Deliverable 1.2

LA Analysis

Author(s): Tarabanis Konstantinos (UOM)
           Tambouris Efthimios (UOM)
           Zotou Maria (UOM)

Editor(s): Tarabanis Konstantinos (UOM)

Responsible Organisation: University of Macedonia (UOM)

Version-Status: V1 Final

Submission date: 31/03/2016

Dissemination level: PU
Deliverable factsheet

Project Number: 562236-EPP-1-2015-1-EL-EPPKA3-PI-FORWARD
Project Acronym: PBL3.0
Project Title: Integrating Learning Analytics and Semantics in Problem Based Learning

Title of Deliverable: D1.2 – LA analysis
Work package: WP1 – PBL_LA educational approach construction
Due date according to contract: 31/03/2016

Editor(s): UOM
Contributor(s): OUNL, BOC, AAU
Reviewer(s): OUNL
Approved by: All Partners

Abstract: This deliverable presents an overview of the Learning Analytics (LA) domain. For this purpose, a systematic review of the relevant state of the art was carried out.

The research areas included in this study contain LA terms and objectives; LA lifecycle steps and data processing on each step; LA frameworks; ethical issues and limitations; existing approaches and practices of LA in educational and training settings; and existing approaches and practices of LA in PBL-designed educational and training settings.

Keyword List: Learning Analytics, Problem Based Learning, lifecycle.
## Consortium

<table>
<thead>
<tr>
<th>Role</th>
<th>Name</th>
<th>Short Name</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Coordinator, academic partner</td>
<td>University of Macedonia</td>
<td>UOM</td>
<td>Greece</td>
</tr>
<tr>
<td>2. Technology enhanced learning expert</td>
<td>Open University of the Netherlands</td>
<td>OUNL</td>
<td>Netherlands</td>
</tr>
<tr>
<td>3. PBL expert</td>
<td>Aalborg University</td>
<td>AAU</td>
<td>Denmark</td>
</tr>
<tr>
<td>4. Semantic and Learning analytics expert</td>
<td>University of Alcala</td>
<td>UAH</td>
<td>Spain</td>
</tr>
<tr>
<td>5. IT partner</td>
<td>BOC</td>
<td>BOC</td>
<td>Austria</td>
</tr>
</tbody>
</table>
Revision History

<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
<th>Revised by</th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>01/03/2016</td>
<td>UOM</td>
<td>Circulation to all partners for comments and contribution</td>
</tr>
<tr>
<td>02</td>
<td>09/03/2016</td>
<td>OUNL</td>
<td>Submission of comments and contributions</td>
</tr>
<tr>
<td>03</td>
<td>15/03/2016</td>
<td>BOC</td>
<td>Submission of comments and contributions</td>
</tr>
<tr>
<td>04</td>
<td>16/03/2016</td>
<td>AAU</td>
<td>Submission of comments and contributions</td>
</tr>
<tr>
<td>05</td>
<td>21/03/2016</td>
<td>OUNL</td>
<td>Submission of contribution on LA in MOOCs</td>
</tr>
<tr>
<td>06</td>
<td>28/03/2016</td>
<td>UOM</td>
<td>Synthesis of all received feedback</td>
</tr>
<tr>
<td>1</td>
<td>31/03/2016</td>
<td>UOM</td>
<td>Final edition of the deliverable, Submission to EACEA</td>
</tr>
</tbody>
</table>

Statement of originality:
This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Disclaimer:
This project has been funded with support from the European Commission. This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.
Table of Contents

DELIVERABLE FACTSHEET ..................................................................................................................2
CONSORTIUM .................................................................................................................................3
REVISION HISTORY ..........................................................................................................................4
TABLE OF CONTENTS .......................................................................................................................5
LIST OF FIGURES ............................................................................................................................8
LIST OF TABLES ..............................................................................................................................9
LIST OF ABBREVIATIONS ................................................................................................................10
EXECUTIVE SUMMARY ..................................................................................................................11

1 INTRODUCTION ............................................................................................................................13
  1.1 SCOPE ......................................................................................................................................13
  1.2 AUDIENCE ..............................................................................................................................13
  1.3 STRUCTURE .............................................................................................................................13

2 METHODOLOGY ............................................................................................................................15

3 LEARNING ANALYTICS OVERVIEW .........................................................................................18
  3.1 LEARNING ANALYTICS DEFINITIONS ....................................................................................18
  3.2 LA DIRECTIONS .........................................................................................................................20
    3.2.1 Learning Analytics on an academic level .............................................................................21
    3.2.2 Learning Analytics on an individual level ...........................................................................21
  3.3 LEARNING ANALYTICS FRAMEWORKS .................................................................................22
    3.3.1 LA framework by SOLAR .................................................................................................22
    3.3.2 LA framework by the OU of the Netherlands .................................................................23
    3.3.3 LA Reference Model .........................................................................................................25
    3.3.4 LA Framework for Educational Virtual worlds ..............................................................27
  3.4 LA STAKEHOLDERS ..................................................................................................................28
    3.4.1 LA Practitioners ................................................................................................................28
    3.4.2 LA Decision makers .........................................................................................................29
  3.5 LA METHODS .............................................................................................................................30
  3.6 LA LIFECYCLE ...........................................................................................................................33
    3.6.1 LA Model by Siemens ......................................................................................................34
    3.6.2 LA lifecycle by LTEE .......................................................................................................36
    3.6.3 LA model by Elias ............................................................................................................37
    3.6.4 LA process model by Verbert et al. ..................................................................................38

4 EDUCATIONAL REFERENCE PROCESS FRAMEWORK ..............................................................40
  4.1 THE INTERNATIONAL REFERENCE PROCESS FRAMEWORK ..................................................40
    4.1.1 The structure of the Reference Process Framework .........................................................40
5 LA TOOLS

5.1 GENERAL PURPOSE LA DASHBOARDS

5.1.1 GLASS tool

5.1.2 CourseVis tool

5.1.3 Google Analytics

5.2 LA STANDALONE TOOLS

5.2.1 SNAPP

5.2.2 LeMo application

5.2.3 StepUp! LA tool

5.2.4 LOCO-Analyst

5.2.5 Netlytic tool

5.2.6 eLAT toolkit

5.3 LMS-BASED TOOLS

5.3.1 Gismo Moodle add-on

5.3.2 MOCLog Moodle add-on

5.3.3 Lae-R Moodle add-on

5.3.4 SmartKlass Moodle add-on

6 EXISTING APPLICATIONS OF LA

6.1 LA APPLICATIONS IN HIGHER EDUCATION INSTITUTIONS

6.1.1 Open University of the United Kingdom (UK) case study

6.1.2 Nottingham Trent university case study

6.1.3 University of Bedfordshire case study

6.1.4 Purdue university case study

6.2 LA CASE STUDIES IN BUSINESS

6.3 LA CASE STUDIES IN MOOCs

7 EXISTING APPROACHES ON USING LA IN PBL-ENVIRONMENTS

7.1 LA AND PBL IN MATHEMATICS AND STATISTICS

7.1.1 Methodology

7.1.2 Data collected and LA performed

7.2 DOCTORS TRAINING WITH ANALYTICS AND PBL

8 ETHICAL ISSUES AND CHALLENGES

8.1 ETHICAL ISSUES

8.2 LIMITATIONS AND CHALLENGES

8.2.1 Limited LA practices

8.2.2 Objectivity of LA interpretations

8.2.3 LA software

9 LA CONCEPTS IN PBL3.0

D1.2 LA Analysis
D1.2 LA Analysis

9.1 DATA GATHERING .......................................................................................................................... 77
9.2 DATA PROCESSING ......................................................................................................................... 80
9.3 DATA ANALYSIS AND VISUALIZATION ....................................................................................... 80
9.4 INTERVENTIONS / RECOMMENDATIONS ...................................................................................... 81

10 CONCLUSIONS ................................................................................................................................. 85

REFERENCES ........................................................................................................................................ 87
List of Figures

FIGURE 2-1 GOOGLE TRENDS ON LEARNING ANALYTICS .......................................................... 15
FIGURE 2-2 D1.2 OVERVIEW ................................................................................................. 16
FIGURE 3-1 MULTIDISCIPLINARY NATURE OF LA ...................................................................... 19
FIGURE 3-2 LA DIRECTIONS ................................................................................................. 21
FIGURE 3-3 LA INTEGRATED ENGINE FRAMEWORK BY SOLAR (SIEMENS ET AL., 2011) ............ 23
FIGURE 3-4 OU NETHERLANDS LA FRAMEWORK (DRACHSLER & GELLER, 2011) ..................... 24
FIGURE 3-5 LA REFERENCE MODEL (CHATTI, 2012) ................................................................ 25
FIGURE 3-6 FRAMEWORK FOR LA IN 3D EDUCATIONAL VIRTUAL WORLDS (FERNÁNDEZ-GALLEGOA ET AL., 2013) ............................................................................................. 27
FIGURE 3-7 LA STAKEHOLDERS .............................................................................................. 28
FIGURE 3-8 KNOWLEDGE CONTINUUM MODEL (BAKER, 2007) ............................................. 34
FIGURE 3-9 LA MODEL (SIEMENS, 2013) ................................................................................. 35
FIGURE 3-10 LA LIFECYCLE (DIMITRACOPOULOU, 2015) ....................................................... 36
FIGURE 3-11 LA MODEL (ELIAS, 2011) .................................................................................... 37
FIGURE 3-12 LA PROCESS MODEL (VERBERT ET AL., 2013) .................................................. 39
FIGURE 4-1 THE ADAPTATION MODEL IDEAL ......................................................................... 42
FIGURE 4-2 THE IDEA PHASES ............................................................................................. 45
FIGURE 5-1 GLASS TOOL USER INTERFACE .......................................................................... 47
FIGURE 5-2 VISUALIZATION OF DISCUSSION THREADS FOCUSING ON THE STUDENTS WHO HAVE INITIATED THE THREADS (MAZZA AND DIMITROVA, 2007) .................................................. 48
FIGURE 5-3 GOOGLE ANALYTICS ......................................................................................... 49
FIGURE 5-4 SNAPP TOOL VISUALIZATIONS .............................................................................. 50
FIGURE 5-5 LEMO TOOL VISUALIZATIONS .............................................................................. 51
FIGURE 5-6 STEPUP! TOOL INTERFACE (SANTOS ET AL., 2012) ............................................ 52
FIGURE 5-7 LOCO-ANALYST USER INTERFACE (ALI ET AL., 2012) .......................................... 53
FIGURE 5-8 NETLYIC LA TOOL ............................................................................................... 54
FIGURE 5-9 ON-LINE-BASED LA PROCESS (DYCKHOFF ET AL., 2012) .................................. 55
FIGURE 5-10 GISMO OUTPUT – LEARNERS’ ACCESS TO A COURSE ........................................ 56
FIGURE 5-11 MOCLOG TOOL ACTIVITY REPORT ................................................................... 57
FIGURE 5-12 DETERMINING ASSESSMENT CRITERIA WITH LAE-R (DIMOPOULOS ET AL., 2013) .......................................................................................................................... 58
FIGURE 5-13 SMARTKLASS DASHBOARD (KLASSDATA, 2016) ................................................ 59
FIGURE 6-1 LA DASHBOARD AT NTU ..................................................................................... 62
FIGURE 6-2 SES ENVIRONMENT AT UNIVERSITY OF BEDFORDSHIRE .................................. 63
FIGURE 6-3 COURSE SIGNALS SYSTEM AT PURDUE UNIVERSITY (EDUCAUSE REVIEW, 2010) ............................................................................................................................. 64
FIGURE 6-4 FORMAL TRAINING MODEL (SALAS ET AL. 2012) ............................................... 65
FIGURE 6-5 LA MAPPING TO THE TRAINING MODEL (JAGER ET AL., 2014) ......................... 66
FIGURE 6-6 THE MOOC LEARNING ANALYTICS INNOVATION CYCLE (MOLAC) .................. 67
FIGURE 7-1 ANALYTICS WORKBENCH (GÖHNERT ET AL., 2014) ....................................... 71
FIGURE 8-1 ORGANIZATIONAL CAPACITY FOR ANALYTICS (EDUCASURE REPORT, 2013) .... 75
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>MACRO-LEVEL LA</td>
<td>29</td>
</tr>
<tr>
<td>Table 2</td>
<td>LA DIMENSIONS AND EXAMPLES</td>
<td>31</td>
</tr>
<tr>
<td>Table 3</td>
<td>THE PROCESS MODEL OF ISO/IEC 19796-1</td>
<td>41</td>
</tr>
<tr>
<td>Table 4</td>
<td>INFORMATION PRESENTED IN COURSEVIS</td>
<td>47</td>
</tr>
<tr>
<td>Table 5</td>
<td>LA FREE &amp; OPEN SOURCE TOOLS</td>
<td>59</td>
</tr>
<tr>
<td>Table 6</td>
<td>DATA GATHERING FOR PBL3.0</td>
<td>78</td>
</tr>
</tbody>
</table>
## List of Abbreviations

The following table presents the acronyms used in the deliverable in alphabetical order.

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>EDM</td>
<td>Educational Data Mining</td>
</tr>
<tr>
<td>GLASS</td>
<td>Gradient’s Learning Analytics System</td>
</tr>
<tr>
<td>ICT</td>
<td>Information and Communication Technology</td>
</tr>
<tr>
<td>LA</td>
<td>Learning Analytics</td>
</tr>
<tr>
<td>LD</td>
<td>Learning Design</td>
</tr>
<tr>
<td>LMS</td>
<td>Learning Management System</td>
</tr>
<tr>
<td>LS</td>
<td>Learning Semantics</td>
</tr>
<tr>
<td>MOOC</td>
<td>Massive Open Online Course</td>
</tr>
<tr>
<td>PBL</td>
<td>Problem Based Learning</td>
</tr>
<tr>
<td>PLE</td>
<td>Personal Learning Environment</td>
</tr>
<tr>
<td>TEL</td>
<td>Technology Enhanced Learning</td>
</tr>
<tr>
<td>VLE</td>
<td>Virtual Learning Environment</td>
</tr>
<tr>
<td>WP</td>
<td>Work Package</td>
</tr>
</tbody>
</table>
Executive Summary

The overall aim of the PBL3.0 project is to enhance Problem Based Learning (PBL) with Learning Analytics (LA) and Learning Semantics (LS). Therefore PBL3.0 will produce a new educational paradigm and pilot it to produce relevant policy recommendations.

WP1 is responsible for the needs analysis of the project, providing state-of-the-art input on the PBL strategy as well as the LA field that will guide the project in constructing the PBL3.0 educational approach. In particular, it aims to:

- Identify all educational data that are generated for each step of the PBL strategy.
- Analyse processes and techniques that transform educational data into meaningful, multi-modal information.
- Identify all intervention mechanisms that could be put into practice based on all LA feedback during PBL3.0 courses’ design, implementation and delivery.
- Construct the PBL_LA educational approach.

The present deliverable is the second deliverable of WP1, D1.2 – LA Analysis. Its purpose is to present all preparatory work done to analyse the LA domain as well as all the results derived, such as all the LA components that are required to be employed in order to effectively raise the quality of education and training. Information that is included in this deliverable is valuable to all partners for ensuring the appropriate development of the PBL_LA educational approach and the LA modules that will be employed in the pilots.

More specifically, this deliverable covers issues such as the LA stakeholders, existing frameworks that aim to describe the LA domain and depict all related concepts, as well as existing mechanisms regarding the LA lifecycle, e.g. data gathering, data processing, data analysis, data storage, data cleaning, visualization and interpretation as well as interventions or recommendation methods that are currently available based on the relevant literature. Indicative examples of models identified in the study carried out include behaviour modelling, discourse analysis, sentiment analysis, natural language processing, association rule mining, knowledge domain modelling etc.

Furthermore, the report provides information on available tools that have been developed in order to perform specific LA functionalities. Indicative examples of technologies identified include:

- General-purpose LA dashboards, which are solutions that provide information and visualizations around different types of data for multiple stakeholders. Such tools include CourseVis, GLASS and Google Analytics.
- LA standalone tools, which provide specific information and be used in different platforms, such as SNAPP for social network analysis, LeMo for detecting learners’ navigation patterns, StepUp! for allowing reflection etc.
• LMS-based tools, which have been developed, usually as plugins or add-ons, to support LA functionality within LMS environments, such as Gismo for evaluating the social, cognitive and behavioural activities of learners in the online courses, MOCLog for analysing all interactions that can be recorded in an online course etc.

The deliverable also covers ethical issues and limitations and studies the classification of the LA ethical issues, i.e. location and interpretation of data, Informed consent, privacy and the de-identification of data, and the management, classification and storage of data. Additionally, the limited available LA practices, the objectivity in interpretations of LA results, and the proper and effective usage of relevant software are some of the challenges and limitations identified in regards to the LA effective application in real world settings. Finally, the deliverable elaborates (based on existing practices of LA application in real world) educational and training settings in general and usage of the PBL strategy in particular.
1 Introduction

The aim of this section is to introduce the background of the work pursued with Task 1.2 “LA analysis”. The scope and the objective that the current document has set out to achieve are presented in sub-section 1.1. The intended audience for this document is described in sub-section 1.2 while sub-section 1.3 outlines the structure of the rest of the document.

1.1 Scope

The present document is the Deliverable 1.2 “LA Analysis” (henceforth referred to as D1.2) of the PBL3.0 project. The main objective of D1.2 is to provide an in-depth overview of the LA domain and investigate the integration of LA mechanisms in PBL-oriented settings. These results will feed into the development of the PBL_LA educational approach and will guide the identification, configuration and application of LA tools in the project’s pilots.

1.2 Audience

The intended audience for this document is the PBL3.0 consortium, the European Commission, and the public interested in investigating the LA domain.

1.3 Structure

The structure of the document is as follows:

- Section 2 presents the methodology followed in this deliverable.
- Section 3 provides an analytical overview of LA, including the different LA definitions and detailed information on the different LA directions and existing frameworks. The Section also presents the LA stakeholders as well as the most commonly employed LA methods and techniques. This aims to improve the project’s in-depth knowledge on the LA field. The section also describes the related work on the LA lifecycle by elaborating on existing LA models and presenting each step within the LA lifecycle. This aims to provide information on all the LA steps that will later on help the project design the PBL_LA educational approach.
- Section 4 presents the Educational Reference Process Framework and the adaptation model IDEAL, and deals with the tasks and potentials of modelling and measuring educational processes and looks for appropriate instruments.
- Section 5 presents LA tools that exist to support the LA processes for education and training. This aims to provide information on all the existing technological solutions that can be exploited and / or configured for the purposes of the project during WP3.
- Section 6 describes some interesting existing applications of LA in different real world contexts, such as higher education institutions, MOOCs, business trainings etc. This aims to
provide empirical evidence on methodology, ways of implementation and lessons learned that can be exploited in the project during WP4 and WP5.

- Section 7 presents some interesting existing applications of LA in PBL environments. This aims to provide insights on existing ways LA is integrated with the PBL strategy and to detect good practices, limitations, gaps and opportunities for the project to put to practice and/or to address and solve.

- Section 8 presents the ethical and privacy issues related to the LA domain. This aims to raise awareness on the issues that need to be taken into consideration when designing the PBL_LA educational approach, configuring the LA tools, and applying LA related approaches in real world contexts.

- Section 9 provides an initial synthesis of the studied literature in regards to all the components that are relevant to the project’s objectives. This aims to provide a solid initial foundation for the work to be carried out in Task1.3 and in the following WPs.

- Section 10 concludes the document.
2 Methodology

In order to perform a thorough and in-depth review of the LA domain, a collection and structured review of the relevant State of the Art was carried out. The review included the following research fields covered by PBL3.0 in order to identify main terms, concepts and components of the LA domain that will guide the development of the PBL_LA educational approach. Such research fields include:

- LA terms and objectives;
- LA lifecycle steps and data processing on each step;
- LA frameworks;
- Ethical issues and limitations;
- Existing approaches and practices of LA in educational and training settings;
- Existing approaches and practices of LA in PBL-designed educational and training settings.

We started by searching the major research databases of computer science, i.e. ACM Digital Library, IEEE Xplore, SpringerLink, ScienceDirect and Google Scholar using keywords such as learning analytics, learning analytics lifecycle, learning analytics problem based learning, learning analytics ethics, learning analytics methods etc. We preferred publications dated from 2011 (as according to Google Trends’ search and news reference volume data the term ‘Learning analytics’ started becoming popular in March 2011).

![Google Trends on Learning analytics](image)

*Figure 2-1 Google Trends on Learning analytics*
The references of the selected papers were checked and additional papers were found. Electronic articles written in blogs such as Learning and Knowledge Analytics\(^1\), elearning space\(^2\), Effective Learning Analytics challenge by Jisc\(^3\) etc, referring to LA were also reviewed. Finally, initiatives coming from research institutes, universities, training organizations, companies and funded projects were also included in the survey.

This resulted in a collection of more than 150 publications that included (a) conference, workshop and symposium papers, (b) journal articles, (c) electronic articles and (d) technical reports and white papers. Around 110 publications were finally selected as the most relevant. Furthermore, a total of 35 sites were selected, which discuss initiatives from researching organizations/institutes, research communities, standardization bodies, consortia, and funded projects.

Figure 2-2 presents the overall overview of the work carried out in this deliverable and the corresponding main concepts addressed and studied.

---

2. [http://www.elearnspace.org/blog/](http://www.elearnspace.org/blog/)
More specifically, the overall aim of studying and analysing the LA domain led to the identification of the sub-sections that are relevant to the study and that combined provide a holistic overview of all LA components.

The thorough study and analysis of the LA domain led to the identification of the main concepts that are relevant to the PBL3.0 project and that will later on feed the work done in D1.3 “PBL_LA educational approach”.
3 Learning Analytics Overview

This section reports on the generic overview of the LA concept. Section 3.1 presents all the available definitions of the term and comments on the commonalities of the existing approaches. In Section 3.2, main concepts relevant to the LA domain are presented, while Section 3.3 elaborates on the existing techniques and methods of LA processes. Finally, Section 3.4 presents frameworks that have been constructed to describe the LA domain.

3.1 Learning Analytics definitions

The increased usage of learning technologies such as learning management systems, web-based learning environments, Web 2.0 tools, social media etc, has led to the generation of a large variety of different and multimodal educational data (Ferguson, 2012). Thus, an important question is how we can gather and make sense of all these data in order to improve learners’ performance, teachers’ role and course’s design.

Up until 2010 there was still limited exploitation of such data from educational institutions and companies, leading to increased numbers of dropouts and delayed actions taken to enhance performance (Long & Siemens, 2011). As a promising solution, the concept of LA has started gaining increasing attention from the end of 2010 to the beginning of 2011, when the 1st International Conference on Learning Analytics and Knowledge (LAK, 2011) took place.\footnote{Proceedings of the LAK 2011 - 1st International Conference on Learning Analytics and Knowledge Banff, AB, Canada — February 27 - March 01, 2011} The main purpose of this new field is to try and make sense of learning related data and provide insights on how to enhance the learning experience for both learners and teachers (Ferguson, 2013). Following the growth of the interest on LA, a variety of different, yet also similar definitions is available. Some of the most representative definitions are as follows:

1) “LA is the **measurement, collection, analysis** and **reporting** of data about learners and their contexts, for purposes of **understanding** and **optimizing** learning and the environments in which it occurs.” (Long & Siemens, 2011)

2) “LA refers to the **interpretation** of a wide range of data produced by and **gathered** on behalf of students in order to **access** academic progress, **predict** future performance and **spot** potential issues. The goal of LA is to enable teachers and schools to **tailor** educational opportunities to each student’s level of need and ability in close-to-real time.” (NMC Horizon H.E. Report, 2012 & EDUCAUSE)

3) “The field of LA focuses on tracking learning activities and the context in which these activities occur, to promote awareness and reflection through algorithmic analysis (in educational data mining) or information visualisation.” (Duval et al, 2014)
Another interesting definition that applies to the business sector is the following:

4) “LA is an educational application of web analytics, a science that is commonly used to by business to **analyse** commercial activities, identify spending trends, and predict consumer behaviour.” (Horizon Report, Schools 2014)

All aforementioned definitions share an emphasis on collecting and analysing learning data and translating them into meaningful actions for improving the learning experience for all involved stakeholders (Chatti et al., 2012).

The field of LA has emerged from and is closely connected to multiple and different research fields and interests related to analysis, such as business intelligence, statistics, web analytics, academic analytics, data mining, social network analysis as well as research interest in the field of learning sciences such as pedagogies, Technology Enhanced Learning (TEL), cognitive sciences etc. LA is finally strongly related to the learning technologies that have become an integrated part of the learning experience through the technology enhanced learning paradigm. Such technologies generate large amounts of educational data, and range from cognitive tools to more sophisticated and complex environments like Learning Management Systems (LMSs), Virtual Learning Environments (VLEs) and the recent Massive Open Online Courses (MOOC).

![Multidisciplinary nature of LA](image)

Some of the fields that are more inter-connected with LA include:

1. **Business intelligence.** Process established in the business world for generating informative decision making capabilities through the analysis of data (Elias, 2011).
2. **Academic analytics.** Process established in the academic world for applying tools and methods from the business intelligence domain to academia (Golstein and Katz, 2005).

4. Constructivism. Theory according to which humans generate knowledge and meaning from an interaction between their experiences and their ideas (Piaget, 1967).

The LA domain can thus reinforce education and training through providing feedback based on generated data and allowing an in-depth understanding of the learning experience. This can be done by accumulating as much educational data as possible and enabling learners and educators/trainers to comprehend the information provided and make decisions in regards to the learning process, the learning processes, learners’ knowledge and skills as well as their weaknesses and misconceptions, the assessment’s efficiency etc. All these insights can then underpin successful personalized and adaptive learning that improve all aspects of education and training.

3.2 LA directions

LA can be viewed as focusing on two main directions, namely academic analytics and activity analytics. The first category refers to the academic level of LA’s influences while the second refers to the individual level, where LA’s methods and results affect education and training processes on a macro- and meso-level.
Figure 3-2 presents some examples of the two main categories, and more details are provided in the following sub-sections as well as in Section 3.4

3.2.1 Learning Analytics on an academic level

LA provide methods and techniques that contribute to the continuous improvement of academic institutions. Also, they provide a personalized learning experience, by exploiting available data so that academic institutions and educators can address learners’ needs (Dietz et al, 2013).

LA on an academic level provide information focused on the administration of the academic sector, underpinning operations such as improved decision making, efficient resources allocation, representative view of the successes and challenges of the institution and increase of the organizational productivity (Dietz et al, 2013).

3.2.2 Learning Analytics on an individual level

LA are also useful for individuals within the academic institutions, as they provide information on the learners, such as at risk cases, and can suggest intervention methods to improve these learners’ pedagogical approaches.
The students’ performance evaluation in a complex computer supported learning environment is a tiring and time-consuming process for educators who must take into consideration a large amount of parameters. Students’ assessment should also be connected with their participation, the spirit of collaboration, and the quality of engagement and creativity in the group assignments (Dyckhoff et al., 2012). These parameters are difficult for an educator to assess without using sophisticated analytics methodologies and tools. Such tools can also exploit data recorded in an LMS and allow the educators to create a learner’s model of successful behaviour, so that they can design their future courses based on this model and increase the probability of enhanced learners’ performances (Dietz et al., 2013).

Apart from support for educators, LA also support the learners, by providing them knowledge on their own performance by allowing them to reflect on their progress, view their ongoing engagement in the learning process and ask for scaffolding when needed.

### 3.3 Learning Analytics Frameworks

The proper, efficient and beneficial exploitation of LA in education and training requires the investigation of all dimensions that should be taken into consideration during LA’s application in real world settings. Relevant research on this matter has led to the construction of multiple frameworks and models that describe the basic dimensions related to the LA domain.

#### 3.3.1 LA framework by SOLAR

The Society for Learning Analytics Research (SOLAR), was the instigator in this research, proposing integrated toolsets through the development of four specific tools and resources:

1. A Learning analytics engine, a versatile framework for collecting and processing data with various analysis modules
2. Adaptive content engine
3. Intervention engine: recommendations, automated support
4. Dashboard, reporting, and visualization tools
Figure 3-3 LA integrated engine framework by SOLAR (Siemens et al., 2011)

Figure 3-3 shows all the components such an integrated solution comprises. According to the SOLAR team, the Analytics engine is the central component of the engine, as it comprises of a framework that focuses on gathering and processing data from multiple different sources, e.g. LMS environments, social media websites, dashboards, and even physical-world data such as attendance lists, usage of learning materials etc.

For example, the LA framework would include gathering data from a discussion forum of an LMS, which would include the scope of the forum, the topics and posts submitted, the interactions, replies, quotes etc., and then performing processing with various methods and techniques such as data mining, social network analysis, natural language processing, prediction models design etc.

3.3.2 LA framework by the OU of the Netherlands

The Centre for Learning Sciences and Technologies of the Open Universiteit (OU) in the Netherlands designed an LA framework (Drachsler & Geller, 2011) which consists of six main dimensions, as shown in Figure 3-4.
These dimensions represent the main components that are strongly connected with the successful application of LA in real world settings.

**Stakeholders.** The framework recognizes that the stakeholders are an important part of the LA process and suggest four main categories of interested parties, namely institutions, teachers, learners and parents.

**Objectives.** This dimension refers to the specific target outcomes of the analysis of educational data. As examples, the framework includes reflection and prediction. Reflection information can regard ongoing progress on learners’ performance, common mistakes, most active and inactive learners etc., while prediction involves the identification of learners at risk, final outcomes etc.

**Educational data.** This dimension refers to all data that is generated during learning. The framework divides this data into open and protected, emphasizing the importance of privacy and sensitive data.

**Technologies.** This dimension refers to the different tools and methods that allow proper and successful gathering, processing and analysis of the educational data. Closely related concepts and terms include educational data mining, recommender systems and statistical analysis. These tools and methods will also provide the means to properly visualize the target objectives, by enabling the development of reflection dashboards and prediction models.

**Constraints.** This dimension refers to all the limitations and issues to take into consideration when applying LA in real world settings. Such limitations include privacy and ethics issues, as well as ownership and legal protection in regards to the educational data generated and analysed.

**Competences.** This dimension refers to all the knowledge and skills required for the correct and effective application of LA in real world settings. The proper and contextual interpretation of the LA
outcomes into robust actions to improve the learning experience for all stakeholders requires competences such as critical thinking, interpretation skills, ethical understanding and self-directedness.

### 3.3.3 LA Reference Model

An LA reference model was suggested by Chatti et al. (2012), which consists of four main dimensions, namely the questions “What?”, “Why?”, “How?” and “Who?” regarding an LA process, as shown in Figure 3-5.

![LA Reference Model](image)

**Figure 3-5 LA Reference Model (Chatti, 2012)**

The dimensions of the reference model are as follows:

**What?** This question refers to the actual data and the environments in which data is generated and analysed. LA requires that the data to be gathered, processed and analysed is educational, and thus collected through environments that are used in learning (e.g. LMS, VLE, social media, ePortfolios, Wikis, MOOCs, sensors etc.). Also, it is important to determine which of the educational data is available and / or can be made available, as well as which of these datasets can actually provide meaningful insights when analysed (Pistilli et al., 2014). It is considered that the utilization of data from multiple data sources can provide more accurate insights on the learning process, as they analyse different types of aspect of the learning experience.

Apart from the content and the source, another concerning data is the amount of data gathered. Technology enhanced learning data, i.e. data that derive from different tools applied for learning can
be quite large, especially when technologies such as MOOCs are utilized. These big data require specific operations, such as increased storage availability, fragmentation etc.

**Why?** This question refers to the specific objectives and motivation behind utilizing LA for learning and teaching. Each objective requires a different set of indicators and metrics that determine which data will be gathered, what processing will be carried out and what kinds of visualizations / actions will be available to achieve the set goals. As shown in Figure 3-5, seven main objectives are listed that cover the majority of the reasons why stakeholders use LA.

These objectives include:

- monitoring and analysis for making decisions depending on the on-going progress;
- predictions for performing interventions when needed, tutoring for providing scaffolding mechanisms;
- assessment for evaluating the learning progress during each step and not only in the end and providing feedback to learners;
- adaptation for allowing flexibility according to the learners’ needs;
- recommendations for transferring the control to the learners and letting them decide on how they will learn;
- reflection for allowing on-going overview of the learning progress to both teachers and learners.

**How?** This question refers to all the LA methods and tools that are available for performing all the steps of an LA lifecycle. The methods that are more commonly used include clustering, social network analysis, predictive models, machine learning, statistics etc. The outcomes of these methods are then depicted in various types of visualizations such as dashboards, graphs, networks etc.

**Who?** This question refers to the stakeholders, i.e. the people that are closely connected to the LA domain. These can include students, teachers, trainers, trainees, researchers, institutions, system designers etc. Each stakeholder’s interest in LA differs depending on what objectives they aim to achieve by using LA. For example, learners want to be able to reflect on their progress, get recommendations and adaptive learning pathways that will help them improve their performance, teachers want to be able to monitor the learning process more proactively, make sense of the generated data, scaffold their learners and design their courses more effectively, institutions want to be able to make more efficient decisions based on what works and what does not etc.

This dimension involves two other main concepts, namely competences and constrains. The exploitation of LA by each stakeholder requires a specific set of competencies that allow efficient critical thinking, decision making and proper interpretation of the LA outcomes. Furthermore, the nature of the data presents issues such as privacy, ethics, data management etc. Appropriate
decisions and actions need to be taken by the stakeholders in regards to these issues before any data analysis is carried out.

### 3.3.4 LA Framework for Educational Virtual worlds

Fernández-Gallegoa et al. (2013) propose an LA framework focused on virtual learning environments as shown in Figure 3-6 and includes concepts such as process models, educational world, learning platform, educational data and process mining system.

![Figure 3-6 Framework for LA in 3D educational virtual worlds (Fernández-Gallegoa et al., 2013)](image)

This framework shows that the educational world is consisted of the stakeholders involved in LA include the teachers, the students and the instructors as well as the learning resources and services that operate during learning.

This concept is inter-connected with the process models, which refer to learning flows constructed based on the IMS Learning Design (LD) specification. The process models represent the actual processes that are simulated within the virtual world and are realized by the teachers and students inside the environment.

The process models receive input from the event logs of the virtual learning platform and the learning flows are designed in accordance to the data processing carried out by the Process Mining System.
3.4 LA stakeholders

LA can be targeted at and be interesting to a different number of stakeholders, each of which has different reasons and expectations from the analysis of educational data. The literature review provides different suggestions on who the LA stakeholders are and what is their connection to the LA domain. Figure 3-7 presents a holistic representation of the most commonly identified interested parties.

![LA Stakeholders Diagram](image)

The LA stakeholders can be divided into two main categories, namely: (1) the practitioners, i.e. the parties that work with and engage with LA, and (2) the decision makers, i.e. the parties that do not actively participate in the LA processes but they make decisions based on the LA outcomes and assess their benefits, limitations, opportunities for their respective sector (e.g. academia, business, public sector).

3.4.1 LA Practitioners

LA can be done at different levels and characterised accordingly in macro-level, meso-level, and micro-level analytics (Buckingham Shum and Ferguson, 2012; MacNeill et al., 2014).

The LA practitioners represent the micro-level LA which involves learners, teachers and trainers, researchers, technologies and data analysts. All of these stakeholders have a similar and yet
different perspective and expectations of the data (Dyckhoff et al., 2013; Romero & Ventura, 2013; Greller & Drachsler, 2012).

**Learners.** Students and learners of all other domains (e.g. private and public sector) will be mainly interested in using LA in order to increase their performance and reflect on their own progress and competencies.

**Teachers/trainers.** Teachers and trainers will mainly use LA in order to make sense of all the data generated during learning, support their learners, enhance their monitoring abilities, identify problems more quickly and improve the learning experience for them and the learners. This can be done by identifying at risk learners and providing helpful interventions and recommendations as well as providing learners with visualized insights of their learning progress. Furthermore, teachers and

**Researchers.** Researchers are mainly interested in LA in order to identify how this new concept benefits the learning experience for all stakeholders. Moreover, researchers will study ways to improve the field’s aspects by applying LA in different contexts and investigate the impact of this application on education and training.

**Technology developers.** Technology developers are interested in the tools and technologies that can support and reinforce the LA lifecycle. These stakeholders actively engage in the development of technologies that can improve, visualize and automate the different steps of LA.

**Data analysts / providers.** Data analysts and providers refer to all stakeholders that deal with providing and analysing data. These stakeholders are interested in the nature, formats and different analyses that can be carried out on the data generated during education and training.

### 3.4.2 LA Decision makers

The macro- and meso-levels of LA involve mostly stakeholders that do not usually engage with LA actively, but can make decisions around and based on LA objectives and outcomes. LA in these levels is usually carried out with the aim to understand the opportunities and weakness of learning at an institutional, regional and national/international levels so that productivity and decisions making can be improved within the educational institutions.

**Table 1 Macro-level LA**

<table>
<thead>
<tr>
<th>Scale of Analysis Objective</th>
<th>Who benefits</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Institutional</strong> : learner profiles, performance of knowledge flow</td>
<td>Administrators, funders, marketing</td>
</tr>
<tr>
<td><strong>Regional</strong> : comparison between systems</td>
<td>Funders, administration</td>
</tr>
<tr>
<td><strong>National and International</strong></td>
<td>National governments, education authorities</td>
</tr>
</tbody>
</table>

**Institutional leaders.** Authorities within institutions are strongly interested in LA so that they can be
provided with feedback on the different learner profiles, analytics information on the learners’ progress across the departments and years, comparison across different systems and institutions etc. All these LA-generated outcomes can help improve administrative decision making as well as resource allocation within the institutions. Furthermore, institutional leaders can identify different patterns that emerge as well as the quality of teaching, the institution’s reputation etc, and interpret these insights into meaningful actions.

**Policy makers.** Policy makers are interested in LA outcomes in order to transform and reform existing policies, mainly in education and training. These policies can then allow and underpin the growth of novel educational paradigms such as new learning approaches, strategies, curricula, technologies etc, that will reinforce the field.

**Product vendors.** Product vendors refer to all stakeholders that can exploit LA into developing and supplying new products that generate significant added value in the business domain, e.g. innovative dashboards, learning workflow modelling tools, analytics tools etc.

**Instructional designers.** All stakeholders that design and develop courses for education and training are interested in LA because the generated outcomes increase the learning designs’ effectiveness and success and allow them to understand how the learning resources are being used by the learners. These insights provide them with the necessary knowledge to better design and refine their courses in the future so that the learning experience will be adaptive and beneficial for all different learner styles.

**Research / standardization bodies.** Standardization and research bodies are interested in LA because the field strongly affects the way we deal with and understand educational data. The field of learning semantics has also emerged in the last few years, which has led to the release and publication of new standards that allow easy educational resources annotation.

### 3.5 LA methods

According to the co-founder and president of SOLAR, LA methods can be categorized in two main components, namely techniques and applications (Siemens, 2013). Techniques refer to the specific models and algorithms that are used for carrying out the analysis of the educational data, while applications refer to all the different ways we can exploit the techniques in order to improve teaching and learning. However, the distinction between the two terms is not absolute as they can sometimes overlap. Representative examples of techniques are provided by Baker and Yacef (2009) who list five primary areas of this dimension:

- Prediction
- Clustering
- Relationship mining
The techniques dimension focuses more on the technical aspect of LA, by using mostly machine learning methods, network analysis, statistical analysis etc. The information that can be gained through these processes can then provide meaningful insights into different kinds of applications such as user behaviours, interventions on learning contents, recommendations, predictions etc. The aforementioned techniques can also design models that will identify students’ progress, predict their performance and identify parameters of similarity amongst group of students.

On the other hand, Bienkowski, Feng, and Means (2012) list five main areas of the application dimension:

- Modelling user behaviour, knowledge and experience
- Creating user profiles
- Modelling knowledge domains
- Trend and patterns analysis
- Personalization and adaptation

These application areas can guide educators into enhancing their courses (e.g. by personalizing learning content in accordance to the learners’ profile and knowledge level in the specific knowledge area) and developing new curricula based on viewing insights on the knowledge domain or the trend analysis (Herskovitz et al., 2013). Furthermore, educators can study the social networks formed within a classroom and determine how these are formed and affected from the tools and learning strategies used during learning. This can be carried out through social network analysis (SNA), which analyses the interactions carried out during learning during their group collaborations, information transfer etc. and then uncovers any identified patterns or structures of relationship networks that can provide insights valuable to the learning process (Bakharia & Dawson, 2011).

Table 2 provides some representative examples of some of the most commonly used techniques and applications (Siemens, 2013).

<table>
<thead>
<tr>
<th>LA dimension</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Techniques</td>
<td></td>
</tr>
<tr>
<td>Modelling</td>
<td>Learner modelling</td>
</tr>
<tr>
<td></td>
<td>Behaviour modelling</td>
</tr>
<tr>
<td></td>
<td>User profile design</td>
</tr>
</tbody>
</table>
Relationship mining

- Sentiment analysis
- Discourse analysis
- Association rule mining

Knowledge domain modelling

- Natural language processing
- Assessment by matching learner’s knowledge with knowledge domain

Applications

Personalization & adaptation

- Adaptive content to learners
- Recommendations on content, activities and interactions

Prediction and trend analysis

- Changes in learner behaviour, identification of errors
- Early risk identification
- Interventions
- Classification

Structure discovery and analysis

- Social network analysis
- Information flow analysis
- Semantic analysis
- Clustering

**Modelling** techniques focus on designing models for learners in order to better structure the learning experience. The existence of behaviour, user and learner profiles allows educators to formulate their curricula and learning activities based on the needs of their students. Data analysis can detect different learning styles depending on for example the types of content and activities each learner prefers (e.g. pictures – visual, podcasts – aural, writing – verbal, forums – social etc.), thus allowing more adaptability and flexibility during learning and teaching.

**Relationship mining** techniques focus on methods that detect any forming relationships between specific variables and investigate how strong these relationships are (Baker & Siemens, 2014). Such variables can include types of discourse (e.g. interactions, cognition, text etc.), different sentiments (positive, negative, neutral), patterns of behaviour (e.g. what behaviour can be linked with high performances) etc. For example, discourse can show how learners are thinking collectively. The way in which each learner participates in a discourse can indicate how each learner a) handles and
address others’ ideas, b) compares others’ ideas with his own and c) communicates his ideas to others. (De Liddo et al., 2011).

Knowledge domain modelling techniques focus on structuring and mapping the knowledge domains of each learning subject in order to personalize and adapt learning for specific students. Appropriate domain modelling can lead to the gathering and exploitation of data about the learners, their profiles, the content, data trails on their interactions etc. The analysis of all these data in correlation with the specific concepts of the knowledge domain can bring forth informative insights on learners’ knowledge in relation to the domain’s structure as well as predictions, adaptations and personalization of the learning processes.

Personalization and adaptation applications focus on providing flexible learning pathways during the learning process. This can be realized through the availability of adaptive content according to the analysis of the generated educational data, as well as through recommendations provided to both educators and learners on different learning resources, activities and interactions that could lead to higher performances.

Prediction and trends analysis applications focus on providing insights on future developments based on educational data analysis. Predictions can involve various variables such as learners’ scores, performances, risk of dropping out etc., and they usually are realized by analysing other predictor variables, such as frequency of participation, level of interactions and group work, progress on assignments, number of content accessed etc. (Baker & Siemens, 2014).

Structure discovery and analysis applications focus on performing analysis on the educational data in order to detect any structures that are not yet visible or identifiable. Such structures can include the detection of specific groups of students that are characterized by particular features through clustering or social network analysis (e.g. similar learning styles, similar skills or knowledge levels etc.), the identification of learning contents or activities characterized by particular features through information flow or semantic analysis (e.g. supporting similar learning objectives or similar skills, relevant to similar learning subjects etc.).

The specific techniques and applications that will be used in real world settings heavily depend on the context of each learning and teaching environment (e.g. offline / blended / online, higher education / corporate trainings / public sector trainings, technology enhanced / no tools etc.) as well as on the kinds of data that can be collected during learning.

3.6 LA Lifecycle

The LA lifecycle represents the LA process, i.e. all the different steps that lead from raw data to meaningful information that has significant added value to the enhancement of education and training. There are many LA lifecycle models that have been designed since the concept emerged, and while they are different in some ways, they share many similarities as well. The majority of them
are based on the Knowledge continuum concept which goes back to the 1800s but was re-designed by Baker (2007), as shown in Figure 3-8.

![Knowledge Continuum model (Baker, 2007)](image)

**Figure 3-8 Knowledge Continuum model (Baker, 2007)**

The bottom layer of the Knowledge continuum refers to all the raw data that is generated and available for accumulation. Once meaning is attached to this data, it is transformed into information that can answer specific questions. The next level represents the knowledge, which is generated once the information is analysed and synthesized using particular methods. The top level refers to the actual application of knowledge, which leads to wisdom, as by applying the created knowledge we can get design and achieve our objectives, get lessons learnt for future high quality knowledge.

It is important to study all the different steps that are essential for proper LA utilization, because these steps will guide the process, will determine the types of tools and methods to be used as well as the different LA outcomes that can be generated. The following sub-sections present some of the most representative LA models that have been designed to describe the LA lifecycle and steps.

### 3.6.1 LA Model by Siemens

The LA Model (LAM) proposed by Siemens (2013) consists of seven components, as shown in Figure 3-9. This model aims to provide a more system-based approach to the LA process, where educational data can be exploited not just for individual monitoring of progress, common mistakes etc., but also for more high quality and complex insights such as specific actions that can optimize the learning experience, e.g. by providing recommendations or suggesting predictions etc.
The LAM components are as follows:

**Collection & Acquisition.** The actual gathering of all the data that will be used for analysis is the initial step of the LA lifecycle. The decision on what data we will process always depends on the purpose leading our choices. The data types, sources, formats etc. will change, if our goal is to enhance learning (e.g. students’ scores, resources more commonly accessed etc.) or to facilitate administration (e.g. log data from a Student Information System - SIS). Thus, this step is closely connected with the different data sources available for collection and acquisition; these can include the different LMS environments used during learning, the SIS used by the institution, any sensors employed during labs, tests or courses, manual entries etc.

**Storage.** The gathered data should be stored in databases that will allow their easy retrieval and processing. If the data is retrieved from multiple data sources, then an issue to be examined is the storage of different types of data and their reconciliation.

**Cleaning.** The format and content of the gathered data is also important, as they determine whether or not pre-processing operations are needed. If the data is already structured and only the information useful for analysis is included then no further cleaning might be required. However, if the data is not structured and there are empty or duplicate records and nonsensical content, then it is important to clean it to make sure it is accurate and valid by using methods from the data mining field, such as data cleaning, data integration, data transformation, data reduction, data modelling etc. (Romero and Ventura, 2007).

**Integration.** This step refers to the combination of data residing in different sources and having different formats for the provision of a unified view.

![Analytics Model](image-url)
Analysis. The analysis of data can be carried out with multiple LA methods, such as social network analysis, predictive models, at-risk identification, recommendations etc. Specific techniques and tools are available to support these kinds of methods that can process raw data and produce valuable insights.

Representation & Visualization. The analysed data can be represented and visualized in ways that make sense to the stakeholders. These visualizations are usually in forms of dashboards, graphs, tables, annotated text etc., and they allow for a comprehensive representation of all the valuable information retrieved from the analysis step.

Action. The visualizations will allow for stakeholders to take specific actions in order to achieve the set goals. Such actions include interventions in the learning process that will improve learning, optimization techniques, improved course design and evaluation, increased scaffolding mechanisms, ongoing monitoring etc.

As shown in the above Figure, this LA lifecycle is a data loop, since the outcomes of the Action step also feed into the Collection & Acquisition step. This is done so that the lessons learnt from every LA process can improve future endeavours.

Siemens (2013) also highlights the importance of a data team, which consists of many different types of professionals that possess a specific set of skills and knowledge that are required for the successful operation of all the LAM steps.

3.6.2 LA lifecycle by LTEE

Dimitracopoulou (2015) presents a LA lifecycle that consists of a loop with six main steps, as shown in Figure 3-10.

![Figure 3-10 LA lifecycle (Dimitracopoulou, 2015)](image)

The steps represented in this lifecycle are as follows:
Data Selection. During this step, the data that is generated during learning and that is relevant for our objectives are selected. This step is closely related to data capture, where all different kinds of data are collected. Such data can include explicit data (e.g. posts on assignments, access to learning resources, answers to quizzes etc.), tacit (e.g. time on pages, social media posts, replies to posts etc.), physical (e.g. location of resources, sensor-based data etc.) and other.

Data storage. During this step, all selected and captured data is stored for further analysis.

Data analysis. During this step, the stored data is analysed according to the available LA methods, depending on the types of analysis we aim to perform.

Data operationalization. During this step, we try to identify clear and measurable data variables that can help measure concepts that are complex, fuzzy or difficult to measure. This way, we can understand these concepts in terms of which of the analysed data is of the wider concept and which is not. For example, it is challenging to directly measure someone’s leadership skill. For that reason, we identify data that can help measure this concept, such as the number of times this person has initiated activities, made decisions etc.

3.6.3 LA model by Elias

Elias (2011) presents an LA model that consists of a cycle with three main phases and four components that support all LA steps aimed to constantly improve learning and teaching, as shown in Figure 3-11.

![Figure 3-11 LA model (Elias, 2011)](image)

The three phases comprising the LA model by Elias (2011) are as follows:

Data gathering. This step refers to the selection and capture of the data that we want to analyse.
**Information processing.** This step refers to all the different methods used to analyse and process the gathered data. Such methods can include aggregation of multi-resource data, prediction models etc.

**Knowledge application.** This step refers to all the different actions taken for the contextual application of all the knowledge derived from the information processing. Such actions include the usage of the knowledge for improved courses and learning activities, the refinement of learning resources and course designs etc.

The four components that support all the LA phases and steps are as follows:

**Organizations.** This component refers to the organizational capacities of the institutions applying LA in their processes. It is essential that committed and skilled leaders are in place to ensure improved performance and corresponding cultural change.

**Computers.** This component refers to the different technologies that support each LA step. In regards to data gathering, it is essential to utilize tools that allow the seamless collection of data and the integration of data from diverse sources. During information processing tools are available that can develop prediction models and aggregate the data. During knowledge application, tools are available that support processes such as data visualization, decision trees, dashboards etc. The visualization of the analytics results is essential for appropriate decision making, as it allows large amounts of information to be displaying in a comprehensive way (Baker 2007).

**Theory.** This component refers to all the pedagogical theories and practices that can support LA, by underpinning collaboration, student motivation etc. However, there are no available guidelines for educators on which LA parameters may be useful on a pedagogical level (e.g. what types of data can indicate student motivation (MacFayden and Dawson, 2010), and how can the LA results be used to better structure the courses based on pedagogical theories.

**People.** This component refers to the importance of people’s involvement in the LA process. Even though tools perform the majority of the data analysis and visualization, the actual knowledge application should be realized by people, i.e. educators, instructional designers, trainers etc. These people are required to check the accuracy and appropriateness of the LA results for their specific contexts and make the final decisions whether they will apply the knowledge to achieve their set goals. These actions, however, call for people with a specific set of skills and knowledge, i.e. problem solving, decision making, leadership, critical thinking etc.

### 3.6.4 LA process model by Verbert et al.

Verbert et al. (2013) proposed an LA process model that consists of four main stages that indicate the level of impact of the LA results in learning and teaching, as shown in Figure 3-12.
The four stages are as follows:

**Awareness.** This stage refers to the actual data, where stakeholders are made aware of their content through different overviews and visualizations.

**Reflection.** This stage refers to the investigation of the data by asking questions about their nature, usefulness, validity, added value etc.

**Sense making.** This stage refers to the process of answering the questions posted in the reflection stage and having a more clear and holistic understanding of the analysed information.

**Impact.** This stage refers to the process of making decisions and allowing change when it is appropriate according to our set objectives.
4 Educational Reference Process Framework

This chapter deals with the tasks and potentials of modelling and measuring educational processes and looks for appropriate instruments. The answer on the question: "How to implement and improve Learning Analytics in learning, education, and training?" leads to the support which can be provided by international educational quality standards. Therefore we will introduce international standards as an appropriate means for structuring Learning Analytics. The first ISO quality standard for learning, education, and training called "RFDQ" (ISO/IEC 19796-1, currently under revision with the new numbering ISO/IEC 40180) will be explained and the adaptation model IDEA will be introduced. In summary, the implementation and adaptation of the international process standard demonstrate its support for establishing Learning Analytics in learning, education, and training.

4.1 The international Reference Process Framework

Quality does not exist in a simple manner as we have shown before (Stracke 2010). First, all stakeholders have to define their own understanding what the term “quality” is standing for in relation to the given context, here: Learning Analytics. Then these different perspectives and opinions about quality have to be combined, to be brought into consensus and transferred into practice. The specification of relevant aspects and criteria to define quality as well as the application of these criteria into the given context of the organisation are quite abstract by itself. For this purpose a common reference framework is needed. The standard RFDQ (ISO/IEC 19796-1), the first international quality standard for learning, education and training, is providing such a common reference framework for educational processes and will be explained in the following.

4.1.1 The structure of the Reference Process Framework

The reference process framework of ISO/IEC 19796-1 called RFDQ is the integration of the following two main reference models (cf. ISO/IEC 2005) that will be described in detail below:

- the generic process model and
- the generic descriptive model.

The reference process model covers the whole lifecycle of learning, education, and training in general including e-Learning and blended learning. Therefore it can be used to describe any learning scenarios as well as any educational and vocational training product and learning solution. It is important to note that the reference process model does not include any regulations about the sequence of the processes or interdependencies between them as well as it does not give any instructions on its specific implementation in detail as a prescription or regulation. The reference process model serves as an open descriptive framework that always needs the adaptation to the organisation, the learning context, and the given situation (cf. Stracke 2006, 2010).
### 4.1.2 The Process Model of ISO/IEC 19796-1

The reference process framework is based on the generic process model that is divided into seven process categories containing in total 38 processes. It is described by the following table:

**Table 3: The process model of ISO/IEC 19796-1**

<table>
<thead>
<tr>
<th>ID</th>
<th>Category</th>
<th>Description</th>
<th>Processes</th>
</tr>
</thead>
</table>
| NA | Needs Analysis         | Identification and description of requirements, demands, and constraints of an educational project | NA.1 Initiation  
NA.2 Stakeholder Identification  
NA.3 Definition of objectives  
NA.4 Demand analysis |
| FA | Framework Analysis     | Identification of the framework and the context of an educational process    | FA.1 Analysis of the external context  
FA.2 Analysis of staff resources  
FA.3 Analysis of target groups  
FA.4 Analysis of the institutional and organisational context  
FA.5 Time and budget planning  
FA.6 Environment analysis |
| CD | Conception / Design    | Conception and Design of an educational process                              | CD.1 Learning objectives  
CD.2 Concept for contents  
CD.3 Didactical concept / methods  
CD.4 Roles and activities  
CD.5 Organisational concept  
CD.6 Technical concept  
CD.7 Concept for media and interaction design  
CD.8 Media concept  
CD.9 Communication concept  
CD.10 Concept for tests and evaluation  
CD.11 Concept for maintenance |
| DP | Development / Production | Realization of concepts                                           | DP.1 Content realization  
DP.2 Design realization  
DP.3 Media realization  
DP.4 Technical realization  
DP.5 Maintenance |
| IM | Implementation         | Description of the implementation of technological components               | IM.1 Testing of learning resources  
IM.2 Adaptation of learning resources  
IM.3 Activation of learning resources  
IM.4 Organisation of use  
IM.5 Technical infrastructure |
| LP | Learning Process        | Realization and use of the learning process                               | LP.1 Administration  
LP.2 Activities  
LP.3 Review of competency levels |
The process model (table 5-1) structures the lifecycle of learning processes, but it does not contain any prescriptions on the structures or procedures of how to deal with the stated processes. Thus, it can and has to be defined and adapted for any purpose including Learning Analytics (cf. Stracke 2010, 2014b).

4.2 The Adaptation Model IDEAL

The ISO/IEC 19796-1 reference process framework is a generic model: This means that it cannot simply be implemented and used as it is, but instead it has to be adapted to every specific context of usage. In this chapter we will describe the process of implementing and adapting the reference process model of the standard in practice based on first gained experiences introducing the adaptation model IDEAL as a helpful instrument (cf. Stracke 2013, 2014a).

The following figure presents the overview of the IDEAL model:
In the implementation process of quality development based on the reference model of the quality standard ISO/IEC 19796-1, an individual selection of processes, which are applicable, has to be made and each of the selected processes has to be specified according to the current situation. During this adaptation, the specific requirements and objectives of the current situation are considered and thus, become part of the model.

Since the process model covers any learning processes, it is applicable to any application scenario. Each scenario has got specific characteristics and focal points. In the planning phase of a learning opportunity (product or solution), the model provides valuable support especially for the analysis of the needs and the requirements. The reference process model supports customers defining a call for biddings as well as providers customising corresponding learning opportunities. In the development phase of learning contents, the model can be helpful for the design of a learning opportunity as well as for selecting and implementing an appropriate infrastructure. Moreover, the model also supports the production, implementation and realisation of learning opportunities as well as the continuous evaluation just from the beginning (cf. Stracke 2010).

To achieve a holistic quality development the needs and requirements of all stakeholders of the current learning scenario have to be considered (Feigenbaum 1986; Ishikawa 1985; Soin 1992). This perception is also valid for the adoption and introduction of the reference process model: A strong procedure systematically planned is needed for adapting the reference process model of the standard ISO/IEC 19796-1 to a specific organisation including all stakeholders. Therefore simple to use quality tools as the adaptation model IDEAL with its IDEA Phases can deliver helpful support.

The abbreviation IDEA stands for the four main tasks to introduce quality development:

1. **Initiate!**
2. **Do!**
3. **Evaluate! and**
4. **Act!**

These four tasks and their phases and steps were developed according to the Deming cycle and the ISO standard family ISO 900x (cf. Stracke 2006a). They have to be fulfilled for the implementation of quality development in LET: These phases and steps can be realized and applied according to the specific needs and the given situation of the organization including feedback loops, individually adapted sequences and parallel implementation.

The phases and steps of the four tasks can be described and defined as follows (cf. Stracke 2013):

1. **Task: Initiate!**

First the raising of the awareness of all stakeholders and their full involvement and participation is needed. In transparent procedures the vision for the quality development should be defined based on a common and shared understanding of quality for learning, education, and training and the innovations that should be achieved by the implementation
of quality development. The long-term policies and strategies will be discussed and agreed upon the approved vision.

- 2. Task: Do!

For the starting implementation the quality model has to be selected (here in our case: RFDQ) and to be adapted as an application profile. Concerning RFDQ that means the selection of the appropriate and relevant processes in a quality profile. Then the specific definitions and success criteria has to be described, for RFDQ those are the attributes of the descriptive model, mainly the methods, actors, metrics and criteria. The preparation of the implementation will include all needed decisions and developments for the organizations. And finally the realization of all planning for the quality development and its integration into organizational processes is completing this task.

- 3. Task: Evaluate!

The evaluation will focus on three distinctive objects: First, the realization of the implementation of quality development itself as the main outcome of task 2. Second, the adaptation of the quality model selected and adapted at the beginning of task 2. And third, the evaluation of the initiation task 1 including the revision of the vision, strategy and policies for the quality development.

- 4. Task: Act!

The fourth task is dedicated to the sustainability and long-term impact of the adaptation and implementation of the quality development. The communication and further discourse with all stakeholders will guarantee the ongoing debate and consensus building on the quality definitions and common understandings. Based on the evaluation results, the adaptation of the quality model will be revised and the vision and strategy will be reviewed. Finally it should lead the whole organization to the establishment of a continuous improvement cycle for the quality development related to all tasks and steps from the other phases.

The following figure shows the IDEA phases of the IDEAL adaptation model in an overview:
Figure 4-2 The IDEA Phases

The adaptation model IDEA is presented here by using the example of introducing the ISO quality standard RFDQ (ISO/IEC 19796-1). It is an appropriate model for establishing a continuous improvement cycle based on the principles of ISO 900x and the Deming cycle, in particular for the implementation of quality development in learning, education, and training including Learning Analytics. For the PBL3.0 objectives related Learning Analytics the IDEAL model can be used for adapting RFDQ, the unique ISO quality standard for LET, to the specific organizations and given situations to introduce and implement Learning Analytics in problem-based learning (cf. Stracke 2010, 2013, 2014a, 20014b).
5 LA tools

The availability of tools that are suitable for LA has increased in the last few years, adopting existing technologies from similar disciplines such as data analytics and data science. Currently there is a wide variety of tools that can support and implement LA processes towards improved learning and teaching. This is also the case due to the gradually increasing application of technology enhanced learning, where technologies are actively used in order to enhance and support learning (Siemens, 2013). This brought forth the easy recording of educational data, a process that is extremely challenging to perform manually on a traditional learning setting (Picciano, 2012).

Today, the online and blended courses can capture and analyse a wide variety of information, including clicks, navigation patterns, times spent on an activity or a learning resource, social networks, types of resources accessed by each learner etc. (West, 2012). Such courses can also adopt success stories from the business world, e.g. by using dashboards to view LA results in a more visualized and comprehensive way. Such functionalities support all types of stakeholders that are interested in improving learning and teaching.

Existing tools that can be applied in education and training can be divided into three main categories as follows:

- General purpose dashboards
- Ad-hoc tools
- LA standalone tools
- LA plugins

The following sub-sections provide examples for each of these categories of interesting LA tools.

5.1 General purpose LA dashboards

General purpose LA tools include solutions that provide information and visualizations around different types of data for multiple stakeholders, i.e. students, teachers, researchers, institutions etc. These visualizations can be in the form of graphs, tables, charts etc., and aim to assist stakeholders in getting a more holistic overview of course activity realized in a specific platform. The following sub-sections present a few examples of this LA tools category.

5.1.1 GLASS tool

GLASS (Gradient's Learning Analytics SyStem) is a platform that visualizes events realized in a learning environment (Leony et al., 2012). The architecture designed for the development of this platform includes four main components:

- Sensors that send events of actions taken to the database.
- Database that stores the information of the captured events.
• **GLASS viewer** that represents the information.
• **User** that understands and interprets the environment’s visualizations.

The user interface of the tool is shown in Figure 5-1.

![GLASS tool user interface](image)

**Figure 5-1 GLASS tool user interface**

The types of visualizations available include user activity, user statistics, activity evaluations and group statistics. The users can choose one or multiple visualization views for the captured data and thus explore a multimodal dashboard that will help them make beneficial decisions to improve learning and teaching.

**5.1.2 CourseVis tool**

CourseVis (Mazza and Dimitrova, 2007) is a system designed for helping educators become more aware of what is happening during learning and make more informed decisions. This system includes information such as:

<table>
<thead>
<tr>
<th><strong>Social aspects</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Participation in group work</td>
</tr>
<tr>
<td>Participation in discussions</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Cognitive aspects</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall performance in the course</td>
</tr>
<tr>
<td>Level of knowledge achieved for each domain concept of the course</td>
</tr>
<tr>
<td>Students having difficulties with a concept</td>
</tr>
<tr>
<td>Comparison of a student’s progress with that of other students and the whole class</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Behavioural aspects</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Accesses to the course</td>
</tr>
</tbody>
</table>
Reads of course materials
Performance on evaluation proofs
Progress with the schedule

A representative visualization is shown in Figure 5-2, where educators can identify the learners that have initiated threads in a discussion forum. This can provide insights as to who are the students with a more active engagement and leadership skills and who are the students that require scaffolding so that they can start participating more during learning.

![Figure 5-2 Visualization of discussion threads focusing on the students who have initiated the threads](Mazza and Dimitrova, 2007)

This solution aims to be used for tracking generated data within Course Management Systems (CMS). Currently it is only supported by WebCT.

5.1.3 Google Analytics

Google Analytics⁵ may not be developed specifically for LA purposes; however it can be very beneficial for courses that utilize different websites.

---

⁵ Google Analytics: http://www.google.com/analytics
Google Analytics can track any website traffic and feed back to the users information such as statistics about the users accessing the websites, real time overview, types of visits etc.

5.2 LA standalone tools

LA standalone tools have been developed so that they can provide specific information and be used in different platforms. The following sub-sections provide some representative examples of such tools.

5.2.1 SNAPP

The SNAPP\(^6\) (Social Networks Adapting Pedagogical Practice) tool was developed to extract all user interactions especially in discussion forums from various commercial and open source LMS environments, including BlackBoard, WebCT, Moodle and Sakai. This tool performs real-time social network analysis and presents the discussion forum activity on a network diagram, as shown in Figure 5-4. This visualization derives from the analysis of data such as:

- Who posted and replied to whom
- What were the topics of the discussions
- How spread-out were the discussion

---

\(^6\) SNAPP tool: [http://www.snappvis.org](http://www.snappvis.org)
The information presented in the network diagram can help educators detect more easily learners that do not participate, measure the level of engagement for each student and provide scaffolding features to students in need. The social network diagrams can also help educators identify:

- Isolated learners
- Groups malfunctions or conflicts
- Learners that serve as information brokers

5.2.2 LeMo application

LeMo is a web-based LA application that aims to provide informative insights on learners’ navigational patterns and to identify ways to improve and re-design the courses (Fortenbacher et al., 2013). The intended target groups are the educators, researchers and the content providers.
The LeMo application employs various LA methods such as statistics, sequential pattern mining etc., in order to retrieve data from various databases and log files. This data analysis will be analysed and the outputs will be visualized with different visualization techniques, as shown in Figure 5-5. The stakeholders can then study the outputs and assess the learners’ behaviour and the quality of the courses. An important benefit of the LeMo application is that data analysis can be carried out across different learning platforms and LMS environments, allowing the comparison of learners’ behaviours in each environment.

5.2.3 StepUp! LA tool

The StepUp! Tool was developed with an aim to underpin reflection features for educators and learners (Santos et al., 2012). To this end, the tool gathers, analyses and visualizes data such as time spent on a course, usage of resources, usage of social media etc., as shown in Figure 5-6.
The Figure shows that each row describes a specific learner, and learners are clustered together in the groups they belong to. The table also shows the group and each learner’s number of posts, comments, hours spent on the course, number of tweets etc.

StepUp! tracks these types of data, such as posts and comments of learners on blogs, tweets using a specific course related hash-tag etc., through RSS feeds from the blogs and social media’s APIs.

5.2.4 LOCO-Analyst

LOCO-Analyst is a semantic web application that aims to provide feedback to educators in online learning environments, as well as information regarding the quality of the learning content (Ali et al., 2012). The feedback provided by the tool is developed in three different levels:

1. All activities that were designed by the learners during the learning process.
2. The usage and understanding of the learning content prepared by the educators.
3. The social networks developed amongst the learners in a VLE.
The tool was developed as an extension of the Reload Content Packaging Editor, an open source tool for course design, aiming to facilitate educators by using the same tool for the design of their courses, the retrieval of analytical information and feedback and the configuration of their courses.

![LOCO-Analyst user interface](image)

**Figure 5-7 LOCO-Analyst user interface (Ali et al., 2012)**

The analysis carried out in LOCO-Analyst is based on the concept of the Learning Object Context (LOC), which refers to the interactions of a learner with a learning content through a specific learning activity towards a specific goal. Information provided from the analysis can include more commonly discussed topics, time spent on a module, page revisits, quiz scores etc.

LOCO-Analyst also exploits semantic annotation by using the annotation facilities of the Knowledge & Information Management (KIM) platform in order to link different learning artefacts that have certain relations between them, such as lessons, activities, content etc.
5.2.5 Netlytic tool

Netlytic⁷ is a cloud-based social network analysis tool aiming to analyse textual data and social media interactions realized in different tools, such as Twitter, blogs, forums, chats etc.

![Netlytic LA tool](image)

**Figure 5-8 Netlytic LA tool**

Netlytic can detect learners’ engagement levels as well as how active each learner is within a group relationship.

5.2.6 eLAT toolkit

Dyckhoff et al (2012) developed the exploratory Learning Analytics Toolkit (eLAT), aiming to reduce limitations they encountered during learning and teaching, such as the collection, integration and analysis of raw data from log files of the VLEs used in courses. eLAT’s main goal was to improve the online teaching methods depending on personal interests and observations.

---

⁷ Netlytic: https://netlytic.org/home
Figure 5-9 Online-based LA process (Dyckhoff et al., 2012)

A typical LA process starts with the collection of data. This stage accumulates all the learners’ activities from a Virtual Learning Environment (VLE), an LMS or a Personal Learning Environment (PLE). The next stage is the export of the processed data through visual representations, which educators can draw conclusions on their teaching’s efficiency and effectiveness and take appropriate actions.

The success application of the eLAT software required the following features:

- Usefulness
- Use of use
- Extensibility
- Reusability
- Real time operation
- Data privacy

In order to cover all the above features, eLAT comprises of three main components: the Indicator framework, which generated assessment reports, a mining database and a visualizer application, which visualizes the generated reports.
5.3 LMS-based tools

A typical form of online learning involves the use of an LMS system like Moodle or WebCT. These LMS environments require that educators constantly adapt their courses (both the structure and the content) to ensure more comprehensive, high quality and effective learning. The successful application of an LMS for learning requires that educators are made aware of what is happening and how engaged their learners are. This brings forth the need for ongoing feedback (Ali et al, 2012). All LMS environments by default record various data indicators that can be gathered and processed, such as time spent on a page, access time, number of posts, navigation paths etc. and provide reports that present some of this data in a more comprehensive way. However, proper LA outputs that will underpin a higher level of decision making to improve learning and teaching requires additional data gathering in regards to learners’ behaviours, ongoing progress, social networks etc. A number of tools have been developed, usually as plugins or add-ons, to support LA functionality within LMS environments (West, 2012).

The following sub-sections describe a few representative examples of such tools.

5.3.1 Gismo Moodle add-on

Gismo is a graphical interactive monitoring add-on for Moodle, which aims to assist educators in evaluating the social, cognitive and behavioural activities of learners in the online courses (Ali et al, 2012). Gismo can be easily installed in Moodle and provide diverse visualizations regarding learners’ activities, as shown in Figure 5-10.

![Gismo output – learners’ access to a course](image)

Some examples of insights provided by Gismo include:

- Access of learners to a course.
- Access of learners to the resources included in a course.
- Submission of assignments by learners.
- Quizzes overview.

5.3.2 MOCLog Moodle add-on

MOCLog is a toolset that was developed based on GISMO and aims to analyse all interactions that can be recorded in an online course (Mazza et al., 2012). To this end, MOCLog tools create statistical reports and visualizations depending on each user’s role in the system and allow a comprehensive overview of all available data.

![MocLog tool activity report](image)

**Figure 5-11 MocLog tool activity report**

Figure 5-11 shows an activity report created with MOCLog, which presents the usage of the available tools within an online course. Relevant information includes number of usage for every type of tool, comparison of tools’ usage levels etc.

5.3.3 LAe-R Moodle add-on

LAe-R (Learning Analytics Enhanced Rubric) is a Moodle add-on that was developed as an upgraded assessment method and is an improved version of the existing Rubric add-on for Moodle. LAe-R is used for the assessment of learners’ performance based on the analysis of the data gathered during online courses. Educators can determine specific criteria that are to be used for assessment, such as collaboration, scores in assignments, learning content access etc., as shown in Figure 5-12 (Dimopoulos et al, 2013).
In regards to the collaboration, LAe-R analyses and visualizes data related to the forum posts, such as new messages, replies, quotes, number of files attached to forum posts etc. Furthermore, learners’ behaviours are assessed using the number of ideas learners present in the available learning assignments. Educators assess the learners’ performances through these two criteria in a quantitative manner, while a qualitative assessment is carried out through the grading of previous assignments by determining how much learners comprehended the learning material.

5.3.4 SmartKlass Moodle add-on

SmartKlass\(^8\) is an LA dashboard that analyses behavioural data in order to present insights on learners’ performances.

---

\(^8\) SmartKlass: http://klassdata.com/smartklass-learning-analytics-plugin/
The different types of visualizations available can help educators detect at-risk learners and send them warnings and assistance, see the overall progress of a course etc., and can help learners reflect on their progress, compare their performance to that of their group, receive personalized feedback from their teachers etc.

The following table provides a representative list of available free and open source LA tools, frameworks, applications etc.

**Table 5 LA free & open source tools**

<table>
<thead>
<tr>
<th>LA TOOL</th>
<th>TYPE</th>
<th>DATA MANIPULATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dragon ToolKit 1.3.3</td>
<td></td>
<td>Data analysis</td>
</tr>
<tr>
<td>(2008-01-16)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOCO-Analyst3 (2009-12-20)</td>
<td></td>
<td>Data extraction, Data visualisation</td>
</tr>
<tr>
<td>Learning Analytics</td>
<td>MOODLE add-on</td>
<td>Data analysis</td>
</tr>
<tr>
<td>Enriched Rubric 1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R 3.0.2 (2013-09-09)</td>
<td></td>
<td>Data transformation, Data analysis, Data visualisation</td>
</tr>
<tr>
<td>DataMelt 1.2 (2015-07-10)</td>
<td>Application software</td>
<td>Data transformation, Data analysis, Data visualisation</td>
</tr>
<tr>
<td>Gephy 0.8.2 (2013-01-01)</td>
<td>Application software</td>
<td>Data visualisation</td>
</tr>
<tr>
<td>Iramuteq 0.6 alpha 3</td>
<td>Application software</td>
<td>Data extraction, Data transformation, Data analysis, Data visualisation</td>
</tr>
<tr>
<td>Juxta</td>
<td>Application software</td>
<td>Data analysis, Data visualisation</td>
</tr>
<tr>
<td>Application software</td>
<td>Data transformation, Data analysis, Data visualisation</td>
<td></td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>--------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>KEEL 1.0 (2014-01-29)</td>
<td>Data transformation, Data analysis, Data visualisation</td>
<td></td>
</tr>
<tr>
<td>KH Coder2.Beta.30</td>
<td>Data analysis, Data visualisation</td>
<td></td>
</tr>
<tr>
<td>KNOT</td>
<td>Data analysis</td>
<td></td>
</tr>
<tr>
<td>Knime 2.9.1 (2014-01-15)</td>
<td>Data extraction, Data transformation, Data analysis, Data visualisation</td>
<td></td>
</tr>
<tr>
<td>LightSide</td>
<td>Data analysis</td>
<td></td>
</tr>
<tr>
<td>Orange Textable</td>
<td>Data extraction, Data transformation, Data analysis, Data visualisation</td>
<td></td>
</tr>
<tr>
<td>Apache Mahout</td>
<td>Data analysis</td>
<td></td>
</tr>
<tr>
<td>Apache OpenNLP1.5.3</td>
<td>Data extraction, Data transformation, Data analysis</td>
<td></td>
</tr>
<tr>
<td>Apache UIMA2.5.0</td>
<td>Data extraction, Data transformation</td>
<td></td>
</tr>
<tr>
<td>Lingpipe 4.1</td>
<td>Data analysis</td>
<td></td>
</tr>
<tr>
<td>Orange 2.7 (2014-02-14)</td>
<td>Data analysis, Data visualisation</td>
<td></td>
</tr>
<tr>
<td>Gismo 3.2 (2013-10-18)</td>
<td>Data visualisation</td>
<td></td>
</tr>
<tr>
<td>KoRpus</td>
<td>Data extraction, Data transformation, Data analysis</td>
<td></td>
</tr>
<tr>
<td>SNAPP 2.1</td>
<td>Data extraction, Data analysis</td>
<td></td>
</tr>
<tr>
<td>SmartKlass</td>
<td>Data analysis, Data visualization</td>
<td></td>
</tr>
<tr>
<td>LAe-R</td>
<td>Data analysis, Data visualization</td>
<td></td>
</tr>
<tr>
<td>MOCLog</td>
<td>Data analysis, Data visualization</td>
<td></td>
</tr>
<tr>
<td>eLat</td>
<td>Data extraction, Data analysis, Data visualization</td>
<td></td>
</tr>
<tr>
<td>Netlytic</td>
<td>Data extraction, Data analysis, Data visualization, Social network analysis</td>
<td></td>
</tr>
<tr>
<td>LOCO-Analyzer</td>
<td>Data analysis, Data visualization</td>
<td></td>
</tr>
<tr>
<td>StepUp!</td>
<td>Data analysis, Data visualization</td>
<td></td>
</tr>
<tr>
<td>LeMo</td>
<td>Data analysis, Data visualization</td>
<td></td>
</tr>
<tr>
<td>Google analytics</td>
<td>Data extraction, Data analysis, Data visualization</td>
<td></td>
</tr>
<tr>
<td>CourseVis</td>
<td>Data analysis, Data visualization</td>
<td></td>
</tr>
<tr>
<td>GLASS</td>
<td>Data analysis, Data visualization</td>
<td></td>
</tr>
</tbody>
</table>
6 Existing applications of LA

The gradually increasing research on LA has led to multiple interesting applications in real world settings. The investigation of related work indicated that there is still limited evidence of wide LA applications that can provide meaningful empirical evidence of improved teaching and learning (Baker, 2014). Two main reasons of that are the lack of available technical infrastructure in the organizations (Wastiau et al., 2013) and the lack of ICT skilled personnel that can employ LA in their settings (Ferrari et al., 2014).

6.1 LA applications in higher education institutions

Higher education institutions that exploit LA aim to improve students’ learning experience, enhance their performances, reduce the number of dropouts and underpin teachers in improving their courses’ success. Other reasons include allowing students to receive feedback on their progress and reinforcing relationships between students and academic staff. Finally, a higher level motivation is the availability of more informative decision making on an institutional level regarding course structures, curricula, assessment methods etc. The following sub-sections provide some short case studies of higher education institutions that have applied LA in their courses.

6.1.1 Open University of the United Kingdom (UK) case study

Open University (OU) of the United Kingdom (UK) utilizes LA for multiple purposes, namely for improving learners’ performances and underpinning educators’ roles, but also for performing macro-level analytics, e.g. helping their internal staff to understand students’ characteristics. This is carried out through various activities such as:

- Provision of access to data regarding students’ progress and demographic profiles
- Provision of internal and external surveys on students’ feedback
- Provision of institutional data
- Support of internal and external review processes
- Provision of points of students’ withdrawals

The above information can be exploited by all academic members in order to re-design their courses, change their assessment techniques or involve new learning strategies that will improve students’ support system (Prinsloo et al., 2012).

OU UK continues to work on capturing educational data in order to better understand why students behave the way they do and how can learning processes be improved. Towards this goal, OU UK has invested in exploiting the Moodle system installed for their courses in order to track the generated data and get insights on students’ behaviours and actions in the system. The ultimate goal of the OU UK team is to provide a single “fits all” student support model that will allow scalability and
automation in its support to the students, while still allowing educators the power to reconfigure its features when deemed necessary (Jisc report, 2014).

6.1.2 Nottingham Trent university case study

Nottingham Trent University (NTU) aims to enhance students’ academic performances by increasing the information received from learning and exploiting it for improved decision making. From past experiences, educators of the university were trying to support students they thought were in need but were not made aware of others that needed their help more. This can be remedied with LA, since they can help identify at-risk students and allow educators to intervene in time.

NTU teams collected data such as logs from the VLE used, data from their SIS, data from their assessment systems, smart cards, library activities etc. For the analysis of these data, NTU chose a commercial LA vendor to process the fragmented data and provide integrated and interoperable visualizations. These visualizations are available in a dashboard that aggregates all different data sources and provides a holistic overview of students’ engagement levels, as shown in Figure 6-1.

![Figure 6-1 LA dashboard at NTU](image)

These engagement levels are calculated by tracking attendance, usage of learning resources, assignment submissions, assessment grades, library loans, card swipes and academic history. Based on these visualizations, the educators can contact students with a low engagement level, alert them on their status and provide them with guidance (Jisc report, 2014).

6.1.3 University of Bedfordshire case study

The University of Bedfordshire utilizes LA in order to enhance students’ experience by monitoring and improving student engagement and student retention (Jisc report, 2014). Another main goal is
to reinforce the institution’s decision making processes by examining and consulting evidence-based information.

To this end, the university has utilized main concepts from the Business Intelligence (BI) field and incorporated respective technologies into their systems to perform enhanced LA. This initially involves collecting data from various sources, such as attendance records, library (e.g. number of visits, resources checked out etc.), student ID cards records, student information helpdesk data, data from their VLE (Blackboard), login information to the university’s network, assignment submissions, exam scores etc.

The processing of the above data is carried out through a number of LA methods, such as statistical analysis, predictive analysis, and the results are viewed on an improved Student Engagement System (SES) developed, as shown in Figure 6-2.

Figure 6-2 SES environment at University of Bedfordshire

The above Figure visualizes the engagement levels of a specific student in comparison to their peers. Other types of visualizations within the SES dashboard include risk alerts for at-risk students, student engagement activities, student engagement measurement indexes etc.

Some of the most prominent issues and lessons learnt from the university’s usage of their LA environments include challenges related to:

---

9 https://www.flickr.com/photos/jiscinfonet/sets/72157632977073177/
data collection devises, managing the increasing volume of data generated and ensuring accuracy and reliability;

- managing the database systems for possible bugs and overload;

- raising awareness regarding the importance of the LA’s application in order to involve the staff and more senior stakeholders.

6.1.4 Purdue university case study

A very commonly referred to case study on LA application is Purdue university, who aims to improve student’s performances and detect potential problems arising during learning. To this end, the Course Signals system was developed, which is based on the Blackboard Learn system (Caulfield, 2013).

The Course Signals system collects and analyses data such as historical grade information, attendance, scores etc., and determines possible problems that may occur during learning. While analysing this data, the system marks each student with a specific status for the respective course using green-yellow-red labels, where each label shows how at-risk the student is, as shown in Figure 6-3.

Some of the lessons learnt from the system’s application in the university’s courses include an improved retention by 20% and an increased completion of the degree by 4% (Norris and Baer, 2013).
6.2 LA case studies in business

Many companies and associations perform trainings on their employees in order to ensure that the staff will be constantly skilled and knowledgeable on the organization’s interested fields and any new developments that can improve the organization’s goals. This also agrees with the developing concept of lifelong learning, which has become a basic requirement for economic growth and personal development in Europe (European Commission, 2013).

The need for lifelong learners in all professional fields has also caused an increased interested of Human Resources in organizations about data generated for learning. Some key questions that can be answered more quickly, accurately and effectively with the application of LA in trainings include (Nielson blog, 2013):

- What components of the training programs are the most and the least effective?
- What topics are more effective for each employee?
- At what parts of the trainings are the employees most and less engaged?
- How are the assessment of the employees relate to their performance?

The European Association for Practitioner Research on Improving Learning (EAPRIL) along with Maastricht University carried out a project that resulted in the European Learning Compass, which is a systematic analysis and overview of reports on LA in companies of Europe (Jager et al., 2014). This project gathered data (e.g. documents, websites, reports, publications etc.) from 126 different associations and companies across Europe regarding their training programs and their applications of LA. Their study relies on a theoretical training model that describes the process of formal training in companies, shown in Figure 6-4.

![Figure 6-4 Formal training model (Salas et al. 2012)](image-url)
The model presents the different phases before, during and after training as well as the processes carried out within each phase.

The regular evaluation process in training was mostly based on employees’ feedback (Elias, 2011). However, this method does not provide a comprehensive and overall insight on training’s effect and success. To this end, LA has started becoming used in work trainings for every phase described in the training model. The study performed by EAPRIL gathered data regarding the number of different LA being used for each phase of the training process. The results showed that the investigated companies use 52 different LA in the different phases of the training model, as shown in Figure 6-5.

![Figure 6-5 LA mapping to the training model (Jager et al., 2014)](image)

An interesting insight is the fact that the majority of the LA (25 out of 52) are linked to the before training phase. More specifically, the training needs analysis seems to be the stage that most interests companies in regards to gathering data and getting informative feedback. This can be the case as this stage affects the entire training modelling, since the materials, processes, activities and evaluation techniques need to be in accordance with the training needs.

### 6.3 LA case studies in MOOCs

In an effort to bring together the different domains, objectives, levels of analysis and processes for learning analytics and MOOCs into a joint picture the MOOC Learning Analytics Innovation Cycle (MOLAC) has been developed (Drachsler, and Kalz, 2016). The cycle works on three different levels. At the micro level, data from a single course is collected to foster predictions and reflection for individual learners or teachers. At the meso level, educational institutions combine several MOOCs and enable the sharing and analysis of data beyond a single course via metadata standards. The
combined data from different MOOCs can be used for classification of learners and contributes to the heavily debated notion of learner types and learning styles in a more informed and data-driven approach Figure 6-6. At the macro level, the analysis is conducted across MOOC providers and curricula and data is shared between providers via a data repository. This type of cross-institutional learning analytics targets the identification of interventions that contribute to the innovation of learning and teaching for the individual institution but also for a wider group of stakeholders like the learning science community at large.

Figure 6-6 The MOOC Learning Analytics Innovation Cycle (MOLAC)

In order for learning analytics to be useful in the context of MOOCs different approach needs to be followed since different pedagogies are deployed due to the massive number of students. Currently OUNL is involved in two MOOC projects were LA have been deployed: EMMA\textsuperscript{10} and ECO\textsuperscript{11} projects.

According to Brouns (2014) value is derived from the use of MOOCs when use of LA is carefully designed and a model is defined first in order to know what and why needs to be known. Based on

\textsuperscript{10} http://platform.europeanmoocs.eu/
\textsuperscript{11} http://project.ecolearning.eu/
that the proper indicators to be measured can be selected. Among indicators that have been selected to be measured in these projects are: Performance, Mastery (of knowledge), Progress, Engagement, Effort, Satisfaction and Social affiliation.
7 Existing approaches on using LA in PBL-environments

This section aims to present a few case studies that have attempted to integrate LA in PBL environments. The relevant literature is still limited on this subject, as there seems to be a lack of available knowledge on how to utilize LA when the courses are structured using the PBL strategy.

7.1 LA and PBL in mathematics and statistics

Tempelaar et al. (2013) carried out a research study for learning and teaching of mathematics and statistics in a blended learning environment.

7.1.1 Methodology

The methodology used includes a face-to-face part of the learning process which followed the PBL strategy based on the Maastricht PBL model and an online part which is optional. Within PBL, students were required to form small groups, where each group had a specific content expert and mentor. The online components are three digital environments, namely the Blackboard LMS and MyStatLab and ONBETWIST used for test-directed learning. Students’ engagement with the online technologies is optional as this is more in line with the Maastricht model, where a) students are responsible for their own learning and educational choices and b) students have different prior knowledge and thus don’t always gain knowledge and skills by using the same environments.

The course begins with a question, and students are encouraged to attempt to answer it. When students find difficulties answering the question, they are given two options to solve the problem:

- Help Me Solve This: students ask for help on how to solve the problem using a step by step method.
- View an Example: students ask for a holistic example that shows how similar problems have been solved.

Once students acquire new knowledge from this process, they are provided with a new problem that they try to solve based on what they have learned so far. Additionally, students take quizzes every two weeks for formative assessment of their progress. These are taken in the online environments, which encourages students’ participation in the online tools and allows them to reflect on their scores and compare their progress to that of their peers.

7.1.2 Data collected and LA performed

The data collected during the learning process included:

- Performance scores
- Frequency of using the practice tests
- Time spent on practice tests
- Number of attempts at solving a problem
- Variables that demonstrate prior knowledge
- Student profile
- Groups formed
- Engagement levels
- Learning styles

The analysis of the gathered data provides insights to students that help them monitor and reflect on their progress, their level of comprehension and get feedback from the teachers on their performance. The teachers are the facilitators in this process, and thus view LA outputs that help them guide and scaffold the students when needed (e.g. low performances, isolated students within groups, at-risk students, common mistakes frequently made etc.) as well as perform interventions through the overview of predictive results (e.g. predicted performances on future maths exams).

The case study concluded that the usage of the online environments as a complimentary tool to PBL proved to help students, as they helped in the support of self-direction, reflection and decision making. Students of PBL are usually new at this learning model, where they hold the majority of the responsibility to gain knowledge, and it seems that visual feedbacks on how the self-directed learning is proceeding has made them more confident in this control shift.

### 7.2 Doctors training with Analytics and PBL

Göhnert et al. (2014) have developed a web-based Analytics workbench that is generic and easily extensible and supports various analytics processes and visualizations. The workbench allows multiple types of analytics results as separate modules, which educators can choose and retrieve insights from, as shown in Figure 7-1.
Some representative examples include:

- Community detection
- Networks of interactions amongst learners and course elements (resources, activities etc.)
- Log statistics
- Workflows
- Identification of central actors
- Diverse visualizations with graphs
- Activity statistics

The workbench has not been developed specifically for education and training, however the authors believe that it can easily be employed in such settings, and perform three different case studies to determine the validity of their assumption. One of the case studies involves a project called KOLEGEA, which provides trainings for doctors practicing in family medicine using the PBL strategy. Medicine is one of the educational domains where PBL is most commonly used, and in this case, doctors were required to generate cases and solve these cases in a PBL context, i.e. they can
collaborate, share and discuss the cases in small groups, while the trainers mentor and guide the process.

The integration of the platform used in KOLEGEA with the Analytics workbench required the gathering of data from the platform so that it can be processed and analysed within the workbench (e.g. log files, user profiles, data on course elements etc.). The LA results available after this integration include graphs or data tables with results similar to the modules supported by the workbench. No specific information is provided in the study regarding the effectiveness of the solution and actual LA outputs from the doctors’ PBL practices.
8 Ethical issues and challenges

This chapter provides information regarding some of the most commonly cited ethical issues, limitations and challenges concerning LA.

8.1 Ethical issues

One of the most commonly cited definitions for ethics with focus on the digital world describes them as “the systematization of correct and incorrect behaviour in virtual spaces according to all stakeholders” (Pardo & Siemens, 2014). The specific ethics that should be taken into consideration for any domain are usually quite complex to locate, decode and make sense of (Willis, 2013). Documents that can be studied and consulted include the Nuremberg Code, the UN Declaration etc. provide information regarding protection, safety, human rights, minimization of harm etc. (AoIR, 2012). The principles available in such documents are easily applicable in the LA field, as data handling is relevant to all the aforementioned ethics principles.

The LA field is still quite new and, despite its many benefits, the research and educational world is still sceptical about the actual application of the concept in real world settings. The most crucial issues that concern the stakeholders interested in LA involve all the ethical and legal issues around the collection, analysis and visualization of data that may be considered personal or sensitive. Addressing such issues would involve proper anonymization and cleaning of learner-related data in accordance to national or European legal requirements or data policies (Greller & Drachsler, 2012).

Furthermore, consent is essential for proper LA application. Learners should always be aware that their actions are being recorded as well as to who has access to their data, how their data is being used and to always be allowed to ask the termination of their data’s recording.

Apart from the above, other ethics issues are involved in many levels, such as follows:

- **Accountability.** Creation of LA outputs with validation processes and accuracy verification.
- **Clarity.** The purpose as well as the scope of the data gathering and analysis should be publically and openly stated.
- **Transparency.** The educational data gathered should be open and all interested stakeholders should be aware of the LA processes being carried out and the ways in which the LA results will be exploited.
- **Privacy.** Specific mechanisms should be put in place for handling sensitive and out of scope data (e.g. cleaning, machine learning, transformation etc.).
- **Access.** This refers to the different data formats, and different types and levels of access to data by each stakeholder.
- **Validity.** The gathered data should be educational in nature, relevant to the educational goals set and accurate.
Furthermore, it is important to investigate the amount of data gathered and analysed; sometimes more data does not mean better results. This requires proper LA tools that can harvest the data correctly and more importantly, educators with strong decision making and critical thinking skills who can determine which and how much of the available data can and should be exploited for analysis and to what end. During this process, educators will also decide whether the available data is also ethical and in line with the ethics principles (Dimitracopoulou, 2015).

### 8.2 Limitations and challenges

#### 8.2.1 Limited LA practices

Even though LA is a newly developed domain, there is already significant research being done on all its dimensions. However, this is not the case with the actual application of LA in real world settings, where “The work of researchers often sits in isolation from that of vendors and of end users or practitioners” (Siemens, 2012, p.5). To this end, limited evidence-based results are available to address LA challenges that hinder the successful transformation and improvement of learning and teaching.

This is also reinforced by Figure 8-1, which shows the different domains that use Data Analytics and how prepared the organizations of the domains are to properly exploit these analytics. Based on the Figure, education is one of the least prepared domains for data analytics, especially regarding ease of data capture, IT usage, data-driven mind-sets and overall data availability (Educause report, 2013).

The limited existing practices are usually focused on gathering and analysing simple educational data and getting informative visualizations; however, no promising progress has been made in finding out and implementing ways to translate the LA outputs into actions that address learning and teaching issues. Furthermore, the availability of LA tools does not necessarily mean that they are being used properly or sufficiently in education and training. Studies indicate that teachers and trainers are still not appropriately knowledgeable on LA and do not fully understand how to use the available tools in order to exploit the potential benefits for improving learning and teaching (Swan, 2012).
Studies indicate that the future research on LA should focus on the learning sciences dimension so that it will be easier for learners and educators to transform the learning processes based on the LA outputs (Dyckhoff et al., 2013; Ferguson, 2012).

8.2.2 Objectivity of LA interpretations
A commonly expressed concern regards the possibility that educators will rely exclusively on the LA results for the assessment of their students and do not try to make sense of the information.
presented, review it and then determine which parts of the LA outputs they will exploit. This strongly diminishes educators’ abilities to think and observe critically and interact with the students. Furthermore, sole usage of LA for assessment can cause incorrect assumptions for students, since the assessment process requires a more holistic approach that takes into consideration other elements along with LA outputs, e.g. learners’ profiles, specific contexts of the course, particular conditions etc.) (Siemens & Long, 2011).

8.2.3 LA software

The proper and effective usage of relevant software is essential in the LA domain. Higher education institutions and training organizations use a variety of diverse technologies during education and training for their courses. This significantly affects the LA process and the data quality, as the data generated is distributed across multiple sources and platforms and can be accessed in different formats (Reyes, 2015).

Furthermore, each LA tool has been developed for specific purposes, and usually access and process specific types of data. This, in turn, means that the majority of LA software will not be able to record the complete set of the learning experience or holistically assess learners’ behaviours.

The correct usage of LA software is usually limited due to incorrect application by the educators. Educators need to determine the learning objectives at the beginning of the course design phase, and correspondingly decide on which data indicators and measures are relevant for analysis before choosing and applying specific LA software.

Another significant challenge regards the time and effort needed for successful LA process and software application. Very few educators will find the time to properly design their courses and incorporate the correct LA software. This is also connected to the usual lack of collaboration amongst educators, the IT staff, the admin staff and the decision makers. Such collaborations can lead to more holistic overviews on the educational and training needs and to long term decisions on how to address these needs through effective LA mechanisms.
9 LA concepts in PBL3.0

The LA analysis carried out in the previous sections of the deliverable aim to provide a holistic overview of the field. This overview will provide informative insights on which of the field’s concepts are relevant to the project’s objectives and what are the different aspects related to each concept that should be taken into consideration during the work done in the next Task of the current WP as well as in the following WPs. This investigation covers issues such as the study of all mechanisms regarding:

- **Data gathering.** We aim to answer questions such as: what kind of data can and should be gathered during learning, in what state, in what formats, where does it get stored and how etc.
- **Data processing.** We aim to investigate and analyse the processing that can be carried out, e.g. data mining, aggregation, clustering etc.
- **Data analysis and visualization.** We aim to identify the information can be drawn, e.g. learners’ engagement, social networks, risks of flight, behavioural interactions, group collaborations etc., as well as the ways in which this information is visualized.
- **Ethical issues and limitations.** During this investigation, we will follow the classification of the LA ethical issues suggested by Slade & Prinsloo (2013), i.e. location and interpretation of data, informed consent, privacy, and the management, classification and storage of data.
- **Intervention / recommendation methods.** We aim to identify and examine all intervention and recommendation mechanisms currently available, e.g. recommendations on what activities students should participate in based on their progress and performance, detection of students at risk of failure and corresponding notification etc.

Towards this end, this section describes and elaborates on the most prominent aspects regarding each of the above concepts that are also strongly connected with the other two main components of the project, i.e. learning semantics and PBL.

9.1 Data gathering

The first step of LA refers to the selection and gathering of data. The PBL strategy involves active participation of learners, which means that they will generate large amounts and multimodal data. The most important aspect of this data is that it is educational in nature, as the success of LA in education and training always relies on what types of data is gathered and analysed. This also is strongly dependant on the types and kinds of data that is available for gathering and analysis and that can provide meaningful information that will improve education and training (Pistilli et al., 2014).

Dyckhoff et al. (2013) carried out an extensive review of LA state of the art and listed 200 different indicators that are used for LA and mapped the different data sources in six main categories, i.e.
student generated data, academic profile, performance, evaluation, course metadata and context
data. The study carried out indicates that the majority of the data that is exploited by existing LA
tools regards basic data about student activity and performance and it usually comes from a single
source, i.e. an LMS. This limits the level of insights available to the educators for improved decision
making; thus, it is essential that the educators themselves decide on which indicators to be used for
data gathering based on their specific objectives.

The following Table includes a list of possible data sources and types that would be beneficial to be
gathered in a PBL based environment during education or training. These data may come from
either the learners or the teachers/trainers or from the learning technology itself (e.g. log data,
records on movements, traffic data etc.). It may also be structured (e.g. scores per learner, log data)
or unstructured (e.g. forum posts, Facebook posts) (MacNeill et al., 2014; Van Harmelen &
Workman, 2012).

### Table 6 Data gathering for PBL3.0

<table>
<thead>
<tr>
<th>Data type</th>
<th>Data source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assignments submitted</td>
<td>Learner – generated data in LMS, VLE, MOOCs etc.</td>
</tr>
<tr>
<td>Courses enrolled</td>
<td></td>
</tr>
<tr>
<td>Times spent on each unit of learning</td>
<td></td>
</tr>
<tr>
<td>Access to learning resources</td>
<td></td>
</tr>
<tr>
<td>Navigation patterns</td>
<td></td>
</tr>
<tr>
<td>Frequency of log ins</td>
<td>Learner – generated data in LMS, VLE, MOOCs etc.</td>
</tr>
<tr>
<td>Activities accessed / used</td>
<td></td>
</tr>
<tr>
<td>Posts on forums</td>
<td></td>
</tr>
<tr>
<td>Number of participants per group</td>
<td></td>
</tr>
<tr>
<td>Clusters of students who made specific mistakes</td>
<td></td>
</tr>
<tr>
<td>Contributions to shared documents</td>
<td></td>
</tr>
<tr>
<td>Facebook posts and interactions (replies, shares, tags)</td>
<td>Student – generated data in social media and online</td>
</tr>
<tr>
<td>Twitter posts and interactions (replies, retweets)</td>
<td></td>
</tr>
<tr>
<td>Blog posts and interactions (comments, ratings)</td>
<td></td>
</tr>
<tr>
<td>LinkedIn posts and interactions (networks,</td>
<td></td>
</tr>
</tbody>
</table>

---

Page 78 of 92
All the above data is generated in large amounts in student-centred learning strategies such as PBL, and it is usually stored in the respective databases and not exploited for decision making and enhancing learning and teaching (Siemens, 2013). PBL also usually involves and allows multiple environments for student engagement, which causes a significant challenge in regards to gathering data from distributed learning environment for analysis (Ferguson, 2012). However, the exploitation of data from multiple sources can provide a more holistic and accurate insights of the learning experiences and thus reinforce learning and teaching (Papamitsiou & Economides, 2014). Thus, it is important to address the integration challenge and find innovative ways of exploiting multi-source data.

Another important issue that is closely related to the project’s objectives is the determination of appropriate educational goals, which will guide and strongly influence the identification of the data to be gathered. The PBL strategy relies on the definition of a problem, which determines the way the learning process will proceed. Thus, the selection of the corresponding data indicators from the beginning of the PBL implementation will influence the type, quality and accuracy of the LA outputs (Van Harmelen & Workman, 2012).

In principle, two main processes are carried out during data gathering, as follows:

- **Data selection.** This process involves the initial definition of the objectives, i.e. what are the desired outcomes of the educational or training experience (Elias, 2011). This will in turn raise the questions of which data is available that is in line with the set objectives, where they are generated and in what formats.

- **Data acquisition.** This process involves the actual collection and storage of the selected data, which may include one or multiple data sources. This is strongly connected with the
pedagogical strategy employed in every educational or training setting, i.e. PBL, inquiry based learning, flipped classrooms etc. (Pardo, 2014).

9.2 Data processing

The data processing phase in LA regards all the operations carried out to the captured data so that they are transformed into appropriate formats for further analyses. This process is more complicated in a PBL setting when multiple data sources may be utilized; in these cases, operations like cleaning, aggregation etc. aim to remove unwanted or irrelevant data and locate attributes that can allow interoperability and integration of similar datasets for more homogenous analysis. The creation of a unified format of educational data is considered one of the most important challenges of LA (Pardo, 2014). This issue will be investigated within the project through the research of the LS field and all the available ways to provide common meaning to educational artefacts (Cooper, 2014).

9.3 Data analysis and visualization

The data analysis and visualization phase in LA regards all the operations that transform raw data into valuable information for conducting improvements in education and training. The level of analysis and the complexity of the representation / visualization of the LA outputs depend on the available tools and methods, the goals set by the educators and the types of data available.

The most common visualizations are reports and dashboards that report back the analysed data to the stakeholders, i.e. educators and learners. These dashboards provide a holistic and comprehensive overview of the LA outputs and are used for monitoring and reflection of the learning process (Pardo, 2014; Elias, 2011).

Student-centred strategies such as PBL require constant visualization of the learning process in order to empower the learners’ active role and support their paths towards solving the set problems. This will reinforce their 21st century skills and help them become less dependent on the educators and more self-directed (Brown, 2013; Buckingham, 2011). On the other hand, educators become facilitators and thus are in an increased need to have an overview of the learners’ progress and provide assistance whenever needed. Visualizations that would help in both of these cases include dashboards that show:

- Student activity
- Student navigation pathways
- Student interactions
- Student interim scores
- Student engagement levels
- Comparison amongst groups progress
- Hints, alerts and other scaffolding mechanisms (e.g. suggestions on learning materials)
• Common mistakes
• Learner profiles in accordance to their learning style, ongoing progress and recorded needs

All the above can underpin a PBL-based environment where educators will feel confident in their supporting role as they will be able to view progress and assist when necessary, and where learners will confident in leading their own learning by doing activities as they will be able to reflect on their work and get help according to their needs.

9.4 Interventions / recommendations

The more complex and demanding phase of the LA lifecycle regards operations related to decision making actions for transforming learning and teaching. These mechanisms are important for the project, as they reinforce adaptive and self-directed learning.

Relevant operations can be categorized into six main categories (Chatti et al., 2012), as follows:

• **Predictions and interventions.** This regards the operations that answer previously asked questions about possible outcomes. Examples of such actions include estimations of a learner’s future scores, potential dropouts or at-risk learners, etc. These actions can be carried out through a number of LA methods such as statistical inference and machine learning (Pardo, 2014).

• **Adaptations.** This regards the tailoring of the course materials or learning activities based on a learner’s progress in order to ensure maximum performance (Pardo, 2014). The types and number of adaptive actions to be performed is a decision made by the educators depending on the learning context and the specific learner.

• **Mentoring.** This regards all the actions that are performed by the educators in their facilitating roles, especially in student-centred learning strategies. In these cases, educators study the LA outputs and provide guidance and advice based on the LA information.

• **Assessment and feedback.** This regards the support of formative and self-assessment actions during the learning process as well as the provision of informative feedback to both educators and learners regarding the ongoing progress. Continuous assessment is an essential objective of PBL3.0 and such mechanisms will be studied during the project.

• **Recommendations and personalization.** This regards the provision of suggestions to educators and learners based on specific learner-based LA information.

• **Reflection.** This regards the provision of visualizations and information that show educators and learners the ongoing progress, current achievements and mistakes made, comparison between learners and groups, across courses etc.

All above intervention mechanisms are in line with the project’s objectives as they support and reinforce student-centred and self-directed learning strategies such as PBL. More specifically, the
project aims to focus on intervention methods regarding continuous assessment, feedback and reflection, actions that support PBL and allow collaboration amongst the educators and the learners.

The project aims to focus on three specific dimensions regarding LA, as follows:

- Enhanced continuous assessment
- Adaptation of the course’s implementation process
- Adaptation of the course’s design
- Understanding and capturing of learning resources

More specifically, it is essential for PBL3.0 to determine all the ways in which LA can reinforce and support ongoing assessment in all the phases of the learning process, by using innovative modelling and meta-modelling techniques and tools (e.g. ADOxx environment, assessment design models etc).

To be more specific, the Evidence Centered Activity and Assessment Design (ECAAD) methodology – developed with the ADOxx meta-modelling platform - can be used. The ECAAD is based on the Evidence Centred Design (ECD) methodology, arguing that assessment tasks are - in most cases – designed tasks. In terms of the basic ECD approach, it is argued that assessment is supposed to be interpreted as a top-down view, meaning getting from students’ knowledge and learning to activities in the actual learning environments. (adoxx.org, 2016).

- **ECAAD approach.** The idea is taking NEXT-TELL’s ECAAD approach as a basis, and adapting it - taking into consideration the specificities of the PBL_LA approach. Briefly, in ECAAD the ECD approach is extended through 1. generalization of ECD through involving formative assessment and several ICTs, 2. meta-modelling approach for learning activity sequences and assessment methods, 3. extending assessment designer role to teachers, and full integration of assessment design into activity planning. (adoxx.org, 2016) Usage of NEXT-TELL’s ECAAD Planner tool as a planning approach for activity and assessment design is considered. (adoxx.org, 2016) This is elaborated on and enhanced in more detail in D3.1 and D3.2. The model-based ECAAD approach is going to support “learners and teachers in their knowledge-intense planning, design and specification task for the 21st century, technology-rich classroom to allow for knowledge sharing, collaboration among domain experts and operationalization of the created artefacts.” (Utz, Kossowski, Misiak, 2013).

- **Methodology.** In terms of methodology, the first step is conceptualization, referring to specification and translation of requirements in terms of the LA approach; secondly, implementation/adaptation of the tool using the ADOxx platform; thirdly, deployment, release management and functionality packaging considering different users and technologies. (Utz, Kossowski, Misiak, 2013).

Overview of the ECAAD procedure:
This includes both assessment of the:

- course design phase (e.g. appropriateness and multimodality of learning content prepared, relevance and diversity of learning activities designed, difficulty levels and adaptability of assignments designed, suitability of learning strategy employed etc.),

- course implementation phase (e.g. learners’ performance, degrees of learners’ engagement, development of learning communities, quality of learners’ outcomes etc.),

- course evaluation phase (e.g. number and diversity of evaluation methods employed, appropriateness of evaluation methods, feedback gathered from learners etc.).

Additionally, the large availability of intelligent data (e.g. open data, linked data, semantic data), and other types of data (course data, learner profiles, other learner data etc.), requires that these are capitalized properly for enhanced predictions and interventions during learning and teaching. The project’s future work on LS will also aim to contribute to this endeavour. When we start to
understand the learning resources and annotate them with all of their elements, course adaptation and personalization becomes much easier.

Finally, PBL3.0 aims to tackle issues related to successful course design with the exploitation of LA results. Usually, learning content, resources and learning activities are pre-designed before the courses being. Thus, these predetermined designs can be adapted and reconfigured if they take into consideration the LA outputs regarding learners’ performances and progress throughout the course. When instructional designers have detailed knowledge on when and where learners experienced difficulties, requested feedback, answered the same questions incorrectly, performed really well etc., they can transform these insights into actual design alterations for the next course implementation (Siemens, 2013).
10 Conclusions

The purpose of this deliverable was to present all the work done towards the analysis of LA and its components. These components are then analysed so that they can be later on integrated with the PBL learning strategy towards a novel educational paradigm.

More specifically, this deliverable discussed:

- Different definitions of LA as well as the different principles related to the domain. These definitions provided an introduction to the field as well as the most prominent keywords to be used for an in-depth LA analysis. Such keywords include: data measurement, collection, analysis, prediction, reporting, tracking etc.

- Existing LA frameworks that describe the field as well as all the concepts relevant to LA. Such concepts included:
  - stakeholders that are interested in or involved in LA (e.g. learners, teachers, researchers, instructional designers etc.);
  - objectives that guide each LA component, i.e. the purpose that each LA functionality serves (e.g. reflection, prediction etc.);
  - educational data that are gathered, processed and analysed (e.g. students’ scores, access times, navigation paths, social networks etc.);
  - tools and technologies that are used for each LA operation (e.g. social network analysis, data visualization, data gathering etc.);
  - methods and techniques that are employed to cover different LA operations (e.g. clustering, relationship mining, discovery with models etc.).

- Methods and techniques that are used for carrying out the analysis of educational data and the different ways we can exploit the techniques in order to improve teaching and learning. Indicative examples identified in the study carried out include behaviour modelling, discourse analysis, sentiment analysis, natural language processing, association rule mining, knowledge domain modelling etc.

- The LA lifecycle as portrayed in different LA models, as well as all the different steps identified in each model that lead from raw data to meaningful information that has significant added value to the enhancement of education and training. The models described in this deliverable present different LA steps; however, multiple similarities and overlaps are identified. The most commonly identified steps in the LA lifecycle include data gathering, data cleaning, data storage, data analysis, integration / aggregation, data visualization and interpretation.
• Existing tools that have been developed and can support and implement LA processes towards improved learning and teaching. The deliverable provided a brief overview of such tools, including:
  o General-purpose LA dashboards, which are solutions that provide information and visualizations around different types of data for multiple stakeholders. Such tools include CourseVis, GLASS and Google Analytics.
  o LA standalone tools, which provide specific information and be used in different platforms, such as SNAPP for social network analysis, LeMo for detecting learners’ navigation patterns, StepUp! for allowing reflection etc.
  o LMS-based tools, which have been developed, usually as plugins or add-ons, to support LA functionality within LMS environments, such as Gismo for evaluating the social, cognitive and behavioural activities of learners in the online courses, MOCLog for analysing all interactions that can be recorded in an online course etc.
• Existing applications for LA in real world settings, i.e. in academia, business and in MOOCs. The study identified several higher education institutions that exploit LA in order to improve students’ learning experience, enhance their performances, reduce the number of dropouts and underpin teachers in improving their courses’ success.
• Existing approaches on using LA in PBL-environments, which attempt to integrate LA in PBL environments. The relevant literature is still limited on this subject, as there seems to be a lack of available knowledge on how to utilize LA when the courses are structured using the PBL strategy.
• Ethical issues and challenges related to gathering, processing and analysing educational data. Such issues include anonymization and cleaning, consent, transparency, privacy and validity. Additionally, the limited available LA practices, the objectivity in interpretations of LA results, and the proper and effective usage of relevant software are some of the challenges and limitations identified in regards to the LA effective application in real world settings.

Finally, the deliverable provides an overall overview of all LA concepts identified that are relevant to the PBL3.0 project and that can be utilized for integration with the PBL strategy and with LS techniques. This overview provides information on the four main steps of LA, i.e. data gathering, data processing, data analysis and visualization and interventions / recommendations and identifies the most prominent components of each step, i.e. the specific educational data gathered, the different processing methods employed, the types of analysis and visualizations performed and the ways of intervening during the course and making recommendations based on LA results.
References


Cooper, A. (2014). Learning Analytics Interoperability-The Big Picture In Brief.*Learning Analytics Community Exchange*.


Markham, A, & Buchanan, E. (2012). Ethical decision-making and Internet research 2.0: Recommendations from the AoIR ethics working committee.


