Learning Semantics

for learning dashboards supporting the transition from secondary to higher education

Co-funded by the Erasmus+ Programme of the European Union

STELA Erasmus+ project (562167-EPP-1-2015-1-BE-EPPKA3-PI-FORWARD)

"The European Commission support for the production of this publication does not constitute an endorsement of the contents which reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein."

Contents

Co-funded by the
Erasmus+ Programme
of the European Union

Chapter 1

Introduction

The semantic web could create a universal medium for the exchange of data. It would allow to smoothly interconnect personal information management, enterprise application integration, and the global sharing of commercial, scientific, and cultural data[1]. One point is especially key in the semantic web: the explicit definition of the meaning.

Also in the domain of teaching and learning, semantics is important as learning is not “isolated” but takes place in a larger learning ecosystem [5]. Students are doing activities in the online learning environment, they are assessed and obtain grades, etc. Therefore learning data and learning traces should have “meaning” also outside the particular context. The activity of the virtual learning environment recorded by the technology provider should be “interpretable” for the teacher offering the course. The “grade” of the student should make sense outside the context of the institute of the student, especially in the global world where the student will study and work in different countries. The domain of learning is however still far from the ideal of the semantic web. The learning ecosystem mostly consists of isolated islands, each operating rather independently [5]. As a result the learning traces and data often do not make sense outside the particular environment.

Gasevic highlighted that ontologies still have to be developed for the domain of learning analytics. He also points out that semantic technologies have the potential of making learning analytics more meaningful and ubiquitous [5].
An additional challenge is the application of learning semantics outside e-learning as most of the work on learning semantics and ontologies is focused on e-learning. The STELA project however also focuses on using ‘small’ data that is readily available at any higher education institute, so outside of the e-learning context.

Learning analytics standards are key in the development of learning semantics. Learning analytics standards support the transferability and reusability of learning analytics. The purpose of standards is to facilitate search, evaluation, acquisition, and use of learning traces both for human stakeholders as for automated software processes. Specifically, in order to achieve interoperability of data from learning environments, and including both small and big data, e-learning standards exist that make it possible to collaborate between different learning platforms. The use of data standards offers many advantages for the collection and exchange of data [4]. The "Learning Tools Interoperability" (LTI) standard (IMS Global Learning Consortium[7]), the AICC learning standard (Aviation Industry Computer-Based Training Committee) or Scorm (Sharable Content Object Reference Model, Advanced Distributed Learning (ADL)), supported by the US Ministry of Defense) make it possible to integrate external learning applications in a Learning Management Systems (LMS) or other learning environments.

[1] Semantic Web Activity Statement `https://www.w3.org/2001/sw/Activity`

Co-funded by the Erasmus+ Programme of the European Union

Two new standards for registering experiences within learning environments within the framework of learning analytics are currently gaining worldwide importance. When a student performs activities in a learning environment, they are logged on the basis of a standardized data format as events with a timestamp. Currently two different open standards, the IMS Caliper specification (developed by the IMS Global Learning Consortium) and on the Experience API (or xAPI) from ADL are under development. These open standards, defining rules for logging student experiences as activities, have one major advantage over closed standards: data collection and data sharing can be more easily verified or controlled. One of the current challenges is to include the open learning standards into learning applications (virtual learning environment and learning management systems) of higher education institutes. However, so far the open standards are not fully incorporated in the learning technologies used within higher education institutes. Higher education institutions can take a stronger position towards commercial suppliers of learning technology regarding the use of open standards if they take a common standpoint.
An additional challenge is that the current standards, and the xAPI standard in particular, are so far " lacking specific support for any student profile information" [4], which is a strong limitation within the context of the project.

The IEEE Standard for Learning Object Metadata (IEEE LOM standard) is a multi-part standard that specifies learning object metadata [9]. The standard is based on a data scheme that defines the structure of a metadata instance for a learning object. In the IEEE Standard for Learning Object Metadata a learning object is defined as "any entity—digital or non-digital—that may be used for learning, education, or training" [9]. The IEEE LOM standard was developed by the e-learning community to support interoperability of technological solutions and e-learning platforms in particular [10]. Jovanovic et al. concluded that IEEE LOM standard focuses on learning objects only and that that it does not yet provide sufficient support to reuse learning objects over different contexts [10].
On top of modelling learning objects as with the IEEE LOM standard, one could also focus on modelling the learning process itself, for instance by using the Educational Modelling Language. An example is the IMS Learning Design Specification (IMS-LD), which aims at interoperability and transferability of Learning Designs (LDs) [8]. But again, as stated in the goals of the specification ("The IMS Learning Design specification supports the use of a wide range of pedagogies in online learning."), it again has a particular bias towards online learning.
The standards on learning objects and learning designs fail to capture sufficient information on the context [10, 12]. Such contextual information is however required for the personalization of the learning process, which would require knowledge on the learners' needs, preferences, learning styles, etc. [10]. Therefore, research has been performed regarding ecological approaches connecting the models of learners to the learning object they interact with [10] and regarding learning object context ontologies [10] as an addition to the ontologies for learning objects and learning design for the learning object **context**. One example is LOCO-cite [10], which adds the required information for contextualization, such as user characteristics and user evaluation, to both learning objects and learning designs. Again, both the ecological approaches and the learning context ontology focus on the e-learning domain.

An important factor in learning analytics standards is the flexible use of verbs and learning objects, which make it possible to log a wide range of learning traces from an equally wide range of learning environments into a natural language. The **JSON data format** is a format that supports the readability of the statements in the data for both machines and humans. Using such a shared, easy readable format, makes it possible to compare learning experiences in different applications. Other advantages of the use of standards are a more efficient data collection (less work on the clearing of data), flexibility in access to the own data (these can be stored in a learning recored store that is kept in-house). Also academic research at higher education institutions would profit from interoperability between different learning solutions and standardized data that can be shared and compared.

Co-funded by the Erasmus+ Programme of the European Union

1.1 Focus of the project

From the beginning, the project took the approach to use data that is already available at most higher education institutes. The project additionally focused on sustainability and transferability of learning analytics solutions. One of the main goals was to develop learning analytics interventions that could be easily integrated within the existing practices in higher education institutes.
Two important choices from the onset of the project influencing the approach for learning semantics were:

- **Use the available data:** data should be used that is readily available at any higher education institution, and therefore using the data standards that are available.
- **Integrate with existing technologies**: the learning analytics applications developed should integrate with the existing technologies of the higher education institute to promote sustainability. A modular software design supports this integration, while still supporting transferability to other contexts.

1.2 Approach of the project

As the introduction of this report elaborated the current ontologies are focus on e-learning and therefore online learning traces rather than on the data typically available at any higher education institute. Moreover, learning semantics and ontologies are not yet fully integrated within university systems. While progress is made regarding the use of e.g. standards such as the xAPI standard for online learning traces, more work has to be done for the use of learning semantics for small data such as questionnaire output and study progress.

As the focus of the project was on the use of available data and the integration with existing technological solutions at higher education institutes, the project focused on the portability of the data (Chapter 2).

Co-funded by the Erasmus+ Programme of the European Union

Chapter 2

Portability of the data

The STELA project focused on the portability of the data. This portability is achieved in two ways:

1. the use of the JSON data format to enhance readability and therefore interpretability and re-suability of the data,
2. the contextual clarification of the data by including information of the peers. This choice is in line with the 'social comparison' dimension that is underlying all case studies in the project (see Output 6).

The data layer within the LASSI (feedback on learning and study skills), POS (feedback for aspiring students), and REX (feedback on academic achievement) dashboards developed within the STELA project, supports the portability of the data.

(The information of the POS dashboard in this chapter is based on the STELA project publication [2]).

2.1 JSON data format

The data layers of the dashboards developed within the project are separated from the presentation layers. The data layer of a dashboard offers the data of a particular data client in a JSON data format, when a authorized request is received.

Importantly, as visible from the examples below, all data is stripped from personal identifiers and characteristics that would allow for re-identification of the student based on the provided data only. Therefore, even if the data of student is either retrieved 'by accident' or intentionally (e.g. due to a brute force attack), the data can not be linked its owner (see pseudonymization section).

The first box below details an example of how the JSON data format was used in the LASSI dashboard [1, 3], and in particular for the case study transferring the LASSI dashboard to TU Delft. The second box details how the JSON data format was used in the POS dashboard [2], which operates in a challenging context of aspiring students not yet subscribed to a higher education institute.

Co-funded by the
Erasmus+ Programme
of the European Union

JSON data format for LASSI

The code below shows one example of the JSON data format for sharing the data from the LASSI questionnaire [6, 13].

The left side shows the **data of the student at hand**. The student is "identified" by a "hash". The connection between a particular student at TU Delft and the hash can only be made by TU Delft for those students who granted permission to make the connection. The JSON data of the student includes both the raw scores on the five LASSI scales (motivation, time management, (lack of) anxiety, concentration, and the use of test strategies) and the norm scores (A, B, C, D, or E) positioning the student within the norm group.

The right side shows the **group-level information of the peers** of the group that the student at hand belongs to ("id": "stela-tudelft-I"). For each of the five scales, the number of peers in the five norm groups (A, B, C, D, and E) is provided. This information allows to position the student with respect to his/her peers in the program for each of the five measured learning skills. This information supports the interpretation and thus the transferability of the information.

```
"student": {
    "hash": "afde065b34c0fc27cbfa5bf34",
    "score": {
        "motivation": 34,
        "time_management": 37,
        "anxiety": 35,
        "concentration": 37,
        "test_strategy": 36
    },
    "norm": {
        "motivation": "A",
        "time_management": "A",
        "anxiety": "A",
        "concentration": "A",
        "test_strategy": "A"
    }
```

```
"group": {
    "id": "stela-tudelft-I",
    "histogram": {
        "motivation": {
            "A": 35,
            "B": 31,
            "C": 39,
            "D": 21,
            "E": 27
        },
        "time_management": {
            "A": 35,
            "B": 29,
            "C": 32,
            "D": 20,
            "E": 37
        },
        "anxiety": {
            "A": 47,
            "B": 26,
            "C": 26,
            "D": 22,
            "E": 32
        },
        "concentration": {
            "A": 27,
            "B": 22,
            "C": 51,
            "D": 31,
            "E": 22
        },
        "test_strategy": {
            "A": 34,
            "B": 32,
            "C": 28,
            "D": 21,
            "E": 38
        }
    }
}
```


Co-funded by the Erasmus+ Programme of the European Union

JSON data format for POS

The code below shows a (translated) excerpt of the data of one participant of the positioning test in the JSON data format. This data format was used by the POS dashboard providing feedback to participants of the positioning test and the questionnaire accompanying the positioning test. The data shown is the data of one particular participant to the "11-ir-1" positioning test, which took place on the provided date ("dataPositioningtest": "2018/07/02") to the "burgerlijk ingenieur" positioning test ("nameTest").

The **data of the participant at hand** includes both information on the outcome of the *positioning test itself* ("scorePositioningTest": 9, "numberCorrect": 16, "numberTotal": 30, "numberBlanco": 7, "numberWrong": 7) and the information collected through the *questionnaire* (e.g. "scoreMotivation": "average", "scoreTimeManagement": "average", "scoreConcentration": "BLANK", "percentageMath": "65-70%", "adviceSecondaryEducation": "positive").

Moreover, the data of each student includes the **group-level information** of the "peers". In this case the peers are the other participants in the positioning test. The participant at hand is placed, based on the global positioning test score, in the low group compared to his/her peers ("globalGroup": "Low"). The groups are defined in the JSON data format as well, such that the required contextual information is provided ("globalLowLabel": "<10", "globalMiddleLabel": ">=10 en <14", "globalHighLabel": ">=14"). Next, the group-level information is provided, i.e. how many participants are in the different groups ("globalPeersLowNumber": 460, "globalPeersMiddleNumber": 313, "globalPeersHighNumber": 242).

```
{
"sessionGroup": "11-ir-1",
"dataPositioningTest": "2018/07/02",
"nameTest": "burgerlijk ingenieur",
"linkMoreFeedback": "https://www.ijkingstoets.be/content/feedback_docs/11-ir-1.pdf",
"nameStudyProgram": "burgerlijk ingenieur",
"scorePositioningTest": 9,
"scoreMotivation": "average",
"scoreTimeManagement": "average",
"scoreConcentration": "BLANK",
"percentageMath": "65-70%",
"adviceSecondaryEducation": "positive",
"numberCorrect": 16,
"numberTotal": 30,
"numberBlanco": 7,
"numberWrong": 7,
"globalGroup": "Low",
"globalLowLabel": "<10",
"globalMiddleLabel": ">=10 en <14",
"globalHighLabel": ">=14",
"globalPeersLowNumber": 460,
"globalPeersMiddleNumber": 313,
"globalPeersHighNumber": 242,
"globalPastLowAAHighPct": 10,
"globalPastLowAAMiddlePct": 45,
"globalPastLowAALowPct": 44,
"globalPastMiddleAAHighPct": 33,
"globalPastMiddleAAMiddlePct": 39,
"globalPastMiddleAALowPct": 27,
"globalPastHighAAHighPct": 63,
"globalPastHighAAMiddlePct": 25,
"globalPastHighAALowPct": 12
}
```

Finally, the data contains the **group-level information to connect the outcome of the positioning test to academic achievement** of students at the end first year based on the data of earlier cohorts. E.g. "globalPastLowAAHighPct": 10, "globalPastLowAAMiddlePct": 45, and "globalPastLowAALowPct": 44, indicate that for participants of earlier cohorts who were in the low group based on their positioning test score, 10% obtained a high academic achievement at the end of the first year, 45% obtained an average academic achievement, and 44% a low academic achievement.

Rec. 1: Context matters.

As the context is important to interpret the data, data formats should allow to include information about the context. An important part of the context is understanding how a student positions with respect to the peers in the same module, course, or program. The JSON data format used in the project integrates the information of how peers performed at a group level in each of the individual data files. This allows to understand the positioning just from the data of one student, rather than from all individual data files.

2.2 Pseudonymization

The dashboards implement the conceptual framework for de-identification for learning analytics described by Khalil & Ebner [11].

- In case of the **POS dashboard**, test and survey results are collected and processed by a shared organization trusted by the participating universities. Before delivery to data warehouse of the dashboard, these data are stripped from personal identifiers and characteristics that allow for straightforward re-identification. A pseudonymization process replaces the identifiers with a surrogate key of the format `10ir-9c7s-41jn-18` (dummy example), the **'feedback code'**. In addition to a prefix and optional spaces or dashes as chunk delimiters, the key format contains ten random alphanumeric digits (a-z; 0-9), resulting in 36^{10} possible combinations.
- In case of the **LASSI dashboard @TU Delft**, students receive a "hash", which only allows TU Delft to re-identify the students. Next, the questionnaire is administered and processed by KU Leuven, using the "hash" (e.g. e86ba5c13e91d4883eee8ad3d) as only information. The hash consists of 25 alphanumeric characters, where the first character ("e" in the example) identifies the program of the student.

2.3 Authentication and access

In case of the **POS dashboard** [2], a multi-institutional feedback dashboard for aspiring students, a valid authorized request is received when a valid feedback key is provided.
In case of the **LASSI and REX dashboard deployments at KU Leuven**, a request is authorized using the Shibboleth authentication of the university.
In case of the **LASSI** deployment at TU Delft a double approach was used:

- The LASSI questionnaire was administered and processed by KU Leuven. TU Delft students received a unique url to the questionnaire with a "hash" that ensured pseudonymization. The

Co-funded by the Erasmus+ Programme of the European Union

data layer at KU Leuven offered the data using a JSON data format upon reception of a valid hash.

- TU Delft "harvested" the data from the data layer of KU Leuven using the hashes they provided to students. TU Delft connected the hashed data to the particular students. The data was offered to students in their own proprietary environment, when a request authorized using Shibboleth authentication was received.

Data layer for LASSI @TU Delft

The data from the LASSI questionnaire, processed by KU Leuven, was made available using a "hash" trough an online interface. Upon receipt of an authorized hash, the data is made available through a personalized url such as: `https://learninganalytics.set.kuleuven.be/stela-tudelft-lassi/data/afde065b34c0fc27cbfa5bf34.json`.

Data layer for POS

The data from the positioning test and the questionnaire accompanying the positioning test is processed by a a shared and trusted organization, is made available using a "feedback code" trough an online interface. Upon receipt of a valid feedback code, the data is made available through a personalized url such as: `https://feedback.ijkingstoets.be/ijkingstoets-11-ir/data.php?fc=11ireb828705f5`.

Outside of the intended dashboards' scope, this approach provides each of the participating institutions (or any other party) a possibility to request the feedback data for a given student through an automated process, given that this student chooses to share the authorization (feedback code, Shibboleth login) with the institution. This possibility demonstrates future **portability of the feedback data** and how students could be put in control of such portability in a very lightweight manner. The approach allows to give student control to their data. They can decide who (e.g. an higher education institute or an external service provider) with the "key" to their data.

Rec. 2: Provide students the "key" to their data

Students should be the "owners" of their data and should get the "key" to their data. They should be able to decide to share the key to their data (and retract that permission) to for instance an higher education institute or even an external service provider. *(output 5)*

Any higher education institute could implement their own learning analytics instruments consuming the available information, supporting the transferability of the offered learning analytics solutions. This was especially an asset for the POS dashboard, focusing on aspiring students. Although these students are not enrolled at a university yet, the proposed approach still allows each involved university to implement an own feedback instrument.

Remark that the presence of the group-level information in each individual data file is key for interpreting the data and the context around this data. This allows to get a first grasp of the context from a single data file of one individual, without the need to have all single data files of the entire cohort. Including the group-level information will provide additional meaning to the data, even for institutes with whom the data client is sharing their data with in a later stage.

Co-funded by the Erasmus+ Programme of the European Union

Chapter 3

Conclusion

The STELA project took a pragmatic approach regarding semantics as the state of the art has a strong focus on e-learning, rather than regular on-campus education, and the ontologies are not yet fully integrated within the universities daily practices.
To ensure the **portability** of the data, a JSON data format was used that has shown to allow to support two challenging scenarios:

1. the sharing of pseudonymized data between two higher education institutes (LASSI dashboard @TU Delft case study) and
2. the sharing of pseudonymized data of aspiring students, not yet enrolled at any higher education institute, between higher education institutes (POS dashboard case study).

To support the **interpretability** of the data, contextual information, and the position within the peer group, is added inside the data of each data subject. This allows to read the position of the data subject at hand within the peer population (social comparison), without the need to read all the individual data files of all data subjects. This further supports the portability of the data, as this contextual information can be read from every single data file. Therefore, each data client can share the information with other instances (such as a future higher education institute or external technology providers) without loosing this contextual information regarding the position within the peer group.

The approach taken of portability will profit from further inclusion of standardization and learning ontologies provided that higher education institutes further include them into their institutional practices. In addition, further research is required regarding if the existing standards and ontologies, bias towards online learning, provide sufficient power. If not, further research is required to further develop the ontologies and data standards.

Co-funded by the Erasmus+ Programme of the European Union

Authors

Tinne De Laet
tinne.delaet@kuleuven.be
KU Leuven, Belgium

Tom Broos
tom.broos@kuleuven.be
KU Leuven, Belgium

Philipp Leitner
philipp.leitner@tugraz.at
TU Graz, Austria

Martin Ebner
martin.ebner@tugraz.at
TU Graz, Austria

Jan-Paul van Staalduinen
J.P.vanStaalduinen@tudelft.nl
TU Delft, Netherlands

Bibliography

[1] Tom Broos, Lauri Peeters, Katrien Verbert, Carolien Van Soom, Greet Langie, and Tinne De Laet. *Dashboard for actionable feedback on learning skills: Scalability and usefulness*, volume 10296 LNCS. 2017.

[2] Tom Broos, Katrien Verbert, Greet Langie, Carolien Van Soom, and Tinne De Laet. Multi-Institutional positioning test feedback dashboard for aspiring students lessons learnt from a case study in Flanders. In *ACM International Conference Proceeding Series*, pages 51–55, 2018.

[3] Tom Broos, Katrien Verbert, Carolien Van Soom, Greet Langie, and Tinne De Laet. Dashboard for actionable feedback on learning skills: how learner prole affects use. In *Springer Lecture Notes in Computer Science (LNCS) series. (Proceedings of the ECTEL 2017 conference; ARTEL workshop)*, pages 1–15, Tallinn, Estonia, 2017.

[4] A. del Blanco, A. Serrano, M. Freire, I. Martinez-Ortiz, and B. Fernandez-Manjon. E-Learning standards and learning analytics. Can data collection be improved by using standard data models? In *2013 IEEE Global Engineering Education Conference (EDUCON)*, pages 1255–1261. IEEE, 3 2013.

[5] Dragan Gasevic. Semantic Technologies in Learning Environments -Promises and Challe. . . , 2009.

[6] H&H Publishing H&H Publishing Company and Clearwater Florida 33765 Inc., 1231 Kapp Drive. LASSI Dutch version, 2017.

[7] IMS Global Learning Consortium. Learning Tools Interoperability | IMS Global Learning Consortium.

[8] IMS Global Learning Consortium. Learning Design Specification, 2003.

[9] Institute of Electrical and Electronics Engineers., IEEE Computer Society. Learning Technology Standards Committee., and IEEE-SA Standards Board. *IEEE standard for learning object metadata.* Institute of Electrical and Electronics Engineers, 2002.

[10] J. Jovanovic, C. Knight, D. Gasevic, and G. Richards. Learning Object Context on the Semantic Web. In *Sixth IEEE International Conference on Advanced Learning Technologies (ICALT'06)*, pages 669–673. IEEE.

[11] Mohammad Khalil and Martin Ebner. De-identification in learning analytics. *Journal of Learning Analytics*, 3(1):129–138, 4 2016.

[12] Gord McCalla. The Ecological Approach to the Design of E-Learning Environments: Purpose-based Capture and Use of Information About Learners. *Journal of Interactive Media in Education*, 2004(1):3, 5 2004.

[13] Claire E. Weinstein, Stephen A. Zimmermann, and David R. Palmer. *Assessing learning strategies: The design and development of the LASSI.* Academic Press, 1988.

Co-funded by the Erasmus+ Programme of the European Union