Current Issues in Emerging eLearning

Volume 4 | Issue 1 Article 5

6-13-2018

Traversing Across Learning Environments And The Need For A Single Data Standard In Digital Learning Environments

Liam McNamara
University of Edinburgh, liamsmcnamara@gmail.com

TRAVERSING ACROSS LEARNING ENVIRONMENTS AND THE NEED FOR A SINGLE DATA STANDARD IN DIGITAL LEARNING ENVIRONMENTS

Liam McNamara[1]
[1]University of Edinburgh / liamsmcnamara@gmail.com

INTRODUCTION

Big Data is becoming pervasive in society and a hot-topic which can now provide opportunities for new and intricate ways to collect and analyse data. It has become its own science, an industry, and consequently made its way into the educational sector where it has become a beacon for solutions (Boyd and Crawford, 2012). Education has developed into big data and all the sophistication that it must offer, such as recommender systems as well as business intelligence and decision-making for institutions (Slade and Prinsloo, 2017). The inclusion of data has affected every level of education, from the macro-perspective of national and international education, right down to the everyday dynamics of the classroom (Shum, 2012). Though, in education, data is collected from a variety of sources in a Virtual Learning Environment (VLE) however this data comes from a variety of technologies that do not necessarily adhere to a standard. When technologies adhere to standards and data is collected en masse, this creates difficulty in analysing learner behaviour. But also, this collection of data creates an ethical burden. It becomes particularly significant when the lifelong learning paradigm is invoked and data is collected beyond simply a tertiary course or an online workshop. The implications of big data and data standards are to be discussed herein.

DATA

Big data in industry is becoming ever present and is extending into education however it is creating a peculiarity leaving the question "What is the data?" In education, this has mostly been done through the measurement, collection, analysis and reporting of data about learner performance that is designed to reflect their individual performance or the overall performance of their respective institutions. Shum (2012) indicates three levels of data in education: macro-, meso-, micro-levels. This has led to several developments, such as predictive modelling, social network analysis, usage tracking, content/semantic analysis, recommender systems (Clow, 2013). This consequently has led to instruments that are utilised by organisations for data-driven decision-making particularly in higher education institutions (Beer, 2012). However, data does not guarantee success or is as an important component as Shum (2012) suggests, whereby promises that data-driven education can make but there is still much needed critical debate.

Shum (2012) authored a document for UNESCO that outlines several recommendations one which clearly indicates:

> *"Institutions should collaborate on establishing trusted partnerships and robust mechanisms to share student data, analytics techniques and information visualization tools."* (Shum, 2012)

But most interesting from this recommendation is the need for *mechanisms* through which *student data* can be shared. The issue then becomes the relevance of the data and its significant as well as ethical considerations for the reliance on such data. To ensure data governance, several factors must be considered; ownership of data sets, interpretation of data, and decision making (Elouazizi, 2015). However, the importance of interpretation and decision making would depend entirely on the data sets and quality of that data. For data sets to be meaningful and shareable as Shum (2012) in the UNESCO report suggests, then the data itself needs to be relevant so that it can be meaningfully interpreted.

DATA STANDARDS

An example of data standards in practice, medicine has already developed ways to approach sharing and communicating data. Peck (2008) illustrates the usage Digital Imaging and Communications in Medicine (DICOM) which provides diagnosticians with technology that allows them to readily transfer data between clinics for diagnostics and treatment planning. The usage has become a standard in the medical field which provides inspiration for the possibilities in education and learning analytics. It becomes a matter for necessity, that much like medicine, education also needs a standard and, more specifically, data standards.

Del Blanco, Serrano, and Freire (2013) illustrate the usage of data standards for the purposes of collection and transmission because in the current systems used to collect learner data, each system is tailored and consequently different. The comparison between systems refers to typically Massive-Open-Online-Courses (MOOC) which collect vast amounts of learner data through learning management systems. The data standards proposed by Del Blanco, Serrano, and Freire (2013) refer to IEEE Standard for Learning Technology and Experience API, both of which are existing technologies not readily adopted by MOOC providers or embed within learning management systems.

Experience API derives its premise from a less technological and more pedagogical philosophy, that being Vygotsky's *Constructivist Learning* and Silvers's *Activity Theory* because of the reliance on activity based data collection (Kevan & Ryan, 2015). This is inspiring in the sense that no longer is technology leading pedagogy but instead, in this specification, pedagogy is leading technology. The common drawback of virtual learning environments is the inconsistent data standards, especially when each technology implements its own standard. This makes learning analytics difficult because data is captured differently in each system. Experience API is designed to solve this by allowing each technology to record to a Learning Record Store (LRS) using a common framework.

Experience API utilises a Learning Record Store (LRS) which store information such as learning activity streams but also provides greater possibility for learning analytics. As the learner engages or interacts with various objects such as course pages, other webpages, games, and simulations. The learning data is stored in the LRS under the Experience API (Lim, 2016). The ability to collate vast amounts of data under one framework means that data becomes interoperable and transferable. Kevan and Ryan (2015) suggest that through event driven data collection it becomes possible to record learning events across numerous platforms as well as an individual lifelong learning experiences.

This results in a single facility that collects learning data from a variety of technologies that otherwise would not be possible without a uniform data standard. This provides several possibilities that otherwise would not be possible. If, in a typical learning environment, only one of the technologies adequately collects data about the learner and their experience then it is difficult to assess the learner's proficiency. However, if a LRS can be implemented and data is referred to the store from many different learning technologies then the learner's proficiency and engagement is more accurately able to be analysed and reported on.

ETHICAL CONSIDERATIONS

Poeppelman et al (2013) illustrate problems that exist when data in different systems is incompatible making tracking learner performance difficult. However, the usage of technologies, such as Experience API, will allow for capturing of key learning data from multiple systems and storing in a single technology that can be accessed for analysis. Hruska et al (2014) follow on from Poeppelman et al (2013) to suggest that using multiple learning systems that collect data longitudinally need to be further understood. When considering the need for continued development of informational infrastructure for learning analytics, there becomes a concern regarding how the data is stored and used.

However, through the advancement of a centralised data collection of all learning activity from a variety of learning technologies, there becomes the apparent security and ethical implications for data collection and retention. Essentially, data can be collected without the learner being aware of this fact. Prinsloo and Slade (2013a) and Zimmer (2012) suggest that ethical considerations for data, therein learning analytics, should concern: who benefits, consent, de-identification, opting out, vulnerabilities, collection, analysis, and storage. This ethical consideration often becomes institutional based and specifically the policies focus on an academic level, meso-level analytics as denoted by Shum (2012), meaning that the policies do not necessary reflect the demands of learning analytics (Prinsloo and Slade, 2013b). Policy then becomes ever more complex with the sophistication of Experience API and Learning Record Stores where many systems store and transfer learning information.

Virtual Learning Environments (VLEs) comprise many technologies that under Experience API would record learning activity. This configuration becomes known as Personalised Learning Environments (PLE). Wilson et al (2007) refers to a PLE and the culmination of multimodal, many technologies model of learning which better reflects the aspirations of lifelong learning paradigm. However, from a security and ethical perspective, lifelong in the digital world raises concerns because who takes responsibility for the data storage, its security, access, and its usage. Institutional policy then should ultimately provide guidance as how ethical data collection and storage (Slade and Prinsloo, 2013b).

Conversely, data collection and mining already have a history in ethics and several ethical arguments in favour of large scale data collection exist, as van Wal and Royakkers (2004) contends the following:

- Data mining itself does not give rise to new ethical issues
- Many individuals have simply chosen to give up their privacy, and why not use this public information
- Personalisation leads to individualisation instead of de-individualisation

By extension, learning analytics is simply adding another derivative, or application, of data mining, meaning that these arguments could be cast easily. However, that is not to say that PLE and LRS are without further ethical scrutiny. One ethical issue that presents itself for a single data standard for lifelong learning is a dilemma regarding who is ultimately responsible for the storage and the accessibility for such information.

“Just because it is accessible doesn’t make it ethical” (Shum, 2012)

To refer to the medical analogy, doctors have access to medical histories for the purposes of effective and preventative medical care. Without access to historical information, doctors cannot mitigate the risk of maltreatment; in its essence, professionalism. Such a medical record is synonymous with being lifelong. This could readily be applied to education and could assist educators in mitigating risk. However, medical practice is grounded in the rigours of physiology, chemistry, biology, and pharmacology – medical science. Education does not share troves of empirical evidence in which to draw upon to make informed decisions. This is when the various tenets as well as sociological and scientific principles in education are likely to be used for deductive reasoning. This is where the subjectivity as to what principles to rely upon as instruments of interpretation.

INTERPRETATION & CONTEXT

The context and the interpretation of data relies heavily on the data preserving information about context for the former and the interpretation lies with the beholder for the latter. As big data takes hold the potential for problems become exponential and particularly though examples are limited. There could be simple hypothesis about potential problems that could arise from the creation of lifelong data tracings of a learner's activity. Overtime that data will comprise of collections from multiple technologies stored under uniform data standard. This longitudinal data could contain artefacts that do not accurately or authentically portray the learner's ability or competence. In the event of big data and artificial intelligence, the advent of recommender systems means that determinations about a learner may be made inaccurately recorded.

This inaccuracy then has further effects as the activity of the learner through the Experience API is recorded in the LRS, a sequence of redundant learning activities is recorded. The question then relates to the long-term effects this has on the learning analytics and interpretations of that learner. The learner has a lifelong permanent record that includes artefacts of inaccuracies which may mislead interpretation. The effect of this is unknown but such issues highlight the existing problem that unbeknownst to learners, in some institutions their activity is already logged and being used for strategic and business intelligence at the meso-level. It becomes peculiar when, at the micro-level which is user-level data, the analysis and subsequent interpretation would be intended to profile and provide the learner with insight into their own learning (Shum, 2012). However, therein lies the issue with data because despite the insight it may be provide, data is able to be filtered and categorised.

Shum (2012) enunciates "Data is Not Neutral" which refers to the reality that big data imposes bias. In the context of personalised learning environments where data collection is standardised, then there is the inherent simplification. The data consequently loses features and even elements of context that might influence analysis and therefore interpretation. This fault could be mitigated through redundancies though a single data standard, such as Experience API and therein Learning Record Store, may not reduce the variance produced by multiple data standards. However, when it comes to learning analytics there is always human judgement that cannot necessarily be mitigated. It could be limited but in saying that it is limited, it took judgement and decision which does not necessarily reduce the variance. At an institutional level, another issue is technical concerns that often outweigh the analysis and what intelligence it may provide to an institution (Macfadyen and Dawson, 2012). Though these variances and therefore analyses are not benign reporting practices but instead the interpretations inform interventions and call for action.

Knight, Buckingham, and Littleton (2013) illustrates the relationship between learning analytics, epistemology, and pedagogy. Learning analytics focuses on a transaction of pieces of assessment or activities completed by the learner which is a constructivist approach by the instructor through scaffolding of material. Learning analytics also tends to focus on curriculum mastery and therefore pedagogy, but this consequently results in the necessity to measure and assess (Knight, Buckingham, and Littleton, 2013). This creates two approaches to the interpretation of data in learning analytics; on one hand it can be considered a trace of the learning process and the gradual development towards independence. Secondly, it can be used to re-engage a focus on specific curriculum based assessment. Assessment is a major source for data capture, but learning analytics also needs to focus on learning (Gasevic, Dawson, & Siemens, 2014), though even when analysis focuses on learning subjectivity and application can become issues. Perrotta and Williamson (2016) outline that both the political dimension and mathematical instruments used in analysis need to be understood and a critical approach to understanding their usage.

EVIDENCE-BASED EDUCATION

The development of evidence-based practice, particularly in education, has led to the notion that a potential revolution of understanding and praxis is about to emanate (Slavin, 2002). However, from a more semantical perspective there is the question of what constitutes evidence. There is systematic precariousness about what constitutes evidence which is where data becomes essential in the analysis of educational hypotheses (Davies, 1999). The pedagogical approach proves consistent by providing repeatable and reliable results however Davies (1999) outlines the need for systematic reviews of educational research while outlining existing issues with meta-analysis studies. Given these uncertainties surrounding the use of evidence-based practice in educational research, this needs to be extended to elearning and personal learning environment.

Without conflating meta-analysis techniques and data standards, there nevertheless exist issues within current evidence-based practices that could relate to digital learning environments and data capture. Davies (1999) points out the issue of comparability between studies in meta-analyses which might apply to the comparability of data captured between two technologies. For instance, a personalised learning environment envisage through multi-technologies and the usage of a single data standard using Experience API allows for a more nuanced analysis of a learner's success and guidance towards desired outcomes. This would be achieved with resolution that conventional pedagogy could not achieve. This would give greater power to an evidence-based intervention within an educational setting; however, the layers of black-box abstraction that exist between the educator and the raw data analysis creates more uncertainty.

An engineer maybe able to use, in a crude sense, back of an envelope calculations to make a professional assessment of the reliability and validity of software analysis for an engineering problem but such utensils are not necessary afforded to educators. For instance consider the reality of how an educator is to determine the validity of results produced by a recommender system based on data collected through a personalised learning environment. Now to consider evidence-based practice, if an educator is to make a decision to forego intervention based on data captured from educational technology and the outcome is adverse such as failure to meet learning goals or failure to continue in a course for a student then the question is, is the educator responsible or the recommender system.

CONCLUSION

The prevalence of big data has not been out of reach of education, and has resulted in the emergence of learning analytics. Data collection across technologies using one data standard, simplifies many of the issues that occur when every technology develops a standard in an ad-hoc manner. The usage of Experience API and Learning Record Store is designed to resolve this. However, this pervasive, longitudinal collection of learner activity across a learner's formal education and continued through the paradigm of lifelong learning raises ethical considerations and possible consequences. However, individualised and large scaled retention raises the possibility of anomalies within the tracing on learner progress and the consequences of these are yet to be known. Despite the benefits of learning analytics, such as adaptive learning, recommender systems, multi-tiered strategy, and informed pedagogy; the sophistication and complexity of data collection and analysis has the potential for these foreseeable problems. A uniform data standard for educational technology may simplify and improve data collection though it may not elevate or mitigate issues but create them.

REFERENCES

Beer, C., Jones, D., & Clark, D. (2012). Analytics and complexity: Learning and leading for the future, 1–10. - From readings

Boyd, d., & Crawford, K. (2012). Critical Questions for Big Data. Information, Communication & Society, 15(5), pp.662–679.

Clow, D. (2013). An overview of learning analytics. *Teaching in Higher Education*, *18*(6), 683–695. http://doi.org/10.1080/13562517.2013.827653 - From readings

del Blanco, Á., Serrano, Á., & Freire, M. (2013). E-Learning Standards and Learning Analytics. Can Data Collection Be Improved by Using Standard Data Models? *IEEE Global Engineering Education Conference*, 1255–1261. http://doi.org/978-1-4673-6109-5

Elouazizi, N. (2015). Critical Factors in Data Governance for Learning Analytics. *Journal of Learning Analytics*, *1*(3), 211–222.

Gašević, D., Dawson, S., & Siemens, G. (2014). Let's not forget: Learning analytics are about learning. *TechTrends*, *59*(1), 64–71. http://doi.org/10.1007/s11528-014-0822-x - From readings

Hruska, M., Long, R., Amburn, C., Kilcullen, T., & Poeppelman, T. R. (2014). Experience API and Team Evaluation: Evolving Interoperable Performance Assessment. *Interservice Industry Training, Simulation, and Education Conference*, 1–11.

Kevan, J. M., & Ryan, P. R. (2015). Experience API: Flexible, Decentralized and Activity- Centric Data Collection. *Technology, Knowledge and Learning*, 1–8. http://doi.org/10.1007/s10758-015-9260-x

Knight, S., Buckingham Shum, S., & Littleton, K. (2013). Epistemology, pedagogy, assessment and learning analytics (pp. 75–11). Presented at the the Third International Conference, New York, New York, USA: ACM Press. http://doi.org/10.1145/2460296.2460312

Lim, K. C. (2016). Case Studies of xAPI Applications to E-Learning (pp. 1–12). Presented at the 12th International Conference on eLearning for Knowledge-Based Society.

Macfadyen, L. P., & Dawson, S. (2012). Numbers Are Not Enough. Why e-Learning Analytics Failed to Inform an Institutional Strategic Plan. *Educational Technology Society*, *15*(3), 149– 163.

Perrotta, C., & Ben Williamson. (2016). The social life of Learning Analytics: cluster analysis and the “performance” of algorithmic education. *Learning, Media and Technology*, *0*(0), 1– 14. http://doi.org/10.1080/17439884.2016.1182927

Poeppelman, T. R., Ayers, J., Hruska, M., Long, R., Amburn, C., & Bink, M. (2013). Interoperable Performance Assessment using the Experience API. *Interservice Industry Training, Simulation, and Education Conference*, 1–12.

Prinsloo, P., & Slade, S. (2017). An elephant in the learning analytics room (pp. 46–55). Presented at the the Seventh International Learning Analytics & Knowledge Conference, New York, New York, USA: ACM Press. http://doi.org/10.1145/3027385.3027406

Shum, S. B. (2012). Learning Analytics. *UNESCO Institute for Information Technologies in Education*, 1–12.

Slade, S., & Prinsloo, P. (2013a). Learning Analytics: Ethical Issues and Dilemmas. *American Behavioral Scientist*, *57*(10), 1510–1529. http://doi.org/10.1177/0002764213479366

Prinsloo, P. and Slade, S., (2013b). An evaluation of policy frameworks for addressing ethical considerations in learning analytics. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 240-244). ACM.

van Wel, L., & Royakkers, L. (2004). Ethical issues in web data mining. *Ethics and Information Technology*, *6*, 1–12.

Wilson, S., Liber, O., Johnson, M., Beauvoir, P., & Sharples, P. (2007). Personal Learning Environments : challenging the dominant design of educational systems. *Journal of E- Learning and Knowledge Society*, *3*(2), 27–38.

Zimmer. M. (2010). "But the data is already public": on the ethics of research on Facebook. In Ethics and Information Technology, 12 (4), pp. 313-325. doi: 10.1007/s10676-010-9227-5.