Open for learning: using open data tools and techniques to support student learning

Final report 2015

Lead institution: Monash University

Project leaders:
Professor Neil Selwyn
neil.selwyn@monash.edu
(03) 99029076

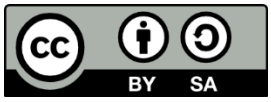
Dr Michael Henderson
michael.henderson@monash.edu
(03) 99052858

Team members:
Dr Shu Hua-Chao
Dr Jonathan Li
Ashan Senevirathne
Thanuja Dharmasiri
Nathan Sherburn

Report authors:
Professor Neil Selwyn
Dr Michael Henderson
Dr Shu Hua-Chao
Dr Jonathan Li
Ashan Senevirathne
Thanuja Dharmasiri
Nathan Sherburn

Project website: http://bit.ly/openforlearning

Support for the production of this report has been provided by the Australian Government Office for Learning and Teaching. The views expressed in this report do not necessarily reflect the views of the Australian Government Office for Learning and Teaching.



Requests and inquiries concerning these rights should be addressed to:
Office for Learning and Teaching
Department of Education and Training

GPO Box 9880,
Location code N255EL10
Sydney NSW 2001

<learningandteaching@education.gov.au>

2015

ISBN	PRINT	978-1-76028-521-0
ISBN	PDF	978-1-76028-522-7
ISBN	DOCX	978-1-76028-523-4

List of acronyms used

ALTC	Australian Learning and Teaching Council
API	Application programming interface
AQDU	Academic Quality & Development Unit
CSV	Comma-separated values (**CSV**) – a CSV file stores tabular data (numbers and text) in plain text.
D3JS, d3.js	A JavaScript library for manipulating documents based on data - **D3** brings data to life using HTML, SVG, and CSS.
HD, D, C, P, PGO	Pass Grades – Higher Distinction, Distinction, Credit, Pass, Pass grade only
HTML	HyperText Markup Language (**HTML)** - standard markup language used to create web pages.
MongoDB	An open-source, leading NoSQL document database.
N, NE, WH	Fail Grades – Fail, Not examinable, Withheld
NoSQL	**Not only SQL** - a database that provides a mechanism for storage and retrieval of data that is modelled in means other than the tabular relations used in relational databases.
NVD3	Re-usable charts for d3.js - without taking away the power that d3.js provides.
PG	Postgraduate
SETU	Student Evaluation of Teaching and Units
UG	Undergraduate

Executive summary

Open for learning: using open data tools and techniques to support student learning

Project context, including the issue it set out to address

Universities generate a mass of digital data relating to students, learning and the environments in which teaching and learning takes place. A wealth of data related to teaching and learning is generated each academic year; usually in disparate, disconnected and restricted forms. These data range from records of student satisfaction, online participation and grades through to room bookings, class sizes and demographics. Thus most data potentially germane to the optimisation of teaching and learning remain inaccessible to many people within university organisations, especially students themselves.

This project was designed as a response to growing calls for better use of data within higher education. In particular it explored the contention that so-called 'open data' principles and practices might provide enhanced opportunities for university students to make informed decisions about their learning. The idea of 'open data' has developed within information and computer sciences over the past twenty years – advocating unrestricted access and use of 'publically acquired' data for as many people as wish to use it. There is now growing interest in areas such as science and government to allow free, unrestricted access to official data-sets. In general terms, the perceived benefits of open data centre on issues of increased efficiency *and* greater empowerment, with the interpretation of data sources by 'end users' seen to enhance the semantic content of data. These principles have clear applicability to higher education. In particular, by combining previously separate sets of already collected and archived data in a university context, students could well be able to 'add value' to the ways in which data can be used.

Aim of the project

Against this background, the present project aimed to explore the benefits *and* challenges associated with the introduction of 'open' access to data related to teaching and learning within universities. In particular, the project focused on open data use amongst students and addressed the following broad challenge:

- To what extent can providing 'open' access to university data allow students to re-configure and re-use data in order to address 'real-world problems' relating to their learning behaviours and educational decisions?

More specifically, the project addressed six distinct questions relating to this broad aim:

1. What data related to teaching and learning exist within universities that could be made openly accessible to students?
2. What forms of open data application can be developed within student communities?
3. How might student communities work with open data applications and practices?
4. What support is necessary from teaching staff and university authorities?
5. What outcomes might be associated with the use of open data amongst students?
6. How might these outcomes be sustained over time with different student cohorts?

Project approach (in brief)

These issues and questions were addressed through a 15 month project (conducted from January 2014 until May 2015). Working in the Monash Faculty of Education with groups of undergraduate and postgraduate students, the project was conducted over four stages:

Stage 1. Audit of potential data sources: A series of 'data audits' were carried out with key administrative and IT staff in order to identify the nature and form of data sources related to teaching and learning. This auditing stage of the project resulted in the identification of data relating to each of the student 'case' cohorts that could potentially be used in an open repository.

Stage 2. Participatory design activities: 'Participatory design' workshops were then conducted with two groups of students. These workshops allowed students to cooperatively design open data services and practices that met self-identified needs relating to 'real world' problems that they felt that they encountered as part of their studies. These workshops resulted in the production of a number of designs for how student groups wanted to access and use the data.

Stage 3. Development of open data applications & practices: The third stage of the investigation saw the project team taking responsibility for 'building' and beta-testing simple versions of the open data applications designed in the workshops.

Stage 4. End-user evaluations: Finally, the applications that were developed successfully were then implemented and evaluated with the student participants during the following academic year, thereby gaining an initial sense of how these open data tools and techniques might 'work' *in situ*.

Key findings

All four stages were completed, open data applications were designed and an application was developed. While the substantive topics that these designs addressed was rather prosaic and bounded in nature, the work nevertheless constituted alternate uses of data to address 'real-world' problems and allow students to consider ways of engaging with data that supported meaningfully oriented behaviour. From this perspective of **proving 'proof of concept'**, the project could be seen as a qualified success. In particular, it is first noting the relative success of the design cycles with both groups of students.

- Once statutory concerns over privacy and duty-of-care were satisfied, university and Faculty managers were able to allow the opening up of selected data-sets;
- There was little overt institutional reticence or resistance to the general notion of open data principles;
- The project saw a range of students collaborating and engaging in determining alternate uses of data that better served their experiences as university students;
- It proved technically possible to design and develop a relatively powerful data-handling application in the short timeframe of the project schedule;
- The open data application and practices that were developed clearly added value and insight to the university's datasets.

However, as is intended in any participatory design project, these activities were as revealing in terms of their 'failures' as any apparent 'successes'. As such, there is much that

our findings highlight the **limitations of the open data philosophy** when applied in the realities of university teaching and learning. In particular:

- The limited 'open data' potential of the datasets that were being produced within the university and faculties;
- Limitations in the data skills and 'data imaginations' of students;
- The limited sustainability of these open data interventions beyond the involvement of the project team;
- The limited 'democratising' effect of the open data design activities and applications.

Project outputs/deliverables/resources
On the basis of these activities, the project has developed a set of concluding guidelines and recommendations highlighting the practical limitations of open data principles and philosophies – i.e. working through institutional concerns over appropriate data use; recognising the likely technical difficulties of such data work; as well as working to develop and sustain student capability and interest.

This deliverable – included in the full end-of-project report - along with details of findings relating to the technical, design and institutional aspects of open data projects have been made available on a project website [http://bit.ly/openforlearning]. Additional dissemination of the project has to date involved the writing and submission of one peer reviewed journal article, one international conference presentation (US) and one institutional presentation to a university-wide 'Educational Excellence' symposium. Now the project cycle has been completed, dissemination of the full findings will continue.

Impact of the project (outcomes to date & projected future impact)
This project was based around conducting a full design cycle – from identification of opportunities and problems, to the design, development and evaluation of an application. As such, we have worked during the lifetime of the project to complete each constituent stage of the cycle, and thereby develop high quality findings that reveal the complexity of any such undertaking. As befits a project of this type, subsequent take-up and impact of these findings will be ongoing now that the full cycle of the project is complete

To date, the project has already begun to impact within Monash University. Based on the project findings, the team has provided expert advice to several university wide strategic initiatives. The implications of these and other discussions has result in the adoption of findings including the need for student participation in designing data applications, and the need for more appropriate collection and collation of data about student experience that is of use to students. A further indicator of impact has been the raising of awareness, and engaging in discussions, at the Faculty and Vice-Provost level regarding risks and privacy concerns relating to student related experience data.

Table of contents

Tables and Figures

Tables

Figures

Project Rationale

Universities generate a mass of data related to teaching and learning. This project addressed the benefits *and* challenges that arise when students are provided with 'open' access to teaching and learning related data that is being collected and collated in their university.

The project worked with students to develop open data tools and techniques relating to their studies on taught undergraduate and postgraduate courses at Monash University. These activities were designed to relate to a set of wider debates concerning the growing need for universities to make better use of data, and the potential of so-called 'open data' practices in extending the use and usefulness of data amongst student populations.

Given the lack of previous work on these topics, this project was exploratory in nature. As such, the most useful and insightful findings relate to current *limitations* of data within university contexts, such as issues of data quality, data awareness and data literacies.

The project provides a useful counterpoint to other OLT work that has been concerned with exploring the promise of learning analytics and 'big data' systems. In contrast, this was an investigation of the realities of digital data and higher education.

Background to the project

The process of higher education generates a mass of digital data relating to students, learning and the environments in which teaching and learning takes place. A wealth of data related to teaching and learning is generated each academic year; usually in disparate, disconnected and restricted forms. These data range from records of student satisfaction, online participation and grades through to room bookings, class sizes and demographics. Thus most data potentially germane to the optimisation of teaching and learning remain inaccessible to many people within university organisations, especially students themselves.

This project was designed as a response to growing calls for better use of data within higher education. In particular it explored the contention that so-called 'open data' principles and practices might provide enhanced opportunities for university students to make informed decisions about their learning. The idea of 'open data' has developed within information and computer sciences over the past twenty years – advocating unrestricted access and use of 'publically acquired' data for as many people as wish to use it (Kitchin 2014). There is now growing interest in areas such as science and government to allow free, unrestricted access to official data-sets. The practical realisation of these principles has been hastened by the development of increasingly powerful and socially-focused 'simple' software applications that allow non-expert users to directly mine, manipulate and interpret data (Longo 2011).

In general terms, the perceived benefits of open data centre on issues of increased efficiency *and* greater empowerment, with the interpretation of data sources by 'end users' seen to enhance the semantic content of data. Some commentators have therefore celebrated the 'open innovation' that can take place through the sharing of data throughout

all stakeholders of an organisation (Boudreau 2010), and a general 'democratisation' as the locus of innovation activities is increasingly moved toward the periphery of organisations (Yoo *et al*. 2012). These principles have clear applicability to higher education. In particular, by combining previously separate sets of already collected and archived data in a university context, students could well be able to 'add value' to the ways in which data can be used.

Of course, the realisation of these potential benefits in universities face a number of likely practical barriers, not least the heterogeneity and dispersion of data sources, the difficulties of providing data in interpretable and usable forms, and the financial, legal and regulatory considerations needed to enable sustained use of data over time (Reichman *et al*. 2011). Open data principles may also run counter to relatively 'closed' institutional cultures (Peled 2011), or face unforeseen 'data divides' such as inequalities between social groups in the meaningful and beneficial 'use' of open data (Gurnstein 2011).

It is also necessary to acknowledge the risks of such open access to data as evinced in recent controversies in the United States of America over allowing third party commercial access to student data. Against this background, the present project was designed to approach the teaching and learning potential of open data in optimistic but circumspect terms. The overarching aim of the project was to move beyond the 'hype' and develop realistic understandings of the opportunities *and* limitations of open data use within student communities *and* the broader context of higher education institutions.

Project challenges

With 'open data' processes and practices now being implemented across many different areas of society, the present project aimed to explore the benefits *and* challenges associated with the introduction of 'open' access to data related to teaching and learning within universities. In particular, the project focused on open data use amongst students and addressed the following broad challenge:

- To what extent can providing 'open' access to university data allow students to re-configure and re-use data in order to address 'real-world problems' relating to their learning behaviours and educational decisions?

More specifically, the project addressed six distinct questions relating to this broad challenge:

1. **What data related to teaching and learning exist within universities that could be made openly accessible to students?** What issues surround the potential 'open' configuration of this data?
2. **What forms of open data access and use can be developed within/for student communities?** What practical improvements need to be made to existing data practices? What new data tools and techniques could be developed to 'fit' the needs of student communities?
3. **How might student communities work with these open data applications and practices?** How might different groups make sense of, evaluate and combine open

data to then generate new learning-related knowledge? What barriers emerge and how might they be overcome?

4. **What support is necessary from teaching staff and university authorities?**
5. **What outcomes might be associated with the use of open data amongst groups of students?** How might student uses of open data support the optimisation of students' learning, the logistics of learning and social/emotional aspects of learning? What limitations and risks (e.g. technical, legal, practical) are evident in the use of open data by students?
6. **How might these outcomes be sustained over time and between different cohorts of students?**

Project approach

These issues and questions were addressed through a fifteen month project (conducted from January 2014 until May 2015). The project was distinguished by its focus on:

(i) assessing the nature of data currently being generated within universities relating to teaching and learning;

(ii) supporting students in 'building-their-own' open data systems and practices;

(iii) evaluating how these open data tools and techniques might be used 'in the wild' of higher education contexts.

The project was conducted over four stages. As a 'seed project', the project focused primarily on one Faculty within Monash University (the Faculty of Education). However, in order to gain a more rounded picture of the university-wide arrangements regarding data use, the first stage of project activities compared the findings from the Faculty of Education with the Faculty of Engineering. Subsequent work with students then focused on two distinct groups within the Faculty of Education:

- Undergraduate students enrolled in Bachelors of Education courses [studying fulltime, on-campus]
- Postgraduate coursework students enrolled in Master of Education courses [studying full and part-time, on-campus]

Working in these Faculties and with these groups of students, the project was conducted over four stages:

Stage 1. Audit of potential data sources

First a series of 'data audits' were carried out in both Faculties in conjunction with key administrative and IT staff in order to identify the nature and form of data sources related to teaching and learning. In-depth individual interviews were conducted with Faculty managers, data integrity officers, e-learning coordinators and other teaching and learning administrators. Individual interviews were also conducted with the senior information

officer and a student data 'lead' from the central office of teaching and learning. Eleven in-depth interviews were conducted in total. Efforts were made to identify archived data from previous academic years and student intakes, as well as data that had been informally collected by teaching staff and through learning management systems. This auditing stage of the project resulted in the identification of data relating to each of the student 'case' cohorts that could potentially be used in an open repository.

Stage 2. Participatory design activities

'Participatory design' workshops were then conducted with one 'working party' of volunteer students from each of the 'case' cohorts (a total of two groups). Each group of four students engaged in a series of workshops. Following established participatory design methodologies (see Sanoff, 2007), these workshops allowed students to cooperatively design open data services and practices that met self-identified needs relating to 'real world' problems that they felt that they encountered as part of their studies.

Each workshop lasted between one to two hours. The workshops were conducted through a participatory design approach drawing on Argote's (2013) framework of sharing, evaluating and combining open data to generate new organisational knowledge. The workshops therefore saw the project team and participants working together through a series of iterative activities:

a) gaining familiarity with the available data-sets;

b) discussing and identifying 'real world' problems that the available data-sets might relate to; and

c) cooperatively designing paper prototypes for 'simple' tools and techniques using these data-sets that might address the identified problems.

These workshops therefore progressed from 'initial exploration and co-interpretation' sessions, through to 'discovery processes' and then 'cooperative prototyping' sessions (Spinuzzi, 2005). The project team systematically collected data from the workshops in the form of observations, recorded discussions and analyses of the artefacts from the design processes (e.g., sketches, plans, storyboards). These workshops resulted in the production of a number of designs for how student groups wanted to access and use the data. The project team acted as scribes for the drawing-up of the final designs.

Stage 3. Development of open data applications & practices

The third stage of the investigation saw the project team taking responsibility for 'building' and beta-testing simple versions of the open data applications designed in the workshops. In order to do this the project team included a consultant and three application developers from the Faculty of Engineering. A representative of the development team participated in the Stage 2 workshops. During Stage 3 the developers and the rest of the project team regularly met over a five month period to work through several iterations of the software development.

To facilitate sustainability, ownership and maintenance, the project attempted to avoid complex application development in favour of the development of software tools that were already licensed, free or open source and cloud/network based hosted. However, as will be discussed, the nature of the data and the prototype design necessitated considerable amounts of application development. This could be seen as not fully realising the principle of any open data tool being easily reconfigurable by users. This stage was also compromised by a substantial amount of development time needing to be spent on the 'cleaning' and synthesising of different data sets to make them operationalisable.

Stage 4. End-user evaluations

Finally, the applications that were developed successfully were then implemented and evaluated with the student participants during the following academic year, thereby gaining an initial sense of how these open data tools and techniques might 'work' *in situ*. Interviews were conducted with both groups of students and university staff to develop a sense of how open data applications of this kind might fare 'in the wild' if they were to be used on a sustained basis across the student population.

Project significance

This was a small-scale, exploratory seed project. As such its scope was limited, nevertheless, the project could be considered to be of genuine significance in a number of ways. In particular, the project has clear parallels with the growing interest in 'big data' and 'learning analytics' within higher education. For example, these issues have been addressed separately in previous OLT/ALTC funded work as well as internationally (for example see Siemens *et al.* 2011). Yet the present project is distinct from this previous work in five significant ways:

- First, the issue of open data has been largely ignored in terms of university teaching and student learning. Conceptually *and* practically, then, the project constituted one of the first empirical explorations of these issues in the world;
- Second, this project was deliberately student-centred and student-driven. Whereas conventional forms of 'learning analytics' and data-mining are used primarily by teachers, administrators and other university authorities on an institutional basis, the present project supported the development of student directed and student focused forms of data use - allowing students to determine their own uses of data;
- Third, this project positioned students as active participants in the re-configuration of the data as opposed to passive recipients of institutionally produced analytics. Even when learning analytics is opened to students, it is usually limited to institutionally defined forms of aggregation, filtering and measures. This project proposed that students also need to be provided with opportunities to bring together personally meaningful sets of data, and apply their own configurations;
- Fourth, the present project was concerned deliberately with the re-use of existing data rather than the collection of new data and/or the monitoring of current learning activity. This project focused on the re-use and re-combination of already

collected and archived data within different parts of the university organisation. The project supported students to combine, aggregate and reconfigure already existing data sets in ways that were meaningful to them;

- Fifth, the present project, as part of exploring the applicability of 'open data' principles, included the goal of developing 'really simple' software techniques and tools in contrast with complex 'analytics engines', 'datamarts' and other data handling applications. The goal was to explore the use of simple applications and tools that students were already using and are familiar with. The focus of this project was on developing and supporting students' effective uses of data, rather than the technical development of software.

Project findings

Insights from the data audit (Stage 1)

The project team conducted a 'data audit' in conjunction with key teaching, administrative and IT staff within the Faculty of Education, Faculty of Engineering and the university's central services. This process was designed to identify the nature and form of internal and external data sources relating to student cohorts in use in each Faculty.

As might be expected, digital data was a prominent feature of both faculties' organisation of teaching and learning. This most often took the form of data relating to the entity of 'the student'. Data was therefore collected and collated throughout "the full life cycle of a student" [Staff interviewee 1]. This involved data relating to initial enquiries from prospective students, formal applications to study, the extensive 'on-boarding' process of enrolling new students, alongside the three (or more) years of progression through courses, eventual graduation and then progression to alumni status.

In this sense, the life cycle of every student at Monash was being captured and represented in a continual process of data collection involving an array of computerised systems. For example, all newly enrolled students were required to enter personal and demographic data into the 'Web Enrolment System': "when a student enrols … they fill out everything that we need" [2]. Much of this data was captured already by an 'Automated Admissions Portal' that collected students' personal details, educational background and course preferences. 'Callista' stores all these data, acting as the university-wide central repository for student history and information relating to admissions, enrolments, completions, assessment and so on. This data was most often accessed via the 'Monash Reporting System' which was described as the '"first port of call" [3] when it came to dealing with student affairs.

Alongside this 'student data', another array of systems was concerned with the administration and management of 'teaching'. This principally involved 'Allocate+' – the university's timetabling system that also recorded students' unit timetable preferences. In addition, any alteration of timetabling information such as venue or student timetable change requests are processed via the 'Syllabus Plus' system. Data on the content of taught units took a number of guises. This commenced with academic staff completing a unit proposal/amendment form, which once approved by the faculty, is entered into 'Cupid';

restricted access administration system. The unit information is then pulled into the annual University course and unit 'Handbook', which allows students to: a) search for information about prospective courses using a public 'Coursefinder' system; and b) search for available unit offerings for their particular course. Select unit data stored in Cupid is also drawn into the Unit Guide Editor - the system which staff use to enter the required Unit Guide specifications and other information. Course statistics were collected by the university's 'Business Intelligence System'. Other systems were used to collect data on teaching staff performance (TSP) and the amount of staff time dedicated to each course – so-called 'load data' (collated on the 'Kronos' system).

These data collection systems were complemented by an on-going cycle of data generation. Much of this was coordinated by the university's 'Academic Quality & Development Unit' that was responsible for surveying students throughout their time at the university. As the AQDU's mission statement put it, this data was intended to "aid in the quality cycle". In practice, this involved surveys being administered regularly to gauge student satisfaction with university provision. These included the Course Experience Questionnaire; the University Experience Survey; the Student Administrative Services Survey; and Graduate Destination Survey. The most frequent of these data generation exercises was 'Student Evaluation of Teaching and Units' (SETU). Here students were asked to evaluate their subjects through a series of rating scores and open-ended comments. This data generation exercise took place in the closing weeks of each teaching period. With Monash hosting nearly 60,000 students all taking a variety of taught courses, SETU represented a substantial undertaking.

The other main source of data generation is derived from Moodle, the university's learning management system. Like most learning management systems, this was configured to collect data relating to 'student profiles' (usually name, contact details, mode of study), user activity, and unit participation. Data were also generated through quizzes, tests, discussion forums, and the submission of assignments. Various "highly localised" [4] additions had been made to different Faculties' configurations of Monash's learning management system, such as the inclusion of the 'TurnItIn' text-matching system to analyse student work for potential plagiarism and malpractice. Taken as a whole, these various systems represented a considerable, but not wholly integrated, volume of data. As a university staff member managing student data acknowledged: "there's a thousand other systems around the edges that have little pockets of data as well" [5].

Clearly the data within Monash (as with any university) is complex, often stored in independent systems rather than interdependent ones, with varying degrees of access and control, and often with no-one person aware of the whole. Therefore working with this data is a challenge in any circumstance, let alone with the intent of being repurposed for an open data project.

Challenges to the access and use of potential open data sources

The second issue addressed in the data audits was the potential for opening up access and use of these data-sets. Of course, as with any aspect of administration and management, the actual implementation of these data procedures and processes was varied and often

restricted. In this sense a number of institutional issues and concerns surrounding the use of this data found in this project are reported in Table 1.

Table 1. *Challenges to open data in Higher Education*

Challenge	Description and quotes
Access/ Compatibility	• Restricted permissions – in particular, data being the ultimate preserve of the university: "most of the data is controlled and owned by central … Central modifies, tweaks and accesses it" [8]. • The "double-handling" of data between different systems that did not "talk" to each other [3]. • The need to manually "export out" data from systems and "modify into the format that meets the requirements of other systems" [7].
Risk	• Concerns over "doing the wrong thing with data" [2]. • The "risk averse" [2] nature of the university data systems. • "High risk for the university … I guess legalities, and not wanting to be sued and all that sort of stuff" [3]. • "The data is high risk and so people are ultra-cautious" [4].
Privacy concerns	• "We are bound by privacy" [1] • "Privacy reasons" [7] • "Privacy legislation" [9] • "Any data that might identify a particular staff member" [11] • "We are not allowed to release any sufficient information that you could identify the individuals from it…if it's a small set, you can't put the stuff there that would lead to saying that's that student" [9]
External reputation	• "Some of it [data] can be really sensitive. [For example,] if you've got courses where the enrolment numbers are very small. You don't want that information getting out into the public … some of it is to do with the marketing [but] releasing raw data, whilst there might be no identifiers in it, it can be misinterpreted … it can be misinterpreted in the marketplace" [11].
Data validity	• Data sets "are not matching. So they come to…different conclusions but using the same set of information" [2]. • "You're working with multiple systems, you're working with multiple people, and you're working in entering data in multiple different places, the opportunity for error is vast" [3].
Lack of staff interest/ incentive	• "If somebody identifies a data set that they wish to say 'can we put this out to the public student body?', if somebody is willing to do that piece of work, the rest of it can happen. If there's a challenge, it's around priorities and all of the other things that are going on and, to be frank, the investment payback" [9].

From the wide range of data collected within the university only four specific datasets relating to teaching and learning were identified by Faculty managers and central university staff as realistically suitable for open access and use (see Table 2). Much of the data that was *not* considered suitable was reckoned to be closed in terms of accessibility, ownership and/or broad concerns over 'privacy'. Other forms of data were reckoned to be unavailable in a raw form.

The suitable datasets related to: student evaluation data; data derived from the Moodle learning management system; data about the unit including assessment regime; and some aspects of the Callista data system (e.g. relating to course choices, grades and progression rates). While hardly constituting a complete picture of teaching and learning activities within the Faculties, these data sets did appear in theory to fulfil basic criteria for open data. For instance, all were machine readable using non-proprietary formats. Most of the data was also consistently structured, and 'cleanly' maintained to allow automatic processing. All the data sets were updated on a regular basis – ranging from an ongoing basis to once a year.

Table 2. *Identified potential open data sources*

Systems	**Examples of data potentially available for open data**
Callista (student database)	• Student demographic details • Subject/Unit grade distribution • Enrolment details, progression and course choice
Moodle (learning management system)	• Online participation • Events • Assessment details, deadlines, submission rates, grades and feedback
SETU (student evaluation data)	• Aggregated Likert scale ratings of student satisfaction • Open ended comments relating to the highlights and what needs improving
Unit Guide Editor (details of the semester schedule, assessment, lecturers, etc.)	• Unit code, title, description, lecturers, assessment details including deadlines and marking criteria, weekly schedule of topics and other activities.

That said, the 'openness' of these four set of data was clearly compromised. These data sets were focused on discrete aspects of teaching and learning, with different levels of measurement and varying forms of labelling. For example, while one dataset reported on individual students with identifiable ID numbers, another dataset reported on a subject cohort, using Subject IDs. Neither set could be linked without a third dataset that revealed student enrolment, and thereby allowing the linking of student ID and course ID. The interoperability of these four datasets was therefore compromised by a lack of clear data schema. This was largely due to the lack of direct database access. This meant that the project team could not get automated 'real time' updates, requiring these datasets to be downloaded manually by Faculty and University administrators rather than being automatically harvested.

The nature of the exports meant that only a proportion of the total fields and the database schema (information about how the data should be organised) was available to us. In addition, while most of the open-ended text data was supplied in complete primary form (i.e. as inputted at source by administrators or students), some elements of the closed numerical data had been modified into aggregate and composite scores. While not compromising the project team's ability to develop open uses of these four data sets, such issues certainly restricted the potential for longer-term practical sustainability of any developed application (see Open Government Data, 2015).

Insights from the participatory design activities (Stage 2)

The discovery stage

During the first sessions of the participatory design workshop cycle the students and project team worked together to identify real-world problems associated with the four university data sets identified as suitable for open data re-use. While the data sets (as just stated) did not constitute 'ideal' examples of open data these participatory design activities still offered valuable insights into the potential re-use of data within the university.

Having been familiarised with the open data sets, the two groups of students discussed and developed a number of 'real world problems' that the data might address. As might be expected, these issues related closely to the respective 'jobs' of being an undergraduate and taught post-graduate student.

Thus undergraduate student suggestions for potential uses of the data focused largely on what was referred to as "survival" [UG participant] as a student. This related predominantly to obtaining satisfactory grades for different units, choosing "easy and fair" units that would not jeopardise grades; gaining a sense of course expectations prior to enrolments and managing workload and deadlines. Much of this information was currently being gathered by students through "word of mouth", sometimes via unofficial online forums such as Facebook groups like 'Monash Stalker Space'. Interestingly, a couple of students were already "documenting [their] own progress" and using tools like Excel to perform simple calculations about grade averages - "we've all got Excel sheets open and trying to see what we have to do not to fail" [UG].

Similarly, the post-graduate group highlighted potential uses of data that were orientated practically towards issues of gaining a better sense of the Monash context and an accurate sense of their own progress. As with the undergraduate participants, the desire to access data relating to performance in terms of assessment grades was stressed throughout the preliminary workshops. As one student reasoned - "doing the assignments is the major thing – assignments are graded … everything else is not" [PG]. Alongside concerns over assessment was the issue of gaining a better sense of context. As students not familiar with Australian higher education these issues ranged from having a better sense of who their teachers and classmates were (e.g. in terms of subject backgrounds, interests); the meaning of the various grade levels used at Monash; and a better overall sense of university expectations:

> "Especially for international students - we were too new to get an idea of what we should be doing. Are we doing the right thing? Everyone was really trying hard, but wasn't sure." [PG]
>
> "An idea about what is a good balance... are we on the right track?" [PG]
>
> "I find it difficult to know what the average grades should be is in the West? I want to know the average, and I am below the average or above the average? … for me the average is OK." [PG]

It was notable how all these areas of potential 'real world' usefulness related to individually-relevant issues of choice making, 'grades' and progression through the university system. Thus the potential of open data for both groups strongly resolved around being more informed about how to succeed in assessment tasks. This orientation towards performance, in contrast with 'learning', was highlighted when both groups of students dismissed any usefulness of 'learning analytics' data that might be available through the learning management system.

Both workshop groups therefore made sense of the four data sets in terms of three main potential benefits. First were expectations of **clarity** – i.e. using the data to provide what was described as "a bit more information" [PG] and "making things clear" [UG]. For example, as these undergraduates described the information that they would most value being given access to when making choices about which units to take:

> "Failure rates, what kind of assessment task, is it group work? … nobody likes group work … does it involve oral presentations?" [UG]
>
> "Yes … at the moment you don't get to see the full assessment task – I'd like to know what they actually are. "60% coursework" could be three short essays or sometimes just one big one!" [UG]

Secondly, both groups raised the idea of the data sets having the potential to provide a better sense of **context** – be it in relative terms of institutional expectations, and their own performance in relation to the performance of others. As was noted with regards to unit-wide grades:

> "the scores are always all over the place … it feels really arbitrary – it would be good to know how other people are doing and how they feel about it" [UG]

Thirdly, participants in both workshops conveyed interest in the potential of data as a source of **insight** – i.e. alerting students to issues that they might not otherwise have seen. Students talked of data being used to provide as a "motivation" that might then "help you know how to go about things" [UG]. For example:

> "[If] you can see your grades *but* also where you're actually ranked … if you found out that out of 90 on your course you are ranked 90 then you might pull your socks up a bit" [UG]

Similarly, in terms of planning and writing assignments, another student talked of the potential benefits of being able to form decisions on the basis of more than their own individual experiences:

> "I don't care if someone last year got a 98% - it doesn't effect me, isn't going to affect how it go … but if I can see [feedback from tutor data] what they did to get a 98% then *that* would be really helpful" [UG]

Clearly, then, students' approaches to the existing university data were shaped by their different circumstances and objectives. All the undergraduate students identified the need to be better informed about their course progression. In the current climate of broad undergraduate degrees that offer students the ability to customise their courses with

optional subjects it was reported that choice-making can quickly become confusing. This is particularly in the case of degrees leading to professional qualifications that may require grade point averages, and a proportion of specific types of subjects. The undergraduate students therefore wanted to know at any time, what options were ahead of them, and what they needed to achieve in order to gain their preferred degree outcome. In contrast, the postgraduate students did not identify course progression as an issue, possibly due to the more focused postgraduate options available to them. However, the postgraduate students identified a desire to find out more about the lecturers and course expectations and norms.

The design stage

Given these broad issues and interests, it was notable how both workshops went on to develop similar sets of designs for data tools and techniques. These related broadly to tools that might distil the data sets for human judgement and/or make use of the data to analyse trends and patterns. Both groups' designs therefore related to tools of **description** and **explanation** of the data, rather than for more complex processes of automated prediction or prescription. While all the workshop sessions were encouraged to be expansive and ambitious in their thinking, there was little/no interest expressed in adding specific analytics, data modelling or the 'combinatorial innovation' (Yoo *et al*. 2012) associated with data 'mash ups' or similarly advanced forms of data work.

As such the workgroups resulted in the development of five distinct forms of relatively basic data application:

Firstly, some participants expressed an interest in designs based around what was termed a '**data dump**'. These were designs to simply present all the data available on an individual, unit or course cohort. The stated aim here was to move beyond the currently 'partial' and 'limited' access to data that individual students felt that they were permitted to have. This 'Wikileaks' approach was described as allowing individuals "to see everything" [UG]. As another post-graduate student put it: "I want to see everything that I can possibly consider on the screen … everything" [PG].

Secondly, for other participants this desire to 'see more' involved designs based around the reduction of data – what was termed a **data 'synopsis'** or '**overview**'. Preferences were expressed for 'quick' orientations towards the data sets – to "get a summary of reoccurring issues" [UG] – for example, along the lines of "a dot point breakdown of 'these are the main things that people said'" [UG]. Here there was a tendency for workshop participants to replicate existing tools from other data applications that they used. Thus designs were developed that included data 'filters' and other familiar forms of data representation. Both groups explored the concept of a "dashboard". Also popular was the simple graphic representation of data – such as bar charts, box-plots and word cloud generators. Particularly within the undergraduate participants, brevity of presentation was considered a desired feature of these data tools.

Thirdly was the design of data tools that offered a sense of '**comparison**'. Here workshops explored ways of presenting correlations between different data fields (for example, grades

and type of assignment task). Other recurring design features included tools to identify 'trends', 'patterns' and changes over time on a whole class basis. As one student suggested:

> "Grade trends would be good – comparing my grades with how many people failed this last year. If I know the average was 92% then I might relax a bit. It's about knowing the context" [UG]

Notably, no designs involved making comparisons beyond the lifecycle of currently enrolled students and cohorts. Thus any interest in comparative design was limited to comparisons either with current classmates or – at best - with the previous year's cohort.

Fourthly, was the design of data tools that offered a sense of '**individual tracking**'. Here workshops explored ways of presenting data relating to the monitoring of individual performance, and expected outcomes:

> "It would be great to see your score as you go along. First assignment is 10%, you got 100% you're sitting on 10. Second assignment is 25% and you get 77% … the [application] will do the maths for you and you are currently sitting on 40 … then you go into the 60% exam" [UG]

> "If I can see if I am on the right track in this semester to be on the right track to carry onto a PhD … if it could say, the average requirement for someone going onto a PhD at this stage would be X, but at the moment your are Y. So now you need to get this to raise your grade average" [PG]

Finally was the more ambitious design of data-based '**notifications**'. For the undergraduate student workgroup, this centred on using data to be informed of expected work, deadlines and other impending events – e.g. "some sort of 'to do' list – a reminder when to do things. That data is all being uploaded to Moodle, so why couldn't it be put together?" [UG]. One recurring feature of these tools was the timely provision of information – what was referred to as the analysis of data taking place in a 'live' form. The postgraduate designs focused on other types of 'push notification' [PG] relating to the activities of peers – for example, being notified when particular readings were being accessed heavily on Moodle, or when coursework was being submitted by other students:

> "Knowing that everyone else had submitted their assignments two weeks early" [PG]

> "Knowing how many people had read something … Like on YouTube … if you find a lot of people have downloaded it then there's a chance it might be interesting. It would be motivating. I really like to know that other people have done things" [PG]

Insights into student understanding of data

As befits the participatory design methodology, these workshops sessions were of additional value in allowing for the further exploration of the relationships that students had with university data. These participatory activities and the discussions behind the development of the prototype designs therefore raised a number of broader salient issues:

First, were clear **differences in the temporal expectations** of how data might be tracked and compared. The conceptualisation of 'tracking' used in the undergraduate group was firmly on a 'live' incident-by-incident basis. Elsewhere, comparisons with the immediately preceding cohort were deemed as appropriate, although more 'historical' data was not valued:

> "You don't want to go *too* many years back … if someone said in 2009 this unit is terrible then its sort of counterproductive. Things change … lecturers can make an effort to change things … maybe just last year's data would be useful. Otherwise you can get the wrong idea." [UG]

Second, were varied preferences regarding the **representation of data**. A split emerged between students wanting traditional graphing of data and/or numeric representation, as opposed to more familiar online visualisation features – such as star ratings and 'thumbs up/down' from popular applications such as Amazon and YouTube. As this workshop discussion illustrates, students' preferences for data representation varied:

> [A]: "Coming from a non-sciencey background I wouldn't make an effort to understand what that [box-plot] means – you're cutting out a bunch of students who don't know what box plots means. Why is that bit skinny, why is that line small?"
> [B]: "Star ratings are good"
> [A]: "… But you'd need star ratings of different bits – how interesting is it, how easy is it. Just one single star rating doesn't tell you what it's rating"
> [C]: "Goodreads gives a number rather than a star … 4.9 I feel is different than a 5 … So lets have a star rating but also a number. Its important to have an image of it (as stars), but the average gives it more meaning" [UG]

Third, was the **influence of commercial and popular applications** in shaping students understandings and expectations of data. Alongside Amazon, YouTube and Goodreads, students' designs also referenced applications such as EBay, Urban Spoon, iTunes and Nike Run. Thus it was notable how the students' designs and configurations often mirrored the lay-out, aesthetics and grammar of familiar applications:

> "Comparing side-by-side, like the Apple website does with your products" [UG]
> "Can you chose the order the data in a way that I want … like in online shopping … filtering and ordering what I actually want to see?" [UG]

Fourth, was the interest in data designs that related to students' **immediate contexts and individualised circumstances** (e.g. data relating to their specific course, or only to peers taking the classes face-to-face rather than online), rather than any wider comparisons (e.g. across whole faculties or institution-wide). As one participant put it: "other people's grades don't matter to me – that is their grades not my grades. I want to know what I have to do to get a better grade" [PG]. As such, the prototype designs tended to include features that allowed users to focus quickly on immediately relevant aspects of the data.

Finally, it was notable through the development of the designs that students were keen not to make uses of **data that they considered invalid or unreliable**. Often this related to data that students had experience of in terms of its generation - such as the student evaluation

data. This was evident, for example, in their concerns regarding the 'unfairness' of negative open-ended comments data supplied by some students:

> "People's anecdotes are completely uninteresting and unimportant. You're only hearing one side of the story. I wouldn't even bother reading that. People with a grudge against the lecturer who didn't give them a good mark" [UG]
> "I'm not judging the teacher based on what other people say about them" [UG]

Similar concerns were raised about the validity of Moodle usage data – reflecting respondents' awareness how some students engaged with the system:

> "People go on Moodle and go 'click frenzy' and clicking randomly so it looks like they reading things and doing the work. You can't tell me that's validated. If you could tell me that people have *actually* read stuff then I might pay attention" [UG]

Paper proto-types

After identifying these sets of real world problems, and exploring their applicability to the available data sets, the undergraduate and postgraduate groups then engaged in a series of exercises in which they described, sketched, and 'stepped-through' how specific data could be used to meet their needs. This was facilitated by the project team, which also included one of the application developers.

Figure 1. Postgraduate design workshop

It is noteworthy that both groups of students eventually focussed on the use of unit climate data (e.g. previous cohort grades and student satisfaction reports) and unit details (e.g., details of teaching staff, assessment criteria, and so on). Despite encouragement to imagine more expansive possibilities in using this data, it became increasingly clear to the students and the project team that the data available was both limited in granularity and scope. This, in effect, constrained the re-purposing of the data. For instance, the undergraduate students' desire to know if they were performing successfully in relation to other students was compromised by the fact that the data available to this project could not give a necessary 'live' and granular detail. Therefore, the eventual designs from both groups had a

number of striking similarities. This was due to the combined effect of similar core ‘real world problems’ and the limited scope of the available data.

While constrained in this manner, both workshop groups produced a number of design sketches that were broadly felt to address some aspect of their identified ‘real world problems’ and preferred forms of data use.

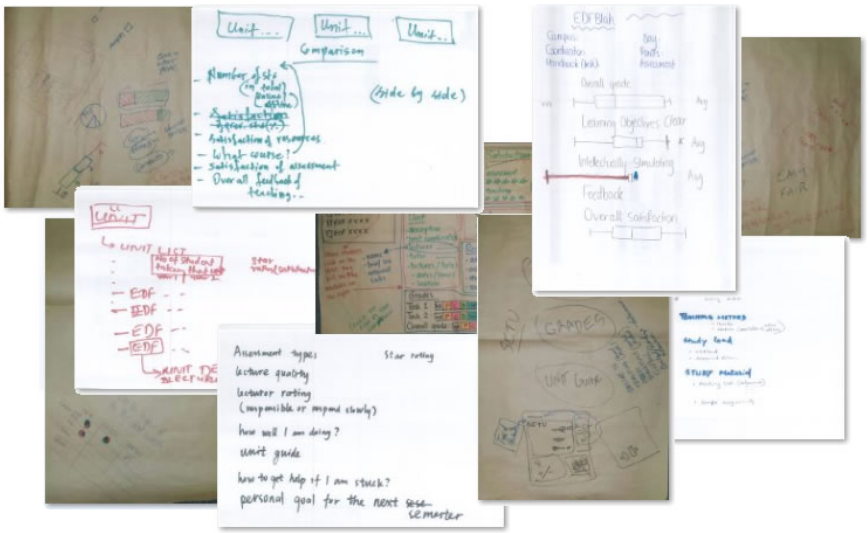


Figure 2. Sample paper designs of prototype components

Due to the similarities across the designs, it was decided that the development team would create a single application that addressed the needs of both designs (rather than dividing resources to create two separate applications). The final paper prototype that combined the design components from the two workshop groups can be seen in Figure 3.

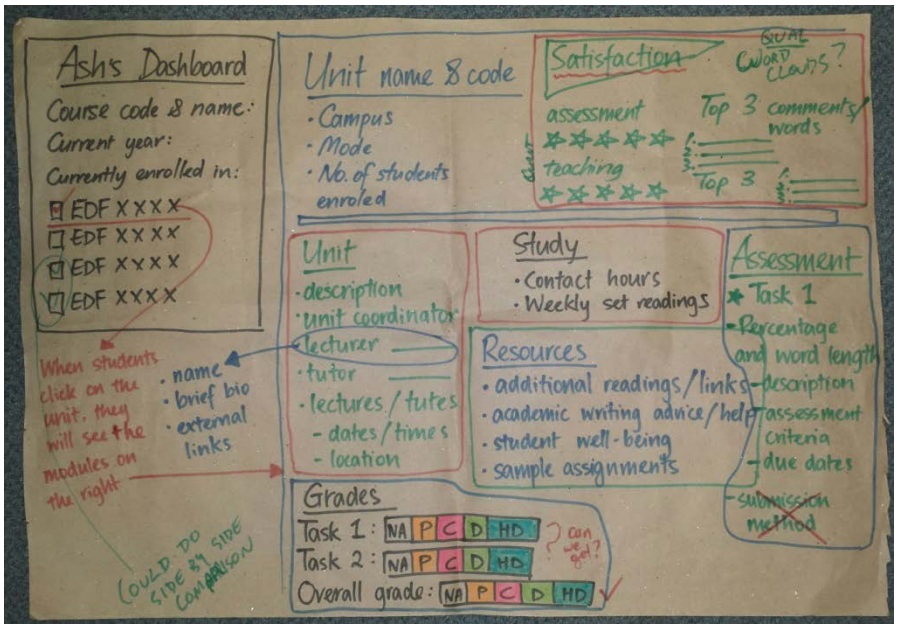


Figure 3. Prototype sketch combining both undergraduate and postgraduate designs

While the paper prototype was successful in articulating ideas between the students and the project team, it was not sufficiently detailed to act as a design brief for the application development team. Therefore, the project team engaged in a process of backward mapping, exploring workshop recordings, sketches, and observational notes to develop an elaborated design brief which contained three application proposals. Table 3 presents an extract of the final design brief provided to the development team. This describes the student 'real world problem' and students' anticipated uses of the data (a complete sample of the design brief is included in Appendix B which contains details of students' preferred form of data interrogation, visualisation and 'blue sky' scenarios).

Table 3. *Student design brief for application development*

Short hand descriptor	**Description**	**The kinds of queries they want to perform / The kinds of questions they want answered**
Trip advisor for subjects	**Students would like to know more about the units before they enrol.** At the moment they only get the unit handbook which does not provide all details about the teachers, the assessment, etc. They want to know about the teachers, the grades, what students say about it, the assessment, the schedule, etc. They want to know about their options. These data will inform their potential unit choices, but also, in cases where they have limited choices these data will orientate them to the 'nature' of the unit (i.e. know what they are letting themselves in for).	1. Who is teaching 2. What is the assessment (type/task, criteria) 3. What are the grades (average, frequency of grade bands, breakdown according to course enrolment) 4. How popular is this course? 5. How is it rated by other students?
Know thy neighbour	**Students would like to understand their position in a course and how they relate (and compare) to others in their course.** They want to know if it is a big/small unit, who the other students are (e.g. are they in the same course as themselves), more about the lecturer, etc. The students felt this was "interesting" and that might give them a faster sense of connection with the class that they don't have – especially in the current online/blended modes. One way to do this (as suggested by students) is to increase communication with others who are doing the same unit/same course.	1. Who else is doing my course? [Not names… but rather a sense of the kind of others … such as numbers of students, grades, current and past courses/units completed, other demographics] 2. Who is my lecturer / tutor? 3. How do I keep in contact / network with those who are studying my course? 4. How do I compare with peers? 5. How do I track my progress in my unit(s) / course? 6. Where can I go next?
Being smart about assessment	**Students want to succeed in their assessment/grades.** They see this directly linked with their assignment performance. Therefore they want to be better prepared for what they need to do for their assignments. **They would like to orient**	1. What is the most important thing I need to do to succeed? [where success is defined by my own criteria, e.g. C, D, etc.] 2. What kind of assessments are in a unit and what are their criteria for

	themselves for the assessment (i.e. know what they are signing up for). What is it? What are the criteria? What are the typical/average/banded grades? What do the students say about it (e.g. in the SETU feedback)? What are examples of past papers? They also want to know their Grade Point Average and what they need to achieve to improve/maintain it.	success? 3. How have other students [preferably those most like me] performed in these assessments?

Insights from development of the open data application (Stage 3)

The next stage of the research involved the development of the open data application. As outlined earlier, the aim here was the development of 'really simple' applications rather than fully working products. This stage acted as a '(dis)proof of concept' in technical terms of data handling and software development, and allowed for exploratory investigations of the likely applicability of the open data philosophy in the realities of the university context. As such, the development of the data application raised a number of salient issues with regards to the technical/programming feasibility of the workshop designs; the limitations of the open data sets; and the likely practicality and use of the data applications.

The development team used their experience of participating in the workshops, the paper prototype, and the design brief to develop the application. The process was lengthy, largely due to the complications of gaining access to some of the data sets in a suitably 'raw' format. These complications related primarily to faculty checks and permissions regarding privacy and risk. In addition, as well as the format of the provided data had significant implications for the development team's ability to respond to the design, and in the functionality of the application.

The challenge of simple applications

Open data projects are best served by the development of 'simple' tools using familiar software such as Microsoft Excel or Google Sheets. Such an approach makes any such tool more likely to be further maintained and extended by the users. However, after considering a range of options, the development team judged that these simple systems were not the best choice for the designed prototype as they would limit what could be achieved with the data - in particular how it could be displayed and interrogated as per the design brief. In order to make the visualisation and manipulation of the data intuitive for students who were not familiar with data modelling in programs such as Microsoft Excel and Google Sheets, a custom interface would need to be programmed. In addition to this, the design briefs involved live searches of media such as YouTube and LinkedIn in order to provide additional details about the subject and/or teaching staff. Finally, the workshops also highlighted the value of a system that could be used on a mobile or tablet. As a result the development team chose to develop a web application that interrogated a database that

was maintainable and scalable through automated procedures, should the project expand to incorporate a larger number of units and faculties.

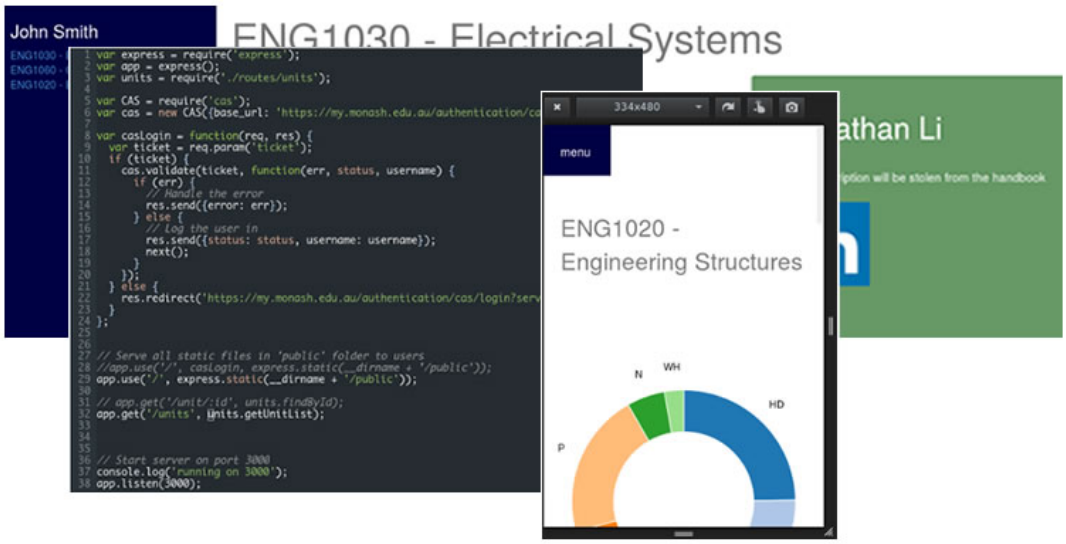


Figure 4. Initial programming designs of prototypes

The challenge of access and parsing of data

The programming team faced a number of challenges relating to the access and parsing of the data. Given its confidential and somewhat personal nature, faculty members were understandably reluctant to make large sets of the data available. This was overcome through open discussion and approaching individual subject leaders for permission to use the data relating to the unit of which they were in charge. This data included open ended student evaluations of subjects, and therefore might potentially include data containing sensitive negative comments about teaching staff, subject, Faculty and university. It is noteworthy that in Stage 1 (data audit) this data was noted by Faculty and central staff as requiring careful handling, but otherwise raised no impediment to its use. During Stage 4 when access to the raw data was sought, the request initiated a series of conversations with senior professional staff who manage university data, Faculty academic leaders and senior university managers. The nature of these discussions clearly indicated that issues of risk and privacy were seen differently across the university.

Once access to data had been finalised, in terms of the technical development of the data applications the primary impediment turned out to be the limited nature of the provided data sets. Given the diverse range of data sources, data for each unit was not provided in a standardised format. As such, it was necessary to parse a number of different CSV and HTML documents using a set of Python scripts. These files were often structured differently between unit groups and so the scripts used to parse the data needed to take into account multiple document formats for each of the document types. Furthermore, unit design differences needed to be taken into account and catered for. For example some units featured three assignments while others featured only one or two. After parsing and normalising the data, it was imported into a MongoDB database (a NoSQL database designed to scale).

While this data normalisation and parsing process was quite laborious, should this data have been maintained in a database and made accessible through an API, the server side development of this application would be greatly simplified.

Finally, in researching ways to display data, the programming team made use of the 'D3 JS' charting library. This JavaScript library - coupled with NVD3 (a library of pre-made D3 components) - provided the means to generate polished, interactive charts with little effort. Moreover, D3 proved to be extremely versatile with respect to what data it accepted as input.

The developed application

The developed application (both client and server) were programmed entirely in JavaScript. The server was written using a number of Open Source technologies including Node.js, Express JS and MongoDB, while the front end was built as a Single Page Application (SPA) using Google's Angular JS framework. The Passport JS NPM package was utilised for security, given that only students and staff with authorisation had permission to access the information. In addition to this, the data was stored on a Monash server in order to keep it safe from external traffic.

Figure 5 illustrates one of the application functions that to allows users to filter and compare Units (subjects) according to a star rating based on student evaluations.

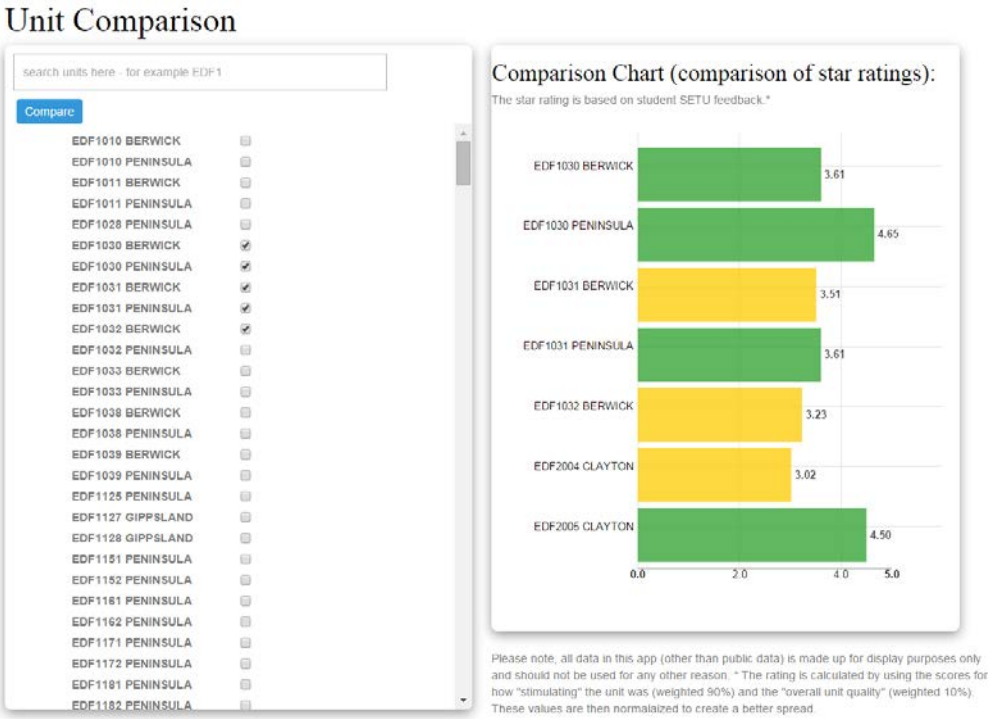


Figure 5. Screenshot of subject comparison function in the data application
Note: data provided in these screenshots are not actual data for the named unit

Figure 6 provides an example of how students could interrogate the data relating to each of the Units (subjects) including the unit description, assessment, grades distribution, student evaluations, and even searches of unofficial sources such as LinkedIn and YouTube.

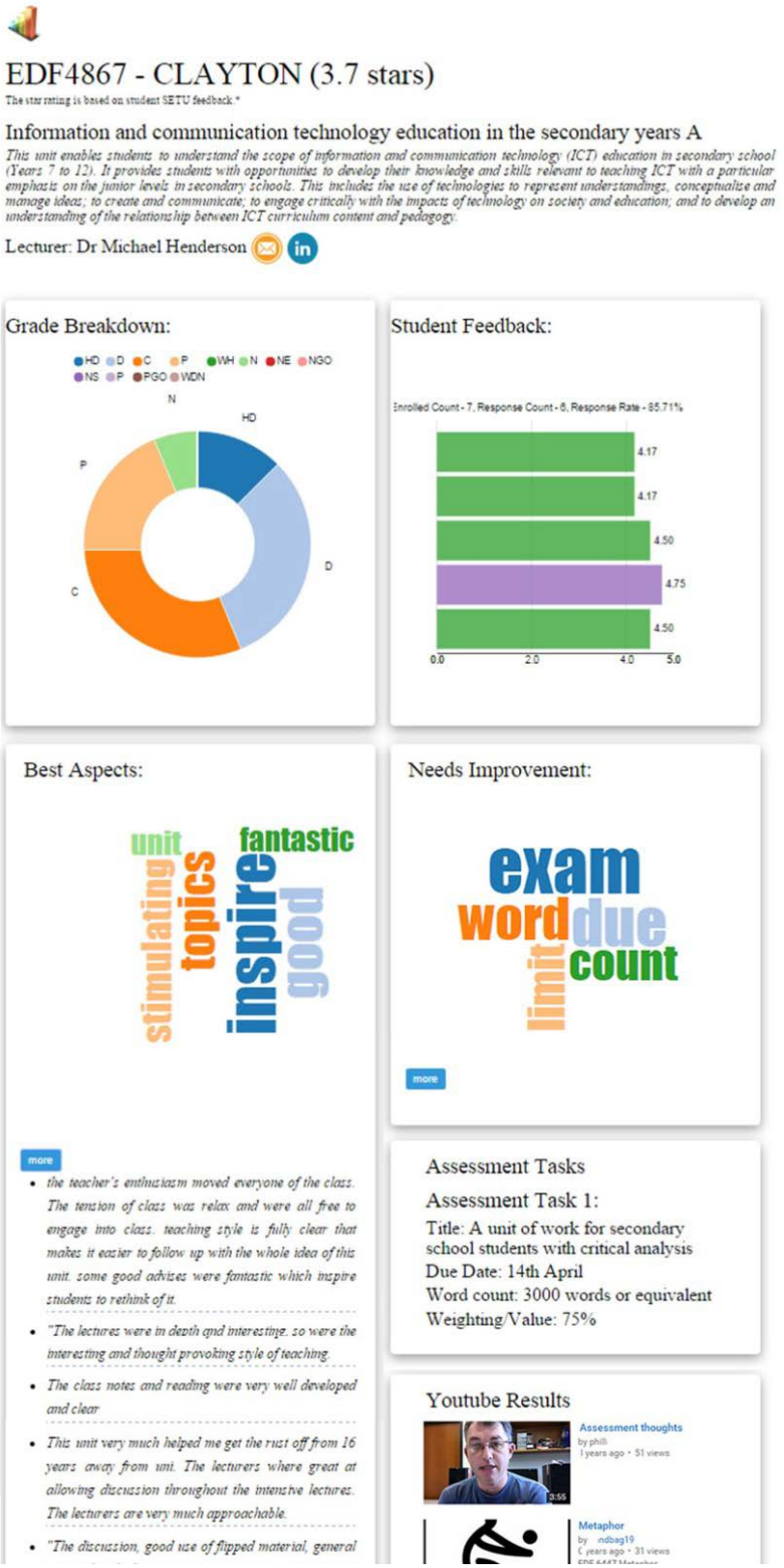


Figure 6. Screenshot of subject level function in the data application
Note: data provided in these screenshots are not actual data for the named unit

Insights from the evaluation of the open data application (Stage 4)

The final stage of the project involved the developed open data application being presented back to the student groups in order to gain a sense of how these open data tools and techniques might 'work' *in situ*. Follow-up testing sessions and interviews were conducted with both groups of students to develop a sense of how these open data applications might fare 'in the wild' if they were to be used on a sustained basis across the student population.

In general, both groups were positive about the developed application in terms of its applicability to their future unit choice making:

> "the information here is really awesome. This is the stuff you want" [UG]
>
> "I can make up my mind with the information" [PG]

The range of data representation options was welcomed as catering for different preferences and numerical abilities:

> "Having the coloured-in stars and the graph so some people that aren't necessarily maths-oriented, the stars will be very, very helpful … this is a very dramatic chart for me in how much people love that unit and don't like that unit."

One mode of data representation included in the application design was the semantic analysis of student evaluation comments. In testing the final built application, however, both groups felt that the use of a word cloud generator provided de-contextualised information that was of little use. For example:

> [A]: The word cloud is a bit unclear, I think. I don't think that's very helpful. Yeah, there's not enough words for that to be meaningful.
> [B]: Well, yeah, too, I think taking those words out of context, it needs improvement limit. Limit what, limit the content, limit the learning time, I don't really know what limit means. [p3]
> [A]: Some of the key words in that word cloud like "good", obviously a lot of people said good. But that does not necessarily tell me anything about the unit. What was good about it?... I'm not sure I like it the way it is. [UG]

Otherwise, as had emerged from the previous design discussions, both groups of students valued the ability to make more detailed comparisons about their own performance in relation to others:

> "knowing most people get credits … that 57% of people get credits. Yeah, that makes me feel better about myself. I like this … because I got a credit too, and initially I thought, oh, that's a bit rubbish … but turns out most people did. So it's a sense of gratification" [UG]

These benefits were seen to outweigh the application's relative 'bare-bones' appearance, and lack of sophisticated presentation. As one student summarised, "this is actually useful so what it looks like doesn't really matter – I'll still use this" [PG]. As such, most of the groups' suggestions for further improvements related to the nature and scope of the data-

sets rather than the data techniques and practices used in the application. For example, having access to the full range of grade data meant seeing grade categories that most students are unaware of. For example:

> "What's the difference between N and NE? And what's WH mean, and PGO? I don't know what that one is. I don't know what half of these mean" [UG].

Other reactions highlighted a perceived lack of data granularity. For example, requests were made for further filtering of the data by particular campus attended, type of student making comments (e.g. international/domestic, part-time/full-time), and comparisons with similar units in other universities. These requests were all clearly beyond the remit of the supplied data sets, but certainly reflect the development of student interest in the usefulness and potential application of such data.

Discussion

While the end-product of this project was clearly limited in scope, it is perhaps worth first noting the relative success of the design cycles with both groups of students. This project was carried out during a period of heightened popular and professional concerns in the US, Europe and Australia over data privacy and the commercial exploitation of student data. Nevertheless, once statutory concerns over privacy and duty-of-care were satisfied, university and Faculty managers were willing to allow the opening up of selected data-sets. At all stages of the project there was little overt institutional reticence or resistance to the general notion of open data principles. Subsequently, it also proved technically possible to design and develop a relatively powerful data-handling application for both student groups in the short timeframe of the project schedule. The open data application and practices that were developed clearly added value and insight to these datasets, and saw a range of students collaborating and engaging in determining alternate uses of data that better served their experiences as university students. While the substantive topic that these designs addressed was rather prosaic and bounded in nature, it nevertheless constituted alternate uses of data to address 'real-world' problems and allow students to consider ways of engaging with data that supported meaningfully oriented behaviour. From this perspective of proving 'proof of concept', the project could be seen as a qualified success.

However, as is intended in any participatory design project, these activities were as revealing in terms of their 'failures' as any apparent 'successes'. As such, there is much that our findings say about the limitations of the open data philosophy when applied in the realities of university teaching and learning. Perhaps most prominent was **the limited 'open data' potential of the datasets that were being produced within the university and faculties**. Of course, the data was clearly working well in terms of the university administration and management. Yet as source material for any potential *open data* project, most of the data being produced in across the faculties and university related to relatively narrow sets of measures, such as officially defined teaching and learning 'performance' metrics and outputs, homogenised indicators of 'well being', and pro-forma unit information. This clearly constrained the possible nature and utility of any reapplications of this data. In short, this was data with little 'radical' potential for re-use and re-purposing.

Most importantly, the quality of these data-sets fell well short of what might be expected from well-organised and well-run data infrastructures – e.g. clean, interoperable data, regularly updated and archived with appropriate meta-data. Instead, the data that was eventually used in the participatory design projects conformed to Kitchin's (2014, p.64) description of the datasets that tend to end up being used in open data interventions – i.e. "low-hanging fruit, consisting of those that are easy to release and contain non-sensitive data that has relatively low utility".

Second, were obvious **limitations in the data skills and 'data imaginations' of the students** who participated in the design workshops. While clearly interested and engaged in the activities, neither of our workshop groups possessed adequate programming skills, statistical knowledge or analytic capabilities to reimagine sophisticated ways of working with the datasets. Similarly, most participants' 'data awareness' (or what could be termed 'data literacy') was confined to very restricted modes of data engagement – most notably following the simplistic forms of data visualisation and data-handling familiar from popular online applications and services. There was little interest and capacity for more advanced engagement with the data practices beyond star ratings, box plot graphs, dashboards and filters. As Noveck (2012, n.p) notes, "the ability of third parties to participate is what makes open data truly transformative". In the case of our participatory design workshops, the ability of students to participate autonomously was clearly limited to approaching open data as a finite product (i.e. the basis of interactions and relationships between specialist data-suppliers and non-specialist data-users) rather than ongoing process of equal participation (Gurstein 2011). Most non-specialists will lack the data and coding skills required to engage fully with open data sets. As such, it is to be expected that attempts to apply open data philosophies in non-specialist settings will slip into 'service' activities such as "providing easy-to-use research tools that negate the need for specialist analytic skills" (Kitchin 2014, p.148).

Third – and continuing this previous point - was the likely **limited sustainability of these open data interventions beyond the involvement of the project team**. It should be noted that completing even the one 'simple' open data application described in this report involved hundreds of hours of organisation, drafting and programming work. Even with the mundane instances of opening access to student satisfaction data, or visualising grade data over time, the labour-intensive nature of our project illustrates the large volume of work that lies behind the abstract promise of 'setting data free'. While the programming aspects of our project was intended to be relatively 'simple', technical development of the working prototype was restricted by constraints relating to the nature and quality of available data-sets. These were *not* straightforward processes of connecting data compiled at different locations together. Instead, the project faced considerable variation in the quality of the data that was accessible. Put simply, the open data application that resulted from our project were sustained by a great deal of 'behind the scenes' technical work by the research team. As such, this was an intervention that is unlikely to be easily replicated without the involvement of the project team. Again, as Kitchin (2014, p.57) observes of open data projects in general:

> "making data open is not simply a case of publishing them in a form held by [an organisation]. Much of the data needs to be repurposed and curated to enable them to be made open (e.g. anonymised, aggregated) and new systems put in place to

enable this to happen. This is not a trivial exercise and … means reallocating funding to pay for this work".

A final set of concerns relates to the rather grandiose claims of empowerment and democratisation that are often attached to open data principles. In contrast, our investigation pointed to the **limited 'democratising' effect of the open data design activities and applications** in universities. Indeed, while our projects could be said to have provided a 'proof of concept' with regards to the technical aspects of procuring and programming open data interventions, it could not genuinely be said to have realised the main claims of open data – i.e. improvements in engagement, empowerment, democracy and accountability. The design and development of our application might have been of practical use within the specific context of the Monash Faculty of Education, but it was hardly challenging the institutional *status quo* within either the faculty or the wider university. Instead, this was a data application that fitted well with pre-existing power relations and conditions of teaching arrangements. This is perhaps the most significant 'non-finding' of our research – i.e. the largely unthreatening, non-controversial, non-transformatory nature of the open data tool that resulted from our participatory design activities. In this sense, applying open data principles in schools would seem likely to result in what Peixoto (2013) describes as *ad hoc*, low-level and ultimately 'toothless' outcomes.

Conclusion

This project marks a first attempt to test the aspirations of open data as a 'social movement' (Halford *et al*. 2013) against the organisational realities of the contemporary university. Clearly any attempt to support more 'open' forms of data access and use within universities faces a number of potential challenges. These include: the quality of the data itself; the congruence between the goals of open data projects and the concerns and responsibilities of the higher education institution; 'external limitations' such as ethical, legal and managerial concerns; and 'internal limitations' such as the data skills and competencies that exist throughout student and educator communities (cf. Greller and Drachsler 2012).

In short, our research would seem to reinforce Rob Kitchin's (2014, p.63) conclusion that "the democratic potential of open data has been overly optimistic". This is not to conclude that university educators and student groups should give up on the ideals of open data altogether. Yet this does suggest that any promises of an imminent 'open data revolution' are acted upon in a critical and circumspect manner, with any attempts to 'open up' data within universities approached with suitably tempered and realistic expectations.

Recommendations: institutional adoption of open data projects

This project has tested the feasibility of open data approaches relating to student use of university data. We are now in the position to offer a concluding guidelines and recommendations highlighting the practical limitations of open data principles and philosophies – i.e. working through institutional concerns over appropriate data use; recognising the likely technical difficulties of such data work; as well as working to develop

and sustain student capability and interest.

Recommendations:

- **An open institutional approach…**
    - recognises that open data projects are necessarily about *students'* real world problems – not about institutional problems;
    - recognises that open data projects involve considerable and ongoing commitment of time and resources;
    - adopts a flexible approach to imagining and designing what data may be used for;
    - is comfortable with a bottom up approach to development;
    - has a realistic expectation of the likely limits of any open data application that is eventually developed;
    - recognises that data access and privacy are key mitigating issues that need to be debated and discussed at all levels;
    - recognises the limited capacity to repurpose data that has been collected for specific institutional purposes.

- **Open data work and application development…**
    - involves a considerable amount of planning to collect and collate data that meets the principles of open data (e.g., raw, live, interoperable, etc.);
    - involves a considerable amount of technical work involved in making data usable and re-purposable;
    - values functionality over form – that is, issues of aesthetics and 'completeness' are outweighed by the drive for addressing a real world problem.

- **Open data projects…**
    - rely on a fully participatory approach – based on partnerships between central data experts, application developers, and students;
    - centrally position student voice in terms of identifying real world problems and meaningful 'solutions';
    - trust students to make sensible and meaningful decisions with regards to their own needs;
    - support students to develop data literacy and data 'imaginations' so that they can be fully involved at all stages.

Impact: dissemination and engagement

This was an exploratory project with the intention that the results have an impact on current and future debates over the role of Open Data in Higher Education. In order to facilitate this, we have concentrated during the lifetime of the project on developing high quality findings that reveal the complexity of any such undertaking. The process of subsequent take-up and impact of these findings will be ongoing. Deliverables and dissemination activities including journal and conference presentations are listed below.

Project deliverables

By combining a participatory design with follow-up evaluations, the project has developed realistic understandings of the opportunities *and* limitations of open data use within student cohorts. This is of clear use to both the higher education community *and* the open data community. We had hoped as part of the original project design to be in the position to offer as part of its deliverables, the development of protocols, applications and technical guidelines for (re)use and (re)appropriation by other Higher Education institutions. However, the practical limitations and challenges across the project life cycle resulted in a much more constrained outcome. This project's main outcomes, therefore, relate to the establishment of valuable groundwork in terms of highlighting constraints and opportunities, and as such feeds directly into our key deliverable, that is, the production of guidance for institutions seeking to adopt open data techniques.

Notwithstanding these limitations, this project has already begun to impact within Monash University. Based on the project findings, the team has provided expert advice to several university wide strategic initiatives. These have included:

- Consultation sessions on 'Analytics as an enabler for improving student experience'
- Office of the Pro-Vice Chancellor Learning & Teaching: "Closing the loop": rethinking the visualisation and use of student evaluation data in the context of other data

The implications of these and other discussions has result in the adoption of findings including the need for student participation in designing data applications, and the need for more appropriate collection and collation of data about student experience that is of use to students. A further indicator of impact has been the raising of awareness, and engaging in discussions, at the Faculty and Vice-Provost level regarding risks and privacy concerns relating to student related experience data.

Impact building will continue with the **on-going dissemination** of the project findings and recommendations. These are hosted on the project website [http://bit.ly/openforlearning].

Dissemination

Peer reviewed journal article:

Selwyn, N., Henderson, M., and Chao, S. (under review). 'You need a system': exploring the role of data in the administration and management of university teaching.

Conference presentation:

Selwyn, N. (2015). Data lessons … making sense of the datafication of education institutions. Paper presented to the *Eastern Sociological Society Annual Meeting*, New York City, February 2015

Other presentations:

Henderson, M. and Selwyn, N. (2014). Open for learning: using open data tools and techniques to support student learning. Paper presented to the *Monash Educational Excellence Research Group (MEERG) Symposium*, Monash University, December 2014.

Endnotes

Staff interviewees are numbered in text as follows:

1.	Faculty of Engineering, Faculty manager – student services
2.	Faculty of Education, Faculty manager – student welfare
3.	Faculty of Education, Faculty manager - teaching
4.	University senior information officer
5.	University 'lead' on student data
6.	Faculty of Engineering, Data integrity officer
7.	Faculty of Education, e-learning officer
8.	Faculty of Education, Data integrity officer
9.	Faculty of Engineering, Course advisor
10.	Faculty of Education, e-learning officer
11.	Faculty of Engineering, Faculty manager - teaching

References

Argote, L. (2013). *Organizational learning: creating, retaining and transferring knowledge*. Berlin: Springer.

Boudreau, K. (2010). Open platform strategies and innovation: granting access vs. devolving control. *Management Science, 56*(10), 1849-1872.

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: a generic framework for learning analytics. *Educational Technology & Society*, *15*(3), 42–57.

Gurstein, M. (2011). Open data: empowering the empowered or effective data use for everyone? *First Monday*, 16(2). Retrieved from http://journals.uic.edu/ojs/index.php/fm/index

Halford, S., Pope, C., & Weal, M. (2013). Digital futures? *Sociology*, *47*(1), 173-189.

Kitchin, R. (2014). *The data revolution*. Los Angeles: Sage.

Longo, J. (2011). #Opendata: digital-era governance thoroughbred or new public management Trojan horse? *Public Policy and Governance Review*, *2*(2), 38-51.

Noveck, B. (2012, July 5). Open data—the democratic imperative. *Crooked Timber.* Retrieved from http://crookedtimber.org/

Open Government Data. (2015). *Eight principles of open government data*. Retrieved from http://opengovdata.org/

Peixoto, T. (2013). The uncertain relationship between open data and accountability. *UCLA Law Review Discourse*, *60*, 200-213.

Peled, A. (2011). When transparency and collaboration collide. *Journal of The American Society for Information Science and Technology*, *62*(11), 2085-2094.

Reichman, O., Jones, M., & Schildhauer, M. (2011). Challenges and opportunities of open data in ecology. *Science*, *331*(6018), 703-705.

Ruppert, E. (2013). Rethinking empirical social sciences. *Dialogues in Human Geography, 3*(3), 268-273.

Sanoff, H. (2007). Special issue on participatory design. *Design Studies*, *28*(3), 213-215.

Siemens, G., Gasevic, D., Haythornthwaite, C., Dawson, S., Shum, S., Ferguson, R., Duval, E., Verbert, K. & Baker, R. (2011). *Open Learning Analytics* (Research Report). Retrieved from Society for Learning Analytics Research website: http://solaresearch.org/OpenLearningAnalytics.pdf

Spinuzzi, C. (2005). The methodology of participatory design. *Technical Communication*, *52*(2), 163-174.

Yoo, Y., Boland, R., Lyytinen, K., & Majchrzak, A. (2012). Organizing for innovation in the digitized world. *Organization Science, 23*(5), 1398-1408.

Appendix A: Certification

I certify that all parts of the final report for this OLT grant provide an accurate representation of the implementation, impact and findings of the project, and that the report is of publishable quality.

Name: Professor Darrell Evans Date: [28/05/2015]
Title: Vice-Provost (Learning and Teaching)

Appendix B: sample of student open data designs

Short hand descriptor	Description	The kinds of queries they want to perform / The kinds of questions they want answered	How they want to manipulate the data	How they want to see the data	Extensions of the goal arising in the workshop which we clearly cannot pursue due to lack of access to data
Trip advisor for subjects	**Students would like to know more about the units before they enrol.** At the moment they only get the unit handbook which does not provide all details about the teachers, the assessment, etc. They want to know about the teachers, the grades, what students say about it, the assessment, the schedule, etc. They want to know about their options. These data will inform their potential unit choices, but also, in cases where they have limited choices these data will orientate them to the 'nature' of the unit (i.e. know what they are letting themselves in for).	1. Who is teaching 2. What is the assessment (type/task, criteria) 3. What are the grades (average, frequency of grade bands, breakdown according to course enrolment) 4. How popular is this course? 5. How is it rated by other students?	**They want to…** *At the unit comparing level:* 1. Sort (the units according to the various indicators – e.g. list the units in order of best to worst) 2. Filter (filter out the units that are not relevant to them, that do not meet their interests) 3. Compare (do a more detailed side by side comparison of key indicators: grade, SETU, other?) 4. Compare the units (electives) they could choose for the following semester or year *At the deeper unit level:* 5. Weight/include/exclude factors (weight the indicators according to their interests, e.g. they don't trust the SETU Qual so they want to weight that less or exclude it from the unit score/ranking, alternatively they value units where students like themselves have a high proportion of D/HD grades.)	• quick "at a glance" visualisations or dashboards. But they want the option of digging down to the more detailed information. • traffic light systems. E.g. "the whole page turns a certain colour" • easy to understand graphs that do not require too much statistical knowledge to interpret. • interpretations of the qualitative data: e.g. word frequencies / tag cloud. However, they want this thematically driven. E.g. what words are used in relation to assessment / lecturer / content. What are the positive / negative phrases used. • the popularity of teaching / units, using a simple rating system (e.g. star rating). • the ability to select data that relates to themselves only • They want the option to customise what data is seen, and still be able to dig deeper, such as view more detailed information.	**NOTE: an extension** [which we cannot address in this project] of this **is** the benefits / skills gained / relevance of doing the unit(s) for my current course / higher degree by research? Also, what previous students have gained from doing the unit?

Short hand descriptor	Description	The kinds of queries they want to perform / The kinds of questions they want answered	How they want to manipulate the data	How they want to see the data	Extensions of the goal arising in the workshop which we clearly cannot pursue due to lack of access to data
Know thy neighbour	**Students would like to understand their position in a course and how they relate (and compare) to others in their course**. They want to know if it is a big/small unit, who the other students are (e.g. are they in the same course as themselves), more about the lecturer, etc. The students felt this was "interesting" and that might give them a faster sense of connection with the class that they don't have – especially in the current online/blended modes. One way to do this (as suggested by students) is to increase communication with others who are doing the same unit/same course.	1. Who else is doing my course? [Not names… but rather a sense of the kind of others … such as numbers of students, grades, current and past courses/units completed, other demographics] 2. Who is my lecturer / tutor? 3. How do I keep in contact / network with those who are studying my course? 4. How do I compare with peers? 5. How do I track my progress in my unit(s) / course? 6. Where can I go next?	They want to… 1. Compare – Situate themselves (their past, current and future studies, interests, grades, demographics, etc.) with the others in their units. 2. Filter - They want to use the various data [e.g. Grades, location, etc.] to refine who they feel closely aligned with. [and potentially be able to contact them] 3. Select - They want to know where they can go after this unit *[overlap with trip advisor theme]*, and what grades they need to achieve.	They want to know the percentage of students [rather than names] who got certain grades in the previous semester / year. Similar forms of data representation with other student data. They want to see which units they have already passed, and what units they can choose in the following semester / year.	**NOTE: an extension** [which we cannot address in this project] of this is the desire to actually connect with students who have been in previous classes with them or who are enrolled in the same course. It was mentioned by one workshop group that if they knew others in the same course they could do group assignments with them, and share relevant readings, etc.

Short hand descriptor	Description	The kinds of queries they want to perform / The kinds of questions they want answered	How they want to manipulate the data	How they want to see the data	Extensions of the goal arising in the workshop which we clearly cannot pursue due to lack of access to data
Being smart about assessment	**Students want to succeed in their assessment/ grades.** They see this directly linked with their assignment performance. Therefore they want to be better prepared for what they need to do for their assignments. **They would like to orient themselves for the assessment (i.e. know what they are signing up for).** What is it? What are the criteria? What are the typical/average/banded grades? What do the students say about it (e.g. in the SETU feedback)? What are examples of past papers? They also want to know their Grade Point Average and what they need to achieve to improve/maintain it.	1. What is the most important thing I need to do to succeed? [where success is defined by my own criteria, e.g. C, D, etc.] 2. What kind of assessments are in a unit and what are their criteria for success? 3. How have other students [preferably those most like me] performed in these assessments?	They want to… 1. View the detail of the assessment including criteria. [see also the extension column for the detail they really wanted] 2. See success of previous students, sorted/filtered according to similarity to themselves.	They want to know the quality of teaching and marking, from the feedback provided by previous students: • Teaching style(s) • Student-lecturer interaction • Assessment tasks, e.g. marked fairly and providing a lot of constructive feedback…etc.	NOTE: this goal around orientation/preparation for assessment was **very high** on the students' agendas. However, their valuing of this idea diminished when we explained that it is unlikely we can get previous examples of assessment (e.g. essays), teacher feedback, or individual assignment marks within a unit but only the overall mark for the unit. I think we could get some of the richer information from Moodle data but that is a different app entirely, and fraught with problems of whether the lecturers use the Moodle gradebook, give feedback through Moodle, etc.