scope is to look at all the datasets generated by all the library services. This amount of data can be too much and it lacks focus on intent of the dashboard.

Since the intent of dashboard is for assessment and demonstrating the Library’s value and contribution to the university, we looked at the library services again and how they are aligned to the library strategic goals, which in turn are aligned to the University strategic goals (Cottrell, 2011). We examined the performance metrics

associated with these strategic goals and used that as criteria for selecting the appropriate datasets. We also looked at existing reports that the teams churned out monthly, either manually or system-generated, to understand what and how the data is being used to report performance metrics. For system-generated reports, we explored what other reports the system could generate that would fit into our criteria.

Based on the above criteria and reports, each of the team representatives, in consultation, with their team leaders prioritised and decided which datasets will be included in the data scope. We started to document the selected datasets along with which systems the datasets reside and categorise them according to the five broad library service areas as mentioned in the introduction above. The resultant data scope documentation was crucial as it was reviewed, modified and referenced multiple times throughout the prototyping iterations.

Based on the sample datasets provided, the IT department's BI team, they further enhanced the documentation to include the technical details like file type, field length and field type that would represent the schema of the dashboard. By the end of the project, we had documented more than 200 fields across more than 20 datasets representing the 5 broad library service areas.

DATA COLLECTION

As mentioned above, the Library has always been generating monthly reports for the purpose of assessment. Teams will either export the reports from systems or collate them manually. These reports will usually be in spreadsheet Excel or Comma Separated Values (CSV) file formats. Some of these reports are usually at a granular level and often teams have to manually summarise the data to present the monthly statistics. With the scoped datasets newly identified for the Library Analytics Project, there were work duplication concerns among staff that they had to export or collate other set of data for ingesting into the QlikView platform.

When we examined the sources of the scoped datasets, they fall under three types, 1) open systems that allow report generation, data export and support Application Programming Interfaces (API), 2) closed systems that are only capable of generating

reports with no API support and 3) individual spreadsheet files that are used for manual data entry and collation. There was an opportunity to automate data collection of some datasets files coming from open systems but datasets that resides in closed systems and individual files will still need to be collected manually. Given that automatic data collection requires another set of developmental effort, the project team decided to leverage on the current work process for report generation for the data collection and preparation of dataset files into QlikView even though the process will still be manual.

The project team decided to examine the current work process for report generation and compared it with the intended process for data collection and preparation. The work processes for both was found to be similar. Basically, dataset files had to be extracted out from a source system and placed into an agreed shared location on the network drive. To address the staff's concern on work duplication, the project team recommended that library teams export only the scoped datasets. The administrative overhead of manually summarising data will be taken care by the QlikView platform programmatically. With the new process, staff only have to export the dataset files from the source system and drop them into a shared folder monthly resulting in a simpler work process.

DATA QUALITY

As with manual handling of data, there is a concern of data quality being compromised. In the data scoping phase, we had ensured that we know exactly what data is being collected up to the file and field level. The schema is then used to programmatically apply data validation rules to ensure that the correct and agreed data goes into the dashboard. In the data collect phase, we have simplified the export and upload process to ensure staff have minimal intervention with the source data. The prototyping iterative process provided checks and balances if we did not get the right data in the first round.

We designed the User Acceptance Test (UAT) by sampling real live data into the dashboard and compare them directly with the source system. Often we found that the data collected was not processed properly, there exists 'dirty' data even after

processing the dataset and even data was incorrect in the source system. This iteration of checking, cleaning and reloading of data was important as an exploratory data analysis for the team to understand their data better and built confidence that the dashboard was accurate and precise.

REFLECTION & CONCLUSION

The data science tasks that the project team went through above have demonstrated some of the potential challenges that a library may experience if they take up a similar endeavour. The solutions often are crafted within the operational capacity and capabilities of the Library and there may be other ways tackling the problem. In the end, what is important is that the staff have a better understanding of the data that they produce and have an increased awareness of the data available across the Library

With better understanding and awareness, Library teams can easily access and refer to the QlikView dashboards, assess their work areas and bring desired action to improve their work areas. This in turn may call for more datasets to be included in the platform and spur on more analysis work that lead to more innovation work improvements.

In the future, with the potential increase of datasets being fed into the QlikView platform, the automation of data collection from our open systems as well as housing datasets of closed systems and individual files into open systems, we hope the platform will grow to be an indispensable tool for effective assessment and decision making in the Library.

REFERENCES

Cottrell, J.R. (2011). What are we doing here, anyway? Tying academic library goals to institutional mission. *College & Research Libraries News*, 72(9), 516-520.

<http://crln.acrl.org/content/72/9/516.short>

Hanken, Tamera (2014). Problem Lah! Learning to Tell the Story of Continuous Assessment and Improvement. Proceedings of the IATUL Conferences. Paper 12.

<http://docs.lib.purdue.edu/iatul/2014/plenaries/12>

Hanken, Tamera & Cribb, Gulcin (2016). Student Engagement and Collaboration: Creative Approaches to an Evidence-Based Learning Space and Management Framework. IFLA World Library and Information Congress 2016: Library Building and Equipment Section. http://ink.library.smu.edu.sg/library_research/67

O'Neil, Cathy & Schutt, Rachel (2013). *Doing Data Science: Straight Talk from the Frontline*, O'Reilly Media.

Singapore Management University (2015). Li Ka Shing Library wins the Outstanding Department Award. <http://library.smu.edu.sg/news/2015/12/17/li-ka-shing-library-wins-outstanding-department-award>