Shared learning data allowing blending of a variety of tools and a central point of data collection

S. Lariccia, G. Toffoli
Sapienza Università di Roma (ITALY)

Abstract

We will present the preliminary results of an investigation on the use of a variety of learning tools loosely integrated by the Next Generation Digital Learning Environment (NGDLE) of the Up2U EU-funded project. Students perform individual and group activities, both formal and informal ones, within a project of fixed duration, to accomplish a common task established by a teacher or chosen by the learners on their own initiative.

Learning-related data are fed to a central store by a popular LMS, by a collaborative platform supporting project-oriented and artifact-oriented learning and by more specialized tools able to track a set of less structured activities. Data are then processed in various ways to analyse meaningful correlations between activities mediated by different tools and are finally compared to information acquired in more traditional ways about some skill levels, before and after the project execution.

The investigation puts a special focus on actions related to the set-up and the coordination of a project group and on the activity aimed at finding educational resources on the web.

The tools in the NGDLE use the xAPI protocol to send statements of user actions. A common Learning Record Store (LRS) is provided by the NGDLE itself. Processing of analytics data is performed both with native functions of the LRS and with external tools being fed with exported data.

Keywords: formal and informal learning, learning analytics, learning with a variety of tools, project-oriented learning, artifact-oriented learning, xAPI, Jupyter.

1 INTRODUCTION

1.1 The Up2U Project

Up2U (acronym for “up to university”) [1] is a project, funded from the EU Horizon 2020 research and innovation programme, whose key objective is “to bridge the gap between secondary schools and higher education & research by better integrating formal and informal learning scenarios and adapting both the technology and the methodology that students will most likely be facing in universities”.

Up2U is a 36-month long project that kicked off in January 2017. It gathers 18 partners from 12 European countries, including national research and education networks (NRENs), traditional and open universities, infrastructure providers and commercial partners. It is coordinated by GÉANT, a second-level data network interconnecting several NRENs across Europe and outside it.

1.2 The NGDLE of Up2U

Up2U makes available a specific Next Generation Digital Learning Environment (NGDLE) that integrates formal and informal learning spaces for secondary school students and teachers who wish to develop and enhance their teaching and learning skills up to the university standards. In order to support universal use, the architecture of the NGDLE is modular, scalable and portable. It relies on open technologies and standard protocols wherever possible and is compliant with the latest GDPR regulations by design.

1.3 Project pilots and impact evaluation

The Up2U workplan includes a series of projects that aim to demonstrate the application of the NGDLE in different educational environments. These projects include the implementation of technopedagogical pilots in many European countries; they are important in promoting the use of the NGDLE, by addressing the unique needs of each country’s education system. By the end of 2019, pilots are expected to involve teachers and students from a thousand schools.
2 THE USE OF LA DATA FOR IMPACT EVALUATION

2.1 General objectives and evaluation methodologies

In order to achieve its objectives, Up2U must carry out monitoring and assessment tasks. Monitoring aims mainly to verify that the activities are progressing according to the planned schedule; it requires, among other things, the continuous collection of data on the access to the Up2U ecosystem and on the use of the deployed tools.

Assessment includes the evaluation of the quantity and quality of the activities done inside the pilots and of their impact; mainly of the impact of the teachers training, of continuous professional development (CPD) projects and of the use of the NGDLE on the quality and effectiveness of the teaching and learning processes; it addresses also the evaluation of the contribution of informal activities of the learners on the development of their skills and on their achievements inside the formal education path.

Tests, surveys and interviews, to be held before, during and after the course of the projects, will allow to perform a first level of assessment. Learning analytics (LA) represents a way of enriching the results of the assessment with a range of indicators, specifically referring to the processes, which can be computed by analyzing the behaviour of the projects’ participants at different aggregation levels.

In a wider and longer perspective, the assessment should benefit from the possibility of accessing the online materials that the learners experienced and the artifacts that they produced as part of their formal and informal learning activity, and to analyse them in automatic way using text analysis and other content analysis techniques and tools.

2.2 Levels of analysis supported by LA

2.2.1 Tracking and global monitoring of ongoing activity

A first objective of LA in Up2U is to collect and integrate in a raw data stream all available tracks of the activity performed by all subjects involved in the pilot projects; the collection concerns also the self-directed learning-related activities performed by individual subjects. This allows, just to start, to monitor the level of the activity going on in each class, each school, each group of schools, each regional or national community.

2.2.2 Getting pictures according to different aggregation dimensions

Next objective is to make possible some analysis of the data based on some broad categorizations, such as distinguishing activities carried out in formal contexts and in non-formal ones, those typical of individual work and of collaborative work, tracks left in the fruition of existing contents and those being evidence of a more active behaviour such as building an artifact. It should be possible to cross this type of analysis with that related to the aggregation of the individuals in communities at different levels.

2.2.3 Identifying activity patterns

A more challenging objective is to identify “patterns” in the activity of individuals and groups. In the first case, a pattern could be matched by a sequence of navigation steps inside a structured educational content or in searching the web, targeted to satisfying an assignment; in the second case, a pattern could be found in a set of actions done by the members of a group while planning and performing work targeted to the collaborative creation of an artifact.

2.2.4 Relating user activities to learners’ progresses

Finally, the most ambitious objective is to put in relation the data collected during the learning process, and the indicators being derived from them, at individual and group level, with the results of the assessment of certain skills which is performed before and after that process.
3 AVAILABLE DATA AND TOOLS FOR LEARNING ANALYTICS

3.1 Sharing LA data with xAPI
How and where should be stored data tracing educational activities and outcomes? During last years, under the pression of the major actors in the LA, xAPI (the Experience API) [2] has established itself as a protocol for the exchange of such data. xAPI specifies how to store statements and how to retrieve a collection of statements by performing a query on a specialized tool, the Learning Record Store (LRS).

3.2 The central store of the data
A common Learning Record Store (LRS) is provided by the NGDLE of Up2U. Currently, it is an instance of Learning Locker (LL), “the world’s most installed Learning Record Store. LL is a commercial product, but its community edition, which is open source, has been deployed. [3]

3.3 The sources of learning-related data
Several tools in the NGDLE use the xAPI protocol to send statements of user actions to the LRS.

3.3.1 Moodle
The Moodle VLE; Moodle keeps in an internal logstore records of user activity; it has a native reporting function that allows you to create reports, i.e. lists of logs (actions performed on the site), filtered and annotated according to some classification attributes; several deployments of Moodle inside the Up2U NGDLE are also equipped with an extension - or plugin - able to send xAPI statements to an xAPI LRS, which in our case is LL.

3.3.2 CommonSpaces
CommonSpaces (CS) [4]; it is a deployment of the CommonS platform, which was developed inside an Erasmus+ strategic action; CS, a platform supporting project-oriented and artifact-oriented learning, has a native analytics functionality, based on the activity stream concept; since its inclusion in the Up2U ecosystem, the CS analytics functionality has been extended in order to exploit the presence in the latter of an xAPI LRS.

3.3.3 Other tools
Some other tools are closely integrated into Moodle as plugins, like H5P, or through an LTI interface, like DSpace; others are loosely integrated by sharing with the other platforms the Single Sign On infrastructure and the same instance of LL as the store of the xAPI statements.

3.3.4 The CS Bookmarklet
The xAPI Bookmarklet [5] allows the user to almost seamlessly create a series of bookmarks while navigating the web in a browser window or tab. The CS Bookmarklet keeps the basic idea but sends page URL and Title to CS; CS adds the information identifying the “actor” and the “verb” of the xAPI statement and sends the basic statement (actor-verb-object) to the LRS. More details can be found in the CS help page Learning Analytics with xAPI in CommonSpaces. The CS Bookmarklet is a very simple tool. Its relevance lies in the fact that it allows to voluntarily trace an informal activity being done outside a VLE or another learning-dedicated platform.

3.4 The tools available for data analysis
As mentioned above, learning-related data are fed to a central xAPI store by a popular VLE, by a collaborative platform and by several other tools able to track a set of less structured activities.

Data are then processed with several tools to analyse meaningful correlations between activities mediated by different platforms and are finally compared to information acquired in more traditional ways about some skill levels, before and after the projects execution.
3.4.1 The LL Dashboards

The simplest way of getting meaningful views of the data collected in the LRS is to use the analytics dashboards provided by the Browser Interface of LL itself, that is by the administrative interface that is accessible through any web browser. The Browser Interface allows to:

- define Queries, to filter the statements based on the values of specific properties (slots) in their structure
- define Visualizations, to present aggregations and multiple series of data with graphical widgets like bar charts, pie charts, line charts, correlation charts, also
- build Dashboards, that is named groups of Visualisations, arranged according to a certain layout, that can be displayed both inside and outside the LL itself.

Very useful features of the Dashboards are the following ones:

- you can make a Dashboard shareable, allowing it to be embedded in external web platforms
- you can modify the URL of a shareable Dashboard, when publishing it in a web page, by adding query parameters, to perform on the fly custom additional filtering of the data displayed.

Moreover, from the browser interface of LL, you can export data, say in CSV format, to process them with external analytics tools; in doing that, you can use a Query to filter the data and a kind of projection operation defined through a simple mapping table.

3.4.2 External analytics tools

With Queries and Visualizations, possibly you can't do with all the processing you need, especially if this involves complex aggregation and correlation operations. In this case, as mentioned above, it is possible to export the results of a Query as a CSV file:

- you can visualize and post-process them with Excel or similar applications
- you could also upload them to a relational database, such as MySql and PostgreSQL, and/or analyse them with a variety of specialized statistical and business analytics applications.

3.4.3 Web services implemented by LL

Besides defining the structure and the semantics of a statement, the xAPI Specifications state that data are transferred via HTTP requests and responses; they define several interfaces, implementing the web services that a the service provider must provide to another application, the service consumer; some of them must be implemented by an xAPI compliant LRS such as Learning Locker. We are interested only in a couple of HTTP interfaces:

- the Statements interface; this is the only interface being (meticulously) described by the xAPI specifications; it deals with the storage and retrieval of the learning records, without addressing filtering strategies and performance
- the Aggregation interface; “the Aggregation HTTP Interface is more advanced ... and allows you to access MongoDB’s powerful Aggregation API for more custom filtration of statements”; this is intended to provide a communication channel of greater flexibility and higher performance.

4 SOME ONGOING EXPERIMENTS

4.1 Using Jupyter Notebooks

As will be illustrated below, in Up2U we did some experimentation of the use of Jupyter [6] as a tool for doing LA. Jupyter is one of the most interesting tools in the Up2U ecosystem; it was integrated in it, with the name of SWAN, together with CERNbox, a cloud data storage service. Jupyter is an interactive environment for exploring ideas and problems in the field of data analysis. It is document-oriented: a Jupyter document is called a Notebook. With Jupyter Notebooks it is relatively easy to process the filtered streams of xAPI statements returned by the HTTP interfaces of LL.

Jupyter uses Python as the default language kernel, but supports many other kernels, each corresponding to a different programming language. Python has been our choice since it is a modern
and elegant object-oriented programming language that provides substantial support to the Internet programming (programmatic web navigation and access to HTTP API).

Jupyter is an ideal environment for training and self-study [7], and the huge number of Notebooks on GitHub reflects its popularity. Some interrelated advantages of the Jupyter are:

- integration of text, code and graphics; it supports narrative, programming and presentation of results; by narrative we intend making explicit the objectives, documenting the code and commenting the results
- modularity; the units of a Notebook are cells, small chunks of text or code, which invite to modularity
- readability; learning Python is relatively easy since its code is readable by design
- sharing; the possibility of editing and saving the code, viewing the execution results and adding documentation allows to record the experiences made and share them with others [8]
- use for evaluation; a Notebook can be provided by a teacher to students as an assignment, that is a specification and/or a track of a task to be performed; a Notebook can be used by a student to develop the cues received, add the code and test it, document the code itself, include the results of the execution and return it to the teacher.

4.2 Doing Learning Analytics as a service

LA is a very vast field: very different data can come from a large set of sources; subjects with different roles and in different contexts can have different goals; there are many tools of very different types available, but their use is not easy and sometimes it is necessary to use a cascade of several tools. It follows that it is not easy to compile a guide suitable for all stakeholders, which enables them to obtain the data of interest on their own and to process them according to their needs.

As a consequence, it has been proposed that the workgroup addressing LA inside Up2U sets up a support service to help the pilot managers and other interested people with the use of LL and of related technologies. We mention below some areas where the support team could be of help.

4.2.1 Design of analytics Dashboards

Defining Queries and Visualizations is not so easy in LL. Moreover, in an educational institution comprising many teachers and learners interested in LA, it could be impractical to give to all of them accounts to access the browser interface of LL and to manage those accounts. The support team can act as a proxy; it can help others to understand the structure of xAPI statements and the role of their components; to be aware of the data available; to design Queries, that is filters for statements retrieval; to design Visualizations; to design and publish Dashboards.

4.2.2 Export of filtered data for further processing

The database of LL can grow very large. The size of the data to be exported for processing by external tools can be reduced by several order of magnitude by proper application of selection and projection operations; selection applies filtering criteria to the set of xAPI statements, while projection cuts-off unused data from inside each statement and performs a mapping from its tree structure to the tabular structure usually required by the export format. The support team can help in the design of both operations.

4.2.3 Development and use of Jupyter Notebooks

Jupyter Notebooks are a flexible solution for accessing LL through a choice of HTTP API. We noticed that they are also an ideal means to share ideas and experiences. They can be used as a bi-directional channel for communication between teacher and students and more in general between people with different levels of experience. A Notebook can be repeatedly exchanged between a “customer”, say a teacher, and a support person in the development of a small customized LA application, starting from a rough description of the requirements and ending with a fully working solution; then it can be executed as many times as desired, possibly with different parameters, and shared with other interested people, which in turn could ask and/or make further improvements or customizations.
4.3 Analyzing some activity patterns

One of our investigation lines puts a special focus on learner actions related to the set-up and the coordination of a project group and on the activity aimed at finding educational resources on the web. This is true especially for the activity done on CommonSpaces.

A typical scenario includes: the creation and management of a collaborative Project; the cataloguing and the vetting of OERs; the development of new OERs, possibly derived from existing ones; the creation of one or more Learning Paths (LP), where its creator can grant to Project members the role of contributor. Users are also allowed to create private LPs in a personal working area.

We are creating a few Notebooks to analyse patterns of activity and collaboration inside instances of said scenario, with reference to:

- the size of the LP, in terms of constituent Nodes, and its structure: Nodes can be organized in ordered sequence or in a tree structure
- the relationship between creation of original content and reuse of materials found on the web
- the work done by the Project members in finding interesting materials to be reused, being self-traced with the help of the CS Bookmarklet
- the degree of contribution of several authors to the same LP
- the relationship among the number of private LPs, LPs private to the Project group and published LPs.

The first provisional results seem to show that often users don’t fully exploit the collaboration opportunities being provided by the environment; but such kind of data could be exploited by Project coordinators to identify a potential problem and possibly to discuss it with the other members.

4.4 Trying different technical approaches

4.4.1 Using the Browser Interface of LL

Possibly we never had ventured to experiment the use of LL without the availability of a fast and easy entry path, like its Browser Interface. Its query functionality has clear limitations and the usability of its user interface is not satisfactory, but it has some strong points:

- the Statements Explore function allows to prototype Queries providing a fast feedback in real-time
- the shareable Dashboards are a great resource to publish sets of analytics visualizations based on always up-to-date data
- the export function comes in cascade with the filtering function and a very valuable projection function, allowing to perform a transformation of the data structures.

4.4.2 Using Postman and the HTTP API

We familiarized ourselves with the Statements Interface of LL (the xAPI interface) using Postman [9]; this is a powerful personal tool for interactively creating and sending HTTP requests. Incidentally, learning to use Postman could be a worthwhile investment for any person interested in exploiting the ever-expanding repertoire of free online tools accessible though HTTP API: some of them could be used to analyse also the context and the results of learning-related activities.

An interactive tool like Postman is intended to be used for testing, one at a time, the functions being exposed by a service provider through a HTTP interface, but it doesn’t provide a way of automatically chaining those requests, using the values contained in each response to set up the parameters of the subsequent request; this is typically a task supported by a programming language.

At this point, to take advantage of the opportunities offered by interfaces such as the Statements API or the Aggregation API of LL, and to overcome also some of their limitations, we moved on to the use of Jupyter. After testing basic requests to LL through Postman, we reimplemented and documented them with a few Jupyter Notebooks, adding a more varied repertoire of filters; a few examples:

- retrieving and analysing statements tracking the creation of artifacts on a certain platform, by a specific user in a certain time span
• retrieving and analysing statements tracking the contributions of several users to the
collaborative development of a specific artifact.

4.4.3 Choosing among different HTTP API
We mentioned above the fact that LL offers several HTTP interfaces to access its database without
relying on the functionality of the Browser Interface. At first, we tried only the use of the xAPI Interface,
due to the fact that it conforms to a standard and is quite simple.
As to the Aggregation Interface, at first, we were intimidated by the idea of using a tool based on a
non-relational database like MongoDB: it is very popular, but we are much more familiar with the
relational technology and SQL. However, we soon realized that it is possible to use these API without
having in-depth knowledge of MongoDB, and that the pipeline architecture on which its query system
is based is very attractive. Moreover, as it was expected, the Aggregation Interface is much more
performant.

4.4.4 Applying external analytics tools to the exported data
The Lithuanian partner of the project has done some interesting experiments in this field, applying
tools they are familiar with, such as Microsoft Power BI, to the impressive amount of data they are
already collecting from a network of schools participating in the Up2U Pilots.

5 FURTHER WORK
In the sections above we have dwelt mainly on work done, focusing many aspects related to the
introduction of the xAPI technology. Higher level issues related to the assessment methodology have
been faced by the project but are still object of revision. We will mention just a few of them.

5.1 Linking LA to the design of learning scenarios
In order to have the desirable amount and quality of data for doing monitoring and assessment, it is
convenient that a choice of activities and tools, able to generate those data, is done while designing
learning scenarios; these are models, inspired by the IMS notion of Learning Design [10], guiding the
design of specific courses and of other structured learning situations, by identifying general context,
main objectives, tools available, phases of activity, roles involved and typical interactions among them.
In Up2U we are defining a set of learning scenario, being proposed for the Up2U pilots and more in
general as a resource for teachers. In doing that, we are trying to enrich these models with the
identification of the user actions that it would be useful to track and of tools able to support the
tracking.
Learning scenarios, as exemplary learning contexts, take us to consider the notion of xAPI recipes; a
recipe can list typical actions (verbs qualified with activity types) for a class of similar scenarios. We
did some review of xAPI recipes proposed by some consortia, and of the vocabularies of verbs and
activity types on which they are based.
We also started comparing the xAPI vocabularies used by the NGDLE tools able to generate xAPI
statements, with the objective of evaluating the coherence of their vocabularies in the face of
comparable activities supported by the different platforms.

5.2 Applying content analysis techniques
The LA data being collected in the LRS as xAPI statements are related to actions carried out by the
learners in the use of training contents, in making potentially formative experiences, in producing
artifacts such as essays and remixed educational contents. But, in practice, they include at most
references to those contents, even if xAPI would be able to handle also many types of attachments.
The assessment should benefit from the possibility of accessing the online materials that the learners
experience and the artifacts that they produce as part of their formal and informal learning activity, and
to analyse them in automatic way, using text analysis and other content analysis techniques and tools.
Moreover, it could be desirable being able to apply similar techniques to analyze the answers of
teachers and students to some open questions aimed at knowing their perceptions of what they have
learned and factors they see as helping or hindering their learning.
We plan to start with simple document classification and with the computation of indices such as lexical richness, syntactic complexity and formal compliance with a standard document template; more challenging goals could relate to topic extraction, document similarity, text cohesion and topic focusing. Low-level text analysis operations will include most of those provided by Voyant Tools [11]. For the implementation, we will rely on the Python library spaCy [12], which is based on a very clean pipeline architecture, some state-of-the-art algorithms and language models covering most of the languages of the Up2U pilot countries.

5.3 Improving the modelling of learning activities with xAPI

One major problem in the application of xAPI lies in the structure and semantics of the statements; the core of an xAPI statement is the triple actor-verb-object; usually, identifying the actor is not a major problem, but

- typically, verbs are very abstract things; they are chosen from a restricted repertoire of words and carry with them only a very vague suggestion on the type of action being modelled
- on the contrary, there is a great deal of freedom in the choice of the object; in fact, the object can represent an educational content, an artifact under construction, a physical object or place, a performance, a social event, or even an individual or a group; it is an instance of an activity type.

As a consequence, in using xAPI often the verb semantics is overloaded (this term belongs to the slang of the object-oriented programming); the verb takes on a precise meaning only in combination with an activity type, in the same way as the “+” operator usually means very different things in combination with integers, strings, lists or sets. Then, which is the usefulness of filtering statements by verb only? Which cues can this provide, besides the fact that a high frequency of the verbs “created” and “submitted” can suggest a more active behaviour than the verbs “viewed” and “experienced”?

Probably, we should use a careful choice of xAPI recipes, as meaningful combinations of verbs and activity types in a certain context, to overcome this problem in designing useful filters for our analyses.

5.4 Coping with user identity and privacy

The Up2U project has dedicated a great deal of effort to create a safe and secure ICT learning environment for all the schools involved in the project pilots, by addressing the GDPR of EU. But, due to lack of experience in the specific field of LA, in planning its application we have not adequately addressed issues relating to privacy, identity and user profiles.

As a consequence, in future we will have to solve some inconsistencies in the tracking of activities with xAPI:

- in the statement format, we haven’t a shared means of identifying the actor in a global but anonymous way, so that currently it is very difficult to correlate activities done by the same user on different platforms; a few possible solutions have been envisaged, but the implications related to privacy must be better investigated
- we haven’t a reliable way of deriving, from the available data, information concerning the user affiliation(s), if any, although most xAPI statements include a lot of information on the contexts in which the actions are being performed, such as courses, projects, discussion forums, aso.

6 CONCLUSIONS

A leitmotiv in the evolution of the Internet is interoperability; often, interoperable systems are shown to be a good alternative to centralized ones. The xAPI specs, as a good example of interoperability protocol, could contribute to assemble the pieces of the puzzle created by the multiplicity and variety of data sources in doing LA.

In the Up2U project we had, at some extent, the opportunity to collect a consistent set of data from an extended experimentation. We evaluated a set of techniques for analyzing them and to present the results of the analysis. We also planned a range of services that can be offered to school principals and teachers in order to familiarize them with LA concepts and practices.

Since LA is a relatively new field - at least it was for us -, we followed in part a bottom-up approach for introducing it in Up2U, in the belief that until you don’t have a variety of available data and are able to
experiment with analytics tools, it can be difficult to get insights on how to exploit them. Now we need
to improve the methodology for using LA in more effective ways, by matching it better to the design of
learning scenarios and to the assessment needs.

We met some difficulties in using the activity tracking model inherent in the format of the xAPI
statements; don’t find around meaningful independent discussion about xAPI advantages and
shortcomings; found relatively low level of support of said protocol in the developer and user
communities of the most popular open source VLE such as Moodle and Open edX.

The overall feeling that we got from a variety of clues, is that in the educational community there is
scarce awareness of the need for the interoperability of learning tools and learning data; this in turn
can depend on the scarce attention being devoted to the exposition of the learners to a variety of
formal and informal learning situations, which could benefit most of such interoperability. As it
happened in the past, it seems that a higher level of interest for the compliance with educational
standards can be found in the producers of commercial tools for job training and in large corporations
that use them.

ACKNOWLEDGEMENTS

This paper is supported by European Union’s Horizon 2020 research and innovation programme
under Grant Agreement No. 732049 - Up2U (Up to University - Bridging the gap between schools and
universities through informal education).

REFERENCES

CHI ’18 Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems,
Paper no. 32, 2018. Retrieved from https://hal.archives-ouvertes.fr/hal-01676633/document on
May 2019.
University of the Netherlands, 2007. Retrieved from
https://pdfs.semanticscholar.org/34c3/69c5e700b9d87d31132eb7620270ff03d234.pdf on May
2019.
[12] Explosion AI, spaCy - Industrial-Strength Natural Language Processing in Python. Retrieved from

6394