Australian Government
Department of Education, Skills and Employment

Learning Analytics beyond the LMS: Enabling Connected Learning via Open Source Analytics in 'the wild'

Final report 2020

Lead institution: Queensland University of Technology

Partner institution(s): University of Technology Sydney, The University of Sydney, University of South Australia

Project leader: Dr Kirsty Kitto

Team member(s): Mandy Lupton, Peter Bruza, Dann Mallett, John Banks, Shane Dawson, Dragan Gasevic, Simon Buckingham Shum, Abelardo Pardo, George Siemens

http://www.beyondlms.org

Support for the production of this report has been provided by the Australian Government Department of Education, Skills and Employment. The views expressed in this report do not necessarily reflect the views of the Australian Government Department of Education, Skills and Employment.

Higher Education Program Management
Governance Quality and Access Branch
Higher Education Division
Australian Government Department of Education, Skills and Employment
GPO Box 9880
Location code C50MA7
CANBERRA ACT 2601

<learningandteaching@education.gov.au>

2020

ISBN 978-1-76051-962-9 [PDF]
ISBN 978-1-76051-963-6 [DOCX]
ISBN 978-1-76051-964-3 [PRINT]

Acknowledgements

This project would not have been possible without inspiration and support from many people across Australia and in the wider international context.

The inspiration for this project stems from the QUT Transform team. Sue Savage's insistence that she needed the learning environments used by real professional learners to be 'knitted together' provided a use case that was both compelling and terrifying. However, people like Sheona Thompson, David Dennis and Karen Whelan all helped us to explore the challenge and support the project to scope out its potential. The support and ideas from numerous Transform Fellows and other professional staff continued to guide the project evolution at QUT, as did the ongoing support of the Awards and Grants team.

Overwhelming thanks go to both Aneesha Bakharia, who provided this project with a firm foundation for its first year, and Kate Davis, who provided ongoing commitment to its outcomes through volunteering her classes as testbeds at QUT. None of the results obtained throughout this project would have been possible without their contributions. Thank you to the many students who have worked on this project throughout its lifetime. Special mention to Sebastian Cross and Koji Nishimoto, who worked on substantial associated sub-projects for the CLA toolkit V1 over a number of semesters. More recently, the contributions and support of the various people supporting the development of the LA-API architecture at UTS cannot be overstated, especially: Radhika Vijay Mogarkar, Gabrielle Gardiner, Moein Ghodrati, Andrei Maslennikov, Aleksandr Gromov, Abishek Arunachalam and Peter Scott.

Internationally, the support provided by the xAPI community to this project has been extraordinary. From early contact with Andrew Downes, who suggested that we had 'a really big recipe' that he would like to help us create, to the ongoing support and guidance provided by Aaron Silvers and Megan Bowe as the topic of data interoperability became increasingly important, this project found an overwhelmingly friendly home among the xAPI community with its associated developers and vendors. The Society for Learning Analytics Research (SoLAR) also provided support to this project through its annually hosted events such as the Learning Analytics and Knowledge (LAK) conference and Australian Learning Analytics Summer Institute (ALASI), where many of the project's outcomes have been communicated. Particular thanks go to the ongoing LAK Hackathon community, which has spent a number of years exploring the potential of xAPI in delivering LA infrastructure. A special mention to Alan Berg and Niall Sclater for their ongoing support, as well as to the Learning Analytics Community Exchange (LACE) data interoperability for learning analytics team, including Tore Hoel, Dai Griffiths and Adam Cooper.

Finally, a special thanks goes to Grace Lynch, who provided far more than an evaluation for this project – supporting it and helping to guide it from its inception through to its conclusion.

List of acronyms used

API	Application Programming Interface
ALASI	Australian Learning Analytics Summer Institute
ASCILITE	Australasian Society for Computers in Learning in Tertiary Education
CoI	Community of Inquiry
EdTech	Educational Technology
GDPR	General Data Protection Regulation
IEEE	The Institute of Electrical and Electronics Engineers
ICICLE	Industry Connections, Industry Consortium on Learning Engineering
LA	Learning Analytics
LA-API	Learning Analytics API
LACE	Learning Analytics Community Exchange
LAK	Learning Analytics and Knowledge Conference
LASI	Learning Analytics Summer Institute
LMS	Learning Management System
LRP	Learning Record Provider
LRS	Learning Record Store
noSQL	No SQL - used in non-relational or distributed databases
QUT	Queensland University of Technology
SIG	Special Interest Group
SQL	Structured Query Language - used in relational databases (RDBS)
UniSA	University of South Australia
USyd	University of Sydney
UTS	University of Technology Sydney
xAPI	Experience API

Executive summary

Over the past decade we have witnessed a proliferation of new technologies involving an increasingly sophisticated use of data. In education, the field of learning analytics (LA) is one such example, seeking to explore how the analysis of student data can bring new insights into the learning process. However, the breadth and diversity of technologies available today and the corresponding wealth of learning data stands in stark contrast to the day-to-day operations and processes that operate in formal education. Large educational systems such as universities, TAFEs and other providers have an essential requirement for technical stability and scalability.

The learning management system (LMS) is an example of a standalone solution that provides stability and scalability. LMSs combine myriad features that can be used in learning and teaching. From the management of courses and enrolment, to discussions, group work, and even assessment and assignment tracking, the modern day LMS provides a seemingly viable solution. Yet there are numerous other tools and platforms that can provide alternate and often more authentic learning experiences – social media and MOOCs, for example. However, an ecosystem of technologies requires significantly more resources to maintain, and to date there have been few solutions for aggregating the data from a learning ecosystem. Despite these complexities, restricting our academic teachers and students to a single institutionally endorsed tool such as the LMS encourages a binary of compliance or non-compliance. As such, tools like the LMS are frequently viewed by teachers and students as barriers to overcome rather than tools to embrace and support teaching and learning practice.

This project explored the dynamic between the need for teaching innovation alongside the need for the formal administration of education technologies. The project sought to identify a solution that would enable educational innovators to teach across platforms and systems using authentic real-world technologies, while recognising the need for quality, privacy, ethics and data control. Project outcomes demonstrate that it is possible to provide rich and authentic learning experiences for students 'in the wild', and still deliver learning analytics to staff and students using interoperable data that is ethically collected and securely stored.

The *Beyond LMS* project aimed to improve the quality of student engagement and learning in collaborative online environments by incorporating analytics developed using the data generated in social media platforms that the majority of students already use. The project explored the possibilities created by the emergence of a new educational data standard, Experience API, (xAPI), investigating its potential for harmonising learning data from a wide array of online environments and then using it to deliver Learning Analytics (LA) to students and staff.

The project created a suite of open source tools that work within closely delimited learning activities, underpinned by connected learning pedagogy. These tools are designed to give students an understanding of, and control over, the data they collect. This both preserves student privacy and helps students to learn about how their social media data is used in analytics systems.

A case study consisting of a series of pilot trials of the Connected Learning Analytics (CLA) toolkit was conducted over a period of 18 months at QUT. This informed a number of key findings for this project (more details can be found in Chapter 2).

Finding 1: To be effective, LA dashboards must be tightly coupled to the learning design and/or assessment regime of a subject. User interpretation of LA results are contingent on the pedagogical context.

Finding 2: Students can make use of LA dashboards and reports to develop an understanding of their approaches to learning beyond the LMS.

Finding 3: Student awareness and training to interrogate data is required. Students were likely to readily accept the LA reports without question.

Finding 4: Tightly coupled LA architectures are difficult to maintain and prone to quick obsolescence.

These findings led to a major redevelopment of the CLA toolkit V2, which remains an active project currently being pursued at UTS. More technical findings and recommendations of this project stem from that redevelopment process (see Chapter 4).

Finding 5: Designers of institutionally scalable LA infrastructure should seriously consider a highly modular architecture that will enable ongoing extensions of the datasets used, modifications of reports, and new integrations.

Finding 6: To be useful in a student-facing context, learning analytics dashboards must be highly configurable, with different reports turned on or off depending upon tools used, learning design, assessment regimes and student data literacy.

Finding 7: The data traces created as students make use of any configurable dashboards are likely to be a rich source of information about metacognition, critical thinking and self-regulated learning. They should be a priority for future LA work seeking to develop 21st century skills.

An important set of outcomes associated with this project concern the concept of data interoperability. As universities enter into more porous relationships with their students it will become increasingly essential that the digital traces those students generate make sense across all of the learning ecosystems with which they engage. It should not matter which educational data standard is supported by the university that a person chooses to attend; all data should be mappable between learning environments as the need arises. This project has tackled the problems of data interoperability and its implications for lifelong data portability (see Chapter 3), which has to date driven its most profound impact.

The major output of this project is an extensive open source codebase. The original primary objective of this project was to deliver the Connected Learning Analytics (CLA) toolkit, but this infrastructure has now been extended with a series of modularised components that serve to build up a Learning Analytics API (LA-API). Thus, a major outcome of this project is a concept and codebase for the delivery of LA at scale. In addition to this codebase and the project webpages, a number of publications, public talks, and informal outputs are highlighted in Appendix B.

To date, the impact of this project has rather unusually centred around the broad and systemic category. The work on data interoperability that has been carried out during this project has been fed into work by two standards bodies (the ISO and the IEEE), changing the ground-rules about what is considered acceptable in educational data interoperability. The Learning Analytics Community Exchange (LACE) project run by the European Union has referred to this project extensively in its work package on data interoperability for LA. It has used the findings of this project to produce recommendations for the field (especially under the emerging environment created by the GDPR legislation).

As the LA–API infrastructure currently being implemented at UTS rolls out, the tools delivered by this project will be used to scale the delivery of LA across an entire university. A number of other institutions have expressed interest in the CLA toolkit, and with the new more modular codebase now being delivered with V2, it will be easy for them to adopt those parts of the infrastructure that suit their needs.

Table of contents

Tables and figures

Tables

Figures

Chapter 1 – Introduction

Learning is an ongoing and continuous process that occurs across formal and informal educational contexts. In short, students can *learn* anywhere and everywhere. However, *education* (as a structured process) is largely confined to formal institutions such as schools, colleges and universities. In such regulated and accountable contexts, the tools and resources used to deliver education are frequently constrained.

There is an interplay in any large organisation between the need for scalable and consistent work processes and the need to evolve and adapt. For instance, within education settings the adoption of the learning management system (LMS) has provided a scalable tool that tends to cater to a wide range of teaching approaches. However, there exists a constant pull towards adapting and evolving teaching practices to better cater to future graduate and industry requirements. This means that established institutional tools are increasingly in discord with the types of teaching models our most innovative staff attempt to employ. There is an end of life for all technologies, and educational technologies are no exception.

At present, much of the technical infrastructure in higher education has been built under the assumption that student learning primarily occurs within the confines of one Learning Management System (LMS), in one institution, using one Student Information System (SIS), for the duration of their degree. Another assumption is that upon graduation the same students transition to a role in a sector in which they will largely remain for their working life. Although these are increasingly erroneous assumptions, the establishment of a singular technical infrastructure simplifies the problem of IT architecture and resourcing for an institution. However, it also diminishes the variety of learning experiences that can be offered to students. The reduced diversity of learning technologies available to staff and students that could be considered as part of an institutional learning ecosystem affects the provisioning of viable Learning Analytics (LA) solutions. To date, LA has tended to almost exclusively focus on institutional data derived from the LMS and SIS. Some Australian universities are now at the point where their underlying data architecture is sufficiently mapped to enable collection of data along a student's learning journey – from enrolment to graduation. But how much, and what data, is missing from this student map?

Students increasingly use apps like Facebook[1] and Whatsapp[2] to talk to one another about their studies; they frequent Slack[3] channels for group projects, make use of sites like Stack Exchange[4] and YouTube[5] to find answers to questions that arise for them, and participate in

[1] https://www.facebook.com/
[2] https://www.whatsapp.com/
[3] https://slack.com/
[4] https://stackexchange.com/
[5] https://www.youtube.com/

MOOCs or other short courses offered by services such as Datacamp[6] to learn more about topics that interest them. Furthermore, students participate in extracurricular activities that can enhance their chances of finding gainful employment upon entering the workforce. They join clubs, coach school sporting teams, undertake internships and mentor peers. They also participate in open source or creative commons communities, generating artefacts in the process that provide strong evidence of their skills, creativity and capability. This learning extends well beyond the boundaries of the university.

An LA system that considers only the data obtained from in-university activities misses this rich and complex set of lifelong learning. Institutionally bound LA frameworks only provide a partial model of a student's overall skills and knowledge. The associated predictions or models about a student being 'at risk' of failure consider only a small subset of relevant data. It is disconcerting that while universities are increasingly concerned with employability, the very artefacts, networks and competencies that students establish 'in the wild', or *beyond the LMS*, are also some of the most indicative of their eventual success. Remarkably, little is being done to make use of this data or to help students to store it as evidence of their learning development over time and demonstration of capabilities. The field of LA must work to provide scalable and intuitive methods for incorporating data generated outside of formal systems to aid the development of comprehensive learning models. However, many technical, social and ethical issues arise when attempting to resolve this challenge. For example:

- How can data be accessed from the wide range of systems in which it is created?
- How can we ensure that the data is consistent across our systems?
- Who should have access to and *control* this data? And how can our students have a say in whether it is even collected?
- What ethical frameworks will help keep students at the centre of decisions made using non-traditional sources of data?
- How are learning data interpreted to enable actionable intelligence?
- How can this data be used in formal education? And *should* it?

The *Beyond LMS* project aimed to improve the quality of student engagement and learning in collaborative online environments by collecting and analysing the data generated in social media platforms that the majority of students already use. The project explored the possibilities created by the emergence of a new educational data standard, Experience API[7] (xAPI), investigating its potential to enable educational data interoperability across the

[6] https://www.datacamp.com/

[7] https://www.adlnet.gov/research/performance-tracking-analysis/experience-api/

widely varying platforms that are increasingly being used in lifelong learning. This chapter introduces the key questions that motivated this project.

Lifelong learning

The emerging age of workforce disruption (CEDA, 2015; WEF, 2016) implies that people will increasingly enter into more porous relationships with the educational system. Rather than completing education in a defined period at the beginning of their lives, people will increasingly need to return to the education sector: for further training; to upskill; or to reskill so that they might transition into new sectors as the need arises. Flexible and modularised offerings will increasingly depend not just on ePortfolios, but also on *data portability*. Students are coming to expect that the digital traces they generate across multiple universities or education systems, are not just collected for auditing or institutional purposes but are also made accessible to them throughout their lifetime (Arnold and Sclater, 2017). In a world reliant upon data and analytics it is essential that we enable people to curate their personal learning data, providing different users with varying levels of access to it and presenting it in such a way that they can highlight their skills and capabilities.

LA systems developed from single stand-alone environments, or which make use of data only from this restricted subset of data, will struggle to provide accurate, or even useful, insights about the skills and capabilities of the individuals that they purport to describe (Siemens, Dawson and Lynch, 2013). It is essential that LA tackle this issue of data interoperability. This project has started to address this challenge for the field of LA (see Chapter 3).

LA for the innovators

A key concern for LA is to better identify its target audience. Much of the work to date emphasises the research needed to understand the learning process for student gains. However, the majority of scalable LA solutions so far developed have been for administrators, or for those instructors who stay 'safe' within the confines of authorised university systems. These are important user groups. However, the early adopters and those staff who are likely to explore new technologies that are not yet provisioned by an institution have been largely uncatered-for. This project addressed this challenge by providing opportunities to bridge data from institutionally established tools with those that sit beyond common solutions. Providing new ways to offer LA 'in the wild', or beyond the LMS, to educational innovators will help to facilitate the development of new approaches and techniques for using data and analytics in ways that are more student-centric.

Ethics: privacy, access, ownership and student-facing analytics

Political changes in Europe as it adopts the General Data Protection Regulation[8] (GDPR) require anyone using the personal data of Europeans (which can include European students in Australian universities) to proactively develop responses to the *Rights of the Data subject* that are described in detail in Chapter 3 of the GDPR regulations[9]. Fortunately, LA has a rich subfield that has considered the issues related to ethics, privacy and consent (Ferguson, 2012; Slade and Prinsloo, 2013; Pardo and Siemens, 2014; Drachsler and Greller, 2016; Prinsloo and Slade 2016). However, less attention has been paid to issues related to student *access* to their LA data, and how data *portability* can be guaranteed. This project tackled these highly problematic issues through two related strands of work:

1. Investigating ways student-facing LA can be incorporated into pedagogical practice to provide learners with *data access* (see Chapter 2).
2. Developing the underlying mechanisms necessary to ensure data interoperability across multiple systems to enable *data portability* (see Chapter 3).

A technical problem: Educational Data Interoperability

In the first year of this project, EDUCAUSE released a position paper (Brown, Dehoney and Millichap, 2015) on the Next Generation Digital Learning Environment (NGDLE). Although the premise of the paper was not a new concept to EdTech (a point that has been noted by a number of other people in the field[10]) the article served to highlight the need for a learning ecosystem that adheres to a common set of standards. The underlying technological infrastructure required to develop such a flexible learning ecosystem had not until that point been well implemented by the standards community. With the release of standards like Learning Tools Interoperability (LTI)[11], Open Badges[12] and various educational data standards, this had changed, providing the chance to construct what Feldstein (2015) termed a Learning Management Operating System (LMOS).

Importantly, this project commenced at an unusual point in time, where one new educational data standard (xAPI) had been very recently published and another (IMS Caliper) was in development. Accordingly, a core aim of the project was to investigate how the data interoperability enabled by xAPI might be leveraged to deliver LA beyond the LMS.

[8] https://www.eugdpr.org/

[9] https://gdpr-info.eu/chapter-3/

[10] See for example posts by Feldstein (https://mfeldstein.com/the-educause-ngdle-and-an-api-of-ones-own/) and the response by Ackerman and Dolphin (2017).

[11] https://www.imsglobal.org/activity/learning-tools-interoperability

[12] https://openbadges.org/

Experience API (xAPI)

The Experience API (xAPI) specification[13] provides a platform-neutral method to collect events occurring in any learning experience. It was released in 2013 as the outcome of an Advanced Distributed Learning[14] (ADL) project that aimed to improve interoperability between e-learning systems that collect and exchange student learning data; and overcome the limitations of SCORM.

Statements are very flexibly defined in xAPI as `actor`, `verb`, `object` triplets, with a syntax defined in an open source technical specification[15] which is freely available on GitHub and developed by the xAPI community. This makes it very easy to create an xAPI Learning Record Provider (LRP), which sends xAPI statements to a Learning Record Store (LRS), for later analysis or reporting. Chapter 3 provides more details about this process, as well as the limitations that were exposed with the xAPI specification during the life of this project. Of particular note, the xAPI specification does not require that terms be used to describe particular `verbs` or `objects` interoperably. This means that anyone implementing a Learning Record Provider (LRP) can generate their own semantics, which can be both a strength, (it is very quick to implement) and a weakness, (there is no guarantee that xAPI statements produced in one environment will make sense in another one). However, the xAPI specification is a transparent community effort, and so has moved a long way during the life of this project, which has played a key part in this evolution. Despite the strengths and openness of xAPI, in 2015 a second competing educational data standard was released.

IMS Caliper

In 2015 IMS publicly released the Caliper specification.[16] IMS Caliper provided an alternative educational data standard. Similarly to xAPI, the Caliper specification creates triples about an `actor`, performing an `action`, on an `object`.

IMS Caliper is open source, but the IMS development model is closed. To influence the Caliper specification it is necessary to sign up for IMS Contributing status, which incurs a significant financial cost. Given the costs involved, the development of this specification has been driven by EdTech vendors and has been tightly controlled. Akin to the approach adopted by the xAPI community, there are strengths and weaknesses associated with IMS Consortium. On the one hand, there is complete alignment on the data models implemented by any vendor implementing IMS Caliper. On the other hand, the data model itself is very tightly aligned with traditional learning ecosystems, and does not readily extend to learning 'in the wild'.

---

[13] https://www.adlnet.gov/research/performance-tracking-analysis/experience-api/

[14] https://www.adlnet.gov/

[15] https://github.com/adlnet/xAPI-Spec

[16] https://www.imsglobal.org/activity/caliper

Two educational data standards

Two educational data standards is a poor outcome for those seeking to implement data interoperability for lifelong learning. It should not matter which environments a student encounters during their educational journey – the data should be implementation-neutral. A likely area of high future impact for the project arises from the solutions that it has provided for LA to maintain educational data interoperability despite the suboptimal situation of the emergence of two standards (see Chapter 4).

Chapter 2 – the Connected Learning Analytics (CLA) toolkit

A key deliverable of this project is the Connected Learning Analytics (CLA) toolkit. The first version of the CLA toolkit is available at https://github.com/kirstykitto/CLAtoolkit. This first attempt at providing LA beyond the LMS has been released under an open source GPL3.0 licence and was used in the trials discussed in this chapter. As an understanding of the problems associated with delivering this tool grew, we identified a need to rewrite the codebase to achieve a more robust and scalable solution. A major rewrite is ongoing, and the codebase for the emerging CLA toolkit V2 was released in December 2018 under an Apache license in a new repository (https://github.com/uts-cic/CLAtoolkitv2).

The CLA toolkit is designed to interface with a variety of social media sources using the APIs that they make available to extract data about a student's participation in defined learning activities (see Figure 1).

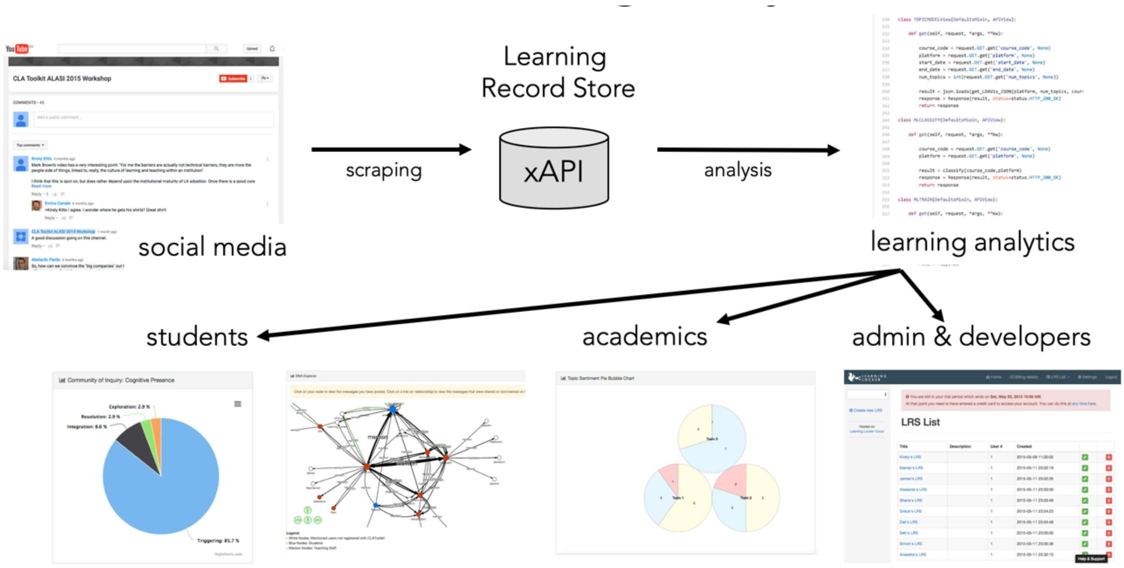


Figure 1: A schematic of the CLA toolkit design.

Privacy

The CLA toolkit is designed to respect student privacy and data control (Pardo and Siemens, 2014). The design emerges from two issues that arise when attempting to incorporate student social media data into a LA system.

1. Universities rarely have access to the aliases associated with a students' social media accounts. While some systems enable students to link social media accounts to various university systems, it is rare to see students voluntarily making use of these

tools. Indeed, there are few motivating reasons to do so, and many reasons why a student may find themselves in trouble if they do link accounts.[17]

2. Student social media accounts contain information that is both relevant and irrelevant to their learning. Distinguishing between the two is a difficult process. Social media can be a platform for knowledge sharing, networking, and developing a modern day professional portfolio can prove highly beneficial when seeking employment. However, the same platforms are also used interchangeably as a place for sharing personal details with friends, or expressing opinions and ideas.

The CLA toolkit responds to these issues by giving students control over the social media accounts they choose to link, and then only collects data for specified activities in those social media environments. It is designed to not only preserve student privacy (as students can opt in and out of any data collection), but also enables student *control* over what subset of social media data is collected by an institution. Furthermore, this tool provides students with *access* to their learning data, thereby helping them to explore the data they generate in social media environments and as a consequence, promotes awareness of, and improves data literacy.

Table 1: The list of social media environments from which the CLA toolkit is capable of collecting data, along with the learning activity, and any restrictions on data collection.

SM platform	Activity	Data restrictions
Twitter	Twitter chats using specified channels defined using nominated hashtags.	Only tweets made by signed-up users and containing the nominated hashtags are collected.
Slack	Group communication using a specified channel.	Only posts made by signed-up users and in the nominated channels is stored.
Facebook	Group communication using a specified private group.	Only posts made by signed-up users and in the nominated group are collected.
WordPress	Blog posting and commenting.	Only posts and comments from signed-up users and in linked blog sites are collected.
GitHub	Commits to codebases, comments, issues, and project management.	Only signed-up user activity in a specified repo is stored.
Trello	Agile project management via a nominated board.	Only activity in a specified board is recorded.

[17] See https://www.edweek.org/ew/articles/2017/06/29/10-social-media-controversies-that-landed-students-in-trouble.html and https://ncac.org/watch-what-you-tweet-schools-censorship-and-social-media for examples of situations where students have faced significant penalties for acts in various social media.

The CLA toolkit was designed to be extensible, and capable of collecting data from a variety of different social media environments (see Table 1). Of note, even when a student links a set of their social media accounts with their CLA toolkit account, it does not collect all of the given student's social media data: only data relevant to specific learning activities is stored, and only if the student has linked that social media account.

Requirements elicitation for the CLA toolkit

Two workshops were run with academics known for teaching 'in the wild' during 2015, the first at UTS and the second at QUT. During this period the team developed an understanding of the types of classes that were teaching 'in the wild' at the two institutions and what types of analytics might prove useful for these situations.

This phase highlighted one of the ongoing problems with eliciting requirements for LA, namely, the general lack of data literacy and algorithmic awareness among staff and students. Staff and students generally lack knowledge of the range of analytical possibilities that are available to them, and so tend to default to asking for analytics that are familiar. Thus, this project has found in numerous elicitation workshops that students usually request a leaderboard if given a voice during the requirements elicitation process. However, this is an analytic that has been shown on multiple occasions to provide no benefit to learning (Khan and Pardo, 2016) and in some cases to actively harm student outcomes (Hanus and Fox, 2015). Care must be taken in responding to stakeholder requests.

The team developed an elicitation process to alleviate this problem. Both workshops followed a structure where a presentation about 'what is possible' in LA helped end users to become aware of the type of analytics that *can* be delivered and prime them towards proposing more creative uses of LA and how it might be used in a student-facing context. A series of paper-based mock-ups were designed by workshop participants and informed the design of a set of prototype dashboards.

Learning Analytics Dashboards

Dashboards for both students and instructors (Bakharia et al., 2016) were made available for use in trials at the end of 2015 (Kitto et al., 2016). Rather than following a careful design-based approach to deliver these dashboards, it was decided to rapidly generate something that could be used in student trials, to help provide insights about the way students would use LA dashboards in class-based scenarios. The dashboards used in Case Study 1 are discussed in Appendix D.

Case Study: LA beyond the LMS (QUT)

The requirements elicitation process unearthed a number of engaged and interested academics working at UTS and QUT. At QUT, one highly motivated academic was identified as a perfect match for initial trials of the CLA toolkit, even in its very preliminary form. Dr

Kate Davis was teaching well beyond the LMS, offering no class support within Blackboard (the QUT LMS). Her courses were run in an inquiry-based and connected learning mode that was designed to help students become practising members of the Library and Information Sciences community. Most of her courses made use of the following two tools:

- A WordPress multisite structure for delivery of subject materials and weekly blogging activities (see http://2016.informationprograms.info/ for an example site).
- Twitter, where the class would participate in weekly chats using defined hashtags.

Dr Davis is a passionate advocate for data and analytics but lacked analytics support in this area, as both teaching materials and instructor/student participation operated outside the LMS. As someone who was widely recognised at QUT as a highly innovative lecturer, Dr Davis was a prime candidate for developing new ways in which LA could be used in the classroom in a student-facing context. For this reason, an ongoing sequence of trials was run in her class over three consecutive teaching periods. More details about this ongoing work can be found in Kitto, Lupton, Davis and Waters (2017). All trials were covered by QUT's Office of Research Ethics and Integrity (OREI), approval number 1500000398.

Trial 1: Look and see (Semester 2 of 2015)

The first trial was conducted with a Masters subject, IFN614: Information Programs, in Semester 2 of 2015. Students were invited to sign up in Week 8 of the subject and explore the student-facing dashboards, using the reports as a process for reflection about their contributions to the subject. Recruitment focused on piquing students' curiosity and played on their interests in data and classification. Of the 34 students enrolled in the class, 12 signed up for the trial, and a few made use of the reports in the CLA toolkit for writing their weekly blog posts. No discernible effect was found upon participants' learning. Although the level of engagement and application can be seen as disappointing, the results led to one of the core insights of this project; that any student-facing LA report must be effectively integrated with the course learning design or assessment. Failure to interweave LA reports and learning design is likely to result in poor uptake, ineffective interpretation, and minimal impact on student learning. Care must be taken to design student-facing LA in such a way that a student can use it to generate insights about a matter that directly relates to them. While it seems plausible that some students should respond to a hook based upon the quantified self and basic curiosity, we are yet to see any studies in the LA field that demonstrate an improvement in student learning outcomes when this approach is adopted. For this reason, Trial 2 made a more deliberate attempt to link the student-facing dashboards to the underlying LD that was being used in the subject.

Finding 1

To be effective, LA dashboards must be tightly coupled to the learning design and/or assessment regime of a subject. User interpretation of LA results is contingent on the pedagogical context.

Trial 2: Designing for student-facing LA (Semester 1 of 2016)

Trial 2 was run in the next teaching period (Semester 1, 2016) for an undergraduate subject, IAB260: Social Technologies. In this case a far tighter integration with the LD of the relevant subject was attempted, using a *do–analyse–change–reflect* learning design pattern designed to assist with the integration of student-facing LA into the delivery of a subject.

1. **Do:** A learning activity is undertaken which generates a data trace. In the case of IAB260 this was a weekly blogging activity that started from Week 1.
2. **Analyse:** The data trace is collected and analysed then returned to the students as feedback for consideration about their participation in the learning activity. In IAB260 students were encouraged to sign up to the CLA toolkit in Week 5, during a face-to-face dashboard elicitation workshop.
3. **Change:** Students are shown reports built on the data trace and prompted to work towards understanding what the LA says about their behaviour, changing their behaviour if they see room for improvement. In IAB260 students were told during the Week 5 workshop that they would need to write an assessed reflection about their contribution to their learning community at the end of the semester.
4. **Reflect:** A final reflective task is used to motivate engagement with the *change* phase. In IAB260, students were provided with a reflective prompt asking them to discuss their contribution to their learning community throughout the semester (see Kitto et al. (2016; 2017) for more details). They could either make use of the LA provided in the CLA toolkit to justify their claims or gather the evidence required to mount a strong argument in their own way.

From a total of 64 students enrolled in the class, 24 signed up for the CLA toolkit by the time the reflective prompt was due. This was a low rate, likely due to generally low participation of the cohort within the entire class. However, 17 students used analytics from the CLA toolkit in their final reflections. A qualitative analysis of the posts submitted by students showed that the majority of students who made use of the tool did so in an appropriate manner, with some performing a sophisticated analysis (Kitto et al., 2017).

Finding 2

Students can make use of LA dashboards and reports to develop an understanding of their approaches to learning beyond the LMS.

The LD patterns developed during this trial were presented in a paper at ASCILITE (Kitto et al., 2016), winning the best full paper award, which led to an invitation to submit a longer form journal paper (Kitto et al., 2017). While Trial 2 provided a sound indication that student-facing LA may lead to improved student outcomes, the sample size of users was small and the class suffered from a generally low level of student engagement. At the same time, the trial highlighted problems associated with the way the dashboards were being rendered and with the sign-up process for the CLA toolkit itself (see below for more details). A further trial was implemented to gather more data about how the tool could be improved.

Trial 3: Consolidation and extension (Semester 2 of 2016)

A final trial was run in the next teaching period (Semester 2, 2016) with the same Masters-level subject that had formed the test bed for Trial 1 (IFN614). This time the student-facing LA delivered by the CLA toolkit was tightly integrated into LD of the subject, following the same do–analyse–change–reflect cycle, but with an additional predict–compare step. The blogging assessment was introduced in Week 1, and a Week 2 guest lecture by the project team introduced students to LA, the CLA toolkit and the Community of Inquiry (CoI) online engagement model (Garrison, Anderson and Archer, 2001). At this point students were required to write a blog post setting goals for the role and activity patterns that they would like to exhibit during the class (*predict*). The standard *do–analyse–change* cycle then ran throughout the subject, before a final *reflect* phase was run in Week 14, when students critically evaluated their engagement in relation to the goals set in Week 2 (*compare*).

Of the 40 students enrolled in the unit, 23 eventually signed up (i.e. the participation rate increased to over half the class), which gives us reason to suppose that the tighter coupling of the LA to the assessment task resulted in a stronger participation rate. Eleven students drew on the CLA toolkit in their Week 14 post (almost 50% of the trial participants).

The most sophisticated reflection looked back at the aims for Week 2 and the actual pattern of behaviour over the semester in terms of the CoI model:

> *In Week 2 I was very aspirational about the role I wanted to play; 'I would like my profile to be professional, respectful, organised, connected and visible. I aim to be an active participant within 'reflection and critical discourse that is the core dynamic of a community of inquiry'. I achieved my aim of being an active participant as I made over 75 comments on my peers' posts, averaging over 5 per week.* ***However I feel I did not participate fully in all 4 phases of the cognitive presence in the Practical [sic] Inquiry Model; triggering event, exploration, integration and resolution – despite having sentence openers taped next to my computer****! Triggering events and some exploration were met by sharing an interesting article relevant to a post I had read and also asking some questions, but I felt a lot of my posts were agreeing with and complimenting upon the erudite musings of my peers. I was definitely wary of confronting differing ideas and promoting a critical discourse.* ***This participation in all cognitive phases needs improving*** *so the sentence openers will remain up!* [our emphasis]

While this post demonstrates that student-facing LA can be used to help students reflect upon and change their approaches to learning, it also highlights some core difficulties in implementing student-facing LA that relies upon machine learning (ML). In particular, these students had been very actively warned that the cognitive presence dashboard being displayed may not be accurate, and that they should participate in the AL^2 activity to think about how their posts were being classified (and potentially misclassified) by the algorithm. As this class consisted of a strong and engaged Masters cohort, we wanted to test the capability of students to challenge ML algorithms that they felt were inaccurate. However, no students participated in the AL^2 activity, and we see in the above post that even the best students believed a report that they had been warned was likely to be inaccurate. While a tighter integration of the AL^2 activity with the LD of a subject can work to improve this problem, it drew attention to the dangers associated with rolling out functionality within the dashboard without a careful consideration of the potential for students to simply accept the LA that it would report, rather than using the LA as a prompt for critical reflection.

Finding 3

Student awareness and training to interrogate data is required. Students were likely to readily accept the LA reports without question.

The sequence of trials highlighted where more support was required to aid student data interpretation. Furthermore, the dashboards generated generally failed to adapt to the context of the course and understanding its learning design: one size does not fit all in LA (Teasley, 2017). For example, student confusion resulted from:

- **The sign-up page**, which enabled students to link all social media accounts for which the CLA toolkit could collect data. This often meant that students would attempt to link accounts that were not of use to the class, which sometimes caused errors.
- **The student-facing dashboard** (see Figure 5) often contained reports that were not relevant to a specific class design. For example, SNA is not useful for a class that is not making use of some form of social learning tool, and a number of content analysis reports were not valid for some contexts in which they were available, but were nonetheless still considered by some students during trials.

The trials revealed a key priority for a dashboard redesign. In short, the design of the reports should be more oriented towards a user-configurable solution, where, for example, instructors could turn reports on or off depending upon the specifics of the LD being implemented in their class. To a certain extent this problem is mitigated by coupling the dashboards with the LD. Chapter 4 returns to this problem and the solution now being developed at UTS.

Embracing imperfection in learning analytics

Many of the insights gained from this phase of the project are distilled in Kitto, Buckingham Shum and Gibson (2018). Extending Finding 3, that paper argues that we should seek to use data, analytics and AI to equip learners with higher order skills and dispositions that underpin citizenship, employability and lifelong learning, such as creativity, critical thinking, agency, curiosity and an ability to tolerate uncertainty. These qualities are typically assessed in *authentic contexts*, often with complex psychological and social dimensions. The generation of analytics in such complex domains will *in principle* have a high degree of imperfection. Nonetheless, this should not deter us from developing automated feedback to provoke productive reflection by learners: (1) about their own progress, or that of their team, or (2) the trustworthiness of the technical infrastructure generating this feedback, cultivating mindful engagement with analytics/AI. This project provided an initial demonstration of how these principles can be actioned in student-facing LA, an important set of results that is now being extended at UTS.

Algorithmic accountability

An outcome from the project relates to a topic of growing concern in recent years. Triggered by the high-profile media coverage of events such as Edward Snowden and Cambridge Analytica, over the project's duration we have seen a shift in societal attitudes to data-gathering at scale, privacy, and algorithmically driven decision-making. An insight drawn from this project has been the need to articulate more clearly what it means for a learning analytics infrastructure to be trustworthy, by identifying the accountability relations that hold between the different stakeholders whose expertise contributes to a design (Buckingham Shum, 2016; 2018).

Ongoing issues with the approach

The case study trials yielded a rich set of user-centric data that informed the project directions. A redesign was necessary to act upon the insights gained. In 2017, CI Kitto relocated from QUT to UTS and additional internal funds were secured to rethink the architecture. The extra UTS funding has enabled a reconsideration of a number of issues raised through the case studies and ongoing project work. These other issues included:

- The volatility of social media, where companies like Facebook, Twitter, Google etc. constantly update APIs, made maintenance difficult. This issue was worsened by the tightly integrated nature of the codebase. **A more modular solution was necessary.**
- Much of the LA initially implemented for V1 was tightly integrated with data scraping. This led to data structures frequently breaking as new social media platforms were integrated. **A more flexible data structure was necessary.**
- The codebase of the CLA toolkit V1 was declared GPL3 in an attempt to conform to the OLT creative commons licensing expectations. However this restricts the potential end user base of the codebase. **A more open codebase was required.**
- Instructors who teach beyond the LMS are very wedded to their (normally unique) technical solutions and learning designs. It became increasingly apparent throughout the project that more flexibility had to be provided for instructors to build up their own reports. Ideally these would couple to sound learning designs, enabling instructors to explore sensible LA 'out of the box' while building up their LA literacy,

and then reconfigure established solutions as they gained confidence. **Configurable LA reports that are easy to link to LD were necessary for wider take-up.**

These problems point to the final finding of this chapter:

Finding 4

Tightly coupled LA architectures are difficult to maintain and prone to quick obsolescence.

Thus, this first version of the CLA toolkit was too tightly coupled to be robust and reusable in all of the contexts in which it might be needed. A new codebase was required to resolve the issues that arose from the tight coupling in this initial prototype. Coupling this finding with the work on data interoperability that will be discussed in Chapter 3 has led to an outcome that is most likely to be of broad global impact beyond the life of this project, the Learning Analytics API (LA-API), which will be discussed in Chapter 4.

The CLA toolkit V2

A second implementation of the CLA toolkit is currently being developed at UTS. The codebase has been released under an Apache licence and the codebase is available at: https://github.com/uts-cic/CLAtoolkitv2. This version follows the basic design as presented in Figure 1, while delivering a more robust access and control mechanism that is core to the design of the tool. A sequence of highly modular sub-programs work to:

1. Allow any user to create a 'classroom' as an instructor
2. Enable other users to sign up to a nominated classroom as students
3. Import the relevant activity data through to a call to a function that collects data
4. Transform the data to a xAPI statement that satisfies the new Profile specification
5. Send the data to an LRS.

Importantly, this solution is making use of GraphQL[18] to import social media data. This helps to modularise the data import and in many cases enables a far more direct integration with social media tools (many of which are now also using GraphQL to reveal user data). It is anticipated that this new approach will lead to more stable data collection, as GraphQL is designed precisely to help maintain a stable interface to datasets despite ongoing development of APIs and user-facing applications. This same principle is being adopted in the LA-API that is now being developed at UTS (see Chapter 4).

Note that no analytics is built into the CLA toolkit V2. Instead, this solution concentrates on collecting social media data and storing it in an interoperable data format for later analytics solutions to use. A second set of tools is being built that is designed to make use of this data and deliver LA to various users. We shall return to this discussion in Chapter 4.

[18] https://facebook.github.io/graphql/

Chapter 3 – Data Interoperability

As one of the earliest Learning Record Providers (LRPs) of xAPI statements that arose in the LA community, the CLA toolkit has helped to generate an understanding of how this data format works, what its strengths and limitations are, and how it might be improved. Indeed, this project has had considerable influence in pushing the boundaries of xAPI and providing a strong use case for why the data interoperability that it delivers is important, which has led to further developments of the specification itself.

The strengths and weaknesses of xAPI

xAPI provides a highly flexible and easy-to-use data format. The base specification (see https://github.com/adlnet/xAPI-Spec) makes it simple for people building LRPs like the CLA toolkit to start sending xAPI statements almost immediately. The base xAPI specification has no defined core vocabulary and a very minimal semantics. Statements made up of `actor`, `verb` and `object` triplets are legal. While each verb and object in an xAPI statement requires a unique identifier that resolves to a URL containing metadata about that concept, very little extra information is required.

This means that the basic syntax of xAPI contains almost no information that is useful for LA, which generally relies on information like timestamps (to generate activity timelines), class details, and free text. While xAPI *can* be used as the underlying data format in various LA applications, care must be taken to ensure that the data captured contains the information necessary for building those applications. The specification itself does not ensure this, which has led to problems for many groups of early adopters, who tended to quickly construct LRPs that generate legal statements, only to discover later on that the data collected in this manner enables very little useful LA. Furthermore, the Learning Record Stores (LRSs) which accept and store xAPI statements according to the specification provide little more functionality than that of a very large and static log file. They rarely provide either full access to the necessary data, or the full reporting capability required by LA groups, and so data often needs to be extracted to a database in order to develop LA solutions.

At the time that this project commenced the specification required communities to work to define and share the structure of xAPI statements and their associated vocabulary as *recipes* specific to a domain. These recipes are analogous to the semantic definitions included in ontologies. Without them, xAPI only provides a set of very loose syntactic rules to compose statements. This is both a core strength of the specification and a significant weakness.

On the positive side, it makes xAPI a highly flexible data format to work with. Statements can be constructed very quickly which adhere to the specification (in contrast to the IMS Caliper specification, which takes a significant effort to implement). This allows for the rapid creation of tools that emit xAPI statements. However, this flexibility of the base specification says very little about the required *semantics* of an xAPI statement, making it impossible to guarantee data interoperability between statements generated by different LRPs.

One of the key global impacts of this project revolves around its influence in driving the xAPI specification towards a recognition of the types of data that are required for using xAPI in LA applications, and in providing sound use cases for extending xAPI with the recently released xAPI Profile specification (https://github.com/adlnet/xapi-profiles).

The Connected Learning Recipe

In the first few months of this project an early paper (Kitto et al., 2015) about using xAPI to combine data from various social media environments was presented at LAK'15. This paper discussed the need for a careful consideration of the vocabulary used in constructing xAPI statements, demonstrating the way in which a wide variety of social media environments overlap in their functionality. Thus, a *tweet* is very similar to a Facebook *post*; the two are semantically related. This means that if care is taken to map the two actions to one unifying vocabulary then it is possible to generate LA that applies to either platform (e.g. Twitter) or to a combination of the two (e.g. Twitter and Facebook). This can lead to far more possibilities in the LA system that is built on these data structures, a feature that was exploited by the CLA toolkit.

The requirement in this project to collect xAPI statements over various social media tools quickly led to the recognition that this mapping would be very important. This was not a widely recognised point at the time, with most LRPs tending to construct statements for their own systems with little thought for their eventual data interoperability. This foundational work was followed by the release of an open source repository on GitHub describing the Connected Learning (CL) Recipe (https://github.com/kirstykitto/CLRecipe). A paper describing the process of developing a recipe that built off this original mapping was presented at LAK'16 (Bakharia et al., 2016).

The CL Recipe had substantial impact in the xAPI community, driving a broader understanding of the need for more complex data structures in the data format. Core to its influence was the way in which the project considered the end use case (providing dashboards to students that would help them to understand their learning) working backwards to consider what these requirements mean for data capture. It is more common for the data to be captured and analysed before the question of how it could be presented to an end user is considered. This leads to data and reports that are not educationally meaningful, termed the 'clicks to constructs' problem during the project.

Interlude: Driving a truce between IMS and ADL

LAK'16, held in Edinburgh, was a turning point for educational data interoperability, and this project played a key influencing role during that event. The Jisc learning analytics project, which has been implemented in xAPI, and an xAPI Camp held in London just before LAK, meant that a large number of xAPI experts attended the LAK'16 hackathon. The dissatisfaction in the LA community with two educational data standards (IMS Caliper and xAPI) resulted in the project lead (Kitto) proposing a hackathon challenge to attempt a mapping between Caliper and xAPI to construct a 'Caliper Recipe'. Unfortunately, the opportunity was lost, as no Caliper experts attended the hackathon (until Anthony Whyte came to present some information about this very new specification in the last session). The closed nature the IMS development model made it impossible to investigate how this mapping could be performed. Frustration at the difficulty in making progress spilled over into the *Edinburgh Statement on data interoperability* (https://github.com/AlanMarkBerg/hack-at-lack16/blob/master/TheEdinburghStatement-Signed.pdf), which was signed by over 50 LA practitioners, vendors and solutions developers. This statement was a key motivator in influencing ADL and IMS to come together and start the mapping between their specifications (see e.g. https://www.adlnet.gov/adl-experience-api-and-ims-caliper-discovery-review/).

This project has also contributed to the ongoing effort regarding Learning Analytics Interoperability, with Kitto providing a guest blog about the topic, contributing expert opinion to the EU-funded Learning Analytics Community Exchange (LACE) subproject on the topic, (Griffiths, Hoel and Cooper, 2016), and attending various ISO/IEC JTC1 SC36 WG8 meetings, (in Edinburgh, Korea and Sydney), to provide expertise from the xAPI community that has fed into the creation of the Learning Analytics Interoperability Standard prepared by that working group. More details are provided in Appendix B.

Recipes to Profiles

A project output presented at LAK'16 (Bakharia et al., 2016) highlighted the need to make recipes machine readable (proposing the use of JSON-LD), and in 2015, at the same time as this paper was being written, a working group formed to start working towards implementing the same goal, publishing a companion specification to xAPI, which after discussions and further funding from ADL eventually resulted in the 2017 release of a new xAPI profile specification (https://github.com/adlnet/xapi-profiles). xAPI Profiles provide stricter rules about how LRPs such as the CLA toolkit should behave, and mappings between related terms in ontologies that meet the requirements of the semantic web. Thus, it has become necessary to update the CL Recipe to meet this new specification, and to test its usability in the process. This work is currently underway at UTS (see Chapter 4).

Outcome: CLA toolkit to OnTask data portability (Sydney & UTS)

Data interoperability enables loose couplings between different LA applications developed by various projects. In particular, generating flows of data between different tools becomes far more straightforward, enabling loose couplings via a common data framework, rather than tight point-to-point integrations that eventually make it difficult to upgrade university infrastructure as new possibilities emerge. With an upgrade of the CLA toolkit to V2 and the emergence of the OLT funded OnTask project (https://www.ontasklearning.org/) it became possible to investigate how data portability between LA applications might be engineered in a flexible manner. Rather than a direct coupling between the two tools, a GraphQL schema is used to facilitate modularity between them, making it possible to keep developing the two independently as long as the schema is adhered to (see Figure 2).

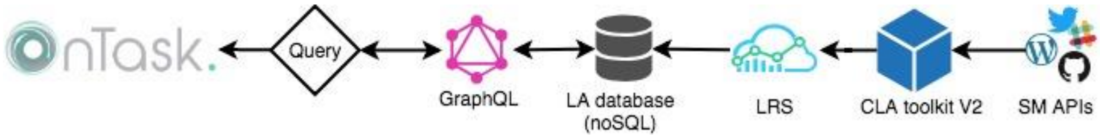


Figure 2: The data flow being implemented using the CLA toolkit to OnTask integration.

Key to this data flow is the GraphQL schema that is used to transform data from the format collected by the CLA toolkit (xAPI) to a form that can be ingested by OnTask (which takes data formatted as a spreadsheet). Chapter 4 discusses a more general framework, the LA-API which has been developed at UTS to extend this mapping.

Chapter 4 – Towards a Scalable University LA-API

The final chapter of this report discusses the work that is now proceeding at UTS to act on the lessons learned throughout the life of this project. An open source architecture is being designed and implemented, which will enable the loose coupling of multiple LA systems and the delivery of LA at scale.

While a number of different universities and collectives are starting to develop enterprise-scale LA solutions, these are commonly offered via a tightly coupled architecture. Data is collected from a student information system (SIS) and LMS, sent to a data warehouse and then used in the generation of predictive models, reports and other applications. This means that an entire LA architecture is built up that relies upon a specific solution. A point-to-point integration (where each system talks directly to the others) means that there is limited flexibility for swapping out solutions as better systems are identified. Even LA solutions built using xAPI have tended to make use of an internal database to generate LA (as did V1 of the CLA toolkit). This solution, while convenient from a perspective of 'getting things done', does not lead to long term or scalable architectures. A blog post[19] by George Kroner in 2015 helped to substantially shape the course of this project. It pointed to a number of highly innovative universities that had developed API models for communication between services and proposed that no IT systems should be directly integrated with one another. Instead, Kroner proposed that if communication between systems was implemented via a series of API endpoints, then universities could move to new solutions as the need arose.

However, LA systems provide an extra barrier to navigate in achieving this type of data architecture, as the data structures used in a storage layer can significantly affect the LA that can be provided in another layer. LA is a rapidly evolving field, which means that flexible data structures are required. In constructing LA for an API-based university we consider it necessary to provide a way of flexibly adding to the data structures served to various applications. The approach delivered by Facebook's relatively new GraphQL specification[20] provides a new way of delivering precisely this type of functionality.

Finding 5

Designers of institutionally scalable LA infrastructure should seriously consider a highly modular architecture that will enable ongoing extensions of the datasets used, modifications of reports, and new integrations.

---

[19] https://edutechnica.com/2015/06/09/flipping-the-model-the-campus-api/

[20] http://graphql.org

GraphQL is a query language that enables an abstraction of server-side API calls under a single neat wrapper instead of to multiple endpoints. This means that GraphQL provides and facilitates precisely the loose coupling that is required for building a modular and adaptable university data and analytics architecture.

UTS's user configurable dashboards

In Chapter 2 we discussed the problems associated with student misinterpretation of dashboards that were not closely aligned to a specific learning design of a class. One of the biggest shortcomings associated with V1 of the CLA toolkit was its inability to tailor dashboards to the specific requirements of a class.

Finding 6

To be useful in a student-facing context, learning analytics dashboards must be highly configurable, with different reports turned on or off depending upon: tools used; learning design; assessment regimes; and student data literacy.

UTS has devoted significant extra resources to tackling this problem of dashboard configurability, and is in the process of delivering a solution to this problem. A configurable dashboard has been designed that (i) restricts the information available to students, and (ii) enables them to choose which analytics they would like to see. Six slots (i.e. html divs) are available, which can be filled up by a user who selects from a set of available widgets and different LA reports. This restriction has been imposed in an attempt to avoid information overload; users need to think about what LA will help them to answer the questions that they are currently most interested in. All users can reconfigure their dashboards by removing a widget to add a new one, forcing a type of rationalisation. A demonstration prototype is served at http://canvasdashboard.utscic.edu.au/ but the full emerging codebase is available at https://github.com/uts-cic/dashboards-la-api. The prototype features the following functionality.

- **Instructor view:** Importantly, instructors can choose to turn off specific reports that they do not feel serve the LD of their subject (see Figure 3).
- **Student view:** Within the restrictions chosen by an instructor, students can choose up to six widgets (see Figure 4). Some widgets take more than one slot, and so the student must weigh the relative utility of using different widgets.

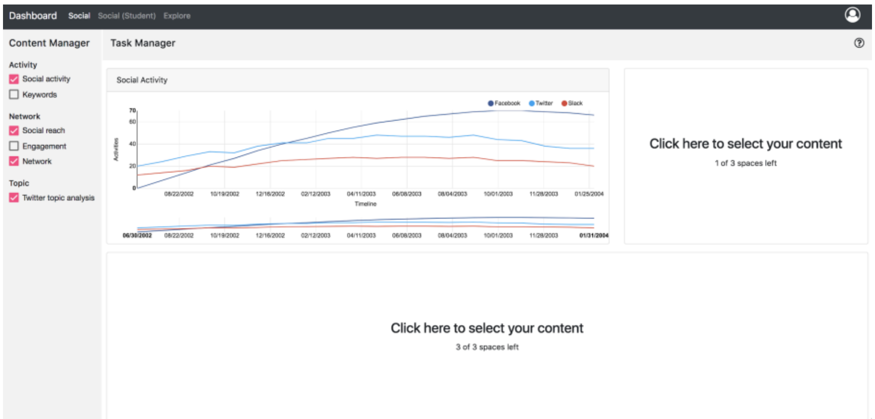


Figure 3: Instructors can choose LA that is appropriate for the LD of their class.

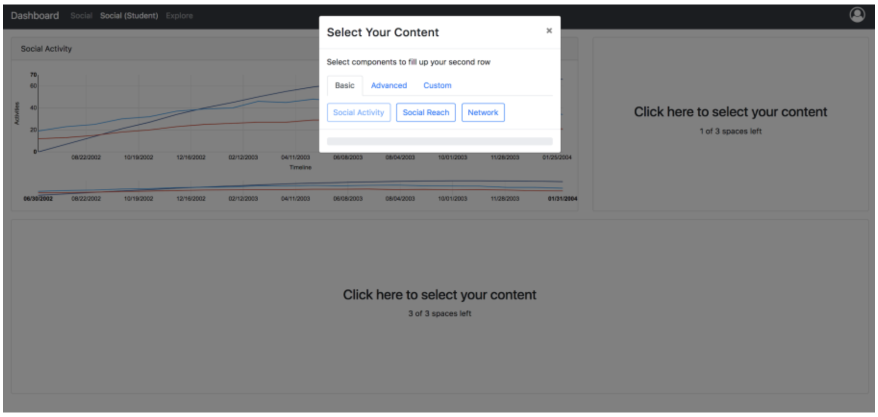


Figure 4: The student-facing dashboard enables students to choose from a set of reports enabled by the instructor.

What a student chooses to show in their dashboard, and how they interface with these configurable dashboards, will become an active area of study at UTS. The trace data that is left as students make choices about what reports to examine, and how they interpret the resulting LA, will give important insights not just about how people make sense of data, but also about how they critically interpret it and what changes in behaviour then result.

Finding 7

The data traces created as students make use of any configurable dashboards are likely to be a rich source of information about metacognition, critical thinking and self-regulated learning. They should be a priority for future LA work seeking to develop 21st century skills.

While the need for this type of user configurability became obvious within the requirements of this project, delivering learner data in the necessary format that is highly non-trivial and has led to the inception of a project outcome that is likely to have significant long-term and wide-scale impact in the learning analytics community.

Outcome: A LA-API

The new dashboards for the CLA toolkit V2 require an architecture that enables them to be not just configurable, but they also extensible, as well as robust to potential changes in code, APIs and other contributing data structures. Furthermore, not all learning occurs in 'the wild'. Many students and educators will remain in the LMS, generating data traces there. To provide students with a more complete understanding of their participation in learning activities throughout their university journey, it will be necessary to develop scalable LA infrastructure that can deliver interoperable data and LA across a multitude of environments. At UTS a solution to this problem is being developed in the form of a LA-API.

Figure 5 provides a representation of the architecture currently in development at UTS CIC. It unifies a number of initial data providers (the CLA toolkit, Instructure Canvas and Articulate storyline), each of which are exposed to various LA applications via a GraphQL gateway. The flexibility of this design is worth highlighting; as back-end data structures and capabilities evolve, the back-end data access of the GraphQL gateway can be updated without breaking the functionality of the front-end LA solutions. Thus, this solution points the way towards *pragmatic data interoperability*, where new data sources can be added and LA services extended, through the GraphQL interface as required. The LA-API thus provides a way in which to pragmatically map LA infrastructure to data that conforms to xAPI, Caliper, and other potential data formats in the future. This will provide a significant advance to the LA community as it struggles to deliver learning data that makes sense over a lifetime.

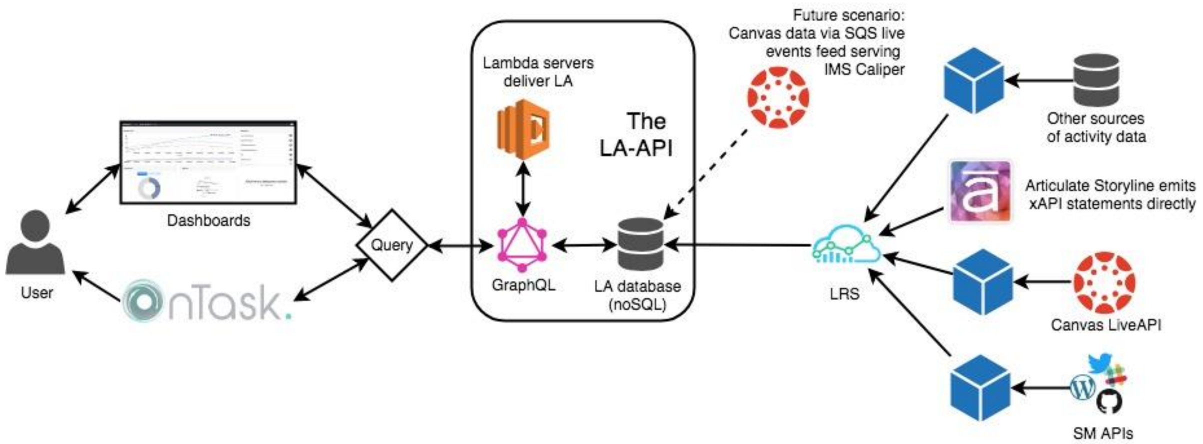


Figure 5: The data flows being implemented in the LA-API being designed at UTS.

This infrastructure consists of the following set of components:

- **An LRS data-holding capability:** A Learning Record Store (LRS) aggregates xAPI data from a gradually increasing number of LRPs. Other sources can be integrated with the LA-API architecture proposed here as it matures.
- **xAPI profiles specify data format:** A number of xAPI profiles are currently being developed to map data interoperably across the contributing LRPs. These can be extended as the list of available xAPI profiles increases.
- **A GraphQL gateway maps the back end to the front end:** A GraphQL interface is used to transform data into the format required to deliver LA to end users. This

approach enables the LRS to collect unanticipated statements that can be used with future modification of the schema. As new data becomes available its relationships can be mapped from new profiles into the GraphQL schema, and thus into the delivery of new analytics solutions.

More technical details about how this architecture can be found in the GitHub repositories where this suite of tools is currently being developed:

- https://github.com/uts-cic/CLAtoolkitv2
- https://github.com/uts-cic/canvas-extract
- https://github.com/uts-cic/lrs-mongo
- https://github.com/uts-cic/graphql-la-api
- https://github.com/uts-cic/dashboards-la-api
- https://github.com/uts-cic/xAPI-profiles

Towards lifelong personalised learning

The design of the LA-API returns us to our original challenge: lifelong personalised learning. With pragmatic data interoperability a way forward presents for delivering portable data across various educational domains at scale. As long as the GraphQL schema has been extended to map to whatever educational data is considered relevant, there will be a way of pragmatically utilising it across various institutional boundaries.

Emerging political pressures make it more urgent than ever to find substantial solutions to problems like educational data interoperability. In particular, the European General Data Protection Regulation (GDPR)[21] will have a significant impact upon how institutions make use of data and this project has provided a way in which to address the issues associated with delivering genuine data portability to our students. This is a significant advance over the current state of affairs for educational institutions and is likely to be an area of substantial future impact that has arisen from the insights developed during this project.

Summary and Conclusions

This project was inspired by the challenge of the shifting societal context for learning — how should this reframe the conceptualisation, design and implementation of LA infrastructure?

- *What are the implications of a future in which citizens need continuous training/upskilling, in formal and informal contexts, mediated via multiple platforms, whose tools and data are owned by diverse organisational entities?*
- *How will Learning Analytics help citizens to evidence their learning and competencies in such an open, 'wild' ecosystem?*

[21] https://gdpr-info.eu/art-20-gdpr/

- *How do we design Learning Analytics to cultivate learning to learn competencies, and critical engagement with analytics and AI?*

This project has investigated these challenges and made both technical and conceptual progress, forging en route new partnerships with the emerging network of practitioners and researchers who share this vision. It has prototyped elements of a flexible learning ecosystem that provides flexible LA ‘beyond the LMS’, in which sub-systems communicate via open data standards, with an emphasis on learner agency and, ultimately, on placing the learner in control of the data that they generate.

References

Ackerman, A. & Dolphin, I. (2017). Open source and the NGDLE. Educause Review. Available at: https://er.educause.edu/articles/2017/9/open-source-and-the-ngdle

Arnold, K. E., & Sclater, N. (2017). Student perceptions of their privacy in leaning analytics applications. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 66–69). ACM.

Bakharia, A. (2014). *Interactive Content Analysis: Evaluating Interactive Variants of Non-negative Matrix Factorisation and Latent Dirichlet Allocation as Qualitative Content Analysis Aids* (Doctoral dissertation, Queensland University of Technology).

Bakharia, A., Kitto, K., Pardo, A., Gašević, D., Dawson, S. (2016), Recipe for success: lessons learnt from using xAPI within the connected learning analytics toolkit. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (LAK16)* (pp. 378–382). ACM.

Brown, M., Dehoney, J., & Millichap, N. (2015). *The next generation digital learning environment. A Report on Research.* ELI Paper/Educause. Available at: https://library.educause.edu/~/media/files/library/2015/4/eli3035-pdf.pdf

Buckingham Shum, S. (2016). *Algorithmic Accountability for Learning Analytics*. Seminar, University College London, April 2016. Retrieved from: http://bit.ly/aala2016

Buckingham Shum, S. (2018). *Transitioning Education's Knowledge Infrastructure: Shaping Design or Shouting from the Touchline?* Keynote, International Conference of the Learning Sciences, London Festival of Learning (June, 2018). Retrieved from: http://bit.ly/icls2018sbs

Committee for Economic Development of Australia (CEDA). (2015). *Australia's Future Workforce?* Retrieved from: http://www.ceda.com.au/research-and-policy/policy-priorities/workforce.

Drachsler, H., & Greller, W. (2016). Privacy and analytics: it's a DELICATE issue. A checklist for trusted learning analytics. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 89–98). ACM.

Feldstein, M. (2015). *The EDUCAUSE NGDLE and an API of One's Own.* Available at: https://mfeldstein.com/the-educause-ngdle-and-an-api-of-ones-own/#identifier_1_5937

Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, *4*(5-6), 304–317.

Garrison, D. R., Anderson, T., & Archer, W. (2001). Critical thinking, cognitive presence, and computer conferencing in distance education. *American Journal of Distance Education*, *15*(1), 7–23.

Griffiths, D., Hoel, T. & Cooper, A., (2016). D7.4 *Learning Analytics Interoperability: Requirements, Specifications and Adoption*. Learning Analytics Community Exchange (LACE). Available at: http://www.laceproject.eu/deliverables/d7-4-learning-analytics-interoperability-requirements-specifications-and-adoption/

Hanus, M. D., & Fox, J. (2015). Assessing the effects of gamification in the classroom: A longitudinal study of intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers & Education, 8*, 152–161.

Khan, I., & Pardo, A. (2016). Data2U: Scalable real time student feedback in active learning environments. In *Proceedings of the International Conference on Learning Analytics and Knowledge* (pp. 249–253). ACM.

Kitto, K., Cross, S., Waters, Z., Lupton, M. (2015). Learning analytics beyond the LMS: the Connected Learning Analytics Toolkit. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge (LAK15)* (pp. 11–15). ACM.

Kitto, K., Lupton, M., Davis, K., Waters, Z. (2016). Incorporating student-facing learning analytics into pedagogical practice. In S. Barker, S. Dawson, A. Pardo, & C. Colvin (Eds.), *Show Me The Learning. Proceedings ASCILITE 2016 Adelaide* (pp. 338–347).

Kitto, K., Lupton, M., Davis, K., Waters, Z. (2017). Designing for student facing learning analytics, *Australasian Journal of Educational Technology*, 33(5), 152–168.

Kitto, K., Buckingham Shum, S., Gibson, A. (2018). Embracing imperfection in learning analytics. *In Proceedings of the 8th International Conference on Learning Analytics and Knowledge (LAK '18)* (pp. 451–460). ACM.

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, *45*(3), 438–450.

Siemens, G., Dawson, S., & Lynch, G. (2013). *Improving the quality and productivity of the higher education sector. Policy and Strategy for Systems-Level Deployment of Learning Analytics.* Society for Learning Analytics Research for the Australian Office for Learning and Teaching.

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioural Scientist*, *57*(10), 1510–1529.

Teasley, S. D. (2017). Student facing dashboards: One size fits all? *Technology, Knowledge and Learning*, *22*(3), 377–384.

World Economic Forum. (2016). *The future of jobs: Employment, skills and workforce strategy for the fourth industrial revolution*. Available at: http://www3.weforum.org/docs/WEF_Future_of_Jobs.pdf

Appendix A: Certification by Deputy Vice-Chancellor

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: ....Derbyshire.. Date: 17-01-19

Professor Suzi Derbyshire

Appendix B: Project Outputs

Publications

Kitto, K., Cross, S., Waters, Z., Lupton, M. (2015). Learning analytics beyond the LMS: the Connected Learning Analytics Toolkit. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge (LAK15)* (pp. 11–15). ACM.

Bakharia, A., Kitto, K., Pardo, A., Gašević, D., Dawson, S. (2016), Recipe for success: lessons learnt from using xAPI within the connected learning analytics toolkit. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (LAK16)* (pp. 378–382). ACM,.

Kitto, K., Bakharia, A., Lupton, M., Mallet, D., Banks, J., Bruza, P., Pardo, A., Buckingham Shum, S., Dawson, S., Gašević, D., Siemens, G., Lynch, G. (2016). The connected learning analytics toolkit. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (LAK16)* (pp. 548–549). ACM. (Demonstration)

Kitto, K., Lupton, M., Davis, K., Waters, Z. (2016). Incorporating student-facing learning analytics into pedagogical practice. In S. Barker, S. Dawson, A. Pardo, & C. Colvin (Eds.), *Show Me The Learning. Proceedings ASCILITE 2016 Adelaide* (pp. 338–347). **(Won best long paper award)**

Kitto, K., Lupton, M., Davis, K., Waters, Z. (2017). Designing for student facing learning analytics. *Australasian Journal of Educational Technology, 33*(5), 152–168.

Kitto, K., Buckingham Shum, S., Gibson, A. (2018). Embracing imperfection in learning analytics. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge (LAK '18)* (pp. 451-460). ACM.

Kitto, K., O'Hara, J., Philips, M., Gardiner, G., Ghodrati, M., Buckingham Shum, S. (2019). The connected university. In Bridgstock, R. and Tippett, N. (Eds.), *Higher Education and the Future of Graduate Employability: A Connectedness Learning Approach*, Edward Elgar Publishing.

Invited presentations and online seminars

- 16/6/2015 : Learning Analytics Toolkit & TinCan/xAPI@Work
- 19/8/2015: Connected Learning via xAPI Learning Café UnConference, Melbourne, Australia.
- 21/7/2015: The CLA toolkit, xAPI Camp, Amazon, Seattle, USA. (Online presentation).

- 11/2/2016: Data interoperability for learning analytics and lifelong learning, xAPI Camp, Autodesk, San Francisco, USA.
- 16/3/2016: Data interoperability for lifelong personalised learning, University of South Australia, Adelaide, Australia
- 22/4/2016: Data pathways for lifelong personalised learning, xAPI Camp, Jisc, London, UK.
- 10/5/2016: Can students learn from imperfect analytics? UNSW, Sydney, Australia
- 20/9/2016: Using Experience API, Learning Analytics Summer Institute (LASI) Asia, KERIS, Seoul, South Korea
- 26/9/2016: The case for imperfect learning analytics as an enabler of student metacognition. Beijing Normal University, Beijing, China.
- 31/5/2017: Student facing learning analytics, Student success and retention summit, Melbourne, Australia
- 21/6/2017: Learner facing learning analytics as an enabler of metacognition and lifelong learning. Learning Analytics Summit, Sydney, Australia
- 27/11/2017: Designing for student facing learning analytics. HERN Symposium keynote, QUT, Brisbane, Australia.
- Central China Normal University: Guest Seminars on LA tools via wolearn run in two consecutive years: 23/5/2017, 13/6/2081 and 21/5/2018, 4/6/2018
- 14/9/2018: Loose couplings and fast development: using xAPI to provide Leaning Analytics beyond the LMS. eLearning Korea, Seoul, South Korea.
- 17/11/2018: Designing for student facing learning analytics, International Forum on Educational Technology, Central China Normal University, Wuhan, China.
- 20/11/2018: Designing for student facing learning analytics, Shanghai Open University, Shanghai, China.
- 29/11/2018: Pragmatic data interoperability for learning analytics. IEEE xAPI and LA SIG. (Online presentation)
- 8/12/2018: Learning Analytics as an Intelligent Personal Assistant for Lifelong Learners. China Annual Academic Conference and International Educational IT Solutions Expo, Beijing
- June 2019: Building Learning Analytics ecosystems. Learning Analytics Summer Institute (LASI), Vancouver, Canada.

Project workshops and demonstrations

- 18/5/2015: From Data to Visualisation, UTS.
- 2/7/2015: Requirements elicitation workshop, QUT.
- 16/11/2015: Learning Analytics for the learner, HERN symposium, QUT.

- 26/11/2015: Down and dirty with data for social learning analytics. ALASI 2015.
- 27/4/2016: The connected learning analytics toolkit. LAK'16 demonstration.
- 27/11/2017: Linking learning analytics with learning design, HERN symposium, QUT.
- 23/11/2018: Using Canvas data for learning analytics! ALASI 2018.

Blogs and informal works

- LAK hackathon. The Edinburgh statement: https://github.com/AlanMarkBerg/hack-at-lack16/blob/master/TheEdinburghStatement-Signed.pdf
- Kitto, K., Why do we need interoperability anyway? Available at: http://xapiquarterly.com/2017/06/need-interoperability-anyway/
- Kitto, K., Towards a manifesto for data ownership. LACE guest blog. Available at: http://www.laceproject.eu/blog/towards-a-manifesto-for-data-ownership/

The Learning Analytics Community Exchange (LACE)

Considerable impact has been achieved via contributions to the LACE workpackage on data interoperability for LA, which has fed into the work on LA standards being driven by the ISO:

- A guest blog about lifelong data ownership and why this necessitates educational data interoperability in the lead up to LAK'16 and the data interoperability hackathon challenge: http://www.laceproject.eu/blog/towards-a-manifesto-for-data-ownership/
- The LACE report on LA data standards and interoperability featured substantial insights provided by this project: http://www.laceproject.eu/deliverables/d7-4-learning-analytics-interoperability-requirements-specifications-and-adoption/
- This work culminated with an invitation to participate in the 2016 LACE Asia tour, with visits and invited talks to both Seoul and Beijing.
- As a part of this ongoing work, the project lead (Kitto) represented the xAPI community by attending a number of ISO meetings (in Edinburgh, Korea and Sydney) for the SC38/WG8 workgroup on Learning Analytics interoperability.

The IEEE ICICLE xAPI and LA SIG

In 2017 the IEEE Industry Connections, Industry Consortium on Learning Engineering (IEEE ICICLE) formed a xAPI and LA special interest group (SIG) to extend the work on using xAPI in LA. A subgroup has convened around the CL Recipe and is working on developing xAPI Profiles for social media using the outputs this project as a starting point. An invited interview about this work was given in November 2018.

Codebase - CLA toolkit V1

- The Connected Learning Analytics Toolkit V1: https://github.com/kirstykitto/CLAtoolkit
- The Connected Learning Recipe: https://github.com/kirstykitto/CLRecipe

Codebase - CLA toolkit V2 and associated software modules

- https://github.com/uts-cic/CLAtoolkitv2
- https://github.com/uts-cic/canvas-extract
- https://github.com/uts-cic/lrs-mongo
- https://github.com/uts-cic/graphql-la-api
- https://github.com/uts-cic/dashboards-la-api
- https://github.com/uts-cic/xAPI-profiles

Table 2: Project impact at completion, and as anticipated for 12 and 24 months post completion.

	Anticipated changes (projected impact) at:		
	Project completion	Twelve months post completion	Twenty four months post completion
Team members	Team members have been recognised via promotions (3 team members), invited presentations, and moves to other universities.		
Immediate students	Pilot trials have been run with the CLA toolkit (V1) in classes at both QUT and UniSA.	Trials of the LA-API infrastructure run at UTS for 2019 with students enrolled in the Masters of Data Science and Innovation. Methods for developing data literacy being used at UTS following on from the Embracing Imperfection paradigm (Kitto et al., 2018) and user configurable dashboards.	
Spreading the word (Contributions to knowledge in the field)	7 publications have been generated during this project, including 1 journal paper, 1 book chapter, and 5 conference papers (one of which won best paper award).	Ongoing work in the emerging LA-API use case at UTS will lead to a new set of publications.	

	A number of code repositories have been produced that can be adapted by other institutions. More details about outputs and public events can be found in Appendix B.		
Narrow opportunistic adoption (at participating institutions)		Tools developed in this project are integrated into 'gold standard' examples of using student-facing LA at UTS and other participating institutions.	
Narrow systemic adoption (at participating institutions)		The LA-API will be available at UTS across all subjects using Canvas, Articulate Storyline, or via the CLA toolkit V2. Integration with other LA tools will be handled by this facility.	The LA-API is seen as valuable and used to integrate and extend LA infrastructure at other participating institutions.
Broad opportunistic adoption (at other institutions)	BeyondLMS was identified as one of the top LMS developments in 2015 (https://edutechnica.com/2015/12/06/year-in-review-top-lms-developments-of-2015/)	The CLA toolkit V2 is being used by various institutions with cohorts that want to teach beyond the LMS. The LA-API presented during the 2019 Learning Analytics Summer Institute (LASI). This leads to opportunistic development and adoption by	The LA-API is being used and extended by various universities to integrate their LA infrastructure and reuse tools developed by other groups. The user configurable dashboards being developed at UTS are being used widely beyond the institutions

		workshop participants.	that participated in this project.
Broad systemic adoption (at other institutions)	The xAPI Profile specification was influenced by use cases investigated during this project. This is now best practice for all xAPI-based Learning Record Providers. This project has influenced the development of the new ISO standard on LA interoperability (ISO/IEC JTC SC36/WG8), and the IEEE working group on LA and xAPI. The LACE work package on data interoperability for LA has made use of this project in defining requirements for data interoperability in the field.	A number of new xAPI Profiles published as outcomes of the LA-API development process. These will influence xAPI solutions at scale. xAPI, IMS Caliper and LA communities appreciate the value of the LA-API solution and starting to investigate potential uses.	Organisations such as Apereo and Jisc understand the value of the LA-API and are helping to extend it. The pragmatic data interoperability offered by the GraphQL solution being used in the LA-API is being used as a standard solution by a number of universities developing scalable LA infrastructure.

Appendix D: Dashboards in V1 of CLA toolkit

Three types of analytics were provided by default in V1 of the CLA toolkit.

1. **Activity reports** that summarise student participation in the defined learning activities from which data is collected (see Figure 6). Two reports are provided, one which looks at total activity in each platform over time, and a second that looks at the activity type (merging across all social media).
2. **Social network analysis (SNA)** provides information about who a particular student has interacted with. The report enables users to explore the patterns of connection exhibited by signed up members of the class and a set of metrics associated with nodes (see Figure 7).
3. **Content analysis** consisting of a representation of topics for the instructor-facing dashboards (see Figure 8). This report makes use of Latent Dirichlet Allocation (LDA) following a methodology developed by Bakharia (2014). The LDA report displays the text most closely associated with a user selected number of topics.

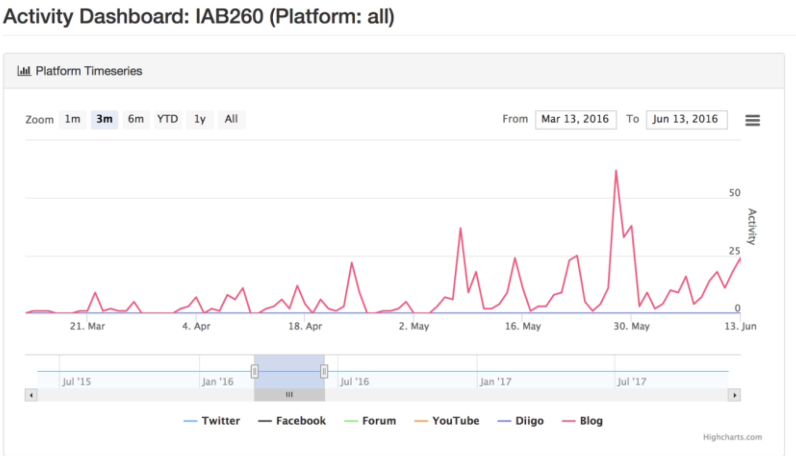


Figure 6: An example activity trace generated by the CLA toolkit for a subject.

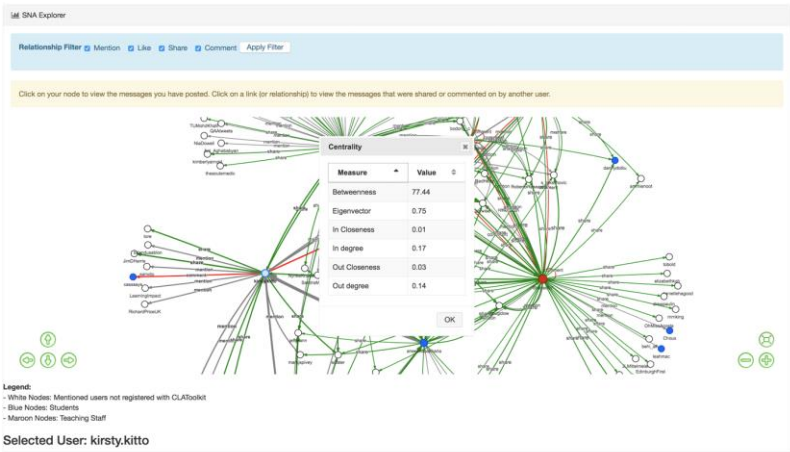


Figure 7: The SNA dashboard available in the CLA toolkit.

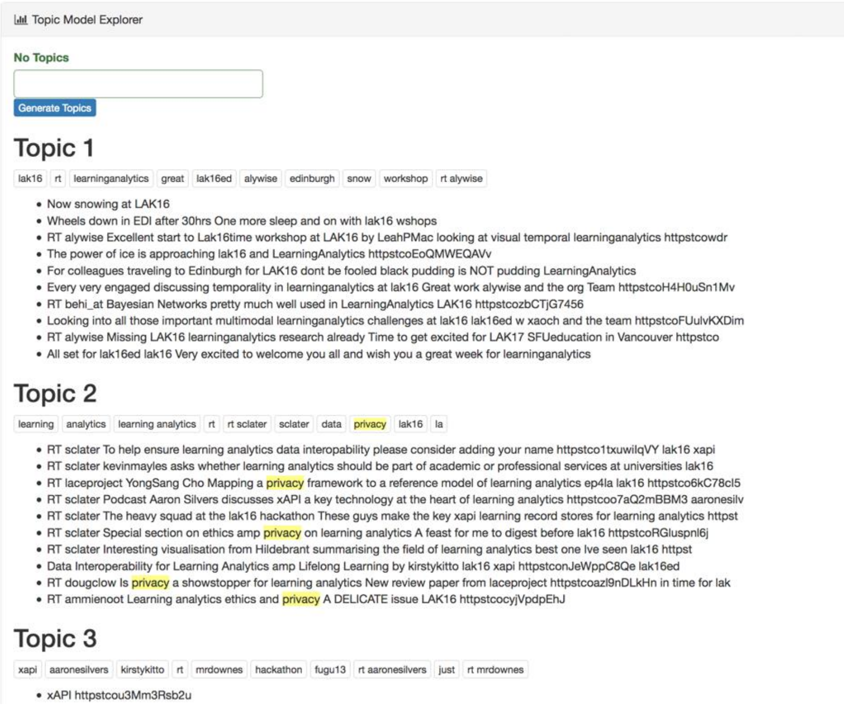


Figure 8: The content analysis dashboard provided for instructors in the CLA toolkit.

Student-facing dashboards

The main student-facing dashboard for V1 is depicted in Figure 9 and includes each of the three broad categories of analytics discussed above. This aggregated dashboard was designed with the intention of helping students to explore the nature of their online interactions during a class to promote data literacy, goal setting and awareness of learning strategies.

Two other student-facing dashboards have also been developed during this project:

4. **Active learning squared**, an activity-based dashboard which teaches students about their participation in a Community of Inquiry. It uses machine learning (ML) to classify a student's posts according to the cognitive presence construct (Garrison, et al., 2001). The student is warned that the ML algorithm is not very accurate, and instructed to reclassify the post as a different phase of cognitive presence if they consider it inappropriate (see Figure 9). As well as teaching students about their behaviour according to some specific educational construct, (the cognitive presence classifier could be swapped out for other ML solutions as they become available), this activity also teaches them about the imperfection inherent in ML approaches (Kitto, Buckingham Shum and Gibson, 2018), and that it is appropriate to challenge algorithms if they consider them wrong (Kitto, Lupton, Davis and Waters, 2017).

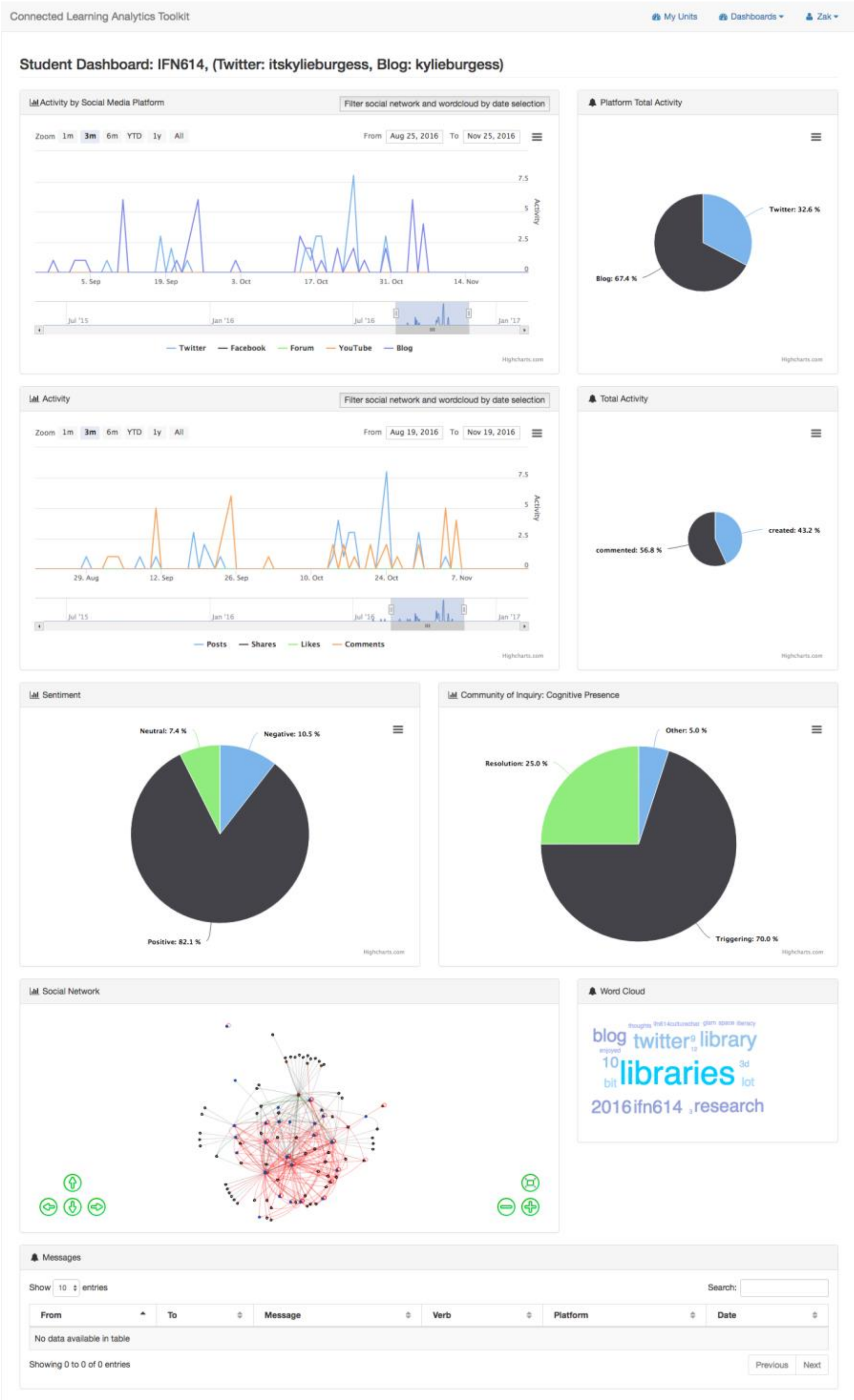
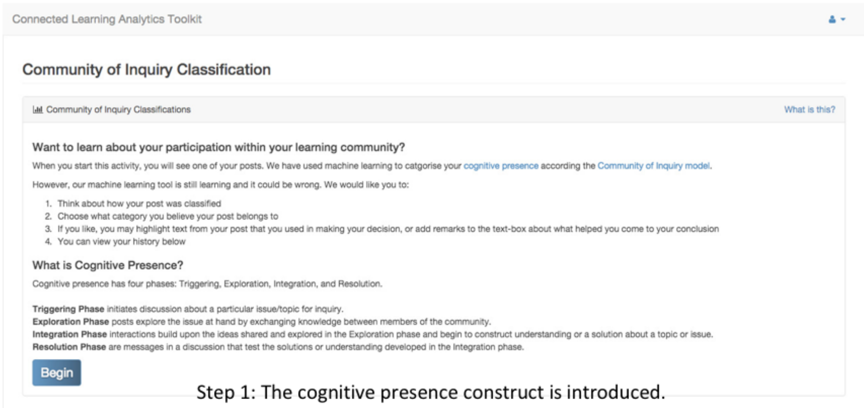


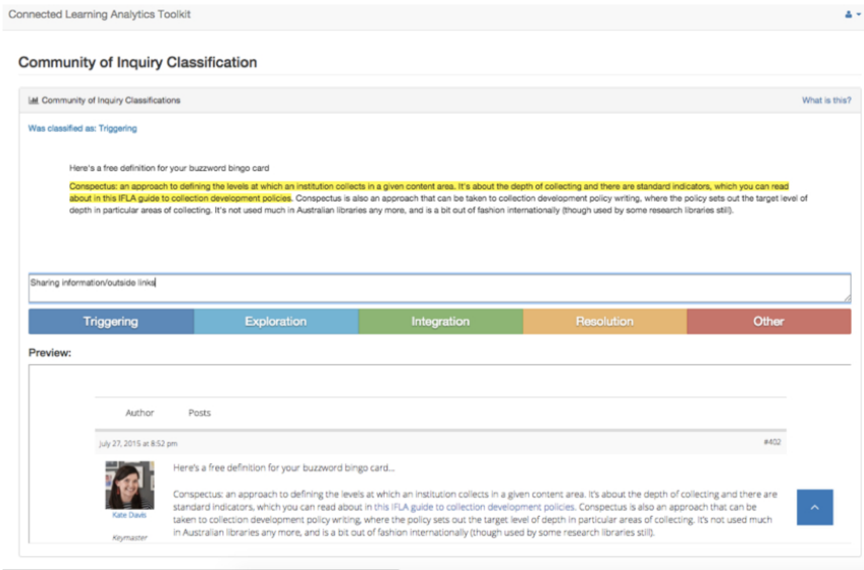
Figure 9: The student facing dashboard created for use in the CLA toolkit V1.

5. **Groupwork dashboard** encourages students to think about their relative contribution to a set of collaboratively developed content or authoring activities (see Figure 11). For every member of a team this report shows the different activities that they and their collaborators have been completing across multiple social media environments. This dashboard is intended to help students understand the different roles inherent in a group work activity, and their different activity profiles, but has not been trialled in a class context because of the potential for this report to lead to

negative group dynamics, anxiety, or more severe emotional responses in the individuals that make up that team. It remains as a proof of concept at this point.



Step 1: The cognitive presence construct is introduced.



Step 2: Student reclassifies posts.

Figure 10: The Active Learning Squared dashboard and its sequence of activity.

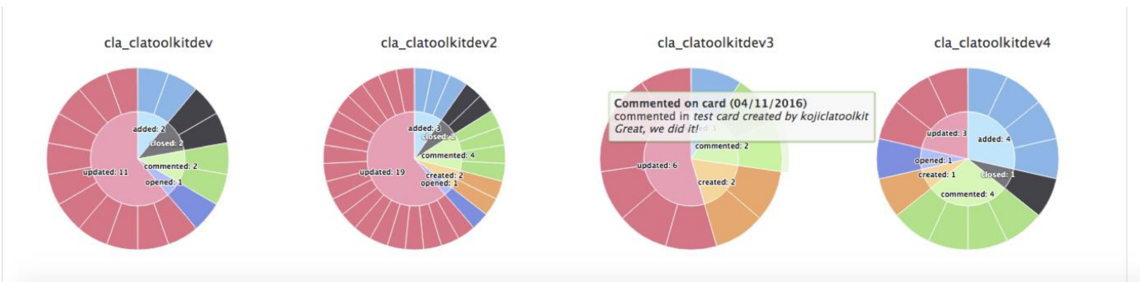


Figure 11: A group work dashboard designed to encourage students to think about their contribution to a team.

Appendix E: External evaluator report

Evaluation Final Report

Learning Analytics beyond the LMS:

Enabling connected learning via open source analytics in 'the wild' ID 14-3821

Lead institution: Queensland University of Technology

Partner institutions: University of South Australia; University of Technology, Sydney; University of Sydney

Project Original Overview

This project evolved and expanded over time. The original project aim was to improve the quality of student engagement and learning in collaborative online environments by incorporating and analysing social media platforms that the majority of students already use. A goal was to create an easy to use and open source Connected Learning Analytics (CLA) toolkit utilising the latest mathematical and computational approaches. The analytic tools delivered by this project were to work within closely delimited learning activities, underpinned by connected learning pedagogy, which will (i) preserve student privacy, (ii) enable academics and students to identify the nature and quality of student connections, and (iii) assist with developing learning analytics systems that have strong pedagogical and technical features.

The **objectives originally of this project** were to:

1. Develop data analytic techniques that can be used to characterise connected learning, drawing on novel models from mathematics and computer science to extend the field of Learning Analytics (LA).
2. Use these new LA techniques to create a Connected Learning Analytics (CLA) toolkit which can be used to both: identify the nature and quality of student engagement in online learning activities; and provide real-time reports to students and academics.
3. Develop a set of protocols for utilising CLA feedback to plan, implement and evaluate real-time interventions for improving the quality of students' collaborative engagement and online learning.
4. Test the utility of the toolkit and protocols for intervening in student learning in specified social media environments, trialling it in selected Test Studies incorporating connected learning pedagogies.
5. Deliver training to the wider Australian higher education community in the underlying philosophy of connected learning, as well as its facilitation using the CLA toolkit.

The **Deliverables** were:

1. Web-based data capture and reporting tools that can query specific artefacts generated through participation in learning activities using designated social media. The web forms will be designed for users without programming skills.
2. An open source CLA toolkit for extracting data from a specified set of social media, analysing patterns of behaviour within it, and reporting it back in an understandable format. This will function in a limited educational context, so ensuring student privacy beyond the designed learning activities.
3. Protocols to guide interpretation of the analytics in the context of the specific learning activities adopted. This includes suggested interventions to facilitate student achievement of learning outcomes.
4. Widespread dissemination, support and training to the Australian higher education sector to aid the adoption and application of the CLA toolkit, via a website, workshops and a community of practice.
5. A report summarising: project activities; toolkit usage in the learning activities designed for social media environments; outcomes of wider training; and future development opportunities.

After the initial project commencement phase the planned phases included toolkit development and testing via action research cycles and evaluation of tools by students and then dissemination of project findings and outcomes. Each action research cycle was to follow the same set of steps:

1. Select a set of connected learning activities
2. Identify the kinds of data needed to describe the nature and quality of the connected learning interactions.
3. Create tools that can run over a specified learning activity
4. Design a reporting capability that interfaces with the dashboard framework proposed
5. Test the proposed learning activities and tools, via (i) surveys and interviews that ask different user groups about their experience in using the tools (ii) analysing these responses with the data generated by those tools, and (iii) correlating both sets of data with external outcomes (e.g. assessment performance).
6. Hold a project team meeting(s) between available participants, reference group members, and the independent evaluator appointed to the project, to use the feedback obtained from users in step 5 to determine the next set of priorities for implementation and testing.

The final phase of the project was to include intensive evaluation of, and wider dissemination about, project outcomes and to finalise the development of project related documentation.

Evaluation Overview

The main purpose of the evaluation approach was to provide both formative and summative evaluation and was directly informed by the ALTC Project Evaluation Resources designed to assist projects in achieving success and to provide a judgement on the overall merit of the projects. The guiding focus of the evaluation was to determine:

- Did the project achieve its stated outcomes?
- Was the project managed and conducted in ways that contributed to project success?
- Did the project achieve as much impact as it should have?
- How could the processes associated with the project be improved and replicated?

Formative Evaluation

A critical element of project success lies in formative evaluation. Not only was this an exciting project to be involved with from the very beginning, but being included as integral team member led to a constant feedback and quality improvement cycle.

The project's formative evaluation processes included:

- Regular team meetings and ongoing monitoring of project management activities among the leadership team (project leaders and managers).
- Inclusion of evaluator in shared team space for access to notes, minutes, timelines, etc.
- Inclusion of evaluator in key reference group meetings
- Inclusion of evaluator in key project team meetings
- Inclusion of evaluator in all reference group communications
- Inclusion of evaluator in all project cluster communications
- Input into analysis of data and evidence collection from surveys and workshops
- Input into data gathered by the tools
- Regular reporting back to project leaders, project team members, reference group
- Planned and ongoing engagement with evaluator around expectations, monitoring of achievements against milestones, risk assessment and strategies for intervention.

Summative Evaluation

The original timelines for activities, deliverables and outcomes is in the table below. The project's timelines were extended not only because project staff were promoted and moved institutions but also because learnings evolved and more work was required. Additional funding for the project was provided by the University of Technology Sydney. In answering the original evaluation questions:

Did the project achieve its stated outcomes?

Yes, most definitely. The CLA V2 toolkit was developed, refined, evaluated and is being adopted into practice for a number of universities. Data interoperability guidelines have been developed. The project findings and outcomes are clearly articulated in the project final report.

Was the project managed and conducted in ways that contributed to project success?

Yes absolutely, strong project leadership and management was provided by Kirsty Kitto and Aneesha Bakharia in the first year. There are always challenges with a multi institutional project and yet this was one of the smoothest managed projects. Kirsty's approach was warm, professional and injected humour when it was needed. There are so many positive outcomes from this project that will continue to influence the face and nature of learning analytics for years to come

Did the project achieve as much impact as it should have?

Yes, again the impact is clearly outlined in the impact appendix. This project has achieved national and international recognition for its outcomes and will continue to grow over the coming years.

How could the processes associated with the project be improved and replicated?

There was nothing more this project could have done. There are always challenges with different universities processes that come to the forefront in a project such as this one. The team met all challenges and moved forward with a highly successful project.

Final words, it has truly been a pleasure working on this project with this team and the project leader. Thank you for being a small part in such ground breaking work.

Phase	Activities	Deliverables/Objectives	Outcomes
Phase 1 – Commencement February – April 2015	• Appointment of staff • Evaluator appointed • Apply for Ethical Clearances • Creation of Google Community • Creation of Twitter Account • Creation of project webpages • QUT hosts first Project Team Meeting • Identify Learning Activities for implementation in first AR Cycle	• Staff recruited and working on project • Ethical clearance obtained • Web pages which can host the web-based data capture and reporting tools for User Group A → D1 • Set of dissemination outlets that can be used by all team members → D4 • Initial CLA data scraping Prototypes developed → D2 • Set of learning activities that will form the basis of Action Research Cycle 1	Project dissemination strategy is started via the project webpages. A community of practice is established for the later dissemination strategy Initial prototypes of CLA data scraping tools made available on webpages for wider exploration
Phase 2 – Development May – October November 2015- April 2016 May - October	First Action Research Cycle • Toolkit development (tools, for scraping analysis and reporting) based upon Learning Activities identified at the end of Phase 1 • Evaluation of tools produced • Project Team Meeting in Adelaide • Identify Learning Activities for implementation in next AR cycle Second Action Research Cycle • Toolkit development (tools, for scraping analysis and reporting) based upon Learning Activities identified at the end of previous AR cycle • Evaluation of tools produced • Project Team Meeting in Sydney • Identify Learning Activities for implementation in next AR cycle • Test Studies at Partner Institutions start Third Action Research Cycle • Toolkit development (tools, for scraping analysis and reporting) based upon Learning Activities identified at the end of previous AR cycle • Evaluation of CLA toolkit occurring at Partner Institutions via Test Studies	Novel data analytic techniques developed to enhance CLA toolkit analysis → O1 Web-based data capture and reporting tools → D1 Tools that will form the basis of the open source CLA toolkit → D2, O2 Protocols for interpreting analytics created by toolkit developed → D3, , O3 Training to the Australian higher education sector in the use of the toolkit → D4, O5 Test Study cohorts identified and Testing of CLA toolkit's utility has begun at Partner Institutions → O4	The community of practice is built up, with different academic User Groups (type A and B) starting to explore its use. Academics are learning how to use and interpret Learning Analytics and the CLA toolkit.
Phase 3 – Wrap up November-January 2017	• Independent evaluation completed • Finalize Toolkit and User documentation • SoLAR LASI • Production of final Report	• Web-based tools fully developed for User Group A → D1 Completed • Open Source Connected Learning Toolkit Created→ D2 Completed • Protocols for interpreting analytics tested and validated by Partner Institutions→ D3 Completed • Training to the Australian higher education sector in the use of the toolkit and LA as a whole via LASI → D4 Completed (but still ongoing due to community of practice). • Project report completed → D5 Completed	**Students** receiving timely and useful feedback on the nature and quality of the connected learning that they display in selected learning activities using social media. **Academics** using toolkit to support and improve student learning towards desired educational outcomes.