Developing an open architecture for learning analytics

Niall Sclater[1], Alan Berg[2], Michael Webb[3]

[1]Jisc, niall.sclater@jisc.ac.uk
[2]Apereo Foundation, alan.berg@apereo.org
[3]Jisc, michael.webb@jisc.ac.uk

Keywords
Learning analytics, systems architecture, communities of interest, Apereo LAI

ABSTRACT

Carrying out learning analytics can involve a complex range of data sources, systems and dashboards, often owned by different parts of the institution. Vendors are quickly moving to try to capture the market and ensure their products are at the centre of the learning analytics landscape (Sclater, 2014b). However the architectures for connecting the various elements are still evolving as new applications for learning analytics emerge. Complicating the picture further, the wide variety of systems in place at different institutions means that architectures are not easily replicable between organisations or sometimes even between departments.

The interest in deploying learning analytics services at the campus level is increasing but there are many barriers to deployment such as: breaking down data silos, understanding the predictive models and interventions, data governance, and the lack of competences within an organisation needed to manage interventions. There is therefore a growing need for guidance to institutions which wish to develop their learning analytics capabilities as to how to integrate multiple data sources and the new systems they need to build, procure or co-develop as part of a wider community.

In the UK Jisc is spearheading an initiative to procure the elements of a basic learning analytics system for higher and further education institutions. This has involved developing an architecture comprising a number of discrete data sources and systems. The model was reviewed by a cross disciplinary team of European experts in Paris in February 2015.

This paper describes the Jisc learning analytics architecture and proposes it as a reference model which organisations can use to help develop their own architectures for learning analytics. The experiences acquired and lessons learned during development have the potential to influence and be influenced by wider international discussions around an emerging Open Learning Analytics Framework such as the Apereo Learning Analytics Initiative. The paper demonstrates how an architectural walkthrough with invited experts taking on discreet roles was used to enhance the architectural model.

INTRODUCTION

There is growing interest in deploying learning analytics services at educational institutions. Stimulating this are a number of successful examples of analytics services that have impacted on students' learning. Course Signals is perhaps the best known (Arnold & Pistilli, 2012). Another example is the Open Academic Analytics Initiative (OAAI) led by Marist College (Jayaprakash, Moody, Lauria, Regan, & Baron, 2014). Both OAAI and Course Signals predict the likelihood of students failing. However, learning analytics has the potential to impact positively on learners throughout their studies, not only through improving student retention, but also through better feedback from recommendation systems and adaptive learning systems. An all-encompassing framework would need to include: the collecting of data; dealing with crucial issues such as data governance and ethics; pre-processing of the data and sharing of the data models; predictive modelling; interventions including dashboards and other strategies, and measurement of their impact on the learning process.

With so many elements for a university to assemble, a practical approach to knowledge dissemination and initial service creation is necessary. JISC's scaled implementation will allow organisations to gain experience of learning analytics using freely-available services.

There has already been a scurry of research activity examining possible frameworks, for example in 2011 an Open Learning Analytics Architecture (OLA) was proposed (Siemens, Gasevic, Haythornthwaite, Dawson, Shum, & Ferguson, 2011). Discussions were then taken forward during a number of OLA framework meetings and summits (e.g. http://www.prweb.com/releases/2014/04/prweb11754343.htm, https://r3beccaf.wordpress.com/2014/12/20/open-learning-analytics-amsterdam/).

The wider discussions have strongly influenced the Apereo Learning Analytics Initiative (https://confluence.sakaiproject.org/display/LAI/Learning+Analytics+Initiative) which integrates a set of interlocking pieces of open source learning analytics software, supported by an international community of higher education organisations. However, a mature conceptual framework supported by a fully functional end to end reference implementation has yet to fully emerge. Jisc's model should form a reference point for further discussion and evolution.

Jisc's proposed architecture for learning analytics (see Figure 1) has been developed in an attempt to conceptualize the end to end elements of a basic learning analytics system which is being procured for higher and further education institutions in the UK. Data comes primarily from the student record system, the virtual learning environment (VLE or learning management system) and a variety of library systems. Institutions are also beginning to use data from other systems such as attendance monitoring and assessment systems.

Data is fed via an extract, transform and load process from the student record system, and from the other systems to a *learning records warehouse* which contains data in both structured and unstructured formats. Data from *existing institutional record stores* can also be integrated, and students may also input *self-declared data* from e.g. wearable technology.

At the heart of the architecture is the *learning analytics processor* where predictive analytics are carried out, and lead to action coordinated by the *alert and intervention system*. Visualisations of the analytics for staff are available in a series of *dashboards,* and a *student app* allows learners to view their own data and compare it with others. Meanwhile a *student consent service* helps to ensure privacy by enabling students to give their permissions for data capture and use.

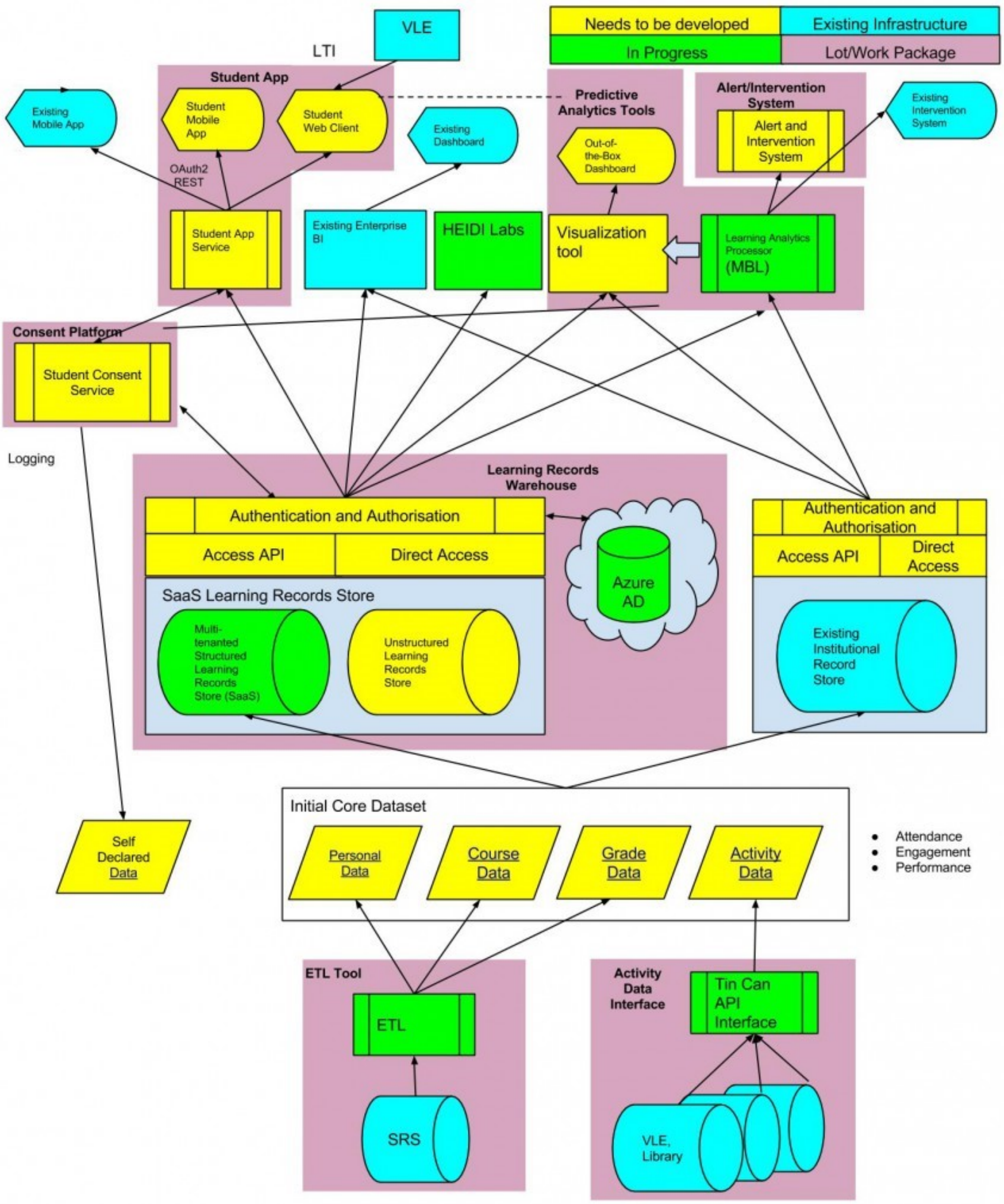

Figure 1: Jisc's architecture for learning analytics

Jisc will procure and provide the components of this basic learning analytics service as a 'freemium' model from the perspective of end user organisations. A core set of functionality will be made available to institutions at no cost to them. The institutions will then be able to purchase additional features if required. The solution will be able to be run as a multi-tenanted or multiple instance solution in the cloud, either hosted by Jisc or by the vendor/public cloud, ideally with the option of being hosted by the institution as well.

The architecture enables the separation of the various elements and allows different vendors to propose solutions for specific areas of functionality. This facilitates healthy competition and the potential for an environment where components can be plugged in or replaced as required. The separation of concerns can be clearly seen by the division of lots during procurement. The bid process includes the following lots:

Lot 1) Predictive analytics tools

Including the learning analytics processor — a tool to provide predictions on student success to feed into the alert and intervention system. Predictive models must be exportable in an open standards compliant format, the Predictive Model Markup Language (PMML). PMML (http://www.dmg.org/pmml-v4-2-1.html) is an XML standard for describing predictive models so that they can be shared easily and then reused across different software vendors' products.

Lot 2) Alert and intervention system

A tool to be used by institutions to allow them to act on alerts and manage intervention activity. The system should also be able to provide data such as methods and success, and feed into an exemplar 'cookbook' on learning analytics. Ideally different interventions could be plugged into the infrastructure.

Lot 3) Student app

The collecting and curating of app requirements are just as important as any software artefact. There have been many dashboards built, but the requirement gathering exercises have often failed to reuse prior efforts. This leads to much reinvention and duplication.

A student app will be built from a specification based on a requirements gathering activity involving students and staff. The app will be delivered on Android and iOS platforms, together with a web app to be integrated within a VLE. Data will come from the learning records warehouse (Lot 4) and the learning analytics processor (Lot 1), and will include appropriate mechanisms for consent and authentication.

Lot 4) Learning records warehouse

A data warehouse solution will be developed to hold learning records. It is anticipated that the solution will require the ability to store both structured and unstructured data, and should scale to hold several billion data items per year.

Lot 5) Standard MIS data extraction, load and transform (ETL) tool

An ETL tool is required to take data from a range of sources and import it into the learning records warehouse (Lot 4).

The exact number of records stored in the learning records warehouse will depend crucially on how much preprocessing of raw activity takes place. For example, activity could be summarised before being stored in the record store. The compression process reduces the need for dashboards and predictive models to have to retrieve large set of datasets to perform similar calculations.

Lot 6) Activity data interface

A solution will be provided using open standards (e.g. the Experience API also known as xAPI (http://tincanapi.com/overview/) to take activity data from a range of learning tools, including VLEs and library systems, and integrate the data into the learning records warehouse.

The advantage of xAPI is that it has market traction and allows the plugging of an increasing number of systems into the infrastructure. Examples of open source xAPI enabled systems enriched under the Apereo Learning Analytics Initiative include: uPortal (https://www.apereo.org/uportal), a widely used portal system; Sakai (http://sakaiproject.org), a VLE used by millions of students every day; and Apereo OAE (http://oaeproject.org), a cloud based next generation collaborative system used by around forty universities across the world.

Lot 7) Student Consent Platform

It is anticipated that students will require the ability to control what data items are released to the various components of the analytics solution, including additional services that are brought online in the future. This platform will provide tools to give that control to students. The issue of consent is complex and potentially expensive to properly enforce. This is an interesting area that needs to reflect the emerging national ethics and privacy policy guidelines. It is no coincidence that the day after an architectural review workshop was held in Paris to critique the Jisc framework, a second workshop on ethics and privacy was held to discuss a taxonomy of ethical and legal issues arising from learning analytics, (Sclater, 2014b, Sclater, 2015).

Lot 8) Expertise in open learning analytics to advise and support

Expert advice to the project is required in specialist areas including learning analytics, machine based learning, predictive algorithms and measuring impact. The infrastructure is by its nature complex. Advice needs to be drawn from a range of experts.

1. REFINING THE MODEL

A number of experts from Europe came together in Paris in February 2015 to review Jisc's nascent architecture for learning analytics. The event at L'Université Paris Descartes was jointly hosted by the Apereo Foundation and Jisc. Delegates included representatives from the Universiteit van Amsterdam, the Pädagogische Hochschule Weingarten in Germany, Surfnet in the Netherlands, University of Michigan in the USA, CETIS and the Lace Project, as well as Jisc and Apereo.

Figure 2: Refining the model at the Jisc / Apereo workshop in Paris

The workshop took the form of an informal architectural walkthrough and involved participants taking on the following roles for the day:

- Oracle, who knows everything about the architecture and the initial requirements
- Student
- Teacher

- Tutor
- Researcher
- Ethical hacker (security expert)
- Software architect
- Privacy enhanced technologist. Roughly speaking, the ethical hacker tries to break into the system while the privacy enhanced technologist tries to keep him/her out.
- Front end developer
- Federative log on expert
- Enterprise service bus and ETL expert
- Data governance expert
- Writer, whose role is to keep honest notes of the day
- Chair, who keeps the walkthrough discussion focused

This was an effective way of obtaining in-depth constructive criticism of the model. Details of key elements of the architecture follow, including the main conclusions from the workshop.

1. CONSENT SERVICE

The student consent platform is likely to be one of the most complex elements of this project, and it is a new and emerging part of the education space. One issue that needs to be solved is that learning analytics may require data to be used in ways that were not initially envisaged at the point it was collected, and therefore consent must be obtained to process it differently. There are two specific use cases to consider:

- The institution may wish to use data from a group of students in a new way, and therefore needs a method to get consent in bulk before it can be used.
- A student may wish to provide their individual data to a new app or service.

To draw parallels with existing services, the former may be similar to users agreeing to new or additional terms and conditions, while the latter may be analogous to a user providing consent for a service or app to access their personal data from another system, commonly seen when apps access social data via OAuth.

For this solution to work a mechanism to authenticate and authorise the student will be required. Federated authentication is available through the Shibboleth SAML based solution across the higher education sector.

It was interesting that almost immediately at the architecture workshop issues relating to privacy were raised. These were of such concern to the Germans present that they believed their students would not be prepared to use a learning analytics system unless the data was gathered anonymously. Once learners had gained confidence with the system they might be persuaded to opt in to receive better feedback. Thus the consent service was confirmed by the group as a critical part of the architecture.

Various issues arise here e.g. what happens if the student has left the institution? Will there be large gaps in the data which diminish the value of the overall dataset?

One participant suggested that students could decide if they wanted analytics to be temporarily switched off – in a similar way to opening an incognito window in a browser. This would allow them to carry out some exploration without anything being recorded. The logistics of building this into multiple systems though would certainly also be complex - and it would potentially invalidate any educational research that was being undertaken with the data.

Students may be worried about privacy, but handling the concerns of teachers was also felt to be crucial. It was suggested that statistics relating to a class should remain private to the teacher of that class; concern was expressed that learning analytics could be used to identify and subsequently fire ineffective teachers. "Could a predictive model allow unintelligent people to make decisions?"

was the way the participant with the "teacher" role summed up the perennial battle for control between faculty and central administrators.

One suggestion to minimise privacy infringements was to use the LinkedIn model of notification when someone has looked at an individual's profile. Certainly every time someone views a student's data it could be logged and be subsequently auditable by the student.

2. STUDENT APP

Putting data in the hands of the students themselves may be a key way to deploy learning analytics in order to improve the chances of educational success. A student app is therefore a crucial part of the proposed architecture (Sclater, 2014c). Functionality may include:

Measuring engagement

Students might find visualisation of their participation in a course of use, measured through a variety of metrics such as VLE access, campus attendance and library use. Comparisons with peers may be helpful. And comparisons with previous cohorts, showing the profile of a successful student might be useful too. These could be presented in a variety of ways, including graphs of engagement over time compared with others. Learners might want to have alerts sent to their device through the app if their participation shows they're falling below an acceptable level or user defined level.

Measuring assessment performance

There is clearly a need to show details of assessments already completed and grades obtained, and the dates, locations and requirements of impending ones. Assessment events transferred to a student's calendar with advance alerts could also be useful. But arguably this is simple reporting and alerting functionality and not learning analytics. A progress bar showing how they are progressing through their modules and qualifications might be helpful. A prediction of their grade if they continue on the same track and what they would need to achieve to improve may also be of interest. Otherwise assessment data could feed into one of the metrics used for measuring engagement.

Module choice

One application of learning analytics is to assist students in making module choices. Analytics can recommend modules where they are most likely to succeed, comparing their profile with those of previous students and presenting them with information such as "Students with similar profiles to you have tended to perform better when selecting xxx as their next module".

Technical Requirements

The initial specification is likely to include an Android and iOS app, along with a web/HTML5 app, potentially to integrate within a VLE via formats such as the IMS Learning Tools Interoperability (LTI) specification (http://www.imsglobal.org/toolsinteroperability2.cfm). The learning records warehouse and the learning analytics processor will deliver data and provide appropriate mechanisms for consent and authentication.

One idea suggested at the workshop was for the student app to use an open API, allowing other student-facing services to be integrated with it. Another issue raised was that most analytics is carried out on data sources which can be fairly "old" however there may be a need for real time learning analytics. And a student app which assessed whether learning outcomes had been achieved could also be very useful.

Real time collection tends to be more expensive than through the ETL layer using a batch process. Not only might there be an impact on the performance of the application that is sending the activity, but the ELT layer affords the opportunity to compress the activity into averages. Compression decreases the amount of data that the predictive models and dashboards need to process later. The play off between the costs and benefits of real time processing will be explored by the freemium service.

One interesting idea mooted at the workshop was whether the most important source for predictive analytics might be "self-declared" data? It may be that some wearable technology monitoring a student's sleep patterns or their exercise levels for example could be mapped onto their learning performance. Or they might want to log the fact that they had watched several relevant youTube videos that they had discovered.

3. LEARNING RECORDS WAREHOUSE

Jisc anticipates that the learning records warehouse will be the central repository for learner activity data, plus optionally as a warehouse for other student data about the student and their courses.

The record store element should take activity data via the Experience (Tin Can) API (xAPI), currently the main standard for learning records. In addition, the overall solution should also be able to take data derived from a more traditional LTI process.

The core requirement for the learning records warehouse is to be able to receive data via the xAPI, and then provide it to the learning analytics processor, the alert and intervention system and various dashboards. xAPI has been chosen at the moment as Jisc believes it to be the most complete and well documented solution available at the time of writing. However, this will be reviewed as other options such as the Caliper sensor API become available.

It is anticipated that the storage of activity data will be a core requirement for most users of the service, in that they will not have an existing equivalent solution. Many will have warehouses in place to hold student data. However, for those that do not there should be a method of integrating this data so that it can be queried by the learning analytics processor.

Concern was expressed around the performance of dashboards when required to process big data. Thus the ETL (extract, transform and load) layer is crucial to determine what data is stored in the learning records warehouse, which data can be compressed to facts in the warehouse and which records should be filtered out as currently irrelevant. The tuning of this process potentially impacts on the accuracy of predictive models, costs of backing up systems and how real time "real time" actually is.

4. ALERT AND INTERVENTION SYSTEM

The alert component comprises the processes and workflow to notify the appropriate person that action is recommended. The intervention management component will allow institutions to manage interventions with students. It is envisaged that this will work in a similar way to a customer relationship management system, but will be focused on learner achievement, including recording communications, interactions and outcomes. This should not only be in place to help those at risk but should also allow the teacher to analyse how well things are going on overall in the class. Interventions might be to congratulate students on their progress as well as to address potential failure or drop-out.

A key requirement of the system is to allow institutions (through sharing of practice) to monitor the effectiveness of different interventions. Reporting ability is therefore of great importance, either directly within the system, or through the use of third party tools.

5. LEARNING ANALYTICS PROCESSOR

Using machine based learning techniques to develop models that predict student success is a new and emerging field. A core element of the learning analytics project is to help the sector engage in this field, to provide evidence of the effectiveness (or otherwise) of using large data sets, and then, if successful, scale-up their use through the open sharing of data models.

The system should produce predictive models based on existing data, and validate the models on new datasets.

This was deemed to be so critical that it should form a separate layer, underpinning the other applications. Meanwhile compliance with the Predictive Model Markup Language has already been specified as a requirement for the predictive models to be used by the learning analytics processor. However one member advised the group to be wary of "gold-plated pigs" - some vendors are excellent at presenting beautiful apps and dashboards which may have unsophisticated or unreliable underlying models and algorithms doing the predictions. Most staff are unlikely to want to know the fine detail of how the predictions are made but they will want to be reassured that the models have been checked and verified by experts.

6. DASHBOARDS

The solution should provide a means to view data from the system in a visual way, with dashboards showing data applicable to the user's individual needs, for example a module overview for a module leader. Institutions may already have some existing portal technology, and it would be advantageous if the visualisations could be integrated. In addition some institutions may have their own visualisation tools which should be able to be used if required.

6. STANDARDS

A key requirement of the learning analytics solution is to provide universities and colleges with tools to improve learner performance, and provide early alert of potential failure or withdrawal from the course. Three techniques are likely to be core to this:

- Application of relatively simple but proven user defined models
- Predictive models based on techniques such as machine based learning
- Insights relating to factors affecting student success through data analysis

Openness is important to many stakeholders within the sector, and there is some evidence that models, both user defined and predictive, can be shared between providers. A solution is therefore required that can output models in open standards. The current requirement is to comply with Predictive Model Markup Language although the solution should be flexible enough to be adaptable if new standards emerge.

The use of technical, preferably open, standards is clearly important for an architecture comprising a number of interchangeable components potentially built by different vendors. The Experience API (Tin Can) has been selected as the primary format for learning records at the moment; it should be relatively easy to convert data from the VLE to this format, and some plugins e.g. Leo for Moodle (http://leolearning.com/blog/leo-releases-major-moodle-xapi-updates/) already exist. However there may be a maintenance overhead every time each LMS is upgraded.

It was suggested that the IMS Learning Information Services (LIS) specification would be appropriate for storing data such as groupings of individuals in relation to courses.

The problem of ensuring universally unique identifiers for individuals (and activities) was also noted. If there is not a unique ID for every user (i.e. a person or another source of activity such as an entrance port to a cafeteria) then predictions will be weakened or the data may need to be cleaned, potentially at considerable expense.

7. SECURITY

The "ethical hacker" (security expert) was concerned that security issues will be complex because of the number of different systems in place. This is because the security issues increase polynomially as the number of components increases. Meanwhile there will be ongoing requirements for patches and version updates for the different VLEs and student information systems involved. Consistent security hygiene is required across all of the systems under different ownerships within and bordering the infrastructure diagrammed in this paper.

8. CONCLUSIONS

The experience gained in a wider ecosphere of communities such as the Apereo Learning Analytics Initiative, LACE, SURF and a number of leading universities was helpful in providing neutral, critical support. Building a cross disciplinary team rather than opening up subscriptions to the workshop ensured a broad range of expert perspectives. The informal walkthrough format was a driver for a focused discussion.

There is much opportunity for reusability of components outside the hosted framework. For example, collecting the lessons learned from a community process around building and interacting with the infrastructure will allow for the creation of advice, cookbooks, and policy documents.

The data transforms that convert and pass on raw data should be relatively easy to share and deploy within individual campuses. The use of xAPI will be an extra stimulus for vendors in the UK to adopt this activity based standard.

Jisc was reassured after the workshop that the architecture, with some minor changes suggested by the group, is robust, and it will be aiming to put in place a basic "freemium" learning analytics solution for UK universities and colleges by September 2015. The infrastructure provides the opportunity for institutions to understand their own requirements and gain experience with learning analytics in a supported, relatively low-cost way. These experiences can be incrementally fed back to improve the technical infrastructure, and help to focus research efforts on enhancing student retention and other aspects of the learning experience.

REFERENCES

Arnold, K. E., & Pistilli, M. D. (2012). Course Signals at Purdue: Using Learning Analytics to Increase Student Success. In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (pp. 267–270). doi:10.1145/2330601.2330666

Jayaprakash, S. M., Moody, E. W., Lauria, E. J. M., Regan, J. R., & Baron, J. D. (2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. Journal of Learning Analytics, 1, 6–47.

Sclater, N. (2014a). Code of practice for learning analytics. A literature review of the ethical and legal issues. Retrieved from http://repository.jisc.ac.uk/5661/1/Learning_Analytics_A-_Literature_Review.pdf

Sclater, N. (2014b). Learning analytics. The current state of play in UK higher and further education. Retrieved from http://repository.jisc.ac.uk/5657/1/Learning_analytics_report.pdf

Sclater, N. (2014c). A student app for learning analytics. Effective Learning Analytics, Jisc. http://analytics.jiscinvolve.org/wp/2014/12/19/a-student-app-for-learning-analytics/

Sclater, N. (2015). A taxonomy of the ethical and legal issues of learning analytics v0.1. Retrieved from http://analytics.jiscinvolve.org/wp/2015/01/30/a-taxonomy-of-the-ethical-and-legal-issues-of-learning-analytics-v0-1/

Siemens, G., Gasevic, D., Haythornthwaite, C., Dawson, S., Shum, S. B., & Ferguson, R. (2011). Open Learning Analytics : an integrated & modularized platform Proposal to design , implement and evaluate an open platform to integrate heterogeneous learning analytics techniques. Knowledge Creation Diffusion Utilization.

AUTHORS' BIOGRAPHIES

Niall Sclater is Consultant and Director at Sclater Digital Ltd, an educational technology consultancy. Previously he was Director of Learning and Teaching at the Open University, responsible for institutional strategy in areas such as tuition, educational technology and learning analytics. More recently he has been leading activities for Jisc in the area of learning analytics, surveying its current use in the UK, commissioning the components of a basic learning analytics system for higher and further education, and developing a code of practice to deal with the legal and ethical issues. He has been involved in the research, development and management of learning and teaching in higher education since 1992, with particular interests in technology-enhanced learning. Further information is available at sclater.com/blog or https://www.linkedin.com/in/niallsclater

Alan Mark Berg, BSc, MSc, PGCE, has been the lead developer at Central Computer Services at the University of Amsterdam since 1998. He is currently working in an Innovation Work Group that accelerates the creation of new and exciting services. In his famously scarce spare time, he writes including authoring a book on recipes for quality assurance (https://www.packtpub.com/application-development/jenkins-continuous-integration-cookbook-second-edition). Alan has a bachelor's degree, two master's degrees, a teaching qualification, and quality assurance certifications. He has also coauthored two Packt Publishing books about Sakai (http://sakaiproject.org), a highly successful open source learning management platform used by millions of students around the world. He has won a couple of awards, including the Sakai Fellowship and Teaching With Sakai Innovation Award (TWSIA).

Alan enjoys working with talent; this forces him to improve his own competencies. This motivation is why Alan enjoys working in energetic, open source communities of interest. At the time of writing, he is on the board of directors of the Apereo Foundation and is the community officer for its Learning Analytics Initiative (https://confluence.sakaiproject.org/display/LAI/Learning+Analytics+Initiative)

In previous incarnations, Alan was a QA director, a technical writer, an Internet/Linux course writer, a product line development officer, and a teacher. He likes to get his hands dirty with building, gluing systems, exploring data, and turning it into actionable information. He remains agile by ruining various development and acceptance environments and generally rampaging through the green fields of technological opportunity.

Michael Webb is Jisc's senior technology innovator, working on a range of projects to develop new digital solutions to support UK Higher Education, Further Education and skills sectors. Current projects include learning analytics, business intelligence, cloud platforms and student led innovation. Michael has 20 years experience of working in the HE sector, including ten years as Director of IT at University of Wales, Newport, and before that leading infrastructure and application development projects at University of Plymouth. He is also an experienced software architect and developer, including mobile development and scalable web based solutions.