Data integration: architecture for learning analytics

Johanna Velander\textsuperscript{1} and Nihad Subasic\textsuperscript{2}

\textsuperscript{1}Linnaeus University, Department of Computer Science and Media Technology, Faculty of Technology, Kalmar, Sweden \texttt{johanna.velander@lnu.se}, \url{https://lnu.se/personal/johanna.velander/}

\textsuperscript{2}KTH Royal Institute of Technology, School of Industrial Engineering and Management, Department of Machine Design, Mechatronics Division, Stockholm, Sweden \texttt{subasic@kth.se}, \url{https://www.kth.se/profile/subasic/}

1. Introduction

Learning Analytics (LA) is an emerging practice at a very early stage of adoption especially in Europe \cite{10}. In particular, LA, which incorporates Predictive Analytics (PA), is dependent on large and rich datasets for accuracy. Most LA solutions use data available from Learning Management Systems (LMS), and Student Information Systems (SIS). As such, the data used is limited to the student’s activities and interactions with these systems. This data might in turn be inadequate for basing analysis and predictions on. Many LA solutions are therefore looking beyond these systems to include data from other sources as well. Data such as social media analysis, library use, student behaviour based on access card swipes and IP address as well as other multimodal data from a variety of sensors etc could be included.

There are several motivations for integrating data such as scaling up LA projects, enabling improvements of new educational technologies, and making more accurate and fine-grained analyses based on a wider set of data. Recent research points to the limited focus on technical details of integrating data and the very limited use of data integration specifications \cite{8}. This study, therefore, aims to help close this gap in current research by identifying limitations and issues in data integration based on previous research efforts to inform and propose an architecture to enable further development and opportunities when scaling up LA projects.

The proposed architecture enables data integration of different data types from various educational technologies used in the learning environment. Drawing on principles of clean architecture \cite{6} the decoupling of business rules is informing the architecture design. As such a component is also independent of other external “layers” such as user interface, frameworks and database.

2. Background

There are several issues when adopting data interoperability specifications such as xAPI or Caliper. Issues such as data format interoperability \cite{8} and protocol compatibility between educational technology providers often limit the usage of data from these providers to their specific platforms. Technology providers that support such specifications adhere to these specifications to varying degrees or interpret these in subjective ways which result in additional work to identify specification non-compliance when attempting to use the data externally. Another issue for educational institutions to consider is their future dependence on a particular specification as a result of vendor lock-in. Future dependence on a particular specification would not be an issue if educational technology platforms were uniformly adopted and adhered to standardized specifications. However, in reality, there are many cases such as emerging and new technologies providing various degrees of support and large vendors often support one or the other of the two competing standards xAPI and Caliper. For example, Instructure Canvas supports Caliper while Moodle supports xAPI.
3 PROPOSED ARCHITECTURE

which in turn places restrictions on the procurement and integration of new technologies to adhere to this standard. [5].

However taking advantage of a specification whilst ensuring a sustainable yet flexible approach for an educational institution to approach new educational technologies and to allow research on data gathered by various methods and technologies requires a flexible approach which does not constrain institutions' choices of educational technology or analytics methods which are moving towards more interactive and user centred solutions[7].

Figure 1: High-level Architectural Overview

3. Proposed Architecture

The diagram in Figure 1 provides a high-level overview of a proposed architecture that deals with the issues of format interoperability and provides a specification-agnostic architecture. Different clients connect via different protocols using different data formats which are then mapped to an internal data model. Other clients e.g. your dashboard clients connect and the domain model is then mapped to the respective client’s required data format/model via whichever protocol they want.

There is a need to map the semantics of an event from the independent domain model. The format mappers will specify the granularity of data and could provide an opinionated object type which denotes the accuracy of data based on the mappability of the collected data. Of course, it should be possible to get the in any format required from this API.

4. A brief discussion and concluding remarks

As education is becoming increasingly dependent on data-driven practices the requirement on the educational institution to facilitate sustainable data modelling and yet be flexible towards adopting different technologies might increase. As such there will be higher demands on “in-house”
data processing skills as is foreseen in many other sectors \[1\]. Development in machine learning methods that are facilitating decision-making processes has recently stirred discourse about agency distribution between humans and AI. Discussions on ethical aspects of AI which are often applied in LA solutions are increasingly focusing attention on the contextual nature of policy, data and algorithms. In light of this more adaptable, interactive and flexible solutions are required to shift analytic power closer to the source, making it reflect the environment, the ethical approach and value base and the requirements of specific learning institutions. The next step is to work on a proof of concept that will test and validate the idea and detail the steps involved.

4.1. Workshop at NLASI2022

This workshop will explore opportunities and challenges of using student data for informing course design. More specifically a scenario based on real course design implemented in a Learning Management System (LMS) Canvas for an electro technique undergraduate course using blended learning has be presented at Workshop (WS) at NLASI2022.

As data is becoming increasingly available through digital learning environments there are ongoing research efforts to establish how this data can be used to improve learning and teaching. Learning analytics is at a very early stage of adoption in European countries \[9\] and especially in the Nordic countries. However data is often collected through LMS and other educational technologies and teachers having access to some of the data through the LMS interface consider how this data can be used to help them improve their course. However studies have shown that student acceptance of the use of their data is dependent on how it is used and with whom it is shared \(3\) \(2\) \(9\) \(4\).

4.2. Our experiments

In two experimental courses (KH1251 ht21 HM1006 vt22 at KTH) we tried to use LA and found a correlation between students’ activities in LMS Canvas and their grades in the final exam. Data from both experimental courses show the need for more detailed information about students’ activities in the LMS to be able to help them individually in the best way to achieve a passing grade at the end of the course.

All course related material like course literature, exercises, quizzes... have been embedded inside the course’s Canvas. Since all material has been accessible through the LMS course page students have not had to search anything outside the Canvas to pass the course. In the current version of LMS Canvas the students’ activities have been operationalized by the number of visits. We do not consider this as enough detailed information.

Our planned Learning Analytics Dashboard (LAD) would show students’ activities based on the number of visits and in cases where there is additional info supplied by other software like Möbius. The variables which are of interest for this project could be but are not limited to:

- Statistical data which is not available today
  - the number of visit for each internal page
  - the number of visit pdf files
  - the number of view of video clip for each clip
  - the number of view of all other pages connected to learning

Alongside the new statistical data that we would like to achieve there are other data which are already available today.

- Statistical data which is available today
  - the number of respond for each quiz
  - resources accessed and activities completed in Möbius.
4.3 Our plans for workshop (WS)

During the WS we would like to discuss following groups of questions:

- **Pedagogical questions**
  - “learning by doing”, individual adaptive learning
  - How are attitudes towards the use of data related to how useful students perceive data?
  - Expected results of completely automatized education — teachers perspective.
  - Early intervention based on available data.
  - How does learning management system affect teaching practice?
  - How can a teacher by course design help students with learning?

Pedagogical issues are closely linked to IT and ethical issues.

- **IT-related and Ethical issues**
  - IT-related technical questions xAPI or SQL — what is possible and who will do what?
  - Who has the right to use statistic data generated in a LMS during a course
  - Data that exists — what can it be used for?
  - Exercise in how to use data in a useful way — opportunities and challenges.

The goal of WS is to find a common language for a project in which we will get LAD as it is planned in our first version of the project. Otherwise we just collect different meanings and publish the theoretical description of the problem.

4.4 Workshop Method

Different types of data from the systems will be represented by “cards” that can be arranged and grouped to achieve certain objectives. Participants will be divided into groups that are encouraged to think about and discuss expected results, opportunities and challenges, consequences (intended and unintended) and ethical issues of aggregating the data. At the end of the session the groups will reconvene and upload their suggestions to a padlet.

The session ends with all participants discussing and reflecting on the outcomes of the exercise.

4.5 Expected results

Results from the workshop are expected to highlight tensions that arise when using student data for purposes of improving learning and teaching and the environment in which it occurs.

It is also our hope that results will contribute with insights and ideas for further discussions involving different stakeholders in the design of LA:

- teachers using LMS Canvas in their courses
  - representatives from KTH and LNU such as educational technologists
  - GDPR-responsible person at both KTH and LNU
  - somebody who has experience from using LA in MOOC-courses — if there is?
  - maybe a student representative (not sure but we can ask the Students Association to make a survey and ask students about their opinion in using LA in education?)
- researchers in LA
- IT-division
References


A. Appendix

Slides from Workshop June 14, 2022

NASI2022
Nordic Learning Analytics Summer Institute
13 - 14 June 2022
KTH, Stockholm
Welcome to LAD workshop

- LAD from Moodle – LNU-experience
- LAD in LMS Canvas
- KTH-experience
  - ht2021 (KH1251)
  - vt2022 (HM1006)
- Dialog – What we would like to know?

Workshop aims

- Discussion between different stakeholders
- Data only stored and not used
- Talking to teachers, educational technologists, students
- What to do with it, different ideas and different opportunities and risks with using data?
Possible implications of data analysis...

- **Feedback**
  - Majority of students appreciate the interactive tasks in the online tools
- **Social contact – discussion between students and teacher**
  - Majority of students engaged in the discussion forum
  - Require new knowledge?
- **What interactive tasks in the online tools give to teachers**
  - more time or increase effectiveness in teaching
  - more information about students' activities
  - more effective ways to support students

Description of courses KH1251 & HM1006

- Both are undergraduate courses in electro-technique
- All course material is in digital form but in different formats:
- All training tasks are:
  - embedded in LMS Canvas (optional, mandatory, exam)
  - in digital form (online) with automatic correction
All assignments can be traced in LMS

Students achieved result during the course
Class’s activities during the course

Re-exam period and Easter holydays

The case of KTH undergrad course in KH1251
**Electrical Measurements, Control Theory and Practice 6 hp**

<table>
<thead>
<tr>
<th>Item</th>
<th>Respondents</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1 Lilström</td>
<td>31</td>
</tr>
<tr>
<td>M2 Mörter</td>
<td>26</td>
</tr>
<tr>
<td>M3 Växelström</td>
<td>23</td>
</tr>
<tr>
<td>M4 Trefas</td>
<td>12</td>
</tr>
<tr>
<td>M5 Regelteknik</td>
<td>28</td>
</tr>
<tr>
<td>M6 Tväpol</td>
<td>8</td>
</tr>
<tr>
<td>M7 Mätgivare</td>
<td>2</td>
</tr>
<tr>
<td>Fyra LAB total</td>
<td>13</td>
</tr>
</tbody>
</table>
Feedback behind each question replaces a teacher in the classroom.

Statistics of a question

The statistical data of the question showed that the question was extremely difficult.
What did students click on?
Which click has the most impact?
Learning Analytics Dashboard – send warning

- Course structure
- Week1/Moment1
- W2 – M2
- W3 – M3
- ...
- ...
- ...

Class’s activities day by day

Would they study more efficiently?
Reflections on data

• Opportunities for improving course design
• See complexity of tasks – e.g. what is the reason for a majority failing a question?
• Opportunities for supporting students e.g. recommending resources based on performance
• Facilitate asynchronous communication e.g. forums
• Feedback for students, interactive features

Themes

Course design  Learning support  Changing teacher role
Themes

Challenges
• Collaboration between different stakeholders
• How will HEIs manage the need for more data science expertise in the future?

Future work
• Proof of concept data integration architecture
• Empirically study how data can support stakeholders in HEI.

Own ideas
•
•
•

Time to discuss!

Thank you!