Learning Analytics beyond the LMS: the Connected Learning Analytics Toolkit

Kirsty Kitto, Sebastian Cross, Zak Waters, Mandy Lupton
Queensland University of Technology
2 George Street
Brisbane, Australia
[kirsty.kitto,mandy.lupton]@qut.edu.au

ABSTRACT

We present a Connected Learning Analytics (CLA) toolkit, which enables data to be extracted from social media and imported into a Learning Record Store (LRS), as defined by the new xAPI standard. A number of implementation issues are discussed, and a mapping that will enable the consistent storage and then analysis of xAPI verb/object/activity statements across different social media and online environments is introduced. A set of example learning activities are proposed, each facilitated by the Learning Analytics beyond the LMS that the toolkit enables.

Categories and Subject Descriptors

H.1.2 [**User/Machine Systems**]: Human factors; H.5.2 [**User Interfaces**]: User-centered design; K.3.1 [**Computer Uses in Education**]: Collaborative learning

Keywords

Connected Learning, xAPI, integration, data ownership

1. ENABLING CONNECTED LEARNING

Connected Learning (CL) which is underpinned by constructivist [4] and connectivist [8] learning theories, is a modern pedagogical approach holding that knowledge and learning is distributed across a social, conceptual network. It claims that when people forge, negotiate and nurture connections for themselves (between people, information, knowledge, ideas and concepts), learning is more powerful and sustainable. But, a question presents: how is connected learning (CL) to be realized? The experimental cMOOC idea pioneered in 2008 by Siemens and Downes [9] demonstrated that social media services can be harnessed to help achieve CL, with different tools combined by individual students to form their own personal learning environments (PLE) and personal learning networks (PLN) [6]. Such approaches allow for the creation of an "open participatory learning ecosystem" [2] which operates outside of the traditional LMS, but they do not come without their challenges. As more democratic, but at the same time more chaotic learning environments [4], cMOOCs provide a pedagogical model which organizations and vendors struggle to support. Traditional siloed approaches to learning analytics are difficult to use as most of the relevant data is created and stored outside of the LMS. Indeed, an analytic approach based only upon the data obtained inside a LMS would be next to irrelevant in a cMOOC as it would only be considering a tiny subset of the available data. This leaves us with a challenge; in order to fully understand learners' engagement in CL environments we must be able to access their data from ubiquitous social media, ethically.

In what follows we will introduce a simple toolkit that allows both learners and instructors to generate records of student learning, using well known social media tools. This toolkit is open source, and respects student privacy, as data cannot easily be captured from beyond the LMS without explicit student co-operation. Furthermore, the toolkit places an emphasis upon returning data to the student directly. Students can analyze their own learning as they desire, but the toolkit will also provide them with more sophisticated reporting capabilities in the future. We conclude by proposing some example learning activities that can assist students to develop data literacy using a case study of immediate and vital relevance; themselves.

2. THE CLA TOOLKIT

The Connected Learning Analytics (CLA) toolkit uses a Learning Record Store (LRS) as specified by the experience API (xAPI).[1] As depicted in Figure 1, the LRS allows for an interface to be developed between two key functionalities of the toolkit, one which sends data to a LRS, and one which extracts it from the LRS, processes it, and sends it to a dashboard which is designed to be OLA compliant [10]. Key to the advance of the CLA toolkit is the extensibility allowed by its use of the xAPI standard; any data can be collected and sent to the LRS if an appropriate scraper is built, but with careful design this extension need not interfere with the functionality of the current suite of reporting tools.

Data Scraping Tools: These interface with the APIs made available by the varying social media resources, or utilize a general scraping tool if no API is available (and this is legally allowed). Data scraping harvests the data relevant to a par-

[1]xAPI was released in 2013 and was released in 2013, and termed *Tin Can* in its development stage: www.adlnet.gov/tla/experience-api/

LAK '15 Poughkeepsie, New York USA

ACM 978-1-4503-3417-4/15/03 http://dx.doi.org/10.1145/2723576.2723627 ...$15.00.

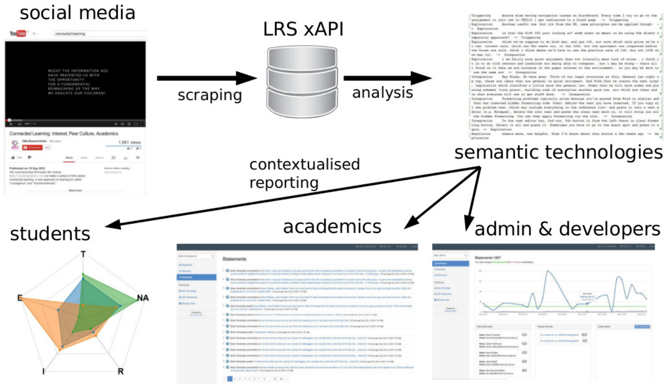

Figure 1: The basic schema of the Connected Learning Analytics Toolkit. ©Kirsty Kitto

ticular defined *learning activity* and sends it to a specified LRS using the required JSON format. A particular advance provided by the CLA toolkit is its provision of a standardized mapping that unifies such statements and so will allow for both cross comparison at a coarse level, or fine analysis at a lower one (see section 2.1.1). JavaScript and PHP based tools have so far been developed to interact with Youtube, Facebook, Google+, Google drive, Twitter, StackExchange and Wordpress. They use the API's of the relevant social media wherever available, and send statements as JSON objects to a nominated LRS.
Reporting Tools: These take data from a nominated LRS and perform analysis upon it. They then send relevant data to contextualized dashboards where it can be examined by learners, teachers, and administrators as appropriate. While different users can be catered to with this reporting functionality, it is important to appreciate that the CLA toolkit is designed with the learner at the center. Each of the learning activities proposed for the use of the toolkit (see section 3) require learners to take control of their data, allowing it to be scraped (or potentially choosing not to participate), and then taking key sensemaking roles in analyzing their data.

2.1 Implementation Details

Here we shall discuss the current form of the toolkit, and motivate some of the design decisions that have been made along the way. First and foremost, the decision to adopt xAPI lies at the center of the CLA toolkit, and we anticipate that this choice will facilitate its extension and interaction with other LA tools as the standard is more widely adopted. The xAPI stores data in an `actor`–`verb`–`object` syntax at its most basic, but can be extended with a wide range of extra information. The data is captured as a JSON object and then stored in a Learning Records Store (LRS). An example xAPI statement for the CLA toolkit is:

```
{"actor": {
  "mbox": "mailto:jeff@example.com",
  "name": "Jeff",
  "objectType": "Agent"
 },
 "verb": {
  "id":"http://activitystrea.ms/../create",
  "display": {"en-US": "Created"}
 },
 "object": {
  "objectType":"Activity",
  "definition": {
   "name": {"en-US":"Posted"},
   "description":{"en-US":"Created a tweet"}
  }
 },
 "result": {
  "response":{"@frank - this book might
              help you: example.book.com"}
 }
}
```

This example is from a function that interacts with the Twitter API[2] to retrieve data, in this case about a Twitter user Jeff. A separate database is maintained in our system which links Jeff's email (which is used as part of his unique identifier in the xAPI system) to his Twitter username. In this statement, we see Jeff tweeting to a peer @frank about a book he has found.

The JSON structure of the statement is apparent, along with its basic `actor`–`verb`–`object` syntax. The `actor` in this scenario is Jeff, and he has *created* (this is the `verb` in the example) an `object` which in this case is of the special type `activity`, which we have subclassified to be of type `posted`. We have also listed the text that Jeff tweeted as a `result` of this activity. All of the statements recorded by the CLA toolkit take this general form, but it is important that a general mapping between the different social media sources be created in order to ensure that the data generated by the toolkit can be integrated in later reports and analysis.

2.1.1 *Mapping Social Media Functionality*

Trying to unify the variety of different types of social media labels into one consistent framework that facilitates reporting immediately raises a number of key implementation issues. Is a post on Facebook the same as one on StackExchange? Is a `retweet` the same as a `share`? This issue has been arising across all serious attempts to implement an xAPI solution, and has led to ideas like registries[3] and recipes[4] which attempt to provide a unified approach. More broadly, statements made in one educational system (e.g. high school) should make sense in another one (e.g. university). Our ability to analyze different xAPI statements is likely to be highly dependent upon the manner in which they were stored. Thus, if a `text comment` is marked as such across a variety of different media objects and activities, then we are likely to be able to use the same semantic technologies in its analysis. A simple classification at this stage will save much effort at a later date. To this end, one of the key advances of this paper is a proposed mapping that can unify xAPI statements across multiple social media platforms, which is depicted in Figure 2. The primary differences between platforms can be captured through reference to what `object` is stored, and what `activity` was being undertaken at the point where it was stored.

2.1.2 *Matching Usernames to Actor IDs*

Student data from beyond the LMS will only be harvested in specific circumstances, and even then, only if a student allows for their social media usernames to be matched against the `actor` identifier that is used in the LRS. In the current implementation, a student must go to a web form and explicitly match their relevant social media usernames to their LRS ID. Thus, social media are not indiscriminately scraped

[2]https://dev.twitter.com/overview/api
[3]https://registry.tincanapi.com/
[4]http://tincanapi.com/recipes/

Verb	Social Media	Object	Activity
Created http://activitystrea.ms/schema/1.0/create	Facebook	Text/URL/Photo/Video/Files	Create
	Youtube	Video	Create
	Google+	Text/URL/Photo/Video/Files	Create
	Google docs	Text/URL/Photo/Video/Files	Create
	Twitter	Text/URL/Photo/Video	Posted
	WordPress	Site	Create
	StackExchange	Text/URL/Photo/Video	Ask
Commented http://adlnet.gov/expapi/verbs/commented	Facebook	Text/URL/Photo/Video	Comment
	Youtube	Text/URL	Comment
	Google+	Text/URL/Photo	Comment
	Google docs	Text/URL	Comment
	Twitter	Text/URL	Comment
	WordPress	Post/Comment	Comment
	StackExchange	Text/URL/Photo/Video	Answer
Viewed http://id.tincanapi.com/verb/viewed	Facebook	Text/URL/Photo/Video	Seen
	Youtube	Video	View
	Google+	Text/URL/Photo/Video	View
	Google docs	Text/URL/Photo/Video	View
	Twitter	Text/URL/Photo/Video	View
	WordPress	-	-
	StackExchange	Text/URL/Photo/Video	View
Shared http://activitystrea.ms/schema/1.0/share	Facebook	Text/URL/Photo/Video	Share
	Youtube	Video	Share
	Google+	Text/URL/Photo/Vide	Share
	Google docs	-	-
	Twitter	Text/URL/Photo/Video	Retweet
	WordPress	Post	Share
	StackExchange	-	-
Liked http://activitystrea.ms/schema/1.0/like	Facebook	Text/URL/Photo/Video	Like
	Youtube	Text/Video	Like
	Google+	Text/URL/Photo/Video	+1
	Google docs	-	-
	Twitter	Text/URL/Photo/Video	Follow
	WordPress	Post	Vote up
	StackExchange	Text/URL/Photo/Video	Useful
Disliked http://activitystrea.ms/schema/1.0/dislike	Facebook	-	-
	Youtube	Text/Video	Dislike
	Google+	-	-
	Google docs	-	-
	Twitter	-	-
	WordPress	Post	Vote down
	StackExchange	Text/URL/Photo/Video	Not Useful
Tagged http://activitystrea.ms/schema/1.0/tag	Facebook	Person/Page	Tag
	Youtube	Person	Tag
	Google+	Person/Page	Tag
	Google docs	-	-
	Twitter	Text/URL/Photo/Video	Mention
	WordPress	Post/Text	Tag
	StackExchange		
Hashtagged	Facebook	Text	Hashtag
	Youtube	Text	Hashtag
	Google+	Text	Hashtag
	Google docs	-	-
	Twitter	Text/URL/Photo/Video	Hashtag
	WordPress	Post/Text	Hashtag
	StackExchange	-	-

Figure 2: A unified naming schema for for xAPI statements in the CLA toolkit. No text indicates that the `action` associated with the relevant `verb` is not available in that social media platform. ©Kirsty Kitto

for data. This feature both avoids many of the data limits on interactions with common APIs and respects student privacy; we only take the data we need. As the student must give explicit permission for the data scraping, they know data is being collected about their activities and are therefore more likely to seek that data. We shall return to this key data ownership point towards the end of this paper.

2.1.3 The CLA toolkit is Open Source

The toolkit has been declared Open Source, under a GPL3 license. This will facilitate the future extension and development of the toolkit by the LA community as a whole. The source for the current version is available via GitHub (https://github.com/kirstykitto/CLAtoolkit).

2.1.4 Choosing a LRS

A LRS is required to use the toolkit. Although more are likely to become available as the xAPI is more broadly adopted, two solutions currently exist that are independent of a LMS: Learning Locker[5], which is an open source LRS championed by HT2, and distributed under a GPL3.0 license.; and SCORM Engine/Cloud & Watershed LRS which are commercial solutions developed by Rustici software.[6]

The Open Source nature of the CLA toolkit has made Learning Locker the obvious choice for a default LRS used by the toolkit. This decision is further supported by the planned development of personalized LRSs in the current Learning Locker roadmap.[7] However, we would like to emphasize that there is no *a priori* reason why Learning Locker must be used by anyone making use of the toolkit. Any other LRS could be used. Indeed, changing the LRS that the CLA toolkit uses is as simple as changing the endpoint for xAPI statements, which can be easily done using web based forms in the current implementation.

3. LEARNER RELEVANT ANALYTICS

Once the data is collected in the standardized format depicted in Figure 2, we see a number of options emerge to present the learners with highly relevant LA that they can use to investigate their own learning processes. It is important to note that analytics can be performed at a number of levels. For example, `verbs` can unify across all `comments`, but a finer grained reporting at the level of individual social media `activity` labels is possible, or data could even be aggregated over `object` types. This section will propose some simple learning activities *in the wild* (i.e. beyond the LMS and released into the WWW) that could be implemented using the CLA toolkit.

3.1 Critical Analysis of Media Contributions

One simple activity would involve a critical analysis of media content generated by a cohort. It could involve instructing students to create a video (as per some set of requirements) and then upload it on Youtube. They would then register the address of their video and their Youtube username on a webpage that would link their LRS account with that username. If students then provided constructive feedback to a subset of the cohort (a requirement for say, 5 comments minimum, could also be stated as a component of the task) then running the Youtube scraper over the list of usernames would pull the data into the relevant student LRSs. Students could then be required to perform an analysis task upon the data generated for their video, comparing these results with reports generated for the full cohort.

[5]http://learninglocker.net/

[6]http://tincanapi.com/lrs-orgs/lrs-for-orgs-home/

[7]http://learninglocker.net/features/roadmap/

3.2 Groupwork

Another prime set of activities that could be facilitated by the CLA toolkit revolve around groupwork. This is always a challenging feature of formal coursework, especially when it comes to the attribution of marks in summative assessment items. Disputes often arise, but students have rarely collected a dataset that they can use to substantiate their claims. Sometimes logbooks or wiki contributions are available, but more often than not these only form part of the picture. We anticipate that the CLA toolkit could be used in a multitude of ways to assist with resolutions in this area. For example, students could be encouraged to work in a Google Drive, which would mean that a count could be kept of how many contributions were made to the submitted documents by the different group members. Students could then be required to submit a reflection which considered their own contributions to the assessment item, when compared to the data obtained about the group as a whole. Thus, if one student was seen to make 70% of the updates to a document that was submitted by a group of 5 then the other students could be asked to justify their apparent lack of contribution.

Such a reflection activity could be made even more powerful if it considered multiple data sources. Thus, if students were required to run their group coordination in a private Facebook or Google+ community, then further data about group contributions could be scraped using the other CLA tools. This would provide a far more nuanced understanding of group dynamics. For example, we would not expect the student who had taken on a Project Manager role to contribute substantially to the final assessment item, but they ought to have been very active in the online community used by the group to coordinate the project. This demonstrates the importance of using the CLA toolkit to encourage student reflection within the context of the learning activity. If the raw counts are analyzed by a data analyst in the back end (i.e. with no reference to student context) then the toolkit is unlikely to facilitate CL; the students must be involved.

3.3 Professional Development and Portfolios

A third application of the toolkit centers around its facilitation of portfolio generation. Students (indeed anyone) can use it to track their impact in fields that pertain to their employability and/or future promotion opportunities. Thus, they might want to track data about their impact in the Twittersphere, or their assistance to others in their community using sites like StackExchange. The CLA toolkit allows them to track these contributions, and the convention depicted in Figure 2 ensures that these events will be stored in a form that allows for comparison across all of the social media platforms that they use. Thus, if someone wanted to demonstrate their role as an instigator of important and relevant conversations, then they might want to report statistics about the number of *triggering* events that they had generated in their wider Community of Inquiry (CoI) [5], and how well these were supported by likes, comments, shares etc. in that community. This would provide individuals with the tools to construct rich narratives about themselves and to place them in public portfolios as desired.

4. PRIVACY VS DATA OWNERSHIP

The potential advantages of LA in helping students to achieve good learning outcomes have been well documented, but likely pitfalls are also becoming more apparent. Two broad perspectives are possible when considering student data: privacy and ownership.

A *privacy* perspective emphasizes the problems involved with prying into the lives of students. It emphasizes security and risk, but often to the detriment of innovation. In contrast, an *ownership* perspective enables students to take control of their data. It teaches them the power of owning their data, understanding what it means, and using it as an enabler of opportunities. The CLA toolkit is designed from within an ownership perspective. We are of the view that the person who generates the data *should* be the one who owns it. While educational institutions will often collect data about students, we consider it essential that they return this data to students in an open form that those students can utilize as they like (and not in an aggregated form that blocks other analytical possibilities). Many of the ethical problems that arise from within the privacy perspective evaporate when students are given full access to their data. Indeed, a recent commentary on the xAPI[8] suggests that the narrative is starting to move from interoperability and towards data ownership, which places the CLA toolkit as a key enabler in this area.

The student control of student data that is facilitated in the CLA toolkit leads to a number of outcomes that are in accord with recent work about ethics in LA. For example, Pardo and Siemens [7] discuss a number of the privacy implications that have arisen in the era of big data, advocating a contextual approach with respect to information privacy: sometimes we want our information to be public, sometimes not. The CLA toolkit, by placing control of data back into the hands of the student (who ideally controls their LRS, but at the very least decides which social media usernames to link with it) facilitates a sophisticated approach to information privacy and future work will consider the manner in which students behave in this type of a data ownership regime.

5. CONCLUSIONS AND FUTURE WORK

Learning beyond the LMS is both desirable and achievable. Furthermore, learning beyond the LMS does not rule out the possibility of data capture and Learning Analytics. In this paper we have presented the Connected Learning Analytics toolkit, which is designed to facilitate both.

The CLA toolkit goes some way towards generating a standardized recipe[9] for storing messages about social media. This is an important contribution. Large scale take up of one unified format of xAPI statements from multiple sources would facilitate the development of OLA, but this is difficult to achieve without a standard expectation as to the form that the data will take. To date we have focused upon tools that extract data from standard social media and send them to a LRS. The next step will involve considering how this data can be returned to students in a way that helps them with sensemaking and the acquisition of data literacy. Immediate candidates for future work include the CoI framework [5] for understanding Communities of Inquiry, Social Network Analysis using something like the SNAPP framework [1] (and extending it with a semantic component), using the ELLI [3] approach to report upon learning power, and using Google Freebase to map out knowledge structures.

We feel it is important to re-emphasize the learner centered nature of the CLA toolkit. For us, adopting a stance that the generator of a data source has a right to access that data has led to a profound re-conceptualization of what LA might become. It has led to a loose coupling of as many tools and learning activities as can be imagined, in a wild and open world; beyond the LMS.

[8]http://makingbetter.us/2014/11/a-glance-back-at-xapi-a-look-ahead-at-data-ownership/

[9]http://tincanapi.com/recipes/

6. ACKNOWLEDGEMENTS

Support (financial, intellectual, and creative) of QUT's Transform program, its core staff, associated project students (especially the INB302 S2 2014 Capstone team: Jarrah Madill, Brian Fernandez, Clinton Redhouse, Nathan Osborne, James Pilkington, Ihsan Mujdeci who developed two early prototypes) and the numerous associated course teams and Fellows is also gratefully acknowledged, as is the ongoing support of the Institute for Future Environments.

7. REFERENCES

[1] A. Bakharia and S. Dawson. SNAPP: A bird's-eye view of temporal participant interaction. In *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*, LAK '11, pages 168–173, New York, NY, USA, 2011. ACM.

[2] J. Brown and R. Adler. Minds on fire: Open education, the long tail, and learning 2.0. *EDUCAUSE Review*, 43(1):16–32, 2008.

[3] R. Deakin Crick. Learning how to learn: the dynamic assessment of learning power. *Curriculum Journal*, 18(2):135–153, 2007.

[4] J. Dron and T. Anderson. *Teaching Crowds. Learning and social media*. Au Press, 2014.

[5] R. Garrison, T. Anderson, and W. Archer. Critical inquiry in a text-based environment: Computer conferencing in higher education. *Internet and Higher Education*, 2(2-3):87–105, 1999.

[6] J. Mott. Envisioning the post-LMS era: The open learning network. *EDUCAUSE Quarterly Magazine*, 33:1, 2010.

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

[8] G. Siemens. Connectivism: A learning theory for the digital age. Technical report, elearnspace, 2004.

[9] G. Siemens. Massive open online courses: Innovation in education? In *Open Educational Resources: Innovation, Research and Practice*, pages 5–15. 2013.

[10] G. Siemens, D. Gašević, C. Haythornthwaite, S. Dawson, S. Buckingham Shum, R. Ferguson, E. Duval, K. Verbert, and R. Baker. Open learning analytics: An integrated and modularized platform. Concept paper, SoLAR, 2011.