Linked Data Generation for
Adaptive Learning Analytics Systems

Sven Lieber
sven.lieber@ugent.be

Ben De Meester
ben.demeester@ugent.be

Anastasia Dimou
anastasia.dimou@ugent.be

Ruben Verborgh
ruben.verborgh@ugent.be

Ghent University – imec – IDLab, Department of Electronics and Information Systems
Technologiepark 19
9052 Zwijnaarde, Belgium

ABSTRACT

According to the Learning Analytics (LA) reference model, LA is used to collect, explore and analyze diverse types and interrelationships of data. Specifications like the Experience API (xAPI) work towards interoperability with respect to interrelationship of diverse learning data. Algorithms for adaptive learning could be improved by incorporation of user-related data, not present in learning activities. Linking these user-related data with learning activity data would fully exploit the potential of interrelationships with data. Conventional solutions, as well as current Linked Data-based solutions focus purely on learning activity data, whereas solutions based on Linked Data could be used to integrate data of different domains. We propose a provenance-aware pipeline to transform xAPI learning activity statements to Linked Data. The integration of learning activities with other user data, provides a more complete set of user data, improving an adaptive learning analytics system. We use the proposed pipeline to build a Linked Learning Record Store based on the Resource Description Framework (RDF). SPARQL queries are used to link data about learning activities, enriched with fine-grained exercise descriptions, with data describing the abilities of users. In this paper, we show how Linked Data can be generated from xAPI statements in a streaming approach, based on existing tools and interfaces. Our solution demonstrates the usage of Linked Data to combine learning activity data with user ability data, to get a more complete set of user data aiming to assist in adaptive learning.

1. INTRODUCTION

Learning Analytics aim to collect, explore and analyze diverse types and interrelationships of data [1]. Specifications, such as the Experience API (xAPI)[1], lead towards analysis of different learning experiences. However, valuable data outside the Learning Analytics context are not considered with conventional solutions. Linked Data offers the functionality to analyze interrelationships between Learning Activity and other data.

Abilities of users adapt over time, while certain abilities might cover different domains, e.g., maths or reading. These user-centric skills offer valuable insights regarding learning

exercises and analysis of users'performance. By combining these kind of user data with learning activity data, adaptive learning is enabled. Developing a self-learning analytics system for adaptive learning is the purpose of the project LEarning analytics for AdaPtive Support (LEAPS, 2016-2018)[2]. Within LEAPS, multiple educational applications were fitted with xAPI logging, to provide a more integrated analysis. Analytical models can be re-used across different use-cases, due to the fact that the learning activities as well as users skills are modelled as Linked Data. Initial data capture has been performed on 18 classes of the first grade in Flemish schools, in 16 sessions and produced more than 2 million xAPI statements.

In this paper, we propose a data processing layer based on Linked Data on top of commonly used Learning Record Stores (LRS). To achieve that, we propose to use an Extract-Transform-Load (ETL) process to (i) read learning activity data expressed in xAPI from commonly used Learning Record Stores (LRS) in a streaming approach, (ii) apply use-case specific transformations to Linked Data, and (iii) load the Linked Data to a triple store for further consumption.

The remaining of the paper is organized as follows: First, we cover different Linked Data approaches regarding Learning Analytics in Section 2. Secondly, we present our integration model in Section 3 and our proposed Linked Data architecture in Section 4. We demonstrate interlinking of learning activity data with user-related data regarding abilities in Section 5. And finally we discuss our conclusion in Section 6.

2. RELATED WORK

Utilizing Linked Data principles for Learning Analytics is not a novel approach. Rabelo et al. [9] proposed a big data architecture using the xAPI ontology and Linked Data. However, they were using it for validation regarding xAPI specification conformance only, and they did not fully exploit the possibilities of Linked Data for Learning Analytics.

De Nies et al. [4] performed the first steps toward semantic interoperability of xAPI learning statements with other domains. They identified learning activities as provenance and

---

[1]https://xapi.com/

[2]https://www.imec-int.com/en/what-we-offer/research-portfolio/leaps_2
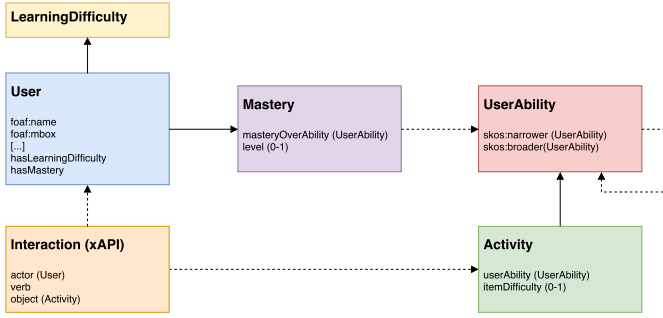
Figure 1: Our proposed integration model across user abilities and Learning Analytics. Available at http://leaps.ilabt.imec.be/specs/
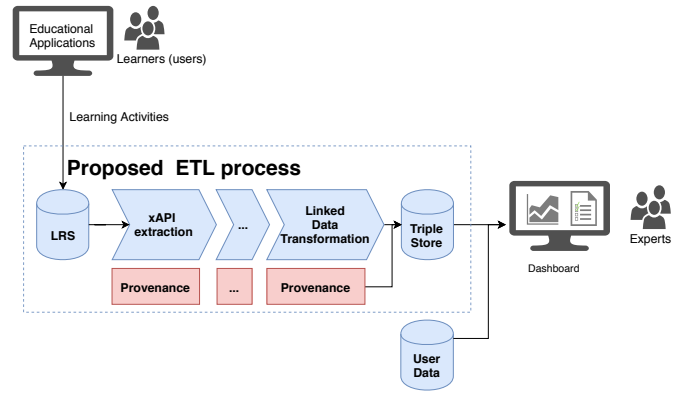

Figure 2: Our proposed ETL pipeline as part of the bigger architecture of the LEAPS project. The Linked Data produced by our proposed pipeline is interlinked with other user data. The xAPI statement extraction component streams newly added statements to one or more transformation components.

proposed the tool TinCan2PROV[3] to generate provenance-related Linked Data from xAPI statements. They relied on JSON-LD contexts to map xAPI terms to an xAPI ontology based on the specification. We are re-using their JSON-LD context within our proposed ETL process as one transformation step.

Anseeuw et al. [2] created an xAPI extension to store more context information regarding the physical environment of the learning activity. They also extended the work of De Nies et al. [4] concerning the xAPI ontology and mainly used it for reasoning. They defined additional services, based on SPARQL queries, which query learning activity data regarding the result and the provided context information. However, their work does not make use of data other than present in the learning statement.

De Meester et al. [3] created the Semantic ExerCise Interchange Format (SERIF[4]) to semantically describe learning exercises. The learning activity statements of the LEAPS project are described using the *SERIF* format as xAPI extension. Interoperable fine-grained description of exercises offer a variety of ways to interlink with user-related data. One transformation step of our ETL process makes use of the provided JSON-LD context of De Meester et al. [3] to generate fine-grained Linked Data describing the exercises. We are then able to properly link user-related abilities with exercises.

3. INTEGRATION MODEL

For this work, we introduce an integration model across user abilities and Learning Analytics (see Fig. 1). Whereas Learning Analytics mostly concerns user interactions with learning activities, we extend to a semantic model that integrates these interactions with user abilities, and is interoperable with Learning Analytics algorithms such as the ELO-rating [5]: as a user interacts with certain learning activities, it can improve its mastery over certain abilities. For example: when a student practices a lot of multiplication exercises, he/she can become better in the "multiplication-ability".

A *User* has – next to its personal data – other data that can be relevant for Learning Analytics, such as having certain *Learning Difficulties* which can be predicted based on

user-related data. This User has *Interactions* with *Activities*, which have a certain difficulty determined by e.g. analysis of data from large amount of students. In our case, these interactions are logged with the xAPI. These Activities are related to *User Abilities*, i.e., they train a User ability on a certain difficulty. User Abilities are general abilities that can be further categorized, for example, the "multiplication-ability" is more specific than the "math-ability". The categorization itself is out of scope of this work, but can be modeled using existing standards such as, e.g., the Simple Knowledge Organization System (SKOS) [8]. A User has a certain *Mastery* over these User Abilities. The level of Mastery can be dynamic based on, among others, the Interactions of the User with Activities. This change depends on the item difficulty of the activity, e.g., a user with mastery level 0.4 successfully solving an activity with item difficulty 0.8 has a larger impact than item difficulty 0.1 [5].

4. ETL ARCHITECTURE

Our proposed solution is a layer of Linked Data on top of commonly used Learning Record Systems. Therefore, we utilize an Extract-Transform-Load (ETL) process, where xAPI learning statements are extracted from LRSs, transformed to Linked Data and loaded into a triple store. Provenance of the pipeline's execution is generated and it is also provided as Linked Data.

The starting point of our proposed ETL process are xAPI learning activity statements (see Listing 3 for an example). They are stored in the open-source LRS LearningLocker[5], from which we extract the data using a streaming approach. Whenever new xAPI statements are inserted to the LRS, we are notified by the underlying storage engine (MongoDB) and start the transformation process. In case of the initial data capture of the LEAPS project, which has already happened, we feed a JSON dump of the learning activities batch-wise to our streaming pipeline.

The transformation process consists of multiple streaming components and therefore new transformations can be

---

[3]TinCan was the former name of xAPI: https://xapi.com/tin-can-experience-api-xapi/

[4]http://edutab.test.iminds.be/specs/serif/

[5]https://learninglocker.net/

plugged in. For the transformation of xAPI statements to Linked Data, we make use of the JSON-LD context of the TinCan2PROV tool introduced by De Nies et al. [4].

Other transformation steps can be added based on the use-case, e.g., if use-case specific xAPI extensions are part of the data. As the case for Anseeuw et al. [2] with their context extension or in other work of us [7], where a privacy usage policy regarding the consented use of the learning data is attached to the xAPI statement as extension.

Additionally to the transformation steps, our proposed ETL process keeps track of provenance, which aims to provide necessary information about the data's origin for later usage. The provenance information per component consists of (i) the name and version of the component; (ii) timestamps of start and end of processing; (iii) used entities (e.g. used JSON-LD contexts, input or output files); (iv) a link to the previous component; and (v) the type of activity (One of the eight sub-classes of the DataActivity class of the GDPRov[6] ontology). The last pipeline step transforms the provenance data to Linked Data, expressing it with the GDPRov[6] vocabulary. At the end of the transformation, the data, as well as collected provenance are available as Linked Data in RDF and are uploaded to a triple store (blazegraph[7]) to be queried with SPARQL. Figure 2 shows our proposed pipeline as part of the bigger LEAPS architecture.

5. THE LEAPS USE CASE

This section demonstrates the interlinking of learning activity data with user-related data, based on the motivating scenario of the ongoing LEAPS project.

Our proposed ETL process transforms learning activity data to Linked Data, with the goal of further data integration. The collected learning activity data are described using the xAPI specification and are using the *SERIF* extension to model the performed exercises. That exercise model allows to describe the difficulty of each exercise.

```
1  @prefix foaf: <http://xmlns.com/foaf/0.1/>.
2  @prefix leaps: <http://edutab.test.iminds.be/specs/datamodel/ontology.owl#>.
3
4  :ben foaf:name "Ben De Meester" ;
5    foaf:mbox "mailto:ben.demeester@ugent.be"@en;
6    [...] ; # more personal data
7    leaps:hasLearningDifficulty :dyslexiaA ;
8    leaps:hasMastery :benMastery .
9
10 :dyslexiaA a leaps:LearningDifficulty .
11
12 :benMastery a leaps:Mastery ;
13   leaps:masteryOverAbility :skillA ;
14   leaps:level "0.6"^^xsd:number .
15
16 :skillA a leaps:userAbility ;
17   skos:broader :math .
```

Listing 1: Linked Data in turtle format, describing abilities of a user regarding *skillA*.

User-related data describing abilities of users according to the integration model introduced in Section 3 (Listing 1), can be linked with learning activity data describing the exercises with the *SERIF* format (Listing 3). SPARQL queries

(such as the one described in Listing 2) can retrieve data which can serve a Learning Analytics algorithm as input.

```
1  PREFIX : <http://example.com/>
2  PREFIX foaf: <http://xmlns.com/foaf/0.1/>
3  PREFIX xapi: <http://semweb.mmlab.be/ns/tincan2prov/>
4  PREFIX lom: <http://data.opendiscoveryspace.eu/lom_ontology_ods.owl#>
5  PREFIX leaps: <http://edutab.test.iminds.be/specs/datamodel/ontology.owl#>
6
7  SELECT ?name ?masteryLevel ?educationalDifficulty
8  WHERE {
9
10   # Get learning activity user by email address
11   ?agent a xapi:Agent ;
12   foaf:mbox "mailto:ben.demeester@ugent.be"@en .
13
14   ?mastery a leaps:Mastery ;
15   leaps:masteryOverAbility :skillA ;
16   leaps:level ?masteryLevel .
17
18   # Get user data by email address
19   ?user foaf:mbox "mailto:ben.demeester@ugent.be"@en ;
20   foaf:name ?name ;
21   leaps:hasMastery ?mastery .
22
23   # Get educational difficulty of all learning activities
24   ?xAPIStatement xapi:actor ?agent ;
25   xapi:object ?object .
26
27   ?object xapi:definition ?objectDefinition .
28   ?objectDefinition xapi:extensions ?extension .
29   ?extension <http://edutab.test.iminds.be/specs/serif> ?serifExtension .
30   ?serifExtension lom:educationalDifficulty ?educationalDifficulty .
31   }
```

Listing 2: A SPARQL query, linking user ability data from Listing 1 and generated Linked Data from Listing 4

6. CONCLUSION

One of the goals of Learning Analytics is analysis of diverse types and interrelationships of data. We proposed a provenance-aware ETL process to transform xAPI learning activity statements to Linked Data, with the aim of linking learning data with other user-related data, to assist adaptive learning analytics. We demonstrated how semantic data regarding user abilities can be linked to learning activities. Linked Data allows to combine relevant user data with learning activity data in a seamless way and enable more sophisticated analytics. Combining learning activity data (which contains personal data) with other personal data might cause trouble in the light of data protection [6]. However, the semantic nature of our proposed solution facilitates also privacy-related tasks, as usage policies could be expressed semantically as well [7]. Additionally the provenance produced by our pipeline offers valuable insights in case of an audit.

7. ACKNOWLEDGMENTS

```
1  { "timestamp": "2017-01-03T12:26:37.450Z",
2    "actor": {
3      "objectType": "Agent",
4      "mbox": "mailto:ben.demeester@ugent.be" },
5    "verb": {
6      "id": "http://adlnet.gov/expapi/verbs/attempted" },
7    "object": {
8      "objectType": "Activity",
9      "id": "http://example.com/LezerGame/1/1.0/a/1-2-3-4-5",
10     "definition": {
11       "name": {
12         "nl-BE": "LezerGame 1-1-0-a, deel '1-2-3-4-5'" },
13       "type": "http://adlnet.gov/expapi/activities/assessment",
14       "extensions": {
15         "http://edutab.test.iminds.be/specs/serif": {
16           "type": "AssessmentItem",
17           "lom:educationalDifficulty": "0.8",
18           "itemBody": {
19             "value": "Kies welk koekje juist is." },
20           "item": [
21             { "type": "ChoiceInteraction",
22               "properties": [
23                 { "key": "maxChoices",
24                   "value": 1  } ],
25               "options": [
26                 { "label": "bes",
27                   "value": "bes" },
28                 { "label": "bal",
29                   "value": "bal"
30                 } ] } ] } } } },
31     "context": {
32       "contextActivities": {
33         "parent": [
34           { "id": "http://example.com/LezerGame/1/1.0/a",
35             "objectType": "Activity"  } ],
36         "category": [
37           { "id": "http://id.tincanapi.com/recipe/assessment/general/1"
38           } ] } } }
```

Listing 3: A xAPI learning activity statement using the SERIF extension. The statement represents one multiple-choice exercise of the LEAPS project.

```
1  @prefix xapi: <http://semweb.mmlab.be/ns/tincan2prov/>.
2  @prefix prov: <http://www.w3.org/ns/prov#>.
3  @prefix foaf: <http://xmlns.com/foaf/0.1/>.
4  @prefix leaps: <http://edutab.test.iminds.be/specs/datamodel/ontology.owl#>.
5  @prefix lom: <http://data.opendiscoveryspace.eu/lom_ontology_ods.owl#>.
6  @prefix xsd: <http://www.w3.org/2001/XMLSchema#>.
7  @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.
8
9
10 # Linked Data describing the whole statement
11 _:b12_b0
12     a xapi:Statement;
13     xapi:actor _:b12_b1;
14     xapi:object <http://example.com/LezerGame/1/1.0/a/1-2-3-4-5>.
15
16 # Actor part of xAPI statement
17 _:b12_b1
18     a xapi:Agent;
19     foaf:mbox "mailto:ben.demeester@ugent.be"@en.
20
21 # Object part of xAPI statement
22 <http://example.com/LezerGame/1/1.0/a/1-2-3-4-5>
23     a xapi:Activity;
24     xapi:definition _:b12_b6.
25
26 # Definition part of the xAPI object
27 _:b12_b6
28     xapi:name "LezerGame 1-1-0-a, deel '1-2-3-4-5'"@nl-be;
29     xapi:extensions _:b12_b7.
30
31 # The SERIF extension part of the statement
32 _:b12_b7
33     <http://edutab.test.iminds.be/specs/serif> _:b12_b8.
34
35 _:b12_b8
36     a leaps:AssessmentItem;
37     lom:educationalDifficulty "0.8"@en;
38     leaps:item _:b12_b9;
39     leaps:itemBody _:b12_b14.
40
41 _:b12_b9 a leaps:ChoiceInteraction;
42     leaps:option _:b12_b12, _:b12_b11, _:b12_b10;
43     leaps:properties _:b12_b13.
```

Listing 4: The result of Linked Data transformation from the xAPI statement from Listing 3 (removed unnecessary triples for readability). The JSON-LD contexts of De Nies et al. [4] and De Meester et al. [3] were used to generate Linked Data.

References

1. ISO/IEC 20748-1:2016. Information technology for learning, education and training – Learning analytics interoperability – Part 1: Reference model, 2016-12.

2. Jonas Anseeuw, Stijn Verstichel, Femke Ongenae, Ruben Lagatie, Sylvie Venant, and Filip De Turck. An ontology-enabled context-aware learning record store compatible with the experience api. In *Proceedings of the 8th IC3K Conference, Porto, Portugal*, pages 88–95, November 2016.

3. Ben De Meester, Hajar Ghaem Sigarchian, Tom De Nies, Ruben Verborgh, Frank Salliau, Erik Mannens, and Rik Van de Walle. SERIF: A Semantic ExeRcise Interchange Format. In *Proceedings of the 1st International Workshop on LINKed EDucation*, October 2015.

4. Tom De Nies, Frank Salliau, Ruben Verborgh, Erik Mannens, and Rik Van de Walle. TinCan2PROV: exposing interoperable provenance of learning processes through experience API logs. In *Proceedings of the Linked Learning Workshop*, pages 689–694, May 2015. URL http://www.www2015.it/documents/proceedings/companion/p689.pdf.

5. Arpad E. Elo. *The Rating of Chess Players, Past and Present*. Ishi Press, May 2008. ISBN 978-0923891275.

6. European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union*, L119:1–88, May 2016. URL http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2016:119:TOC.

7. Sven Lieber, Ben De Meester, Anastasia Dimou, and Ruben Verborgh. Privacy law compliance for learning analytics through provenance. In *International Provenance and Annotation Workshop*, 2018. Currently under review.

8. Alistair Miles and Sean Bechhofer. SKOS Simple Knowledge Organization System reference. W3c recommendation, W3C, August 2009. URL https://www.w3.org/TR/skos-reference/.

9. Thomas Rabelo, Manuel Lama, Ricardo R. Amorim, and Juan C. Vidal. Smartlak: A big data architecture for supporting learning analytics services. In *Frontiers in Education 2015*, pages 1–5, Piscataway, NJ, 2015. IEEE. ISBN 978-1-4799-8454-1. .