Elearning, Communication and Open-data: Massive Mobile, Ubiquitous and Open Learning

Deliverable Title	D2.5 Learning analytics requirements and metrics report
Deliverable Lead:	OUNL
Related Work package:	
Author(s):	Francis Brouns, Marta Elena Zorrilla Pantaleón, Elena Esperanza Álvarez Saiz, Pedro Solana González, Ángel Cobo Ortega, Eliana Rocío Rocha Blanco, Marta Collantes Viaña, Carlos Rodríguez Hoyo, Mariana De Lima Silva, Carmen Marta-Lazo, José Antonio Gabelas Barroso, Pilar Arranz, Luis García, Alejandro Silva, José Manuel Sáez López, Patricia Ventura Expósito, María Jordano de la Torre, Felix Bohuschke, Javier Viñuales
Dissemination level:	PU
Due submission date:	31 January 2015
Actual submission:	27 January 2015

Abstract	In MOOCs, learning analytics have to be addressed to the various types of learners that participate. This deliverable describes indicators that enable both teachers and learner to monitor the progress and performance as well as identify whether there are learners at risk of dropping out. How these indicators should be computed and displayed to end users by means of dashboards is also explained. Furthermore a proposal based on xAPI statements for storing relevant data and events is provided.
Keywords	ECO, sMOOC, learning analytics, learner support, progress, performance, risk of dropping out, xAPI, visualisation, learning analytics dashboard

Co-funded by the European Union

Project funded by ICT Policy Support Programme as part of the Competitiveness and Innovation framework Programme (CIP)

Disclaimer

Versioning and Contribution History

Version	Date	Main modification or revision	Contribution by
Google drive		Initial table of contents, discussion, writing and revision of all chapters	all partners
v0.8	22 December 2014	Compilation of Google docs	OUNL
v0.9	24 December 2014	Last changes from Google and final formatting, submitted for review	OUNL
v1.0	28 January 2015	Incorporated reviewers' comments	UNICAN, UNIZAR, OUNL
v1.1	4 February 2015	Applied new ECO template and CC licence	OUNL

Main contributors to the document:

Partner	Contributors
OUNL	Francis Brouns
UNICAN	Marta Elena Zorrilla Pantaleón, Elena Esperanza Álvarez Saiz, Pedro Solana González, Ángel Cobo Ortega, Eliana Rocío Rocha Blanco, Marta Collantes Viaña, Carlos Rodríguez Hoyo, Mariana De Lima Silva
UNIZAR	Carmen Marta-Lazo, José Antonio Gabelas Barroso, Pilar Arranz, Luis García, Alejandro Silva
UNED	José Manuel Sáez López, Patricia Ventura Expósito, María Jordano de la Torre
Humance	Felix Bohuschke
Geographica	Javier Viñuales

Reviewers to the document:

Partner
Polimi
UMB
FEDRAVE
A. Alfaro

Table of contents

List of tables

List of figures

Executive summary

Learning analytics can, when properly applied, provide valuable support to learners, teachers and course designers. Learning analytics are applied to both provide reflection and prediction. Learners can get insight into their learning process, teachers can be informed about issues arising pointing them towards remedial actions and course designers could find out areas of the course that need attention or modification. In the context of MOOCs (Massive Open Online Courses) the heterogeneity in the learner population asks for personalised learning analytics that are geared towards the intentions of (groups of) the learners. Moreover, ECO promotes social and seamless MOOCs that focus on social and networked learning in authentic and situated contexts that poses additional requirements and challenges for the application of learning analytics.

In this deliverable we explain the value of learning analytics for ECO's social and seamless MOOCs (sMOOC). This summary summarises and highlights key value proposals of the different chapters.

Defining concepts

After a brief introduction, we start in chapter 2 with an overview of Learning Analytics (LA), definitions, uses and applications, justification of its need in MOOCs and challenges, limitations and ethical concerns to be addressed. The chapter provides a review of mathematics and computing techniques which have been profusely used in learning analytics and educational data mining.

LA is able to predict learning issues, identify students or learners at risk and allows intervention while a course is in progress. Personalized learning and adaptive pedagogies and practices show two ways LA is evolving and adapting to education today.

Challenges of MOOCS have to do with learners and teachers. With respect to learners, it is important to note that learning goals and objectives will vary among learners who want to complete the course and get a certificate and those who are only interested in certain parts of the course or just want to "look around". In relation to teachers, its participation is generally limited to important announcements or a 'crisis' situation. Consequently, analytics regarding teacher's participation will change compared to those gathered from LMS.

ECO sMOOCs is about putting the learner central and the learner takes an active role in his learning process. It is also about learning by connecting and interacting with others. In this regard, LA indicators and metrics should be aligned to the learners' personal goals. The metrics displayed to the learner should help him to plan his learning process and therefore these should be shown in a clear and intuitive way.

The diversity of users' intentions and backgrounds and the unconstrained synchronicity of their activities make the MOOC context very different from conventional classrooms. Consequently, ECO MOOCs provides their own definition for the following terms: progress, performance, participation or dropout and categorises learners according to their intentions and learning objectives into completing, auditing, sampling and disengaging learners and proposes a classification based on their activity.

Supporting the learner

Chapter 4 gathers the definition of academic and social metrics and indicators suitable for learners and suggests advanced services which would improve the participants' experience. In ECO, our goal is to provide learners with quantitative and qualitative indicators which allow them to reflect on their way of studying, the resources used and their results, to motivate and stimulate them for the achievement of their learning goals, to positively value their contribution and level of commitment with their peers as well as to compute a social reputation indicator in such way that participants achieve both personal satisfaction and keep engaged in more sMOOCs.

First, we propose a list of learner academic and social goals, a description about what we want to measure and a proposal of issues for their evaluation. Next, we define a set of indicators for learners which will be displayed in dashboards. Learners will see their personal indicators, and have the choice to compare it with the average of 1) the whole group, all the learners enrolled; 2) group of learners with the same ECO profile; and 3) group of learners with the same learning goals in the course.

Supporting the teacher and course designer

ECO's goal is to provide teachers the necessary quantitative and qualitative indicators in order to measure MOOC progress and, if it is convenient, take steps to improve (chapter 5). Learning analytics can also be applied to inform the course designers about the effectiveness and efficiency of their design and can point out how their design affects learners' behaviour.

This chapter proposes a subset of the learners' academic and social goals those which are relevant to the teachers, as well as aspects from course design and usability issues. As for the learners, we describe their metrics and indicators. Similarly to learners, teachers will have the possibility to view relevant indicators for the three groups of learners (1) whole group, 2) group of learners with similar ECO profile, 3) group of learners with similar learning objectives. Additionally these are visualised according to the four types of learners: completing learners, auditing learners, exploring learners and disengaging learners.

Visualisations and dashboards

Metrics and indicators need to be presented to participants, learners and teachers in a meaningful and initiutive way. That is commonly done by presenting the data in a visual manner supported on the use of various types of graphs. The presentation of these visualisations are commonly framed in dashboards, although some indicators, in particular, those which reflect the effort and global satifaction about the course could be also displayed in the context of the course. Dashboards capture and visualize traces of learning activities, in order to promote awareness, reflection, and sense-making, and to enable learners to define goals and track progress toward these goals.

We recommend for the indicators defined for learners and teachers, the type of control or visual element that could be used to show each value. The key metrics defined in the categories: academics and social issues are visualized through of different graphs or charts as gauge controls, bars, progress bars, donut charts, stacked radars and other types of graphs (chapter 7). The dashboards should provide a variety of indicators to capture knowledge levels, allow monitoring and reflection, and thus should not only show static data, but also changes over time. Importantly, the dashboards should visual social activity, social networks and communities.

Issues for implementation

Learning analytics within the framework of ECO project must consider boundary and pedagogical conditions.

The boundary conditions are related to: 1) inclusive and accessible courses for large number of participants and to wide diversity of citizens; 2) the creation of collaborative learning opportunities; and 3) interaction among peers (including some but limited interaction with academic staff).

Pedagogical conditions are: 1) the pedagogical learner centred approach, not only of associative learning but connectivist, social constructive learning and situated practices; 2) the support of adaptive learning strategies; and 3) to enable the possibilities to adapt to the changing intentions of participants during the course which implies that ECO platform must know which are the learner's goals in each moment. Another important aspect to take into account is the emphasis of ECO on the social perspective. Thus, LA should include academic indicators as well as social indicators.

There are several sources of data and several types of data that can be traced and recorded, and there are different ways to persist these tracking and logging data, but to be able to make use of these data it pays to standardise data formats. Chapter 6 gives a brief description of several metadata and paradata formats are available, in addition to specifications developed to record activities and events.

Chapter 8 gathers a list of data needed to answer useful questions for learners, teachers and academic institutions. Furthermore it suggests some surveys which are necessary to better understand the participants' behaviour and the results of each course. Some data are gathered through a questionnaire put to learners at registration and enrolment, prior to starting the course, mainly to gather insights into intentions and learning goals.

We present an overview of different scenarios for implementation illustrating the complexity ECO faces in implementing learning analytics (chapter 9). In this chapter we first describe how two of the ECO platforms, OpenMOOC and Logi Assist currently are tracking learner progress and performance (without learning analytics being implemented). Next we give some scenarios to illustrate how the major metrics for progress, performance and the group of learners at risk of dropping out as defined in this deliverable could be implemented and visualised. For one of the MOOCs of the first pilot we indicate how learners' indicators can be calculated and highlight the data that should be provided by teachers and platforms. To illustrate the complexity ECO faces when having to deliver learning analytics that meet the learner's objectives, various alternative ways in visualising progress, performance and drop-out metrics are sketched.

Ethical concerns, privacy and confidentiality

Finally, we summarise legal, privacy and ethical concerns and the actions to be taken by the ECO project. The ECO platforms and learning analytics services and dashboards make extensive use of personal data of learners and staff like teachers and tutors, including tracking their behaviour and actions in the platforms.

Consequently, these platforms and services have to be designed and operated in a legally compliant manner. ECO has to ensure that only those data are logged and recorded that are needed to calculate the indicators as described in this deliverable, and to the purpose thereof. ECO must treat any data as confidential and handle data with due diligence. The learning analytics service has to ensure that no identifiable data are made available to anybody other than the owner and anonymise data for indicators. Theoretically, learner data could be gathered from applications external to the ECO platforms. This creates additional issues, because learners might not be aware that ECO would be gathering these data.

Because the ECO platform consists of distributed environments, ECO project needs to decide whether to adhere to a national or EU framework.

1 Introduction and methodology

In this deliverable we lay out the requirements for a learning analytics service that is geared towards the unique circumstances of MOOCs. MOOCs provide learning opportunities and as such learning analytics can be applied to provide learners with additional learner support to optimise their learning experience and performance, to provide teachers with indications of when additional actions are required and inform course designers about the efficiency and effectiveness of their courses. However, in contrast to conventional educational contexts, MOOCs – being massive open online course – pose additional challenges. There are further requirements that the learning analytics services need to deal with in the context of the ECO project because ECO wants to support social and seamless MOOCs, sMOOCs. ECO argues that MOOCs need to follow a pedagogical model that is based on social, networked learning and provide authentic situated learning experiences. In this deliverable we describe how to apply learning analytics in these conditions and indicate relevant indicators for learners, teachers and course designers that take these particular conditions into account.

The main target group for this deliverable consists of the ECO project partners, in particular work package 3 and work package 4, as well as the wider public.

1.1 Methodology followed

The development of this document started in May, 2014 with a online meeting, organised by the leader of the work package, to arrive at consensus of what the deliverable should contain. We talked and discussed different approaches that should be addressed and a first structure of the document was proposed. Each partner then chose one or more sections to be completed.

Regular online meetings were scheduled that were more frequent in the beginning, weekly or biweekly, and evolved to monthly meetings. The main point in these first meetings was the redefinition of terms to be adapted to ECO MOOCs and the need of knowing the interaction and log data which ECO platforms collected.

Communication was through the forums on the Alf platform, supplemented by mail and skype for internal communication to solve direct issues.

Documents and drafts were hosted in Google Drive. All participants could read, edit and add comments on them. These were discussed and finally solved using the revision system of Google Drive. All decisions were made by consensus.

1.2 Structure of the document

We start with providing a description and definition of learning analytics and the requirements posed by the ECO sMOOCs. Next we redefine the importance concepts participation, progress and performance in light of these requirements and indicate four types of learners that can be distinguished in MOOCs according to their intended learning objectives and their learning behaviour.

Next we lay out the metrics and indicators for learners, based both on academic goals as well as social goals. The same is done for teachers and course designers who need additional indicators as compared to learners.

This is followed by some issues for implementation that need to be addressed, before moving to suggestions and recommendations on how the learning analytics indicators can best be shown to learners and teachers through the use of visual analytics and dashboards.

In the next chapter we list the data that need to be gathered for the metrics and indicators and suggest relevant xAPI statements.

We provide several scenarios for implementation, illustrating how progress and performance are currently implemented in two of the ECO MOOC platforms and indicating how our proposed learning analytics framework could be applied. Alternative scenarios are provided to indicate the challenges that are posed.

Finally, we conclude by pointing out the ethical concerns and highlight some privacy and confidentiality issues that the ECO project needs to deal with when implementing learning analytics.

2 Defining learning analytics

'Analytics' is a broad term that encompasses all kinds of uses of data, statistical and quantitative methods, and also explanatory and predictive models to allow organizations and individuals to gain insights into and act on complex issues (Oblinger, 2012). In a time when so many new terms related to technology are coined every day, it is essential to get the gist of all of them in order to distinguish one from another and understand their main aim. Analytics is sometimes compared, and sometimes confused, with the term 'big data'. According to the literature, there are three main features that define 'big data': volume, velocity and variety:

- Volume: It refers to the ability of processing massive amounts of information and is one of the main strengths of 'big data' compared with other analytic tools.
- Velocity: To have access to real-time information and the possibility of streaming fast-moving data into bulk storage for later processing allows organizations to gain competitive advantage.
- Variety: Data may come from different sources (mobile phones, computers) and in different forms (images, text, GPS signals). These unstructured data can be turned into ordered meaning.

In the 'NMC Horizon Report: 2013 Higher Education Edition', learning analytics is described as "...the field associated with deciphering trends and patterns from educational big data, or huge sets of student-related data, to further the advancement of a personalized, supportive system of higher education" (Johnson, L. et al., 2013). In the following subsection, we have gathered different definitions in order to have a complete understanding of what 'learning analytics' is.

2.1 Definitions

In 2011, learning analytics (henceforth LA) was already predicted by the NMC Horizon Report as one of the trends in education and technology for forthcoming years, and it fared well, judging by the evolution of this branch of analytics in education today. The definition given in this report was: "Learning analytics refers to the interpretation of a wide range of data produced by and gathered on behalf of learners in order to assess academic progress, predict future performance, and spot potential issues". Also, it mentioned that its goal was "...to enable teachers and schools to tailor educational opportunities to each learner's level of need and ability".

The first international Conference on Learning Analytics and Knowledge *1st Learning Analytics conference in Banff, Canada, 2011;*(LAK 2011:https://tekri.athabascau.ca/analytics/) defined LA as "The measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs". The main goal of LA is enhancing learning effectiveness and providing with insights regarding the educational process through online platforms and interactions between learners and teachers. Norris et al. (2008) highlighted the importance of using educational data to act in a forward-thinking manner in these contexts.

Since these first mentions in 2011, LA has become a popular topic, especially in Higher Education, and more recently, as applied to the latest trend in education, Massive Open Online Courses (MOOC).

After three years, definitions of LA have changed and adapted to recent developments in education. As described in 'NMC Horizon Report: 2014 Higher Education Edition', "Learning Analytics research uses data analysis to inform decisions made on every tier of the education system, leveraging student data to deliver personalized learning, enable adaptive pedagogies and practices, and identify learning issues in time for them to be solved.". Personalised learning and adaptive pedagogies and practices show two ways LA is evolving and adapting to education today.

However, if we take the essence of LA into account, compiling and analysing data and making reports out of the results obtained, we notice this is something that already existed decades ago in traditional learning environments. What's new about LA? First, as it was mentioned before when we explained the difference between analytics and big data, nowadays the volume of data is higher than used to be, especially in massive courses, and it goes beyond human ability of collecting and analysing. But also, LA is able to predict learning issues and identify students or learners at risk and allows intervention while a course is in progress.

2.2 Uses and applications

LA can be used at various levels according to the relevance and usefulness to different members of a learning environment. Thus, there is analytics which may prove useful to students and learners, such as progress and success, whereas instructors might be interested in students' participation and engagement in a course. On the other hand, analytics such as students' or learners' retention or instructor's involvement would be considered more relevant to institutions in order to decide whether 'programs designed to improve retention have been effective and should be sustained (Johnson, et al., 2013).

LA work has been mostly applied to LMS (Learning Management Systems) or VLE (Virtual Learning Environments). Research on this area shows how LA can be used for predicting learners' success. Morris et al. (2005) carried out a study aimed at examining students' engagement in online courses through their behaviour online, defined as frequency of participation and duration of participation. Results revealed significant differences between who they called 'withdrawers and completers" and between "successful completers and non-successful completers' showing thus that variables such as frequency of participation and time spent on tasks are important for successful online learning. Similarly, Macfayden and Dawson (2010) stated that "pedagogically meaningful information can be extracted from LMS-generated student tracking data" and confirmed that analytics tools applied to LMS platforms can identify at-risk students and allow for pedagogical intervention.

However, as ICTs in education continue their evolution, LA research has yielded significant advancements too. For instance, in 2012, CourseSmart, a digital textbook provider, launched CourseSmart Analytics (http://www.coursesmart.com/go/institutions/analytics), an analytics tool that tracked students' activity and interaction with text and provided teachers with analysed data, which could be used to assess students' efforts. In 2013, efforts were put on the development of visually explicit streams of information in real time, about a given group of students. Examples of these digital dashboards can be seen at the United States Department of Education (http://dashboard.ed.gov/dashboard.aspx) or the Ofsted (Office for Standards in Education, Children's Services and Skills) in the UK (http://dashboard.ofsted.gov.uk/).

Today, LA research expands with European projects like LACE (Learning Analytics Community Exchange) (http://www.laceproject.eu/lace/), which is building a community of practice across Europe and beyond to help lay the foundations for the field for the time coming, or the PAR (Predictive Analytics Reporting) Framework, an American provider of LA that brings together institutions to collaborate on identifying points of students loss and improving student retention in US higher education.

2.3 Learning analytics for MOOCs

When the teaching and learning process happens in a centralised environment, such as an LMS, data can be easily extracted and analysed, but what if education happens in a decentralised environment with distributed teaching and learning networks? This is one of the challenges MOOCs (Massive Open Online Courses) present to both educators and institutions today.

There are other challenges rooted in the 'openness' of this kind of online courses. Requirements, such as any previous experience on a subject, are not normally needed to join a MOOC although some general considerations are given to potential participants. This leads to a great variety of learners with different backgrounds all enrolled in the same course, making it difficult to provide accurate and valid results. Similarly, learning goals and objectives will vary among learners who want to complete the course and get a certificate and those who are only interested in certain parts of the course or just want to "look around".

The role of teachers in MOOCs has certainly changed from the role they had in traditional LMS. The massive amount of learners enrolled in MOOCs make it impossible for teachers to involve themselves in the course. The figure of the teacher is indeed there, but its participation is generally limited to important announcements or a 'crisis' situation. Consequently, analytics regarding teacher participation will change compared to those gathered from LMS.

Social learning is one of the key elements in MOOCs and although it is primarily carried out through discussion forums, different learning networks are created outside and around the MOOCs that help learners connect and build and expand their own learning environments. These learning networks are frequently based on learner-learner, and sometimes instructor-learner, interaction through social networks such as Facebook, Google+ or Twitter.

A different subset of LA would be needed to measure this kind of networked interaction, and this is where Social Learning Analytics (henceforth SLA) comes into play.

SLA is "...a distinctive subset of learning analytics that draws on the substantial body of work demonstrating that new skills and ideas are not solely individual achievements, but are developed, carried forward, and passed on through interaction and collaboration." (Buckingham & Ferguson, 2012). This kind of LA focuses on processes in which learners are engaged in social activity, such as interaction with others in social networks ('likes' in Facebook, RT or 'retweets' in Twitter) or in educational platforms that allow others to follow your activity through interactive tools, such as blogs or rating. Instances of SLA applied to MOOCs taking place now are, among others, FutureLearn (https://www.futurelearn.com/), a MOOC platform partly developed by the Open University of the UK, or the Network Assessment Tool (NAT), a tool integrated in a MOOC to visualise real-time discussion activities (Schreurs et al., 2014).

Inside a MOOC platform, massive amounts of data can be stored and analysed by LA, going from instances of videos watched, activities undertaken, discussion posts answered, to much more nuanced information which can lead to uncover learning patterns that in turn can be used to predict, for instance, 'most feared' learners drop out. Concerning personalisation and adaptive learning, LA is expected to provide real-time interventions for learners who need support or a trigger to foster engagement in the course. An example of LA research on MOOCs is the Lytics Lab at Stanford University. This research group "...engage(s) in use-driven research and data-driven design to develop our collective knowledge around improving online learning".

2.4 Challenges for learning analytics for ECO

ECO sMOOCs are seen as part of open education. The "s" in sMOOCs stands for "social", since they provide a learning experience marked by social interactions and participation, and "seamless", since ideally they should be accessible from different platforms and integrate with participants real life experiences. The MOOCs offered by partners are intended to remove all unnecessary barriers to learning and provide learners with a reasonable chance of success in education. This implies 'openness' in the sense not only of no financial cost, but also open accessibility, open licensing policy, freedom of place, pace and time of study, open entry, and open pedagogy.

Following ECO sMOOCs pedagogical conditions, introduced in D2.2 Instructional Design and Scenarios for MOOCs, particularly:

- Support adaptive learning strategies, e.g. using learning analytics.
- ECO sMOOCs enable the possibilities to adapt to the changing intentions of participants during the course.

Learning analytics in ECO sMOOCs thus should collect data which could be used to make predictions about the learner's performance and provide them with assistance and appropriate feedback on a given time during the course. Ubiquity and mobile context

The challenges ECO sMOOCs face, being both social and 'seamless', are having to deal with ubiquity and mobile context, social networks and the variety of learner goals, as explained below.

Ubiquity and mobile context

MOOCs are designed to be displayed on any device, desktop, tablet or even mobile phone, in order to be followed at any time. It is the learner who decides when and what to study at any place/time. Plenty of different activities are being integrated within ECO platforms so that these can be selected in relation to the learning environment given in the study periods of each learner. The sort of activities implemented should be elaborated taking into account the level and nature of contents dealt in any MOOC.

Connected (social) networks

Since MOOCs are conceived to be implemented in open learning environments, most contents might be ready be spread and shared in different social networks by means of "likes", "quotes", CoPs, etc. so that knowledge goes on flowing independently of the platform used to deliver the MOOC. One of the main problems at this point would be until what extent analysing everything happened in external networks would be permitted or not.

One of the most important characteristics common to all ECO MOOCs is its direct connection with social networks. All the modules contained in a MOOC incorporate sharing functionalities to spread the word outside the MOOC.

ECO sMOOCs being social and seamless is about putting the learner central and the learner taking an active role in his learning process. It is also about learning by connecting and interaction with others. The learner might rely on his existing personal network of people, but in the context of the MOOC also interacts with fellow learners. In doing so, the learner builds up a social network. A social network consists of the relations between people. Relations can be directional, and could be categorised into type of relation depending the reason for the relation. The structure of this network can provide important information, both to the learner and to the teacher and course designer. Typically, social network analysis evaluates networks to determine whether there are clusters in a network, whether there are maybe brokers or hubs, i.e. people through whom a lot of information passes, how many connections somebody has, how quickly information passes through the network, etc. The relevance of these measures depends on the objective. Still, it provide an easy visualisation of active learners.

Social network features therefore are important for learners partaking of ECO sMOOCs. Whether those features have to be integral to the platform or whether learners should be directed to external social media and social networks is being discussed in D2.6 Web 2.0 requirements. From the learning analytics perspective it is important to mention here that although there might be solutions to any technical issues in retrieving the data and messages from these external social networks, there are many legal issues and issues around privacy and confidentiality of the data.

Learner goals

Learners enrol in MOOCs with a variety of learning goals. Some even only want to browse around. That complicates very much the way in which proper and suitable learning support can be offered. Only a minor part of participants that enrol in a MOOC go for certification and complete a MOOC fully. That means that most participants have different learning goals than designed by the course designer. Any meaningful support in assisting the learners in completing the MOOC has to follow the intended learning goals of the learner. Consequently LA indicators and metrics need to be able to determine that goal and should be aligned to their personal goal. Some LA indicators that provide information at individual level are shown to the individual learner (Fisher, 2014), but should also be compared to the whole population, learners similar to their profile and background, and learners with the same learning goals. While the LA for teachers and course designers should be based on aggregated data, but again representing the three different groups of learners. LA for teachers also should represent the group of learners who are at risk of dropping out to allow the teacher to consider taking corrective actions for the whole group or consider redesigning part of the MOOC activities or content.

2.5 Learning analytics embedded in the pedagogical model

The metrics in learning analytics have to be connected to the framework of purpose and expectations for the learning activity. It is important to integrate the analytics metrics with the goals of the learning activity. Learners need to understand the purpose of the learning activities, the characteristics of what the instructor expected in relation to their participation in the proposed activities, and how learning analytics provided serve as a representation of this. Learners must be encouraged to set personal goals for their participation and to use the analytics to help monitor these. One of the key attractions of learning analytics is the possibility to support the learner in actively reflecting on and taking action to manage their learning process (Govaerts, et al 2010).

Buckingham Shum (2012) differentiates micro, meso and macro level of learning analytics. Micro-level analytics supports the tracking and interpretation of process-level data for individual learners. Micro-level analytics seems to be most applicable for these metrics, because according to Powell and MacNeill (2012) these assist in:

- Allow individual learners to reflect on their achievements and patterns of behaviour in relation to their peers
- Identify learners who may require extra support and attention
- Help teachers and support staff to plan supporting interventions with individuals and groups
- Enable course teams, to improve current courses or develop new curriculum offerings
- Prove information to institutional administrators on matters such as marketing and recruitment or efficiency and effectiveness measures.

Teachers and learners can benefit differently from learning analytics. Teachers can understand and observe how learners are facing their learning process and the actions performed on the platform supporting the course, and then adapt the course when required. Furthermore, learners' motivation might improve if they are able to compare their individual progress with their peers. In general, learning analytics aim to help us to better understand how we learn and improve our learning (Duval, 2011).

The metrics presented to the learner must serve to plan their learning process. It is also important to show them in a clear and intuitive way. In this regard a key aspect is the design of dashboards to display the Key Performance Indicators (KPIs). The metrics shouldn't be presented simply as a set of numbers; they could be described using charts which have clear meaning in the context of the learning activity. In this design aspects of accessibility and usability should be considered, taking as reference existing standards (WAI, W3C, …)

One important aspect is that the analytics can influence the way in which the learner faces the learning and thus we "become what we measure", even though the metrics only capture some aspects of the overall activity. To minimize this impact, it is necessary to include multiple and diverse measures. However, provide multiple metrics needs to be balanced in order to avoid overload or overwhelm learners unproductively.

2.6 Learning analytics integration into ECO platform

Learning analytics can be deployed in any software application in two folds: directly, observing metrics and indicators that allow users to gain insights about main aspects to achieve their goals; indirectly, by means of software artefacts which collect certain intelligence (e.g. rules) and act according to it.

The first case is used to being implemented by means of dashboards. According to (McFadden, 2010), a dashboard is "An easy to read, often single page, real-time user interface, showing a graphical presentation of the current status (snapshot) and historical trends of an organization's *key performance indicators* (KPIs) to enable instantaneous and informed decisions to be made at a glance." In ECO context, the focus will be addressed towards learners' and instructors' goals.

The term dashboard originates from the automobile dashboard where drivers monitor the major functions at a glance via the instrument cluster. Dashboards thus are unique. The design of each dashboard is driven by the business and its needs and culture. Dashboards are usually a series of graphics, charts, gauges and other visual indicators that can be monitored and interpreted. They can be customised to link these graphs and charts to goals.

On the other hand, it is possible to develop software artefacts that, taking advantages of the data logged in MOOC platforms, provide recommendations to participants such as "try this quiz, 80% participants who did it, achieved the certification", "75% participants who discussed this topic got a good mark in this goal" or "people with a profile similar to yours is now connected, try to get in touch with them"

2.7 Ethical concerns

Data privacy and the use of data are also strong concerns of the use of learning analytics. There are legal and ethical issues in each country and in EU that need to be addressed before ECO can make use of some learner data. Some issues that have to be dealt with are presented in chapter 9.

Regarding activity performed in ECO platform by learners and measured quantitatively, such as time spent carrying a quiz, number of messages sent, frequency of access must be carefully calculated and, perhaps only shown to certain learner profile, e.g. a participant who wants the certification. Furthermore, it must be taken into account that most of the activity could be performed out of the platform and thus, the effort and time involved in each task is impossible to know.

Moreover the ECO project needs to develop a policy on how to deal with data that are external to the ECO platforms. Ubiquitous learning and networked learning are very much promoted in ECO. The ECO pedagogical framework is based on networked learning, connectivism and interaction with others and thus heavily relies on the opportunities offered by social media. Mobile and ubiquitous learning also entails activities outside the platform. On the one hand those learning actions outside the platform can provide valuable data for learning support. At the same time there are many issues concerning privacy and confidentiality that have to be dealt with first.

2.8 Related work

One of the aims of learning analytics is to detect patterns. There are many different techniques and methodologies that are could be applied, ranging from simple statistics to advanced techniques to deal with large datasets. In this section we briefly describe some of the common approaches. ECO will use a selection of these to calculate the metrics that are displayed in the ECO dashboards and other services that are useful for all participants. Given the importance of networked learning and learning through interaction with others, social network analysis will play an important role in ECO learning analytics.

- Visual analytics.
- Descriptive statistics.
- Data mining techniques.
- Social network analytics.
- Text analytics.
- Others.

2.8.1 Visual analytics

According to Thomas et al., (2005) visual analytics is the science of analytical reasoning supported by interactive visual interfaces. Davenport (2013) indicates that, we are living in the period of Visual Analytics 3.0, this new period in the use of analytics is characterized by combining all types of data - large and small, structured and unstructured, internal and external - going beyond just descriptive purposes for predictive and prescriptive purposes. In this context, visual analytics has taken on an increasingly important role. A specific definition of visual analytics would be: "Visual analytics combines automated analysis techniques with interactive visualisations for an effective understanding, reasoning and decision making on the basis of very large and complex datasets" (Keim et al., 2010). The visual analytics process combines automatic and visual analysis methods, characterised through interaction between data in order to discover knowledge from data.

Successful different visual analytic applications in MOOCs can be found in specialized literature. In Xu et al. (2014) an experience developed in the University of Maryland with a MOOC in Coursera is described. They define several visual representations of course success, as measured by the student’s final grade against various types of activity in the course (participation in quizzes, forum or access to lecture pages). The results showed that there are multiple ways to be successful in a course, and it is perhaps not necessary to “do everything” – at least, not for everyone. They also analyzed how students’ self-reported “profiles” interacted with performance. Using traditional statistical analyses they identified three clusters of students (“auditing”, “behind” and “on track” learner’s profiles).

Coffrin et al. (2014) present different visualization strategies to visualize patterns of student engagement and performance in MOOCs. In this work, the authors use alternative learning analytic approaches and visual representations of the output of these analyses. They adopted an approach that involved an iterative analysis of data. In each step of the iterative process patterns are observed in the data and those patterns are used to refine the analysis focus. Using this approach, they were able to meaningfully classify student types and visualize patterns of student engagement which were previously unclear. They used histograms to show cumulative distributions of student performance, and bar charts showing weekly participation divided into three mutually exclusive subgroups: auditors (students who watch videos but do not participate in any assessments), actives (students who participate in assessments) and qualified (met the assessment qualification). This information is mainly useful for the teacher, but learners can also receive feedback on the overall evolution of their group.

State transition diagrams have been also used to analyze logs of students' interactions with an online drag-and-drop learning activity (Judd et al., 2004), or to show student movement between categories of engagement over assignment periods (Kizilcec et al., 2013).

Other examples of the kinds of visualizations that have been generated using MOOC data include:

- Social network diagrams (Kop et al., 2011; Gómez-Aguilar, et al., 2014). The social analysis tool SNAPP can be used to produce online network visualizations of discussions, such as interactions between participants in the course forums.
- Demographic data (age distributions, places of residence, and so on) and professional background of participants (Kop et al., 2011). In this case, Google Map can be a useful tool for visualization.
- Q-Q plots of communication activities in forums (Brinton et al., 2013).
- Log plots of student access to learning activities (Clow, 2013).
- Word clouds presenting summaries of course contents or discussions in forums (Kop et al., 2011). Word clouds can be generated using tools as Wordle.

2.8.2 Descriptive statistics

Descriptive statistics is the discipline of quantitatively describing the main features of a data set. Measures of central tendency and measures of variability or dispersion are frequently used in learning analytics. Measures of central tendency include the mean, median and mode, while measures of variability include the standard deviation (or variance), the minimum and maximum values of the variables, kurtosis and skewness. Another interesting statistic technique is the correlation analysis with the aim of finding relationships between numerical features. These techniques will be used by LA engine directly or indirectly to visualise distributions, to calculate prototypes, to select instances, to delete correlated variables before a mining process is performed and so forth.

2.8.3 Data mining

Data mining is the process of applying data analysis and discovery algorithms to find knowledge patterns over a collection of data (Fayyad et al. 1996). It makes use of a great set of methods and techniques from artificial intelligence, machine learning, statistics, and database systems.

In general, data mining tasks are classified into two big categories: descriptive and predictive (Han, 1996). Descriptive mining tasks characterise the general properties of the data, whereas predictive ones perform inference on current data in order to make predictions. The difference between them is that the former do not need to have a data set correctly classified whereas the latter do it. For instance, if I want to predict if a learner will achieve his certification, I need have a data set with instances which reflect the behaviour of other students in the same course and their grade. In return, descriptive techniques directly work with the data set provided. For example, to discover which participant profiles your course has, only requires selecting the features to use and the number of clusters to build (Zorrilla et al, 2011; Zorrilla et al 2013)

Inside each category, there are a huge number of algorithms developed following different approaches. Most frequently used in educational data mining are association rules, clustering, classification and regression (Ayala, 2014).

Association rules (descriptive method) allow you to discover interesting and frequent relations in large amounts of data. Generally, the patterns are represented in rule-format. For example, learners who perform all self-tests, pass the course. This technique has been used for building recommenders, analysing the design of the course and so on (Romero et al., 2010, Zorrilla et al., 2011).

For instance, an online resource recommender could be built based on the resources accessed by previous learners. Dataset to be mined will be comprised by the set of sessions performed by learners. Each session will have the list of pages and resources used. Then, an association rule algorithm like apriori or yacaree (Balcazar et al, 2011) could be used to discover the most frequent paths. The system will detect the session and pages visited by the user and then it will recommend the resource which most learners have visited as the following URL to visit.

Clustering (descriptive method): The goal of clustering is to find groups that are very different from each other, and whose members are very similar among them. This type of techniques can be used for characterising learners according to the use of resources (forum, self-tests, contents, etc) they use or activities they engage in, demographics features or performance, for instance.

Considering the ECO profile features, a first participant profile could be built from the following data: gender, age discretised in several bins, country, and areas of interest establishing each area as a binary data. Then, a *kmeans* algorithm could be used. The number of clusters should first be calculated for example using NbClust Package available in R package. This cluster could later be updated adding other features collected by means of entry questionnaire. As a consequence of the fact that there are different algorithms to be used and it is not easy to valid the best cluster, R software offers different package which can be applied (available in http://CRAN.R-project.org/)

Another clustering which summarises learners' activity in the course could be built. Data such as number of quizzes performed, passed and failed out of the total, number of resources watched, number of messages written and read from forum and number of accesses per week could be used for this task.

The objective of prediction (Romero et al., 2010) is to estimate the unknown value of a variable that describes the student. In education the values normally predicted are performance, knowledge, score or mark. This value can be numerical/continuous value (regression task) or categorical/discrete value (classification task). Regression analysis finds the relationship between a dependent variable and one or more independent variables. Classification is a procedure in which individual items are placed into groups based on quantitative information regarding one or more characteristics inherent in the items and based on a training set of previously labelled items.

For example, a classification model could be used to identify if the risk of dropout of a student is low, medium, or high for a specific course. This could be built from historical data from the course already finished and the model achieved could be explained to learners and teachers in a rule-format. Features which could be used are: gender, age, background (degree), number of MOOC course finished, number of quizzes performed by week, number of messages written and read in forums by week, number of threads posted by week, social valuation, number of accesses by week, number of resources watched by week and any other feature which can be interesting in the course such as number of videoconferences in which learner participated or metrics such as degree, betweenness, etc. extracted from a social graph. It is very important to highlight that target attribute must be defined, that means, that all instances in the dataset must be the target attributed filled as low, medium or high in accordance with instructors' opinion or deduced from activity data log. The best algorithms to be used are those categorised as based-tree, such as J48, Ramdom forest and based-rules such as NNge or JRIP since they are understandable and perform properly in educational datasets.

Another interesting pattern is the performance model. The attributes could be the same as those previously mentioned changing the target attribute with pass, fail and dropout values. An example of rules which users could read would be the following

if 'Number of quizzes passed '> 7

then student passes

if 'Number of quizzes passed ' < 5

and 'Number of assignments performed ' > 1

and ' student reads forums messages '

then the student passes

if the student neither do quizzes (n_quiz <= 5)

nor carries out assignments (n_assignment <= 1)

then student fails

A very comprehensive review of EDM research can be found in (Romero et al., 2010) and (Peña-Ayala, A., 2014).

2.8.4 Social network analysis

Social Networks Analysis (SNA), or structural analysis, aims at studying relationships between individuals, instead of individual attributes or properties. A social network is considered to be a group of people, an organization or social individuals who are connected by social relationships like friendship, cooperative relations, or informative exchange (Freeman, 2006). Social network analysis brings graph theory from the field of mathematics together with work on interpersonal and communal relationships from the fields of sociology and communication.

Although social networks have been studied for decades (Fortunato, 2010), the emergence of social networking services like Facebook or Twitter has been the cause of the unprecedented popularity that this field of study has now.

ECO project can take several advantages of using SNA. For instances, centrality measures such as degree, indegree, outdegree, and betweennes can be used as attributes to characterise learners and be used to build clusters or classification models as those carried out by Bayer et al. (2012) for the prediction of dropouts and Palazuelos et al. (2013) for prediction of performance. Centrality measures can be achieved by means of *ORA (http://www.casos.cs.cmu.edu/projects/ora/).

Another useful service which could be built is a social graph from the network of forum messages, with tools as SNAPP (http://www.snappvis.org/), with which teachers can assess students' participation in asynchronous discussion forums of online courses. It could also be considered to integrate tools like Meerkat-ED proposed by Rabbany et al. (2014). This same social graph could be used to detect and encourage students at risk (Dawson, 2010) if graphs for different periods are shown and participants whose activity is each time lower are highlighted.

2.8.5 Text analytics

It refers to the process of deriving high-quality information from text and the application of text mining techniques to solve a wide variety of problems.

Text mining is the process of discovering and extracting knowledge from unstructured data; refers to the process of extracting interesting and non-trivial patterns or knowledge from text documents and relates to the analysis of data contained in natural language text. It works by transposing words and phrases (unstructured data) into numerical values which can then be linked with structured data in a database and analysed with traditional data mining techniques. In short, text mining can help make the implicit information of documents more explicit, and comprises three main activities:

- Information retrieval (IR) to gather relevant texts.
- Information extraction (IE) to identify and extract entities, facts and relationships between them.
- Data mining to find associations among the pieces of information extracted from many different texts.

Typical text mining tasks include text categorization, text clustering, concept/entity extraction, production of granular taxonomies, sentiment analysis, document summarization, and entity relation modelling (i.e., learning relations between named entities).

Text analysis involves information retrieval, lexical analysis to study word frequency distributions, pattern recognition, tagging/annotation, information extraction, data mining techniques including link and association analysis, visualization, and predictive analytics. The overarching goal is, essentially, to turn text into data for analysis, via application of natural language processing (NLP) and analytical methods. NLP is being used in a varied of domains and is applied for example in automatic grading of essays and detection of plagiarism (Landauer et al, 1998; Landauer, et al. 2000; Van Bruggen, 2002). It can be used to map learner progress and performance onto a golden rule set by the course designer or determined by that of the group and provide feedback (Landauer et al, 1998; Van Bruggen, et al., 2004; Berlanga et al, 2009; Berlanga et al., 2011; Smithies, et al., 2010).

For example, text mining may be applied to learners' interaction analysis to extract useful knowledge, discover interesting patterns and support decision about learners' activity. Text mining could be useful to extract data from participation in forums or social networks, build a document-term matrix to determine correlation between concepts, find frequent words and associations, create word clouds to visualize important words and text clustering in order to measure learners' participation and performance and its relevance for learning (allowing classifying them).

2.8.6 Tools for learning analytics

Next, we relate web sites which summarise software tools available and, in our opinion are suitable, for developing learning analytics services proposed in this document. In some sections we also mention the name of a toolbox or framework in particular as a result of the knowledge and experience that some members of the group have due to their use in their research field. Furthermore, most of them are referenced in the literature gathered in this document.

Knowledge discovery software

List of software: http://www.kdnuggets.com/software/index.html

Weka: http://www.cs.waikato.ac.nz/ml/weka/

Knime: https://www.knime.org/

Rapidminer: https://rapidminer.com/

R: http://www.r-project.org/

Software for social network analysis:

List of software: http://www.gmw.rug.nl/~huisman/sna/software.html

SNAPP http://www.snappvis.org/

*ORA http://www.casos.cs.cmu.edu/projects/ora/

API to access and measure activity in social networks

GRAPH API https://developers.facebook.com/docs/graph-api?locale=es_LA

TWITTER API https://dev.twitter.com/overview/general

Text analytics software

List of software: http://www.predictiveanalyticstoday.com/top-30-software-for-text-analysis-text-mining-text-analytics/

Apache Lucene: http://lucene.apache.org/

OpinionFinder: http://mpqa.cs.pitt.edu/opinionfinder/

Analytical functions provided by DBMS

http://docs.oracle.com/cd/E11882_01/server.112/e26088/functions004.htm#SQLRF06174

http://technet.microsoft.com/en-us/library/hh213234(v=sql.110).aspx

Visual analytics tools:

Rgraph library: http://www.rgraph.net

Google charts API:https://developers.google.com/chart/.

wxWidgets, open source framework for developing native cross-platform GUI applications (http://docs.wxwidgets.org/trunk/index.html)

2.9 References

Balcázar, J.L.: Parameter-free Association Rule Mining with Yacaree. EGC 2011: 251-2

Bayer, J., Bydzovská, Jan Géryk & Tomáš Obšıvac (2012). Predicting Dropout from Social Behaviour of Students. In: Proceedings of the 5th International Conference on Educational Data Mining. pp. 103-109.

Berlanga, A. J., Brouns, F., Van Rosmalen, P., Rajagopal, K., Kalz, M., & Stoyanov, S. (2009, July, 6-7). Making Use of Language Technologies to Provide Formative Feedback. Paper presented at the AIED 2009 Workshops Proceedings Volume 10. Natural Language Processing in Support of Learning: Metrics, Feedback and Connectivity. Proceedings of the 14th International Conference in Artificial Intelligence in Education, Brighton, United Kingdom.

Berlanga, A. J., Kalz, M., Stoyanov, S., Van Rosmalen, P., Smithies, A., & Braidman, I. (2011). Language technologies to support formative feedback. Educational Technology & Society, 14(4), 11–20.

Brewe, E., Kramer, L. & Sawtelle, V. (2012). Investigating Student Communities with Network Analysis of Interactions in a Physics Learning Center. Physical Review Special Topics. Physics Education Research 8(1), 010101.

Brinton, C. G., Chiang, M., Jain, S., Lam, H., Liu, Z. & Wong, F. M. F. (2013). Learning about social learning in moocs: From statistical analysis to generative model. CoRR, abs/1312.2159.

Buckingham Shum, S., & Ferguson, R. (2012). Social Learning Analytics. Ed. Tech. & Society, 15, 3, 3–26.

Charlton, P., Mavrikis, M., & Katsifli, D. (2013) "The Potential of Learning Analytics and Big Data", Ariadne Issue 71 http://www.ariadne.ac.uk/issue71/charlton-et-al

Clow, D. (2013). Moocs and the funnel of participation. In Proceedings of the Third International Conference on Learning Analytics and Knowledge, LAK '13, pp. 185-189, New York, NY, USA. ACM.

Coffrin, C., Corrin, L., de Barba & P., Kennedy, G. (2014). Visualizing patterns of student engagement and performance in MOOCs. In Proceedings of the Fourth International Conference on Learning Analytics And Knowledge (LAK '14). ACM, New York, USA, pp. 83-92. DOI=10.1145/2567574.2567586.http://doi.acm.org/10.1145/2567574.2567586

Crespo, P. & Antunes, C. (2012). Social Networks Analysis for Quantifying Students' Performance in Teamwork. In: Proceedings of the 5th International Conference on Educational Data Mining. pp. 234-235.

Cuéllar, M. P., Delgado, M. & Pegalajar, M. C. (2011). Improving learning management through semantic web and social networks in e-learning environments. Expert Syst. Appl. 38(4), pp. 4181–4189.

Davenport, T. (2013). State of the Art Practice with Visual Analytics. Harvard Business Review – Key Learning Summary.http://www.sas.com/en_us/offers/13q4/state-of-art-practice-2287192/overview.html

Dawson, S. (2010). Seeing the learning community: an exploration of the development of a resource for monitoring online student networking. Br. J. Educ. Technol. 41(5), pp. 736–752.

Dawson, S., Tan, J.P.L. & McWilliam, E. (2011) Measuring creative potential: using social network analysis to monitor a learners' creative capacity. Australas. J. Educ. Technol. 27(6), pp. 924–942.

Fayyad, U. M., Piatetsky-Shapiro, G. & Smyth, P. (1996). Knowledge discovery and data mining: Towards a unifying framework, in: KDD, pp. 82-88.

Fischer, G. (2014) Beyond hype and underestimation: identifying research challenges for the future of MOOCs, Distance Education, 35:2, 149-158, DOI: 10.1080/01587919.2014.920752

FOAF: Friend of a Friend http://www.foaf-project.org/

Fortunato, S. (2010). Community detection in graphs, Physics Reports, 486(3–5), pp. 75-174, ISSN 0370-1573.

Freeman, L. (2006). The Development of Social Network Analysis. Empirical Press.

Gómez-Aguilar, D. A., García-Peñalvo, F. J. & Therón, R. (2014). Analítica visual en e-learning. El profesional de la información, 2014, mayo-junio, 23(3). ISSN: 1386-6710.

Han, J. & Kamber, M. (2000). Data Mining: Concepts and Techniques, Morgan Kaufmann.

Judd, T. S. & Kennedy, G. E. (2004). More sense from audit trails: Exploratory sequential data analysis. In Beyond the comfort zone: Proceedings of the 21st ASCILITE Conference, pp. 476-484.

Johnson, L., Smith, R., Willis, H., Levine, A., and Haywood, K., (2011). The 2011 Horizon Report. Austin, Texas: The New Media Consortium.

Johnson, L., Adams Becker, S., Cummins, M., Estrada, V., Freeman, A., and Ludgate, H. (2013). NMC Horizon Report: 2013 Higher Education Edition. Austin, Texas: The New Media Consortium.

Johnson, L., Adams Becker, S., Estrada, V., Freeman, A. (2014). NMC Horizon Report: 2014 Higher Education Edition. Austin, Texas: The New Media Consortium.

Keim, D., Kohlhammer, J., Ellis, G. & Mansmann, F. eds. (2010). Mastering the information age solving problems with visual analytics. Ed. Eurographics Association, Germany.

Kizilcec, R. F., Piech, C. & Schneider, E. (2013). Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In Proceedings of the Third International Conference on Learning Analytics and Knowledge, LAK '13, pp. 170-179, New York, NY, USA, 2013. ACM.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to Latent Semantic Analysis. Discourse Processes, 25, 259-284.

Landauer, T., Laham, D., & Foltz, P. W. (2000). The intelligent essay assessor. IEEE Intelligent Systems, 15, 27-31.

Macfadyen, L., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. Computers & Education, 54(2), 588-599.

McFadden, P. (2012) *"What is Dashboard Reporting"*. Retrieved: 2012-05-10.

Morris, L.V., Finnegan, C., & Wu, S. (2005). Tracking student behavior, persistence, and achievement in online courses. Internet and Higher Education, 8, 221-231.

Norris, D., Baer, L., Leonard, J., Pugliese, L. and Lefrere, P., Action Analytics: Measuring and Improving Performance That Matters in Higher Education. EDUCAUSE Review, 43, 1, (2008), 42-67.http://www.educause.edu/ero/article/action-analytics-measuring-and-improving-performance-matters-higher-education

Oblinger, D.G. (2012). Let's talk analytics. EDUCAUSE Review, July/August, 10-13.

Palazuelos, C., García-Saiz, D. & Zorrilla, M. E. (2013). Social Network Analysis and Data Mining: An Application to the E-Learning Context. International Conference on Computational Collective Intelligence Technologies and Applications, pp. 651-660.

Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works, Expert Systems with Applications, 41(4), Part 1, pp. 1432-1462, ISSN 0957-4174.http://dx.doi.org/10.1016/j.eswa.2013.08.042

Rabbany R., Samira ElAtia, Mansoureh Takaffoli & Osmar R. Zaiane (2014). Collaborative Learning of Students in Online Discussion Forums: A Social Network Analysis Perspective, Chapter in: Educational Data

Mining: Applications and Trends, edited by Alejandro Peña-Ayala, published by Springer, Series: Studies in Computational Intelligence, 524. ISBN 978-3-319-02737-1.

REV: RDF Review Vocabulary. http://vocab.org/review/

Romero, C. & Ventura, S. (2010). Educational data mining: a review of the state of the art. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 40(6), pp. 601-618.

Schreurs, B., Van den Beemt, A., Prinsen, F., De Laat, M., Witthaus, G., Conole, G. (2014).Investigating the social configuration of a community to understand how networked learning activities take place: The OERu case study. Paper presented at the 2014 Networked Learning Conference, Edinburgh, UK.

SIOC. Semantically-Interlinked Online Communities. http://rdfs.org/sioc/spec/

Smithies, A., Braidman, I., Berlanga, A. J., Haley, D., & Wild, F. (2010). Using Language Technologies to support individual formative feedback. In P. Escudeiro (Ed.), Proceedings of the 9th European Conference on e-Learning (ECTEL 2010) (pp. 570-577). Oporto, Portugal. November, 4-5, 2010.

xAPI: Experience API http://www.adlnet.gov/tla/experience-api/

Xu, Z.; Goldwasser, D.; Bederson, B.B.; Lin, J. (2014). Visual Analytics of MOOCs at Maryland. L@S 2014, Mar 04-05 2014, Atlanta, GA, USA.http://dx.doi.org/10.1145/2556325.2567878

van Bruggen. (2002). Computerondersteund beoordelen van essays. OTEC Rapport 2002-1. Heerlen, The Netherlands. Open Universiteit, OTEC.

van Bruggen, J., Sloep, P., van Rosmalen, P., Brouns, F., Vogten, H., Koper, R., & Tattersall, C. (2004). Latent semantic analysis as a tool for learner positioning in learning networks for lifelong learning. British Journal of Educational Technology, 35(6), 729-738.

Zorrilla, M., García-Saiz, D., & Balcázar, J. L. (2011). Towards parameter-free data mining: mining educational data with yacaree. In Proceedings of the 4th international conference on educational data mining, pp. 363–364.

Zorrilla, M. García-Saiz, D. A Data Mining Service to Assist Instructors Involved in Virtual Education pp. 222-243. Business Intelligence Applications and the Web: Models, Systems and Technologies Editors: Marta E. Zorrilla (University of Cantabria, Spain), Jose-Norberto Mazón (University of Alicante, Spain), Óscar Ferrández (University of Alicante, Spain), Irene Garrigós (University of Alicante, Spain), Florian Daniel (University of Trento, Italy) and Juan Trujillo (University of Alicante, Spain) . Sep **2011**. IGI Global Publisher

Zorrilla, Marta; Garcia-Saiz, Diego- A service oriented architecture to provide data mining services for non-expert data miners. *Decision Support Systems* 55(1), pp: 399-411. **2013** DOI: 10.1016/j.dss.2012.05.045.

Zorrilla, Marta E.,Diego García-Saiz, José L. Balcázar: Towards Parameter-free Data Mining: Mining Educational Data with Yacaree. EDM 2011: 363-364

3 Redefining concepts frequently used in traditional e-learning settings

3.1 Redefining terms

Since learning analytics rose, the community has defined the main concepts involved in the monitoring and tracking of learners in a very similar but at the same time in a slightly different way, depending on which log data they had got for their calculation. Furthermore, the definitions are mainly addressed towards traditional learning, in all approaches, face-to-face learning and blended and online learning. This is the case of the following terms: progress, performance, participation or dropout.

But, as DeBoer et al. (2014) stated, the definitions until now used to understand and assess the behaviour of the students must be re-conceptualised for MOOC courses. The diversity of user intentions and backgrounds and the unconstrained synchronicity of their activities make the MOOC context very different from conventional classrooms.

Consequently, ECO MOOCs must provide their own definition for the following terms. It is important to highlight that sMOOCs aim to satisfy the participant's goals and these can change along the course. Thus, our definitions will always take which are the participant's goals into account. It leads to enquire these goals at the beginning of the course (read methodology section).

3.1.1 Progress

In ECO *progress* is defined as movement towards the achievement of a goal.

According to our methodology, each course goal must be associated to one or more learning modules and, in turn, each module will have several assessable activities to be done. Some will be compulsory and others not. Then, we understand progress as the advance in the development of compulsory tasks according to participant's goals. These tasks can be passed or failed. However, there might be optional activities in the course design. To depict proper progress both optional and compulsory activities have to be taken into account when visualising progress.

Moreover, progress needs to be measured against the intentions and learning goals of the individual learner and there are several possible scenarios when the intentions and learning goals are not known, or where the learner has opted to browse some topics that do not include compulsory tasks.

3.1.2 Performance

Performance is defined as the degree of understanding and mastery of course material.

Performance, should thus be measured as the results achieved in assessable tasks related to participant's goals, such as grades achieved in quizzes, subjective assessments given by peers about group-works, discussions, etc. following a rubric or social valuation of the tasks carried out, for example, by means of votes or a rating.

3.1.3 Mastery

Mastery is related to performance. While performance is expressed as a mark, degree of improvements is captured in mastery.

3.1.4 Participation

Participation is the process during which participants are consulted about or have the opportunity to become actively involved in a course or activity of the course.

Hrastinski (2008) defines online learner participation as the process of learning by taking part and maintaining relations with others. It is a complex process comprising doing, communicating, thinking, feeling and belonging, which occurs both online and offline. This definition emphasizes that students learn both online, e.g., by computer-mediated communication with peers and teachers, and offline, e.g., by reading course literature.

On the other hand, from our point of view (Zorrilla, 2014), *attendance* refers to the possibility of measuring online attendance, which according to Douglas and Alemanne (2007) is one that incorporates active class participation (synchronous classes) and attention during non-active periods (we consider learner autonomous work as well as group work). In other words, we are interested in measuring the degree of learner participation in the activity or course without considering the quality of his participation since, up until the present, this cannot be automatically evaluated without instructor intervention. Measuring quality is a subjective task which requires the reading and understanding of the text and regrettably, nowadays, natural language processing techniques cannot accurately measure whether the contribution of a person to a certain subject is suitable, notable, significant or not. Thus our online attendance indicators are crude measures which do not cover other important aspects that must be taken into account in order to measure engagement or participation, such as interest, effort or motivation (Bulger et al., 2008; Hrastinski, 2008).

Although participation can be measured in both a quantitative and qualitative way, we hereafter only take quantitative perspective into account, since the other one is just considered in performance. Furthermore, we constrain the measure to collaborative activities related to participant's goals.

3.1.5 Drop-out

Participant at risk of drop-out are those participants with a remarkable low participation, progress and performance with relation to those who have got the same goals.

To measure drop-out in this context is very difficult since participants can adapt their pace of learning to their needs, can change their initial goals extending or reducing the goals to be reached. Therefore, we propose the definition of at-risk-of-dropout. The limit to be at risk of dropout could be established as a 20% or 25% of the activity performed in the course by their peers, but only those with the same goals.

3.2 Categorising users

There are two categories of users in ECO: registered and enrolled learners. Enrolled learners are those users that have registered in ECO and have enrolled in a sMOOC. Only the learners in the category enrolled are of importance for learning analytics. Within the enrolled learners we should again differentiate between types of learners. Literature reports several types of MOOC participants. Kizilcec et al. (2013) determined four types of participants by applying k-means clustering techniques on actual data indicative of engagement. Other researchers distinguish similar types of participants.

- Completing learners: who visit every week, read most of the activities and complete most of the assessments
- Auditing learners: who visit every week, read more than average of the activity pages and infrequently complete assessments
- Disengaging learners: who completed some assessments in the beginning but then decrease in engagement
- Sampling learners: who watch lectures for only one or two periods.

The categorisation of learners is an indication of their engagement. In course designs according to the xMOOC model, engagement usually is determined from viewing lectures and submitting assignments (for credits). Several methodologies have been applied to arrive at categories. Most involve using the k-mean algorithm and/or clustering techniques (Kizilcec, et al., 2013; Anderson et al., 2014). Others use predictive modelling (Ramesh et al., 2014). Here the assignment fraction (number of completed assignments divided by sum of viewed lectures plus completed assignments) is useful to divide learners into viewers and solvers. Solvers are learners who submit assignments but do not pay much attention to resources (Anderson et al, 2014). Most literature reports that the largest population (around 50-60% or more) of participants are those that show very little activity (bystanders, lurkers, followed by those who view lectures (around 20-30%), while only very few consistently submit assignments and assessments (less than 5%).

Moreover these studies focused on providing summative information about type of participants instead of looking into how to promote that learners keep engaged with the course.

Care must be taken that most figures reported in literature are based on xMOOC type of courses, i.e. courses relying on video lectures and assessments. Although video lectures can be part of the ECO sMOOCs, there will be a much more varied type of activities for learners to do. So, we need to verify whether in ECO sMOOCs the same types of participants can be determined. Nevertheless, it seems reasonable to assume that there will be similar types of learners in the ECO sMOOCs, but care needs to be taken to involve other metrics to determine type of learners in sMOOCs.

One particular aspect is the extent of interaction and therefore the use of the social tools to connect to the network. Milligan et al. (2013) differentiated active participants as those who develop networks internal to the MOOC and posted actively on Twitter and blogs. Most of the active participants had participated in MOOCs before. Lurkers were actively following the course, but did not engage with other learners, although they would communicate about course content with others in external networks. Lurkers can be passively involved in the internal network by reading postings, but do not actively contribute.

Some information can be gained from time of registration as well. Anderson et al. (2014) found that only 60% registered before the course started, and 18% registered after the course has ended. The remainder registered during the course. Learners who register late are more likely to be auditing or sampling learners or learners who only download content.

It also seems that learners who keep up with the MOOC starting date are more likely to complete than learners who start at a later time. They tend to start with the contents of week they joined instead of from the start. They also seem to have problems getting integrated into community discussion (Yang et al., 2013).

Because participants all have their own objectives in enrolling in a sMOOC, we need to measure their progress, performance and risk of dropping out against their personal learning goals. The types as defined above, however, can only be determined by their actual behaviour. Consequently, they can only be categorised after the MOOC has been running for some time, while the 'completing learners' can only be classified after they have completed the full MOOC.

Thus, it is hard to provide any feedback based on these classifications, certainly when starting a new course. Later on, when historical data becomes available advice to new participants can be based on this historical data. So, in addition to automatically determining learner type, we need to ask the learner about their intentions and their learning goals when the learner starts the course. This would prevent a cold start problem and would allow us to determine three of the four above mentioned categories. Disengaging learners can have the intentions of all three user types, but for some reason stop attending the MOOC.

In the context of ECO sMOOCs where learners have their own learning goals, these categories need to be redefined, based on actual usage and historical data and can only be truly determined after the course has been completed. When these data are not yet available, and LA have to be presented according to type of learner, some assumptions based on common MOOC literature could be as listed below. This should be supplemented by the answers provided by the learner to the entry questionnaire.

- Completing learners: who visit every week, read most of the activities and complete most of the assessments and basically follow the course as designed aiming for certificate. These are also the learners who have indicated to wish to obtain a certificate, and have selected the whole set of learning goals.
- Auditing learners: who visit every week, read more than average of the activity pages and infrequently complete assessments. These learners have indicated not to aim at a certificate and have selected a subset of learning goals
- Exploring learners: who watch lectures for only one or two periods of the measurement interval, or do only one or two modules. These learners have indicated either to follow the MOOC out of interest in the topic or indicated a single learning objective.
- Disengaging learners: who completed some assessments in the beginning but then decrease in engagement.

In the absence of historical data to determine type of learners, we might use the following figures.

- Completing learners: visit at least once every week, access at least 80% of the activities/lessons/modules defined in the course, submit at least 80% of assessments and get a positive score of at least 60% or the score as set in the scoring rubric.
- Auditing learners: visit at least once a week, access more than average of the activities/lessons/modules defined in the course, but complete only 20% of the assessments.
- Exploring learners: do not visit every week, and probably occur in only one or two of the measurement periods as having accessed any of the activities/lessons/modules defined in the course

In addition to these types of learners that are determined on the basis of what they indicated in the entry survey and their actual behaviour, we consider learners to have become active when they have logged in at least once and have accessed at least 5% of all available material.

3.3 References

Anderson, A., Huttenlocher, D., Kleinberg., J., & Leskovec, J. (2014). Engaging with massive online courses. In proceedings of the 23rd International Conference on World Wide Web, WWW'14, pp. 687-698. Republic and Canton of Geneva, Switzerland.

Bulger, M. E., Mayer, R. E., Almeroth, K. C., & Blau, S. D. (2008). Measuring Learner Engagement in Computer-Equipped College Classrooms. Journal of Educational Multimedia and Hypermedia, 17(2), 129-143.

Cooper, A. (2012). What is "Analytics"? Definition and Essential Characteristics. CETIS Analytics Series, 1(5). Retrieved from http://publications.cetis.ac.uk/2012/521

Douglas, I. & Alemanne, N.D. (2007). Monitoring Student Participation and Effort. In D. G. Sampson, J. M. Spector, D. Ifenthaler & P. Isaías (Eds.). Proceedings of the IADIS International Conference on Cognition and Exploratory Learning in Digital Age (CELDA 2007). October 19-21 (pp. 299-302) Lisbon, Portugal. IADIS Press.

Hrastinski, S. (2008), What is online learner participation? A literature review, Computers & Education, Volume 51, Issue 4, December 2008, Pages 1755-1765, ISSN 0360-1315, http://dx.doi.org/10.1016/j.compedu.2008.05.005.

Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. Paper presented at the Proceedings of the Third International Conference on Learning Analytics and Knowledge, Leuven, Belgium.

Milligan, C., Littlejohn, A., & Margaryan, A. (2013). Patterns of Engagement in Connectivist MOOCs. MERLOT Journal of Online Learning and Teaching 9(2), pp. 149-159

Ramesh, A., Goldwasser, D., Huang, B., Daumé III, H., Getoor, L. (2014) Learning latent engagement patterns of students in online courses. Proceedings of the Twenty-Eight AAAI Conference on Artificial Intelligence

Reich, J. (2014). MOOC completion and retention in the context of student intent. Educause Review. Retrieved from http://www.educause.edu/ero/article/beyond-mooc-model-changing-educational-paradigms

Suthers, D., Dwyer, N., Medina, R., & Vatrapu, R. (2010). A framework for conceptualizing, representing, and analyzing distributed interaction. International Journal of Computer-Supported Collaborative Learning, 5(1), 5-42. doi: 10.1007/s11412-009-9081-9

Tobarra, L., Robles-Gómez, A., Ros, S., Hernández, R., & Caminero, A. C. (2014). Analyzing the students' behavior and relevant topics in virtual learning communities. Computers in Human Behavior, 31(0), 659-669. doi: http://dx.doi.org/10.1016/j.chb.2013.10.001

Wang, A. Y., Newlin, M. H., & Tucker, T. L. (2001). A Discourse Analysis of Online Classroom Chats: Predictors of Cyber-Student Performance. Teaching of Psychology, 28(3), 222-226. doi: 10.1207/s15328023top2803_09

Yang, D., Sinha, T., Adamson, D., & Rosé, C. P. (2013). Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses. Proceedings of the 2013 NIPS Data-Driven Education Workshop.

Zorrilla, M., Álvarez, E., & García-Saíz, D. (submitted). A general method for measuring online attendance in e-learning tools. Sent to Journal of technology enhanced learning in November 2014. .

4 Learning analytics for learners

Learning analytics (LA), as previously mentioned, seek to enhance the learning processes through systematic measurements of learning related data and to provide informative feedback to learners and teachers. Track data from learning management systems constitute a wealthy data source for learning analytics, although this can be enriched by means of surveys and additional information provided by instructors or extracted from the course using natural language processing. Mainly, most research in this field has focused on measuring, from an institutional point of view or as a tool for instructors, student engagement, participation, performance, and dropout in face-to-face learning or supported by e-learning platforms (Verbert et al., 2014) such as Moodle, Blackboard or Sakai. Furthermore these have generally been addressed to instructors and educational centres with the aim of they could make decisions in their coursers in a more informed way.

In ECO, our goal is to provide learners with quantitative and qualitative indicators which allow them to reflect on their way of studying, the resources used and their results, to motivate and stimulate them for the achievement of their learning goals, to positively value their contribution and level of commitment with their peers as well as to compute a social reputation indicator in such way that participants achieve both personal satisfaction and keep engaged in more sMOOCs.

Some of these indicators will be able to be also shown to instructors, e.g. academics, but in this case, each one will observe their metrics in the most understandable and friendly way, that means, with numbers and graphs expressively enough that just by observing, they can make decisions.

In order to display these indicators in a friendly way, these will be organised in a dashboard according to different categories that are described later in this chapter. Dashboards, like a dashboard in a car, are used to present the value of several important functions and features to a user in a visual informative manner.

It must be said that, until now, very few works have focused on providing **learners** with a suitable dashboard. In fact, Verbert et al. (2014) performed the analysis of more than twenty five dashboards, they concluded that most tools answer to teacher's perspective. In (Siemens et al., 2011), authors propose 9 concrete benefits of using learning analytics, being only 2 related to learners or learning. Thus, it is urgent to carry out a study to discover and assess what issues learners are interested in and what interpretation they do about them as also stated in Kruse et al. (2012). Thus it would be good to prepare a survey to send participants in first round ECO courses.. As far as we are concerned, only Santos et al. (2013) have recently performed a study about what issues were important for students but once again in a traditional setting and in which only fifty-six students were surveyed. Nevertheless, we can mention some of them: group member that does not work, motivation, to be aware which resource and tools I and others students use, among others.

ECO sMOOCs must provide learning analytics metrics, on one hand, with the aim of avoiding the loss of motivation and engagement as well as minimising that participants feel isolated, frustrated or disorientated in the course hyperspace. On the other hand, it is also convenient to reinforce their ego and recognise their achievements and their contribution to the learning process of their peers.

Consequently, we consider that sMooc participants, in particular, learners only need a small set of metrics, those which allow them to know their progress and performance in the course according to their goals (it is important to highlight that these can change during the time in which learner is enrolled), as well as their social influence and perceived recognition and reputation (karma). As ECO pedagogical design is centred on collaboration, connection and promoting interaction with peers, some information about their colleagues could be relevant for them, such as their level of participation and engagement in the course, expertise, personal interests and so forth. These metrics should be offered in such way they can compare with the rest of the course or, at least, with a certain group of participants, those with similar interests and goals. This would help to promote the participant's reflection.

As mentioned in chapter 1, it is possible to develop software artefacts that, taking advantages of the data logged in MOOC platforms, provide recommendations and support to participants in their learning process. For instance, identifying which participant is suitable to ask him a question because he seems to know the topic (assessments with a good ark or messages in the forum marked as favourite or well-rated) or searching for participants who have similar goals and/or profile (demographics, likes, background, etc.), among others. Some proposals are mentioned in advanced services section.

4.1 Learner academic and social goals

Next we propose a list of learner academic and social goals, a brief description about what we want to measure and a proposal of issues for their evaluation.

Table 1: Categorisation of goals from learner's point of view

Academic	Description	What issues could be considered
Performance	A variety of indicators designed to measure ability, outperform in order to achieve certification	Compulsory assignments completed successfully (passed) Compulsory quizzes passed
Mastery	A variety of indicators designed to measure understanding, academic competence, or improved performance relative to his initial knowledge (when he enrols in the course)	Compulsory and voluntary assignments performed although he does not achieve to pass The increment of improvement in assessable tasks (if it is possible to repeat or review the activity)
Effort	A variety of indicators designed to measure effort, in particular, demanding academic work	Time spent in an assignment in relation to the rest of peers who achieved the same grade

Engagement	A variety of indicators designed to measure motivation and interest in the course	Access frequency Degree of participation in communicative tasks Degree of participation in collaborative tasks
Progress	A variety of indicators designed to measure progress in the course	Number of assignments carried out (without considering the grade)
Satisfaction	Metrics which reflect participant's satisfaction	Personal valuation at the end of the course
Social		
Social affiliation[1][1]	A variety of affiliative academic behaviours, particularly working together with other learners in productive or cooperative ways	Social graph according to communication, collaboration, demographic data, goals, and so on)
Social recognition	A variety of academic behaviours designed to please, or at least attract, the attention of, significant others	Followers Rating achieved in learner's contributions
Social responsibility	A variety of behaviours involved with participation in supportive roles, or increased academic effort due to his commitment in the accomplishment of assignments	Forum and social networks participation Responsibility in peer evaluation

Table 2 tries to relate the issues written in Table 1with metrics which, we understand that, ECO platform should collect. These metrics, raw data will be able to be summarised or calculated for each resource

[1]Affiliation in the sense popularized by David McClelland and describes a person's need to feel a sense of involvement and "belonging" within a social group

(forum, contents, etc.) and/or for each different learner group that be defined later. The following step will be the definition of indicators for learners which will be displayed in dashboards (see chapter 6).

Table 2: Metrics to measure each issue

Academic Issues	Metrics
Compulsory assignments completed successfully (passed)	Number of assignments passed Grade achieved in each assignment
Compulsory quizzes passed	Number of quizzes passed
Compulsory and voluntary assignments performed although he does not achieve to pass them	Number of assignments performed
The increment of improvement in assessable tasks (if it is possible to repeat or review the activity)	Grade achieved in each assignment or quiz Attempts performed for each assignment or quiz
Time spent in an assignment in relation to the rest of peers who achieved the same grade	Time spent in each assignment Grade achieved in each assignment
Access frequency	Number of times and dates that learner has accessed to each resource
Degree of participation in communicative tasks	Number of messages thread (started) in the resources associated with communication Number of messages written or spoken in resources associated with communication Number of times that his contribution was marked as favourite or achieved a valuation

Degree of participation in collaborative tasks	Number of initiated conversations in the resources associated with collaborative tasks Number of messages written or spoken in resources associated with collaborative tasks Number of times that his contribution was marked as favourite or achieved a valuation Number of conversations where the participant has posted a message Number of forums where the participant has posted a message Date and time in which every message was published
Number of assignments carried out (without considering the grade)	Number of assignments carried out
Personal valuation at the end of the course	Data from a personal survey requested by ECO
Social Issues	Metrics
Social graph according to communication, collaboration, demographic data, goals, and so on)	Data from ECO learner profile (gender, country, age and areas of interest) Course goals selected by learner Relationships arise from the interaction in collaborative tools or in communication tools (e.g. network of forum messages)
Followers	Number of people who read or answer to your contributions in forums, social networks, shared documents…
Rating achieved in learner's contributions	Valuation achieved as a result of your contributions in forums, social networks, shared documents…
Forum and social networks participation	Number of messages thread in the forums and social networks Number of messages written in the forums and social networks

Responsibility in peer evaluation	Average time to answer a question address to him (the learner) Number of peer-evaluation tasks carried out on time Valuation given by his peers (question directly requested by ECO in peers activities)

4.2 Learner's academic and social indicators

As a consequence of the fact that there are diverse MOOC platforms and each one provides different resources and log different events and facts, we delegate to the MOOC platform or it would be better to instructor's course, to select which activities, tasks, or whatever educational element must be tracked and which ones are assessable or not.

Similarly communicative tasks will have to be established by each course, among others, forum, videoconferences, chat should be included. The same must be done with collaborative tasks, those which must be carried out among peers. ECO cannot difference between a forum for discussing a topic from another one for introducing oneself.

Regarding grades, each MOOC platform will have to send grades with its scale with the aim of these can be compared and assessed with others from different courses. ECO will have to define mapping tables for each different scale to one defined as ECO standard.

All numeric indicators will be displayed in comparison with the average of each one of these three groups of peers: 1) the whole group, all the learners enrolled; 2) group of learners with the same ECO profile; 3) group of learners with the same learning goals in the course. As courses are always open, we should define which learners choose to perform calculations. Inside period of the course, LA module could consider only people who selected their learning goals in this period and out of this period, LA module could use a pattern built from the last course edition.

Assumptions:

- Access to the full course is open from the start.
- Course content is organised around topics and structured in modules.
- Each module of the course develops one learning goal, in such way, each learner decides which modules are interested in or selects all. We have simplified the possibility of a learning goal was distributed in several modules or a module had got several goals in order to avoid instructors had to described this only for LA module.
- Each module can have mandatory and optional tasks.
- In order to measure progress, LA module must know when each task is finished. Given that the course is always open, the best option would be that learners marked when he has terminated the task. A learner could repeat a quiz or review a document written, or reread a page. If we use dates

defined by instructors, they could only be considered if learner follows the course according to guidelines.

- Indicators should be updated each week.

Performance (PI):

Performance in each module will be calculated as

PI1. Personal grade in module $P_i = w_i G_i$

Being j the number of the module to be assessed, w_i the weight of each assessable task in this module and G_i the grade achieved. Here we consider only mandatory tasks (quizzes included).

PI2. The final grade will be calculated as $P = \sum_j P_j$

ECO dashboard will show performance indicator of each module to every participant but, the final grade only will be displayed to participant who has got the aim of doing whole course. These indicators will be shown along with the figures obtained in the three groups previously mentioned.

Mastery (MI):

Mastery tries to measure how participants are getting knowledge; therefore we consider here all the tasks, optional and mandatory. If the task is assessable then ECO will show the grade, if not a mark of activity done. In this last case, the degree of improvement won't be calculated.

Table 3: Calculating mastery

Module x	Task1	Task2	….	Task n
Grade in first attempt				
Grade in last attempt				
% improvement				
% participants performed the task				

Progress (PrI)

Progress in each module will be calculated as a rating using this equation

$PR_i = \sum_i \frac{t_i}{i}$

being t_i the number of tasks marked as done and i the total number of tasks to be performed in the module j. We consider all the activities proposed (mandatory and optional) but they must be displayed in two separated indicators.

We propose to show a progress bar for each module. These indicators will be shown along with the figures obtained in the three groups previously mentioned.

Engagement (EI)

This category will show several indicators:

EI1. Access frequency = average of days per week which learner accessed to the course (log in)

EI2. Resource use = average of days per week which each learner accessed to each resource (forum, chat, contents, and so on)

EI3. Percentage of participation in communicative/collaborative tasks= nº messages written or spoken / total (per module and global)

EI4. Ranking of valuation of participation in forum = position that each participant occupies with regard to the number of messages marked as favourite with respect to the total messages marked (per module and global)

EI5. Ranking of valuation of participation in forum = position that each participant occupies with regard to the average valuation given to their messages (per module and global)

EI6. Post duration (according Yan et al., 2013): time difference between first and last post calculated in days (in the same conversation and in forum resource).

These indicators will be shown along with the figures obtained in the three groups previously mentioned.

Effort (EfI)

We understand effort as time spent in an activity. Although it is difficult to measure, in our opinion, it is highly valuable by learners.

EfI1. Time per resource calculated as the average time spent in each task per module and compare with the three groups previously mentioned.

EfI2. Effort vs performance. Another interesting indicator for those who want to complete the course is to show if effort and performance is correlated. It could be visualised by means of a boxplot graph with hours in x axis and grades in y axis

Satisfaction (SI):

Set of indicators which summarises the learners' opinion that performed the course previously. These could be displayed using boxplots (see next image, Figure 1, and substitute day of week by question 1, question 2…). These would help learners to enrol in the course or not.

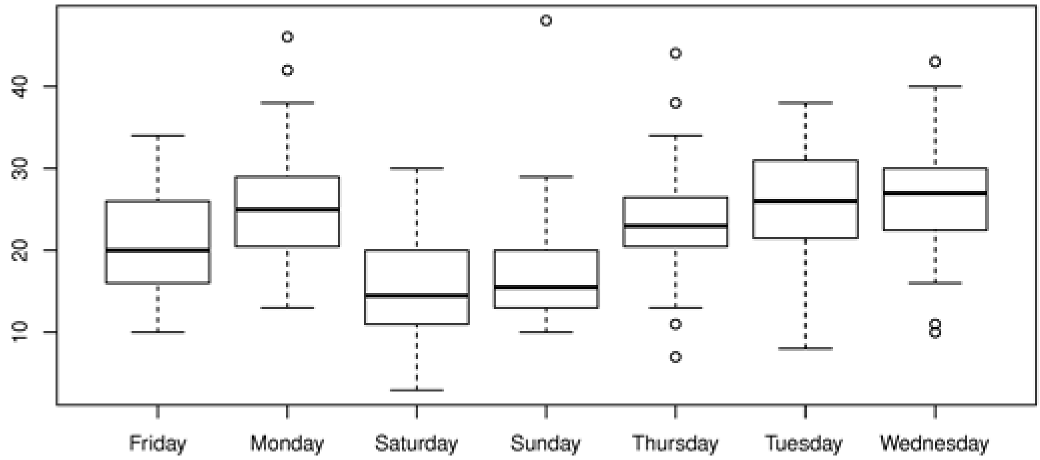


Figure 1: Boxplots assist in making decisions

The survey can collect quantitative and qualitative information taking into account issues such as:

- The effort required.
- Workload for learner.
- Level of difficulty to follow the course.
- Evaluation of pedagogical aspects.
- Adequacy of the activities and resources for learning.
- Course usefulness.
- Evaluation of results in accordance with the personal goals of the learner.
- Degree of overall satisfaction with the course.
- Aspects to be improved in the course

This survey could be extended or integrated with the post questionnaire proposed in Chapter 7. These satisfaction indicators should be displayed at course level, for example, in the main page of the course with the aim of encouraging learners to enrol.

Social affiliation (SAI)

Social graph according to communication messages in forums. It would be possible to build a social graph by each forum (if there is one for module or topic) and all together.

Social networks such as Twitter or Facebook could be included if data is available. Likewise, if platforms offer the possibility to create contact lists, a social graph from these could also be displayed.

The graph will be painted on an earth map showing a point on the country where participant lives, p.eg. a red point if a man or blue one if a woman, and edges connecting people will be thinner or thicker according to the number of messages interchanged. This map will be interactive and allows participant to zoom in and out, filter points and edges according to his interests such as people with the same profile (age, background, …), people with high karma reputation, people with the same learning goals or living in certain geographical area.

Collaboration network is more difficult to discover because we have to use natural language processing in order to select the messages which ask about a topic or help in the development of other learning activities. Furthermore an ontology should be defined or a set of keywords should be provided by instructors.

Social recognition (SRI)

Here we define an indicator called *followers*. It will be calculated as the number of participants who has read his messages in forum. If there are other social networks, then we will define one indicator per social network. If platforms offer the possibility to create contact lists, *followers* will be calculated as the number of people who follow to this participant and *following* the number of people in his contact list.

Additionally we can define another indicator named *contributors* with the number of people who answer to his messages. This contribution could also be measured in other kind of resources such as wikis, blogs or comment pages since the number of people who add texts or comments participate and contribute to their peers' learning process.

Both kind of indicators will be displayed as raw data and shown along with the figures obtained in the three groups previously mentioned.

ECO could also show which social communities have arisen and who are their leaders, for example using FRINGE (Palazuelos et al, 2011). These communities could be characterised using a clustering technique on certain data such as: gender, range of age, geographic area, number of messages read and written in forums, number of tasks performed and participant type (completing, sampling and the rest).

If the platform allows participants to rate each message, then a ranking could be calculated. The valuation of each participant could be computed as the sum of average rating of each message. The indicator will show the position that participant occupies.

Social responsibility (SSI)

Here we define four indicators:

SSI1. Average time to answer a question address to this learner. It will be possible to calculate it only in mail resource since it is clearly addressed to a person. Here the indicator will show his behaviour in ECO or set of courses carried out in each platform since it would be difficult to measure which messages have been received as a consequence of each course.

SSI2.Number of mandatory peer-evaluation tasks carried out on time out of total. Here it is necessary to know which tasks are of this type and their deadlines as well as people who are involved in the same task and their acknowledge of task done. We understand that participants who accept to do this activity have the obligation to correct a number of works. This is the number of mandatory peer-evaluation tasks. But, with the aim of valuing extra effort, we define another indicator, **SSI3** which represent the number of additional peer-evaluation tasks carried out. This could be displayed as a ranking, i.e., the learner's position with respect to the rest of the learners who have performed extra-revision tasks.

SSI4.Valuation given by his peers (question directly requested by ECO in peers activities). This allows learners to know which people work and if they are committed with their peers. It could be calculated as the average valuation along with the number of valuations received. If number and valuation of peer-evaluation tasks are available, ECO could offer another indicator showing number of peer-evaluation according to valuation.

These indicators will be shown along with the figures obtained in the three groups previously mentioned.

Chapter 8 about implementation of scenarios includes a section that shows how ECO learners' indicators can be calculated from an ECO course hosted in an e-learning platform such as OpenMOOC or Logic Assist. Likewise it highlights which data should be provided by teachers and platforms to accomplish learning analytics tasks.

4.3 Course or module level indicators useful for learners

Next we mention some indicators that could be displayed close to the title of each module or the course, rather than in a dashboard.

CLI1. Effort to pass each topic. This can be measured in hours and shown with a boxplot for both the current course and previous versions of the course.

CLI2. Number of learners who passed the module with respect to the type of learner.

4.4 Advanced services

As mentioned in chapter 1, it is possible to develop software artefacts that, taking advantages of the data logged in MOOC platforms, provide recommendations and support to participants in their learning process. Next, we describe those we consider interesting for learners.

- A service that shows the most popular and relevant resources of the course (documents, videos, etc.). It could select the ten artefacts with the highest frequency of access or better valued by learners (social mark) if this facility is provided by the platform.
- A service that helps to identify participants who are experts in certain topic. They could be selected by means of their performance indicator, e.g. those whose mark is in the first quartile. If this is not available, then they could be filtered by their contribution indicator.
- A service which allows participants to search for participants who have similar goals and/or profile (demographics, likes, background, etc.) enrolled in the course o in the platform. This service will use data gathered in ECO profile.
- Other interesting service, once ECO platform has got a considerable number of courses, could be to offer learners the possibility of designing their own course. The service would provide a possible matching between learner goals and course goals (Buckingham et al., 2011).

4.5 References

Simon Buckingham Shum and Rebecca Ferguson. Social Learning Analytics. Technical Report KMI-11-01. June 2011

Corchado, Emilio, Lozano, José A., Quintián, Héctor, and Yin, Hujun. (2014) Automatic Content Related Feedback for MOOCs Based on Course Domain Ontology. In Proc. Intelligent Data Engineering and Automated Learning – IDEAL 2014. LNCS, 8669, 27-35

Jennifer DeBoer, Andrew D. Ho, Glenda S. Stump, and Lori Breslow. Changing "Course": Reconceptualizing Educational Variables for Massive Open Online Courses Educational Researcher March 2014 43: 74-84, first published on February 7, 2014 doi:10.3102/0013189X14523038

Khe Foon Hew, Wing Sum Cheung, Students' and instructors' use of massive open online courses (MOOCs): Motivations and challenges, Educational Research Review, Volume 12, June 2014, Pages 45-58, ISSN 1747-938X, http://dx.doi.org/10.1016/j.edurev.2014.05.001.

Camilo Palazuelos, Marta E. Zorrilla: FRINGE: A New Approach to the Detection of Overlapping Communities in Graphs. ICCSA (3) 2011: 638-653

Santos JL, Charleer S, Parra G, Klerkx J, Duval E, Verbert K (2013) Motivating students in an open learning environment through open badges. In: Proceedings of the eighth European conference on technology enhanced learning (ECTEL 2013), 14 p (to appear)

George Siemens, Phil Long. Sept. 2011. Penetrating the Fog: Analytics in Learning and Education (EDUCAUSE Review). Available http://www.educause.edu/ero/article/penetrating-fog-analytics-learning-and-education

Tobarra, L., Robles-Gómez, A., Ros, S., Hernández, R., & Caminero, A. C. (2014). Analyzing the students' behavior and relevant topics in virtual learning communities. Computers in Human Behavior, 31(0), 659-669. doi: http://dx.doi.org/10.1016/j.chb.2013.10.001

Verbert, K., Govaerts, S., Duval, E., Santos, J.L., Van Assche, F., Parra, G., Klerkx, J.2014. Learning dashboards: An overview and future research opportunities, Personal and Ubiquitous Computing , (6) , 1499-1514

Wilson, T., Hoffmann, P., Somasundaran, S., Kessler, J., Wiebe, J., Choi, Y., Cardie, C., Riloff, E., & Patwardhan, S. (2005). OpinionFinder: A system for subjectivity analysis. In Proceedings of HLT/EMNLP on Interactive Demonstrations.

Yang, D., Sinha, T., Adamson, D., & Rosé, C. P. (2013). Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses. Proceedings of the 2013 NIPS Data-Driven Education Workshop.

5 Learning analytics for teachers and course designers

The pedagogical framework of ECO project intended participants carry out their own learning process. In this sense, teachers have to assume facilitator roles. To achieve this, it is essential to exercise an analysis of behaviours taking place in MOOC environment, caused by the participants in their interactions with the platform and between themselves. The combination of their individual activities with the flow of information generated by the interaction of all participants generates an "emergent collective, which may be seen as a distinctive individual in itself, both greater and lesser than the sum of its parts: it is a socially constituted entity that is, despite this, soulless, a reflection of the group mind that influences but does not engage in dialogue" (Dron & Anderson, 2011).

However, as Terry Anderson (2009) from University of Athabasca says, "distance education educators like to assert that the pedagogy alone defines their distance learning designs", the truth is that "it is only in a complex dance between technologies and pedagogies that quality distance education emerges". Learning Analytics is the only way for teachers to meet the learning needs of the participants, provided in a collective sense and never individually. "It may be virtually impossible to provide individual support and tutoring to every single student" (Zounek, J. & Sudický).

ECO Project goal is to provide teachers the necessary quantitative and qualitative indicators in order to measure MOOC progresses and, if it is convenient, take steps to improve. As for the participant, Learning Analytics are for the teacher not a goal but a means to achieve the two basic objectives of Reflection and Prediction. Teacher's aim is virtually parallel to learners' aim, that is, to finish the course with learning success. Main difference between them is that learners must take care of their own learning, while teachers must extend these concerns to the whole course. As it is said at Chapter 3 about Learning analytics for learners, "in particular, learners only need a small set of metrics, those which allow them to know their progress and performance in the course according to their goals". Because of that, it is obvious that MOOC teachers will need to have access at least to the same figures and statistics of learners. As we say, *at least*: addition of individual pieces of information will return a general view about the MOOC progress.

Learning analytics can also be applied to inform the course designers about the effectiveness and efficiency of their design and can point out how their design affects learners' behaviour. The design of ECO sMOOCs can be considered to be advanced and innovative, but therefore also quite complex to evaluate. After all it relies on a design that has to engage the learner into taking an active role and producing content and obtaining and creating knowledge by interacting and connecting with others. This networked learning in an online environment means that participants have to collaborate using various tools and technologies. This is not something that will happen automatically in online environments and we have to explore how that affects learning. In an online environment participants can perform a variety of actions with a variety of tools and it is not always possible to determine how this relates to learning. In this respect any analytics need to take into account the order and sequencing of processes and time and duration of activities and processes, because learning, interaction and collaboration can change over time or can depend on design of activities and tasks.

The interpretation of the learning analytics and visualisations of metrics is tightly connected with the design of the course (Lockyer et al., 2013; Thompson, et al, 2013). In order for the teacher/designer to be able to properly interpret and understand the visualisation, it must be known to what type of activity/task/resource the visualisation points. For example a visualisation of interactions between learners can turn out completely differently for a forum used to answer question than for a collaborative task.

Lockyer et al. (2013) neatly point this out by referring to 'checkpoint' analytics and 'process' analytics. Checkpoint analytics determine whether learners have accessed all relevant resources and can be dealt with by simpler metrics and visualisations. Process analytics is needed to get some insight into learning processes such as information processing and knowledge application and creation. This requires more advance analysis and techniques, such as social network analysis to depict interactions or even some form of content analysis.

5.1 Teacher academic and social goals

Next, table 1 gathers the list of the learners' academic and social goals that are relevant to the teachers, and about course design and usability issues, a brief description about what we want to measure and a proposal of issues for their evaluation.

Table 4: Categorisation of goals from teacher's point of view

What issues could be considered	Description	Academic
·Compulsory assignments ·Compulsory quizzes	Indicators designed to measure participants' on the way of MOOC completion	Performance
·Compulsory and voluntary assignments and quizzes performed ·The increment of improvement in assessable tasks (if it is possible to repeat or review the activity)	A variety of indicators designed to measure understanding, academic competence, or improved performance relative to initial knowledge of participants (when they enroll in the course)	Mastery
·Time spent in an assignment among whom passes - Time spent in an assignment among whom fails	A variety of indicators designed to measure effort, in particular, demanding academic work	Effort

What issues could be considered	Description	Academic
·Accesses to the platform - Forums participation - Resources consumption	A variety of indicators designed to measure motivation and interest in the course	Engagement[2]
		Course design
- Activity area - Adoption rate	A variety of indicators designed to know which course resources/tools are being used most frequently and when	Courseware
- Badges achieved	Metrics which reflect participants' achievements related to badges, defined by course designers	Badges
·Valuation at the end of the course	Metrics which reflect participants' satisfaction	Satisfaction
		Social indicators
·Social graph according to communication, collaboration, demographic (data, goals, and so on)	A variety of affiliative academic behaviours, particularly working together with other learners in productive or cooperative ways	Social affiliation[3]
- Leaders	A variety of indicators designed to detect the most well recognized participants	Social recognition and leadership

[2]Succesful MOOC rates are directly related to: longer time spent in the platform, more number of accesses, more forum participation and more downloading resources (García-Tizinaray, D., Ordoñez-Briceño, O., & Torres-Diaz, J.C., 2014)
[3]As it is described at chapter 3, "affiliation" in the sense popularized by David McClelland and describes a person's need to feel a sense of involvement and "belonging" within a social group.

What issues could be considered	Description	Academic
- Quality participation ·Peer evaluation	A variety of behaviours involved with participation in supportive roles, or increased academic effort due to his commitment in the accomplishment of assignments	Social responsibility and plurality
		Usability indicators
- Way of access - Device of access	A variety of indicators designed to know method of access	Logging

Table 2 tries to relate the issues written in table 1 with metrics which, we understand that, ECO platform should collect. These metrics will be able to be summarised or calculated for each resource (forum, contents, etc.) and/or for each different learner group that will be defined later. The following step will be the definition of indicators for teachers which will be displayed in dashboards (See section 4.2)

Table5: Metrics to measure each issue for teachers

Metrics	Academic Issues
Number of passes in each task Number of fails in each task % participants performed the task Grade achieved in each task	Compulsory assignments
Number of passes in each quiz Number of fails in each quiz % participants performed the task Grade achieved in each quiz	Compulsory quizzes
Number of passes in each compulsory or voluntary assignment and quiz	Compulsory and voluntary assignments and quizzes

Number of fails in each compulsory or voluntary assignment and quiz % Participants performed the task Average of attempts for each compulsory or voluntary assignment and quiz Grade achieved in each compulsory or voluntary assignment and quiz	performed
Number of participants who have needed or decided doing the task more than once Grade achieved in each task at the [Attempt 1, Attempt 2, …. Attempt n] Average of attempts for each compulsory or voluntary assignment and quiz	The increment of improvement in assessable tasks (if it is possible to repeat or review the activity)
Time spent in each passed assignment or quiz (with rates, per minutes) Time spent average in each passed assignment or quiz Grade achieved in each passed assignment or quiz Number of participants who have needed or decided doing the task more than once	Time spent in an assignment among whom passes
Time spent in each failed assignment or quiz (with rates, per minutes) Time spent average in each failed assignment or quiz Grade achieved in each failed assignment or quiz Number of participants who have needed or decided doing the task more than once	Time spent in an assignment among whom fails
Number of accessing participants Number of clicks Average of accesses	Accesses to the platform

Average of clicks	
Number of participants in each thread Number of messages written in each thread Average of characters written by each participant in each thread Average of messages among participants who have participated in each thread	Forums participation
Number of participants Number of new contributions Number of responses to contributions Average of contributions among participants who have participated in each SN	Social Networks participation
Number of participants who have participated in each wiki Average of access of participants (only accesses with at least a contribution) in each wiki Average of characters written by each participant in each wiki	Wikis participation
Participants who have opened a blog Participants who have published at least one post Participants who post weekly Participants who post daily	Blogs participation
Number of participants who have downloaded each resource Number of participants who have accessed each resource	Resources consumption
Metrics	**Course design**
Number of participants accessing course homepage Number of participants accessing resources (learning materials)	Activity area

Number of participants accessing forums Number of participants accessing wiki (if it exists) Number of participants accessing activities (assignments and quizzes)	
Date each resource is available for participants Number of participants accessing each resource	Adoption rate
Number of participants achieving each badge	Badges achieved
Data from a personal survey requested by ECO	Valuation at the end of the course
Metrics	**Social indicators**
Data from ECO participants' profiles (gender, country, age, areas of interest and course goals selected) Number of relationships arise from the interaction in collaborative tools or in communication tools (e.g. network of forum messages)	Social graph according to communication, collaboration, demographic data, goals, and so on)
Top 10 participants with most connections in each communication tool Top 10 contributors with favourite marks in each communication tool Top 10 most well marked participants in each communication tool Top 10 participants with more contributions in each communication tool	Leaders
Number of participants who have opened a thread in the forums Number of messages written in the forums Number of participants who have written at least a message Average of contributions per thread	Plurality and participation

Number of messages in the social networks Number of participants who have opened a thread in the social networks Number of messages written in the social networks Number of participants who have written at least a message Average of contributions per thread Number of contributions in wiki Number of participants who have contributed Number of characters per participant who has contributed	
Number of participants who have passed the task Number of participants who have failed the task Number of participants who are waiting for being evaluated Distributions mark Time between uploading the task and the complete evaluation by peers	Peer evaluation
Metrics	**Usability indicators**
Number of participants accessing platform from homepage Number of participants accessing platform from newsletter Number of participants accessing platform from warning email (new response, task evaluated and other programmable alerts)	Way of access
Personal computer / Laptop iPad / Tablet Mobile Others	Device of access

5.2 Indicators

As it is established at chapter 2 Learning Analytics, all numeric indicators will be displayed in comparison with these three types of learners:

- Completing learners.
- Auditing learners.
- Exploring learners.

Moreover indicators should be also calculated for the whole group, for learners with similar profiles and for learners with similar learning objectives.

Assumptions:

- Each course has got all its modules opened, thus it is convenient to analyse the course by modules.
- Some figures must be analysed by modules as well as by weeks. These figures are related to the whole course and not to specific resources, such as: Accesses to the platform, Forums participation, Activity area, Way of access, Badges achieved, Quality participation and Responsibility in peer evaluation
- Each module can have mandatory and optional tasks.
- In order to measure progress, LA module requires to know when each task is finished. Given that the course is always open, the best option would be that learners marked when he has terminated. A learner could repeat a quiz or review a document written, or reread a page. If we use dates defined by instructors, they could only be considered if learner follows the course according to guidelines.
- Indicators will be updated each week.

5.2.1 Performance

Performance indicators measure how close or far participants are completing course. Because of this fact, performance in only measured on mandatory assignments and quizzes and avoids voluntary ones.

If the task is assessable then ECO will show the grade, if not a mark of activity done. In this last case, the degree of improvement won't be calculated.

Table 6: Calculation of performance on tasks

	Task1	Task2	....	Task n	Total
Number of passes					
Average grade of achievement					

% participants performed the task					

We will use the same table in order to calculate quiz data.

5.2.2 *Mastery*

Mastery tries to measure how participants are getting knowledge; therefore we consider here all the tasks, voluntary and mandatory. If the task is assessable then ECO will show the grade, if not a mark of activity done. In this last case, the degree of improvement won't be calculated.

Average improvement is calculated among participants who have needed or decided doing the task more than once and only taking into account the difference between the first and the last attempt.

Tasks [1, 2, ... n] should be coloured or signed differently, relying on their mandatory or voluntary character.

Table 7: Calculation of mastery and improvement

	Task1	Task2	....	Task n	Total
Number of passes					
Number of fails					
Number of participants who have needed or decided doing the task more than once					
Average of attempts					-
Average grade of achievement					
% Average improvement					
% participants performed the task					

5.2.3 Effort

We understand effort as time spent in an activity.

Average time spent among passes must be calculated taking into account uniquely participants who have passed. Participants who have failed and who have not tried yet will be excluded. Average time spent among passes also take into account the whole time spent a participant has needed in order to pass even he or she has needed one or more attempts.

It is important to take into account that the time invested is uniquely collectible when participants face quizzes, such as multiple choice. It is not possible, if the teacher asks students for preparing a document DOC or PDF, to be uploaded.

Average time spent among fails must be calculated taking into account participants who have failed and who have tried doing the task although they have not finished it. Participants who have passed will be excluded.

Another interesting indicator is to show how correlate Effort, Mastery and Performance. It could be visualised by means of a boxplot graph with hours in x axis and grades in y axis

Table 8: Calculation of effort

	Task1	Task2	....	Task n	Total
Average time spent among passes					-
Average time spent among fails					-
Number of participants who have needed or decided doing the task more than once					
Average of attempts					-
% participants performed the task					

In relation with indicator "Participants who have passed and have needed [m-n] minutes": "p" and "q" refers to ranges of time, such as "5-10 minutes" or "15-30 minutes", in order to classify participants in groups by their time spent in each task. We follow the same instructions in the case of "Participants who have failed and have invested [p-q] minutes".

Table 9: Time spent per task for learners who pass or fail

	Task1	Task2	....	Task n
Participants who have passed and have needed [p-q] minutes				
Participants who have failed and have invested [p-q] minutes				

5.2.4 Engagement

Engagement figures try to predict if participants are engaged to the course. Consequently, we can also measure the risk of dropout or abandon. A research of García-Tizinaray, Ordóñez-Briceño and Torres-Diaz (2014) concludes engagement rates are directly related to: longer time spent in the platform, more number of accesses, more forum participation and more downloading resources.

Accesses to the platform

Number of accessing participants and number of clicks on a timeline (per day, per week and global).

Average of accesses and average of clicks per learners groups defined at chapter 2.

Forums, social networks, wikis and blogs participation

About forums participation, we consider "Number of messages thread in the forums and social networks", "Number of messages written in the forums and social networks" and "Average of messages per participant".

Table 10: Determining engagement via forum contributions

	Thread1	Thread2	....	Thread n
Number of participants				
Number of messages written				
Average of characters written by each participant in each thread				
Average of messages among participants who have participated				

About social networks (SN), we consider "new contributions" as messages that participants create or post without responding another previous comment.

Table 11: Determining engagements through social network contributions

	SN1	SN2	....	SN n
Number of participants				
Number of new contributions				
Number of responses to contributions				
Average of contributions				

About wikis participation, we consider "number of participants who have participated" and "average of characters written by each participant"..

Table 12: Determining engagement through wiki contributions

	Wiki1	Wiki2	....	Wiki n
Number of participants who have participated				
Average of access of participants (only accesses with at least a contribution)				
Average of characters written by each participant				

About blogs, it should be measured as it follows.

Table 13: Determining engagement through blog contributions

	Number	%
Participants who have opened a blog		

Participants who have published at least one post		
Participants who post weekly		
Participants who post daily		

Resources consumption

Teacher should take into account this data and board:

Number of participants who have downloaded one resource on a timeline

Table 14: Determining use of resources

	Resource1	Resource2	....	Resource n
Number of participants who have downloaded it				
Number of participants who have accessed to it				

5.2.5 Courseware

Courseware figures aim is to determine how participants consume learning materials.

Activity area

Activity area should show how participants move on in the platform and is directly related to timeline. A timeline board should group together the following figures:

- Number of participants accessing course homepage
- Number of participants accessing resources (learning materials)
- Number of participants accessing forums
- Number of participants accessing wiki (if it exists)
- Number of participants accessing activities (assignments and quizzes)

Adoption rate

The adoption rate deals with the time span from uploading a selected learning material to the time of access by students. Although in case of MOOCs we are talking about open and not straightforward, adoption rate can indicate us how participants are consuming the course and, consequently, in which order they are choosing to course it.

5.2.6 Badges

Badges system encourages participants to increase their relationship with the platform: forums participation, task activities… In line with this, knowing how many participants have achieved some badges indicate us how many participants are achieving some goals. If we relate this data on a timeline, we could also know how fast these situations are taken place.

Table 15: Determining achievement of goals via badges

	Badge1	Badge2	….	Badge n
Number of participants achieving each badge				

5.2.7 Satisfaction

Teachers will find useful knowing participants' satisfaction which summarizes the participant's opinion. Participants' opinion should be displayed as an average of the opinions received.

The survey can collect quantitative and qualitative information taking into account issues such as:

- The effort required.
- Workload for participant.
- Level of difficulty to follow the course.
- Evaluation of pedagogical aspects.
- Adequacy of the activities and resources for learning.
- Course usefulness.
- Evaluation of results in accordance with the personal goals of the learner.
- Degree of overall satisfaction with the course.
- Aspects to be improved in the course

This survey could be extended or integrated with the post questionnaire proposed in Chapter 7.

5.2.8 Social affiliation

Social graphs which show graphically the interaction among participants should be included. These could be built from forums, wiki, blogs and other communication tools. It would be possible, at least, to build a social graph by each forum (if there is one for module or topic) and all together.

Social networks such as Twitter or Facebook could be included if data is available. Likewise, if platforms offer the possibility to create contact lists, a social graph from these could also be displayed.

The graph will be painted on an earth map showing with points on the country where participants live. The more people participate from a same location, the bigger the point will be. Interactions and relationships between participants will be represented with lines. Edges connecting people will be thinner or thicker according to the number of messages interchanged.

About forums, floating figures above the graph will indicate the following data:

Table 16: Determining participation for social affiliation

	Thread1	Thread2	….	Thread n
Number of participants who have participated				
Number of messages written				
Average of messages among participants who have participated				

Regarding social networks and wikis, graphs can represent the data mentioned in “engagement. forums, social networks, wikis and blogs participation indicators.

Teacher should be able to zoom in and out and to filter data with the following filters:

- Completing learners [see “learner category” at chapter 2]
- Auditing learners [see “learner category” at chapter 2]
- Exploring learners [see “learner category” at chapter 2]
- Data profile: Occupation (Working, studying, unemployed…)
- Data profile: Age (ranges)
- Data profile: Experience in other MOOCs

5.2.9 Social recognition and leadership

Regarding graph abovementioned at “Social affiliation” section, it is necessary to add the following items, in case the teachers would like to count on them as supporters:

- Top 10 participants with most connections in each communication tool
- Top 10 contributors with favourite marks in each communication tool
- Top 10 most well marked participants in each communication tool

- Top 10 participants with more contributions in each communication tool

We assume “communication tool” as forums, social networks and wikis. In case of wikis, it is necessary to know if marking contributions as favourite is possible.

As well, floating numbers will represent these figures.

Table 17: Determining social recognition based on rating

	Thread1	Thread2	….	Thread n
Number of favourite marks				
Average of favourite marks among participants who have participated				

5.2.10Social responsibility and plurality

We understand plurality as the fact of counting with different perspectives and voices about a topic. The more people participate in a debate, the more plural this debate is.

Plurality and participation

Regarding forums:

Table 18: Determining plurality based on forum activity

	Thread1	Thread2	….	Thread n	Total
Number of participants who have opened a thread in the forums					
Number of messages written in the forums					
Number of participants who have written at least a message					

Average of contributions per thread					

Regarding social networks (Social Networks = SN):

Table 19: Determining plurality based on social network activity

	SN 1	SN 2	....	SN n	Total
Number of messages in the social networks					
Number of participants who have opened a thread in the social networks					
Number of messages written in the social networks					
Number of participants who have written at least a message					
Average of contributions per thread					

Regarding wikis:

Table 20: Determining plurality based on wiki contributions

	Wiki1	Wiki2	....	Wikin	Total
Number of contributions in wiki					
Number of participants who have contributed					
Number of characters per participant who has contributed					

5.2.11 Peer evaluation

Table 21: Determining effectiveness of peer evaluation

	Task 1	Task 2	….	Task n
Participants who have passed the task				
Participants who have failed the task				
Participants who are waiting for being evaluated				
Distributions mark				
Average time between uploading the task and the complete evaluation by peers.				

5.3 Logging

Logging figures indicate how participants log into the course. It is useful in order to improve ways of engagement and attract participants' attention. Logging figures should be represented on a timeline, with the following data:

- Number of participants accessing platform from homepage
- Number of participants accessing platform from newsletter
- Number of participants accessing platform from warning email (new response, task evaluated and other programmable alerts)
- Others

We also want to know how students access the platform:

- Personal Computer or Laptop
- iPad or Tablet
- Mobile
- Others

5.4 References

Dron & Anderson (2011). Three Generations of Distance Education Pedagogy. Special Issue - Connectivism: Design and Delivery of Social Networked Learning. The International Review of Research in Open and

Distance Learning. Available on the Internet: http://www.irrodl.org/index.php/irrodl/article/view/890/1663

Halawa, S. (2014). MOOC dropouts – What we learn from students who leave. University World News. Available on the Internet: http://www.universityworldnews.com/article.php?story=20140708163413797

Dyckhoff, A. L., Zielke, D., Bültmann, M., Chatti, M. A., & Schroeder, U. (2012). Design and Implementation of a Learning Analytics Toolkit for Teachers. Educational Technology & Society, 15 (3), 58–76.

García-Tizinaray, D., Ordoñez-Briceño, O., &Torres-Diaz, J.C. (2014), Learning Analytics to predict dropout of distance students. At Campus Virtuales, nº01, v.III, Revista Científica de Tecnología Educativa. Loja, Ecuador.

Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing Pedagogical Action: Aligning Learning Analytics With Learning Design. American Behavioral Scientist, 57(10), 1439-1459. doi: 10.1177/0002764213479367

Thompson, K., Ashe, D., Carvalho, L., Goodyear, P., Kelly, N., & Parisio, M. (2013). Processing and Visualizing Data in Complex Learning Environments. American Behavioral Scientist, 57(10), 1401-1420. doi: 10.1177/0002764213479368

Yang, D., Sinha, T., Adamson, D. and Rosé, C. P. (2013). Turn on, Tune in, Drop out: Anticipating student dropouts in Massive Open Online Courses. In workshop of NIPS. Available on the Internet: lytics.stanford.edu/datadriveneducation/papers/yangetal.pdf

Zounek, J. &Sudický, M., 2013, DisCo 2013, New technologies and media literacy education. 8th conference reader DisCo 2013. Available on the Internet: http://disconference.eu/wp-content/uploads/2013/8thDisCoReader2013_New%20tehhnologies%20and%20media%20literacy%20education.pdf

6 Issues for implementation

6.1 Learning analytics requirements

Learning analytics requirements within the framework of ECO project must consider the following boundary conditions: i) inclusive and accessible courses for large number of participants and to wide diversity of citizens; ii) the pedagogical approach (learner centred approach) not only of associative learning but connectivist, social constructive learning and situated practices; iii) the creation of collaborative learning opportunities; iv) interaction among peers (including some but limited interaction with academic staff).

Learning analytics (hereinafter LA) for MOOCs should provide metrics and indicators which meet the needs of both learners and instructors. Furthermore, these should be designed in order to satisfy the following ECO MOOCs pedagogical conditions: i) support adaptive learning strategies ii) enable the possibilities to adapt to the changing intentions of participants during the course which implies that ECO platform must know which are the learner's goals in each moment. Another important aspect to take into account is the emphasis of ECO on the social perspective. Thus, LA should include academic indicators as well as social indicators.

In this respect it is important to note that LA only can calculate indicators based on available data. Whenever a course makes use of tools and applications outside of the platform, for example by using Twitter, Youtube, Facebook, these data are lost to the LA service, making it hard to arrive at a proper estimation of the indicators. Consequently LA and dashboards will only provide approximate data.

Based on the conditions outlined, the following ECO learning analytics requirements are detailed:

- Use of xAPI in addition to conventional platform logging.
- Implement common indicators for progress, performance, drop-out.
- Learning analytics must be relevant to the participant. This means that indicators support the learner in meeting their individual learning goals.
- Learner indicators must be calculated for 1) the whole group, all the learners enrolled; 2) group of learners with the same ECO profile; 3) group of learners with the same learning goals in the course.
- Detect learner type automatically from action and behaviour and historical data.
- To overcome cold-start present questionnaire when the learner starts the course to ask the learner about their intentions and their learning goals.
- Indicators should be designed in a flexible and general way and each platform must provide the mapping of use of resources with indicators. Each platform will inform participants what it offers and how, i.e., based on which activity, indicators must be calculated.
- ECO user profile should include the courses enrolled.
- Performance has to be measured against learner's objectives but can only be measured when learners engaged in activities that can be assessed in some way. Course designer must indicate how

activities are assessed: a numeric mark, a percentage over the population, by means of a scoring rubric and which their weights are in order to calculate the final mark.

- Regarding the frequency of indicator updating, we suggest that some indicators such as academic ones should be calculated weekly but others, such as social should be dynamically calculated, for instance, who are now connected with my same goals.

It should be noted that each of the current ECO MOOC platforms differ in the features they offer, making it difficult to present coherent learning analytics across all MOOCs and all platforms. At the minimum all platforms and learning analytics dashboards should provide information about progress, performance and risk of dropping out. These should be shown both to individual learners and, in at least in aggregated form to teachers. However, as ECO provides sMOOCs some indication of social aspects, such as contacts, interactions should be presented as well. As not all platforms provide features for social media and interaction, this is an area that might require additional effort.

The ECO platforms might already track and log several user actions and activities, and can continue to do so. As previously mentioned, each platform must provide the mapping of use of resources with indicators, each platform will inform participant what it offers and how the indicator is calculated, ie, based on which activity. Furthermore it is highly recommendable that all platforms implement xAPI and present these data as xAPI statements in order to synchronise across the platforms and MOOCs. This implies agreement on the choice of activities and verbs. Each platform is responsible for a correct mapping to the xAPI verbs and activities.

On the other hand, learning analytics should develop indicators that are necessarily aligned with the course goals as well as the learners' own goals. It must be remembered that the latter can be changed along the course and so Learning Analytics engine should be aware of it. These indicators will be calculated from the activities performed (forums participation, watched videos etc.), the badges achieved and quizzes and peer evaluation activities carried out. Furthermore, we also are interested in detecting learners at risk of dropping out during the courses, which will allow teachers to assess what extent the design, pedagogical model and course activities are appropriate to the learners’ interests and expectations.

Although there is a user profile schema in sMooc, this information is not sufficient to offer a suitable dashboard to both learners and instructors. It is therefore necessary to extent this schema to allow ECO platform to collect or, at least, know goals and skills to be acquired in each course, assignments to be carried out, scoring and evaluation methods used in each assignment and other important parameters such as certification dates, minimum grade in each assessable activity to achieve the certification, etc. that must be defined by the instructors. Furthermore, it is extremely important that ECO platform stores or, at least, can access to log platforms in order to read events and facts about the activity carried out by each participant (time spent in each resource, resources accessed, quizzes performed, grades, messages sent to the forum and so on). And finally, ECO platform must have the information about learner’s goals and intentions. Therefore it is highly necessary to ask learners about their goals in the beginning of the course and facilitate their change. When learners start the course, the MOOC platform should display the course goals and skills to them and encourage the learners to establish and indicate their own learning goals.

Furthermore, learners should indicate their interest in the course. According to (Foon et al. 2014), the reasons why students sign up a MOOC mainly are: they want to learn about a new subject or to increase their knowledge on something they learned before (Agarwal, 2012, Allon, 2012, Belanger and Thornton, 2013, Breslow et al., 2013, Evans, 2012, Kaul, 2012 and Rice, 2013); they are curious about MOOC (Jacobs, 2013, Kirschner, 2012, Martin, 2012 and Young, 2013); for personal challenge (Breslow et al., 2013); they want to get as many course certificates as possible (Young, 2013). Therefore, a question about learners' intentions will help LA module to better classify learners and analyse data from different points of view.

In addition to the static user profile and asking learners about their intentions and learning goals, certain learner characteristics and category of learner can be determined on the basis of historical data, either from the particular MOOC the learner is enrolled in or from MOOCs similar in tasks and effort.

Next we propose the information requirements of MOOC courses which LA module requires to work properly.

6.1.1 Information requirements about course design

As previously mentioned, instructors or otherwise each platform must map the use of resources with indicators. This means that LA module requires to know which activities are mandatory or optional, which forums are assessable and which one not, which quizzes must be performed and which is the minimum mark to be considered passed, and so on. Therefore, in what follows, we specify the information that MOOC courses should provide.

- **Title** of the course.
- **Language** that is taught.
- **Public:** refers to the various public, profiles and stakeholders in the course.
- **Institutions:** centres or hubs participants as designers of the course.
- **Teachers**.
- **Platforms:** The various platforms which will be developed the course.
- **Starting date.**
- **Ending date.**
- **Methodology:** List of learning methodologies used in the course.
- **Goals:** List of course goals.

Table 22: Goals of the course

Goal code	Goal description
Goal.n	Description

- **Contents:** list of modules comprising the course and the activities performed on them. Each module can help to achieve one or more course goals.

Table 23: Modules of the course

Module code	Module title	Module description
Mod.n	Title	Description

Table 24: Modules of the course in relation to goals

Module code	Goals contribute to achieving
Mod.n	[Goal.1, Goal.2, … , Goal.n]

Table 25: Modules of the course and related activities/tasks

Module code	Module description	Activities to be performed
Mod.n	Description n	[Act.1, Act.2, … , Act.n]

- **Activities / Tasks:** List of tasks with detailed information about how they must be performed and measured in order to compute indicators. They may be individual or require collaboration or group-based activities. The activities / tasks may be evaluable or have complementary character. In addition, for the evaluable activities / tasks will be necessary to know: valuation range, minimum grade for approval, number of attempts and timeout, if appropriate among the activities / tasks will include: video, and video-forums, participation and discussion in forums or social networks on articles, news, lectures, case studies or best practices, quizzes, etc.

Table 26: Activities/tasks of the course

Activity code	Activity title	Activity description	Activity type	Activity character
Act.n	Title	Description	Mandatory / optional	Individual / group
	Valuation range	**Minimum grade**	**Attempts**	**Duration**
	Score	Mark	Number of trials	Maximum time

- **Assessments:** it will be necessary to indicate the different evaluations of the course activities and methods, establishing for each activity, if appropriate, the form of assessment (because there may be no measurable activities). Possible assessment methods are: evaluation by peer, evaluation by groups, interaction in forums, participation through social networks, achieving badges, performing quizzes, etc.

Table 27: Assessments

Assessment code	Assessment method	Assessment weight
Assess.n	Method	Weight

Table 28: Assessment of the course activities

Activity code	Assessment code
Act.n	Assess. n

6.1.2 Learner`s goals for the courses

The distinctive features of the MOOC courses and the number and diversity of learners to whom are targeted suggests that many learners may want to achieve all the goals of the course whereas others are only interested in certain topics and thus their progress and performance must be measured according to their goals. That is, it will have to take into account the own goals of learners in relation to those established in the course, which will, in turn, require that the MOOC platform ask the learners at the beginning of the course in which goals are interested. On the other hand, the adaptive learning strategies require the system to adapt to the changing intentions of participants during the course, so the learners should have the possibility, through the MOOC platform using, to change its goals at any time during the course.

According to the above, for the definition of indicators and its measurement, will be required to collect data for each MOOC course about learner`s goals. Table 29shows the data schema required to collect this information. The learner has to indicate his intended learning goals through the entry questionnaire that is presented at the time of enrolment.

Table 29: Learner´s goals

Course code	**Learner identification**	**Learner´s goals**
Course. n	Learner. n	[Goal.1, Goal.2, … , Goal.n]

6.2 Integrating data

6.2.1 Data formats

There are several initiatives that can be used to express actions or activities performed by users on certain objects. To name a few: activity streams, paradata, contextualised attention metadata, Tin Can. Other formats are used to record social data such as relations between people (FOAF: Friend of a Friend), community descriptions and actions (SIOC: Semantically-Interlinked Online Communities), (REV: Review ontology) to express reviews and ratings. These social data but also metadata and paradata could be gathered from an integrated platform or from multiple platforms. In case of the latter, the use of linked open data becomes opportune.

Some of these formats have been used extensively and resulted in specifications, others remained less active. Nevertheless, these initiatives indicate the need for standardisation. Below we briefly present activity streams and paradata, but will advise the use of xAPI specification for the ECO project as xAPI already re-uses verbs and activities from activity streams. To gather aggregated data of groups and activities, xAPI can be supplemented by paradata statements.

Activity streams and paradata

xAPI records learning experiences in the form of activities performed by a learner on an object. This is also known as paradata. While traditionally metadata is used to describe and classify objects, paradata is used to record action and usage data. Paradata is useful when describing or recording how people have used a resource, like clicked on a link to a learning resource. A similar specification is Activity Streams and paradata is built on that. Contrary to Activity Streams that are intended for individual users, paradata allows aggregation of activities, such as actions that multiple people did or multiple actions performed by an individual.

xAPI, Activity Streams and paradata all share a common syntax for statements: actor verb object, recorded in JSON format. Additional information like date or time, context and measurement or results are also recorded in all of these.

Experience API or xAPI

The Advanced Distributed Learning (ADL) Initiative of the US Department of Defense has developed the Experience API (ADL, 2013), also known as xAPI. ADL is known for its learning specifications, such as SCORM the learning object specification. Recently they added xAPI to it. xAPI is a specification to express, record and exchange statements about learning experiences. Assumption is that learners learn by interaction with other people, with content and learning resources, anywhere. This action can result in an event that triggers learner. On the one hand provides the xAPI specification a simple mechanism to express and store these actions as statements, taking the form of noun verb object, or “I did this”. On the other hand xAPI defines characteristics of ‘learning records stores’ (LRS) that store the statements. The LRS can make data stored available to any learning environment or system to, for example visualise relevant indicators to learners and teachers in a learning analytics dashboard. Learning Locker is the open-source reference LRS and is the first choice of use for ECO.

xAPI is particularly suitable to generate personalised learning support, because it is geared towards storing learning experiences of individual learners. It is based on the idea of tracking activity through activity streams that were developed to provide a better way of expressing social media activity. Therefore it is versatile in that it can record any statement as long as it takes the form actor verb object with result, or “someone does an action to/with something”, like “John submits assignment 1 – Describe the ZPD - in learning-activity 2 Vygotsky’s theory”. The context (social ties, groups, and activity duration), timestamp and also semantics and used tags are also part of the tracked learning activities.

Actors can be people, groups of people or systems. Verb describes what happened between the actor and object of the statement. The object is typically an activity, but can be another actor or even statement. Sometimes the distinction between verb and object is not quite clear. The purpose of the data then determines what the verb and what the object are.

xAPI started as the Tin Can project. The project role was to develop the API, but the project also provides the Tin Can API Registry as a place to store and add activities and verbs. Currently the list of verbs includes experienced, attended, attempted, completed, passed, failed, answered, interacted, imported, created, shared, and voided. The list of verbs however is extensive and includes verbs from Activity Streams. The list of activities is extended, and includes, among others, bookmark, game, audio, image, task, video, assessment, etc. A full list can be obtained from the Tin Can Registry at https://registry.tincanapi.com/#home/activityTypes.

xAPI promotes the use of recipes to standardise the way of expressing experiences, because there are many different ways to express that a person has interacted with a resource, say a video. One way might be: Tim Martin played the first 31 seconds of "How to Make Grilled Cheese"; another Tim Martin watched from 0s to 31s of "How to Make Grilled Cheese"[4]. Both statements express the same experience in slightly different manner.

There is a bookmarklet available that can be added to any browser and when linked to a learning platform like the ECO platform could enables individual user tracking with basic authentication. Examples could be an "I think this," "I learned this," "I like this," or "I don't like this" statements that allows self-reporting.

Due to its versatility xAPI is being used quite frequently in the learning analytics domain and therefore the ECO project will make use of it as well. The data required for the learning analytics metrics often extends those that are regularly logged in a learning environment or MOOC platform. Therefore a tracking and logging tool needs to be implemented to capture all relevant data. The xAPI registry provides a list of verbs and activities, some of which are taken from the Activity Stream specification, while others are added by xAPI. This is list is not exhaustive and ECO might need to define new verbs and activities. When it is necessary to express experiences of multiple persons or on aggregation of activities paradata might be used.

6.2.2 Incorporate multiple devices through Tin Can learning record store

xAPI statements are sent and stored in a Learning Record Store (LRS). The LRS is the repository of the learning records. A LRS can be part of a learning management system (LMS) or in our case MOOC platform, or can be set up separately. Data stored in a LRS can be accessed by other LMSs, MOOC platforms, reporting tools, or other LRS.

Because the learning record store act as intermediate to store the data, data not only can be exchanged with other LRS, but also with any other (learning) system that has a need for these data. It also means that the origin of the learning action data are not restricted to the single learning system, but can originate from any environment the learner is using. Additionally, data can be sent using any device. A device is not restricted to PC or mobile devices, but can entail games, simulations, equipment, etc. Therefore it accommodates the mobile learning aspects of ECO.

[4]Example taken from http://tincanapi.com/recipes/

6.2.3 xAPI statements

The xAPI specification is used to track and log relevant actions the learner takes in the MOOC platform. It registers who performs what activity with what object at which time and in what context. xAPI statements take the form *noun verb object. Result, context and timestamp* are also part of the statement.

An example statement taken from the Tin Can website[5] looks like:


```
{
   "actor": {
  "name": "Sally Glider",
  "mbox": "mailto:sally@example.com"
  },
  "verb": {
  "id": "http://adlnet.gov/expapi/verbs/experienced",
  "display": {"en-US": "experienced"}
  },
  "object": {
  "type":"course",
  "id": "http://example.com/activities/solo-hang-gliding",
  "definition": {
  "name": { "en-US": "Solo Hang Gliding" }
  }
  },
  "context": {
  "instructor": {
  "name": "Irene Instructor",
  "mbox": "mailto:irene@example.com"
     },
     "contextActivities":{
  "parent": { "id": "http://example.com/activities/hang-gliding-class-a" },
  "grouping": { "id": "http://example.com/activities/hang-gliding-school" }
  }
  }
}
```


This statement expresses that “Sally took the ‘Solo Hang Gliding’ course, under the instruction of Irene, as part of Hang Gliding Class A, within the context of Hang Gliding School”.

The statement even can contain the outcomes by adding “result”. The following example expresses that “Sally completed ‘Solo Hang Gliding’ with a passing score of 95%”. The “completion” field tells us that Sally is done, and the “success” field tells us that she passed. The “score” field gives us the 95% figure.


```
"result": {
"completion": true,
"success": true,
"score": {
"scaled": .95
}
}
```


[5] Example taken from http://tincanapi.com/statements-101/#actor

xAPI provides lists of definitions of verbs and activities in the Tin Can Registry. Other verbs could be taken from the Activity Stream specification or new ones be defined. When it is necessary to express experiences of multiple persons or on aggregation of activities paradata might be used.

Potentially relevant verbs that are presently in the Tin Can Registry: accept, access, add, attach, complete, create, delete, dislike, favourite, follow, insert, interact, invite, join, leave, like, listen, play, read, receive, reject, remove, remove-friend, request, request-friend, send, share, stop-following, submit, tag, unfavorite, unlike, unshared, use, watch, answered, asked, attempted, attended, commented, completed, failed, interacted, mastered, passed, preferred, registered, responded, resumed, scored, shared, bookmarked, rated, replied to tweet, retweeted, tweeted, viewed, down voted, planned, Log in.

Relevant activity or objects are: alert, audio, badge, bookmark, comment, file, game, group, image, note, page, Question, review, task, video, Assessment, Course, Module, Objective, Performance, Question, Simulation, discussion, Tag.

In chapter 7 we suggest what verbs and activities could represent data to be tracked in ECO.

6.3 Technical architecture proposed

The integration of different heterogeneous platforms as part of the ECO project is itself complex in terms of the integration of the information about courses design of each MOOC platform and the learner´s activity data for learning analytics: i) the data collected by each platform are not homogeneous and ii) there is heterogeneity in the database schemas of the different platforms to gather information on the learners activities.

The information about the learners' activities on the platform should be related to those given by the course designers (explicit information requirements). Therefore, we should establish an ECO architecture that allows the calculation of indicators for both learners and teachers according to the heterogeneity of MOOC platforms and data they record. Thus, the defined indicators and metrics will be applied to the courses offered on each platform according to data that each platform collects (that may be different between platforms) about the learners' performed activities. That means that each platform will have to compute some or all the indicators proposed by ECO project.

6.3.1 Technical architecture

ECO LA suggests the use of xAPI statements to collect the course information (described in section 5.1.1) and the activity carried out by the participants which will be stored in a central Learning Record Store (LRS). LA engine will work on this LRS and display dashboards. Each platform will map the required information about the course design and the use of resources compliant to xAPI syntax.

ECO is comprised of a set of learning environments that already might have some kind of logging and monitoring system. Each platform will be able to continue using its proprietary system as long as it also provides the required data according to the xAPI specification. Indicators will be computed from data stored in the central learning event store, however these could be complemented by others defined by the e-learning platform, in this case, they will be responsible for populating the dashboard directly.

Thus, LA engine will calculate the indicators following the rules that this document determines (see chapters 3 and 4). It could happen that some platforms do not provide data for the calculation of all indicators; in this case these won't be offered. ECO LA Web Service will integrate all the methods and techniques needed to perform these calculations.

The ECO LA Web Service will provide the indicators and visualisations that will be displayed in the learners' and instructors' dashboards. The required data will be a mixture of static data from the design of the course, plus data gathered by tracking and logging about learners' performed actions and events during the run of the course.

Figure 2 shows a architecture to organise, calculate and display indicators in dashboards.

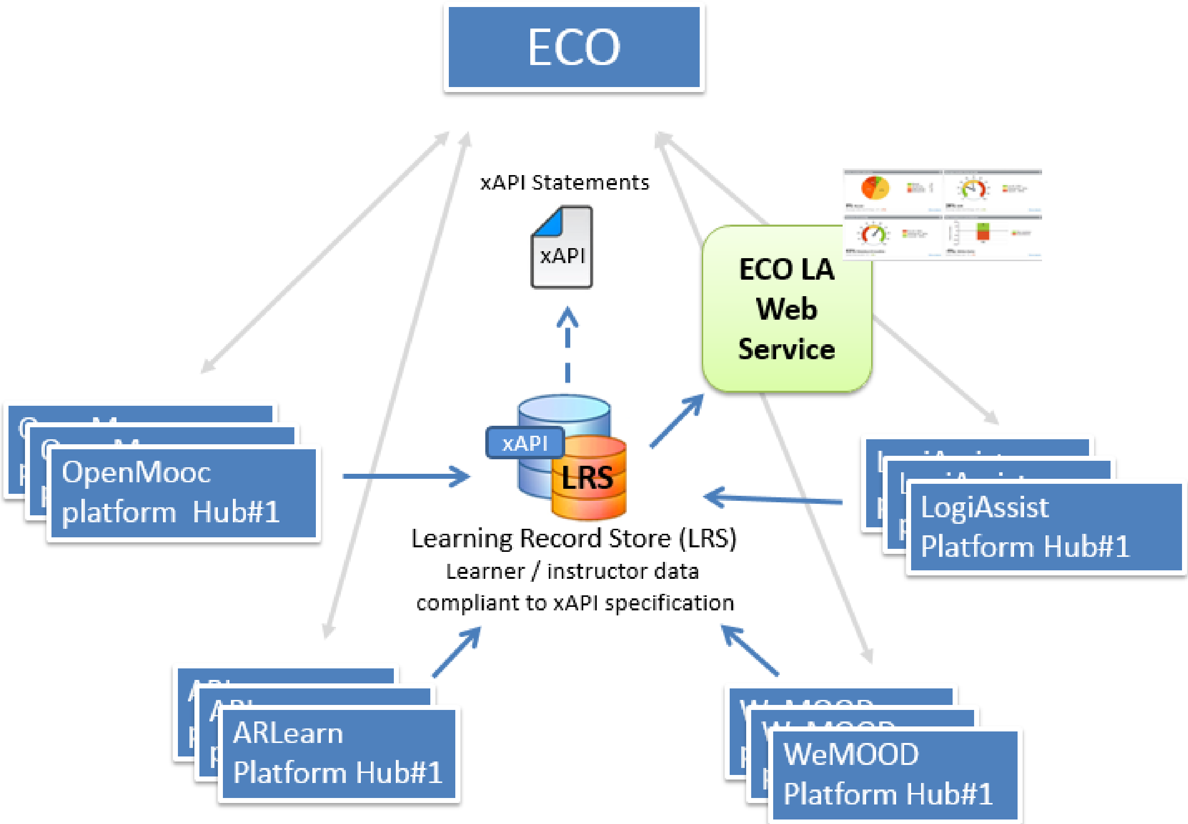


Figure 2: Proposal for architecture

6.3.2 Platform's data

In order to calculate the indicators defined about progress, performance and at risk of dropout for both learners and teachers, it will be necessary to collect data about course design and the activities, tasks and assessments that learners have performed in the course. Data about course design will be contained in each platform integrated in ECO. Each platform will have to map its data according to the common xAPI statements as previously mentioned. Likewise tracking data must be also expressed in xAPI statements and stored in a central learning record store (LRS).

6.4 References

Agarwal, A. (2012) 'Circuits and Electronics', MITx. Chronicle of Higher Education, 59 (6), p. B10

Allon, G. (2012) 'Operations Management', Udemy. Chronicle of Higher Education, 59 (6) (2012), pp. B10–11

Breslow, L., Pritchard, D. E., DeBoer, J., Stump, G. S., Ho, A. D. and Seaton, D. T. (2013) Studying learning in the worldwide classroom research into edX's first MOOC. Research and Practice in Assessment, 8:13–25, 2013.

Crespo, P. & Antunes, C. (2012). Social Networks Analysis for Quantifying Students' Performance in Teamwork. In: Proceedings of the 5th International Conference on Educational Data Mining. pp. 234-235.

Evans, D. (2012) 'Introduction to computer science', Udacity Chronicle of Higher Education, 59 (6) (2012), p. B11 Belanger, Y., & Thornton, J. (2013). Bioelectricity: A Quantitative Approach Duke University's First MOOC.

Jacobs, A.J. (2013). Two cheers for Web U!. New York Times, 162 (56113) (2013), pp. 1–7

Kirschner, Ann (2012, October 1). A Pioneer in Online Education Tries a MOOC, Chronicle of Higher Education. Available: http://chronicle.com/article/A-Pioneer-in-Online-Education/134662/

Kaul, G. (2012) 'Introduction to finance', Coursera Chronicle of Higher Education, 59 (6) (2012), p. B8

Knowledge (LAK '14). ACM, New York, USA, pp. 83-92. DOI=10.1145/2567574.2567586.http://doi.acm.org/10.1145/2567574.2567586

Martin, F.G. (2012) Will massive open online courses change how we teach? Communications of the ACM, 55 (8) (2012), pp. 26–28

Rice, J. (2013) What I learned in MOOC College Composition and Communication, 64 (4) (2013), pp. 695–703

Tin Can API http://tincanapi.com/

Young, J.R. (2013). What professors can learn from 'hard core' MOOC students Chronicle of Higher Education, 59 (37) (2013), p. A4

xAPI: Experience API http://www.adlnet.gov/tla/experience-api/

7 Visualisations

7.1 Dashboards and visual analytics

A key aspect when applying learning analytics is the design of dashboards to display the KPIs (Key Performance Indicators) and other visual elements in a clear and intuitive way. In the context of organizations, Key Performance Indicators, also known as KPI, help an organization define and measure progress toward organizational goals. In the context of ECO project, once the learner has defined his goals, he needs a way to measure progress toward those goals. Key Performance Indicators are those measurements. Interesting trends can be followed up with traditional statistical analyses over the KPI values, but (Xu et al, 2014) argue that visualizations are a faster and more efficient way of "getting started".

Visual analytics is essential in application areas where large information spaces have to be processed and analysed. In particular, MOOCs offer many opportunities for visual analytics techniques. However, visual analytics are more effective when there is a clear purpose and when data can be visualized and communicated in way that is easily understandable. In an educational context, primary purpose of visual analytics is not to create complex graphs or pictures, but to communicate, present data in an effective way and aid decision-making, both to learners and to instructors.

Dashboards typically capture and visualize traces of learning activities, in order to promote awareness, reflection, and sense-making, and to enable learners to define goals and track progress toward these goals (Verbert, et al., 2014). Dashboards are a simple way to organize together and manage multiple charts that share the same underlying data.

Verbert et al (2014) present an analysis of the state of the art in the design of educational dashboards and try to answers to some interesting research questions: what are relevant user actions?; how can data on relevant actions be captured?; how are awareness and self-reflection enabled for different kinds of users through appropriate devices?; how can learning analytics dashboard applications be evaluated?; and, how can the impact of these visualizations on user behaviour be measured?. However, this study is focused on dashboards in traditional e-learning processes using learning management systems, and no reference to MOOCs is included. In the conclusions, they state that evaluation of learning analytics dashboards is often complex. Many researchers have performed usefulness and usability evaluations by asking teachers or learners to perform a set of tasks, however little is known about the usefulness of dashboards to solve real issues and needs of learners. These conclusions can equally be applied to dashboards in MOOCS, with the added difficulty of the existence of learners with profiles, interests and goals that can be very different.

If we want that learning analytics dashboards to be useful for learners and teachers in MOOCs, information needs to be communicated in a way that they may be able to easily see and analyse what is happening, knowing how the calculations have been performed. The basic idea of a dashboard is to enable learners to track their activities, in order to enable self-analysis and comparison with other users, often by aggregating traces into metrics, or by visualizing these activity traces. Comparison with other learners is necessary, however in MOOCs different profiles of learners are presented and each student has their own goals, so the comparison should be made with learners with similar profiles and goals. In Chapter 2 a user categorization based on frequency of visits, and activities performed is proposed. This classification defines four types of learners: completing, auditing, exploring and disengaging learners. Obviously, learner goals may change as the course advances, so it could be necessary to dynamically reclassify learners in the clusters. Once the clusters are identified, any comparative indicator to show the student or teacher would be by reference to learners who are classified in the same cluster.

In the context of ECO project, the designed dashboards have to take into account the following aspects:

- The metrics shouldn't be presented simply as a set of numbers; they could be described using charts which have clear meaning in the context of the learning activity.
- Dashboards should visualize the number of resources (posts, responses to questions, help requests,…) produced by learners in the course and allow both learners and teachers to compare these figures with those achieved by other participants.
- Social interaction graphs can help to identify user communities.
- Visualization of time-spent graphs can help teachers to identify potential students at risk. In addition, learners can compare their efforts with those of their peers.
- Test and self-assessment results that capture knowledge levels should also be used to get a better indication of learning progress.
- Dashboards for tablets and mobile devices should be developed.
- It would be very useful to use technologies that allow instructors and learners configure their own dashboards including different widgets.
- In the design of dashboards aspects of accessibility should be considered, taking as reference existing standards (WAI, W3C, …)
- Heuristic evaluation is a good method of identifying both major and minor problems with an interface; the proposed dashboards should be evaluated according to usability guidelines.

Regarding how each platform should visualise dashboards, we suggest that each learning platform adds a link in main page of each course which invokes the LA service and displays learner dashboard or teacher dashboard depending on the role with which the participant has connected. In chapter 3, some course indicators are mentioned. These should be located in the top left area of each module and course, directly or as well by means of a link which opens a page with this data.

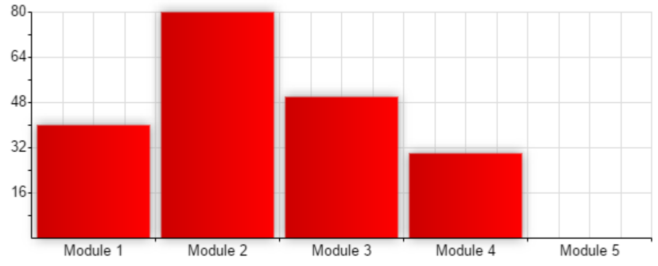
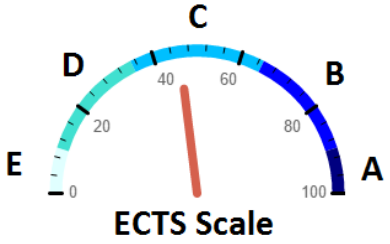
7.2 Learner dashboard

The learner dashboard should have a main panel with items to display key metrics in each of the two defined categories: academics and social issues. From this main panel the learner could have access to secondary panels showing additional metrics and graphs.

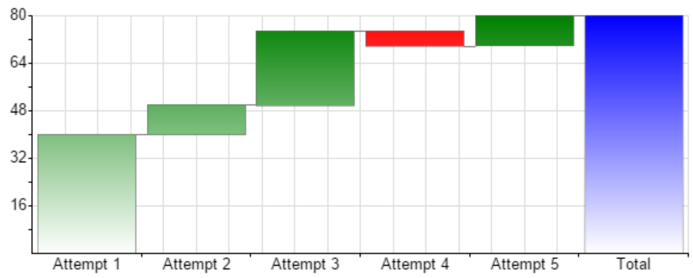
The following table shows the indicators defined in Chapter 3 and the type of control or visual element that could be used to show each value to the learner. Where appropriate indicators are shown for three groups:

- the whole group, all the learners enrolled;
- group of learners with the same ECO profile;
- group of learners with the same learning goals in the course.

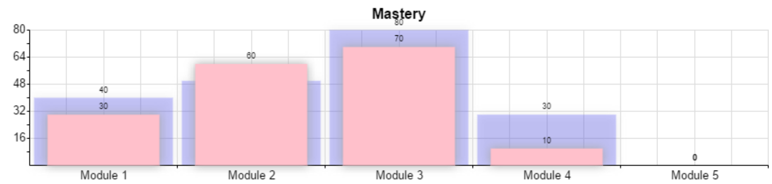
Table 30: Visualising indicators for learners

Indicators	Type of control or chart
Performance: personal grade in modules	We can use text boxes in order to show to the learner the global result in each module and histograms with the values of P_i  The y-axis represents the grading scale used in the module.
Performance: Final grade	The final grade P is shown to the learner, but in order to compare the individual performance with learners in the same group, the ECTS grading scale can be used. The ECTS grading system was defined by the European Commission to make grades more comparable to each other and it is based on the class percentile. The grade shows how a student performed compared to the other students in the same group. Before the evaluation, the results are divided into the two subgroups pass and fail. Therefore, the results are independent from the students who failed a course. The grading system is defined as follows: A: Best 10%, B: Next 25%, C: Next 30%, D: Next 25%, E: Next 10%. The ECTS global grade can be shown to the learner using a gauge control like this: 

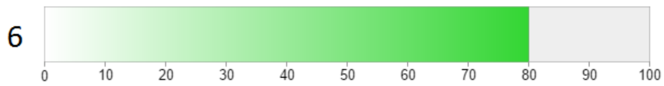
<table>
<tr><td>Mastery: knowledge gain</td><td>
For each assessable task, a waterfall chart can be used to show the improvement or decline in the different attempts. This figure shows an example of this type of graph:



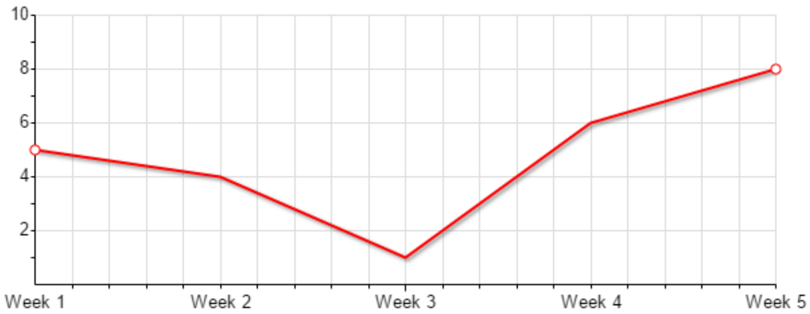
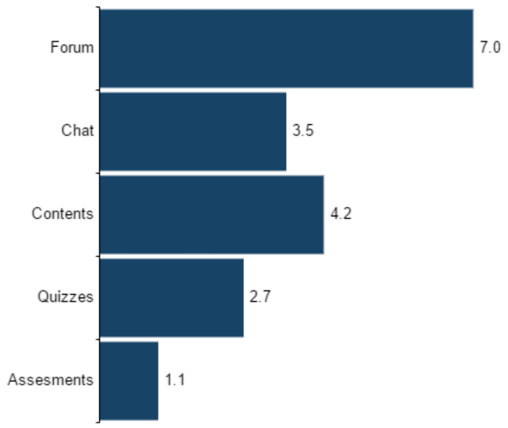
In order to obtain a visual representation of the mastery along the course, a bar chart with bars-in-bars can be useful:

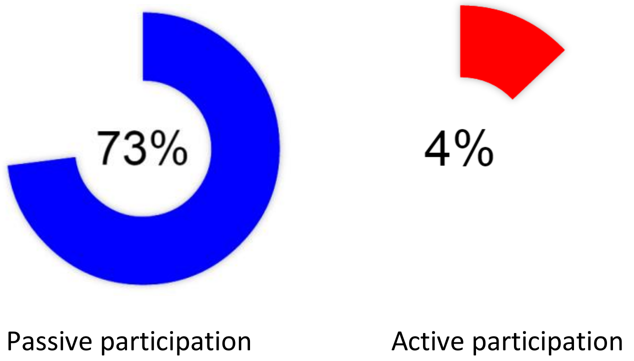
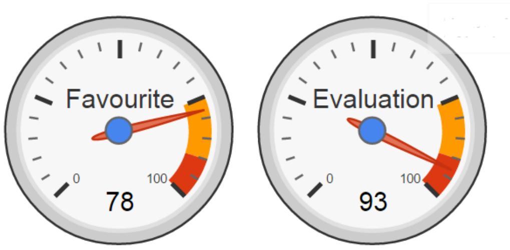


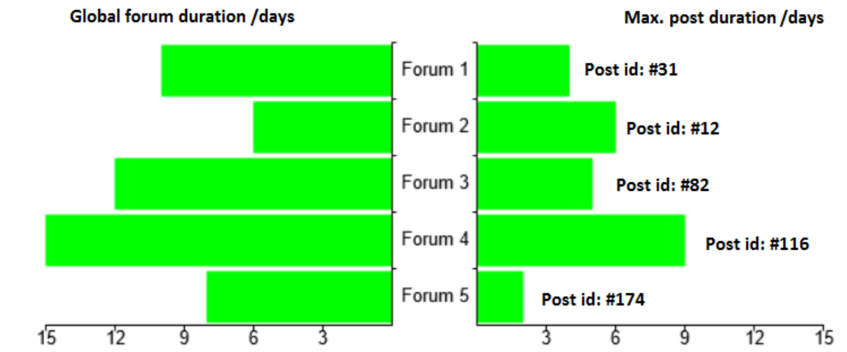
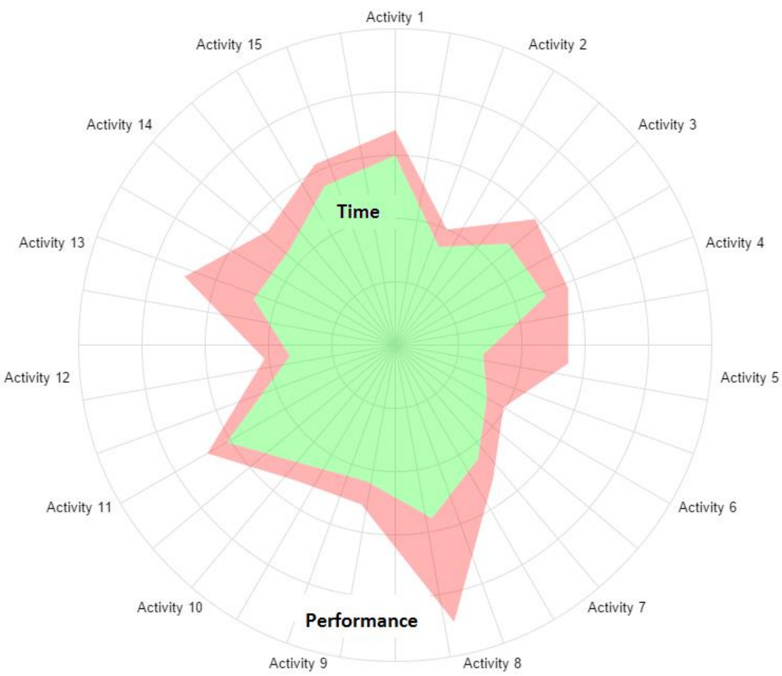
</td></tr>
<tr><td>Progress</td><td>
The progress can be shown using progress bar. In each module a number can indicate the number of tasks marked as done, and a percentage over the total number of tasks to be performed (mandatory and optional) is visualized in the progress bar:

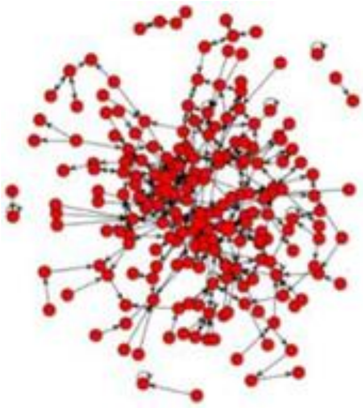
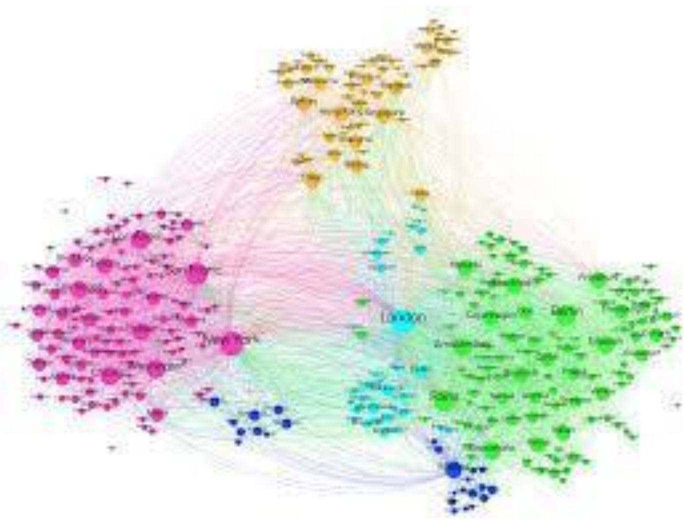
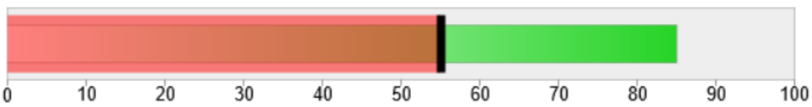


The percentage can be calculated by reference to the total number of scheduled tasks in the course or only the tasks related to the objectives stated by the learner.
</td></tr>
<tr><td>Engagement: access frequency</td><td>A basic line chart can show the number of days per week which learner accessed to the course.</td></tr>
</table>

	 Another alternative is to graph a two-line chart to compare the individual accesses with the group with similar goals.
Engagement: resource use	Horizontal bar charts can show the average of days per week which learner accessed to each type of resource in the course: 

Engagement: participation in communicative / collaborative tasks	The percentage of participation in each communicative activity (forum) can be visualized using two donut charts, the first one shows the percentage of messages read (passive participation), and the second the percentage of messages produced by the learner (active participation). In the second chart we can also distinguish post started and post answered. 
Engagement: ranking of participation in forum	The ranking according to messages marked as favourite or the evaluation received by other users can be visualized using gauge charts.  Another possibility is the use of progress bars. Position that each participant occupies with regard to the average valuation given to their messages can be shown using simple text boxes.
Engagement: post duration	The time difference (in days) between first and last post in the same conversation and in forum resource could be visualized a basic bipolar chart. In the left, the global forum duration is shown, and the right the duration of the largest post in the forum and the post identification are also shown.

	
Effort: time per resource and effort vs performance	The average time spent in each task can be shown using horizontal progress bar. The correlation between time per resource or activity and performance in the tasks or activities can be analysed using stacked radar chart like this: 
Satisfaction	Opinions of learners that performed the course previously can be displayed using boxplots as was proposed in Chapter 3. However, these indicators should be displayed at level course and they shouldn't be included in the learner's dashboard.
Social affiliation	In order to visualize interactions resulting from discussion forum post and replies, we propose the generation of graphs using tools as SNAPP. The network visualisations allow teachers to evaluate and

	identify user behavioural pattern learners and any stage of course progression. The following figure shows an example of interaction graph, each node represents a learner.  The dashboard should allow each learner to be identified in the graph network. SNA (Social Network Analysis) techniques can also be used to community detection in the network. In this case, each learner can identify the members of his community. 
Social recognition	The number of followers and contributors can be visualized using simple horizontal progress bars and in comparison with the user with the highest number of followers. 
Social responsibility	The four indicators defined in Chapter 3 can be shown using a set of progress bars. In the case of the number of peer-evaluation tasks

	carried out can be visualized using progress bars with colour codes indicating the completed, in progress and uninitiated evaluations.

7.3 Relevant information for teachers

The teacher dashboard will be organised differently but will contain very much the same visualisations that are shown to individual learners, but aggregated into several levels: all learners in the MOOC, learners with similar profiles, learners with similar learning objectives. Moreover, the dashboards should allow the teacher to access information about the type of learners identified.

A teacher needs to get a quick overview of general progress and performance and in particular be informed whether there is a group of learners that run the risk of not being able to complete or dropping out. Although the teacher will not deal with individual learners, the teacher could take corrective actions pertaining to the group of at risk learners, or decide to update some of the MOOC activities or content.

In addition the teacher needs information about the number of learners who have enrolled in the course, how many learners have actually started with the course, and how many learners are active at a given moment. Moreover, when visualising progress and performance or information about specific activities, tasks or resources, it should be made clear on what number of learners the indicator is based. Furthermore, indicators should be shown for the whole group of learners, for learners with the same profile, for learners with the same learning objective, as well as for each of the four types of learners that have been identified in chapter 2.

7.3.1 Performance and mastery visual elements

The teacher must be able to visualize using the dashboard the individual results of students in assessment activities, but also overall results and the percent of participants that have performed each task. Visual controls will be similar for mandatory or voluntary activities.

Different visual elements can be used in order to display percentage of completion. These charts have to allow to check the overall level of participation and whether some prefixed objectives have been achieved. These objectives can be fixed using information about previous editions of the course. The following figure shows two examples of controls that indicate the percentage of students who have completed two different tasks

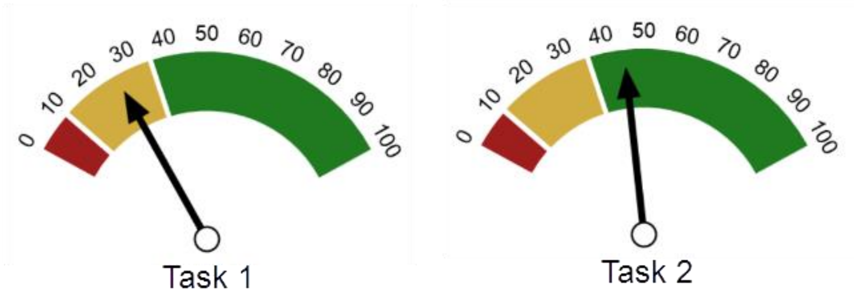


Figure 3: Comparing completion of various tasks

If the task is assessable then teacher's dashboard will show the grade distribution. Firstly, a stacked bar chart can display the number of learners that are in progress, passed or failed in the task. The following figure shows an example:

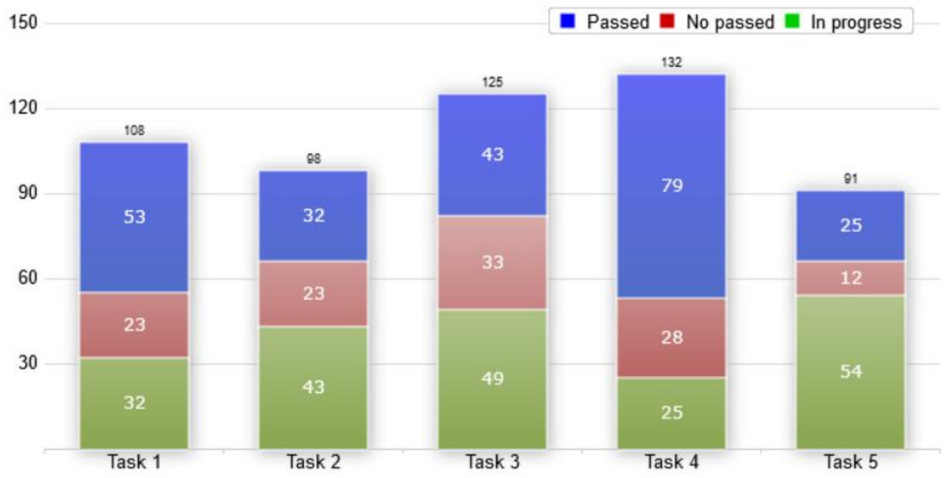


Figure 4: Visualising grade distribution for various tasks

Another bar chart can show the grade distribution in the course. The aggregation of final grades has to include the grades in the personalized scale used in the task, but also the ECTS scale. That is, the graph must inform to the teacher the numeric limit for each ECTS mark according to the percent distribution of passed learners.

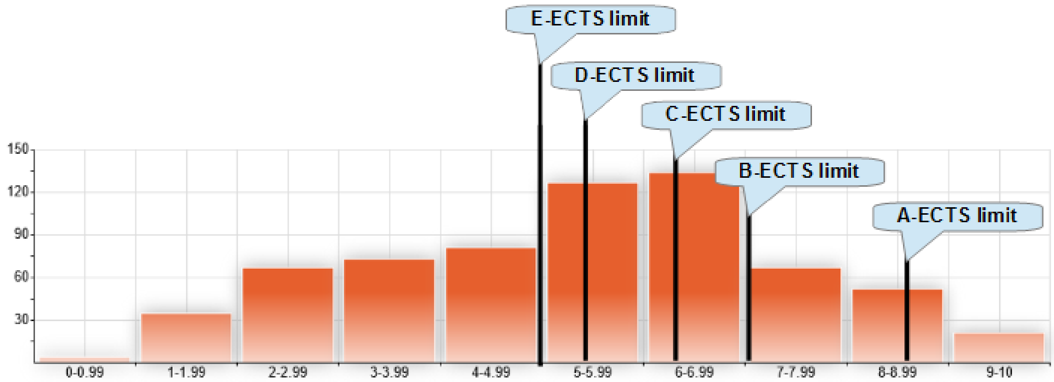


Figure 5: Relating grades to ECTS scale

The equivalence between both scales (personalized and ECTS) can also be shown using a rose chart.

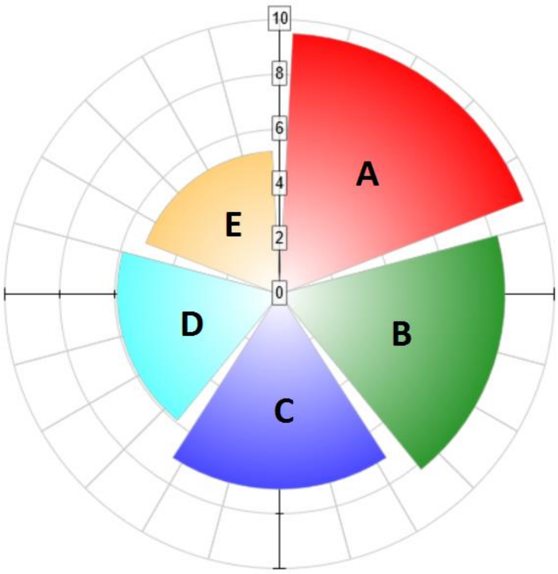


Figure 6: A rose chart for mapping grades to ECTS scale

7.3.2 Effort visual elements

The average time spent in an activity can be represented using two gauge charts. The first one represents average time spent among by learners that pass, and the second one average time spent by learners that fail.

Figure 7: Visualising time spent in a task according to completion status

Other interesting graphs can show the correlation between effort (time spent) and performance (grades). The following is an example of correlation graph:

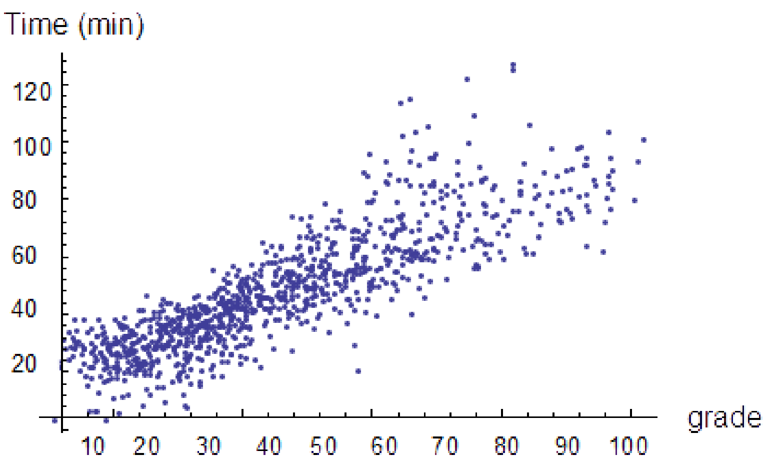


Figure 8: Correlating time spent with grade obtained

7.3.3 Engagement and activity level visual elements

In order to identify learners in risk of dropout or abandon, different activity indicators can be collected and aggregated in a single activity level index. In Cobo, et al. (2014) a multicriteria approach to evaluate and classify the level of interactivity of learners is proposed. In the context of ECO project, indicators as number of accesses to the platform, clicks, written messages, downloaded resources, accesses to each resource, etc. could be aggregated using a weighted average.

Once each learner has an activity level assigned, a plot of interactivity level can be generated, and a cluster algorithm can be used in order to obtain a classification of learners. In this manner groups of very active or collaborative, active, passive and inactive learners are automatically identified. This classification could be aligned with the categorization of completing, auditing, exploring and disengaging learners proposed in Chapter 2.

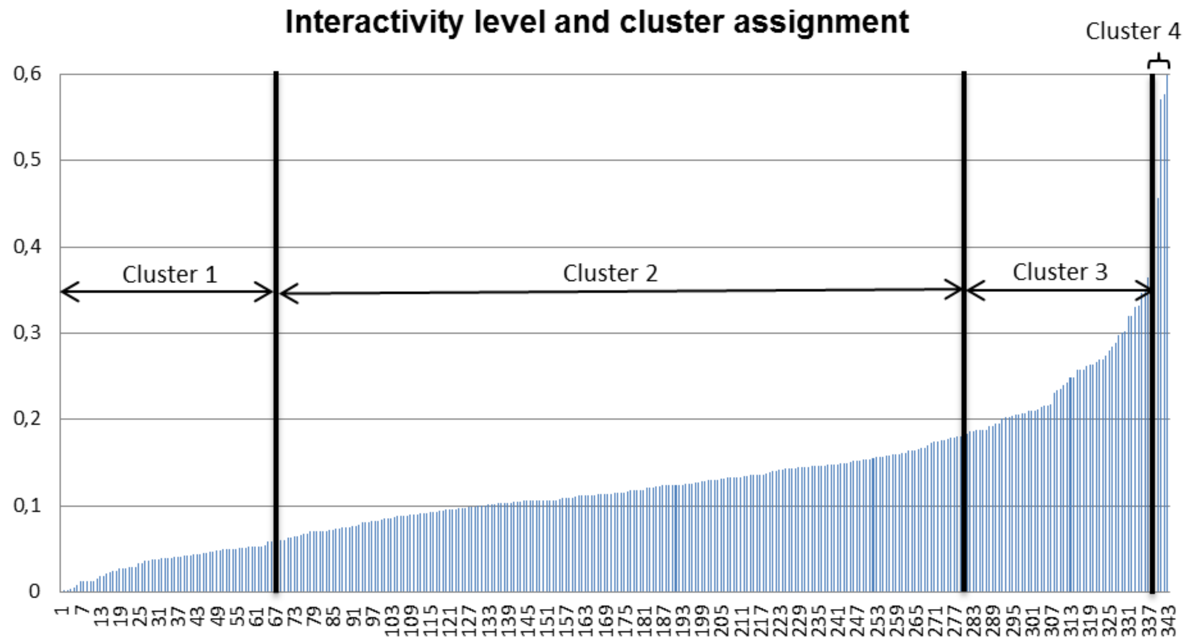


Figure 9: Classifying types of learners

7.4 Teachers and course designers dashboard

The teachers and course designers' dashboard should offer key metrics and graphics, as well as exactly figures through the charts which are developed at chapter 4 of Learning Analytics deliverable.

Teachers and course designers Analytics should have a main panel with items to display key metrics in each of the three defined categories: Academic issues, Course Design and Social indicators. From this main panel the teachers and course designers could have access to secondary panels showing additional metrics and graphs.

The table 31 shows the indicators defined in Chapter 4 and the type of control or visual element that could be used to show each value. Where appropriate indicators are shown for three groups of learners defined at chapter 2: Completing learners, auditing learners, exploring learners.

As well, charts and graphics should give teachers the possibility to define Learner Activity values in order to differentiate these groups:

- Very active learners.
- Active learners.
- Inactive learners.

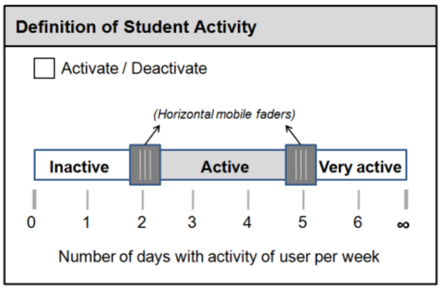


Figure 10: Defining learner activity

Learner Activity is defined by average days the participant logs into the platform. A graphical box titled "Definition of Student Activity" will allow teachers to define how many days a student should log into the platform in order to be considered "inactive", "active" or "very active". The teacher will be able to move horizontally two independent mobile faders, in order to define exactly this data. By default, it could be defined following the following values:

- Very active learners: > 5 days
- Active learners: 2< = x = > 5 days
- Inactive learners: > 5 days

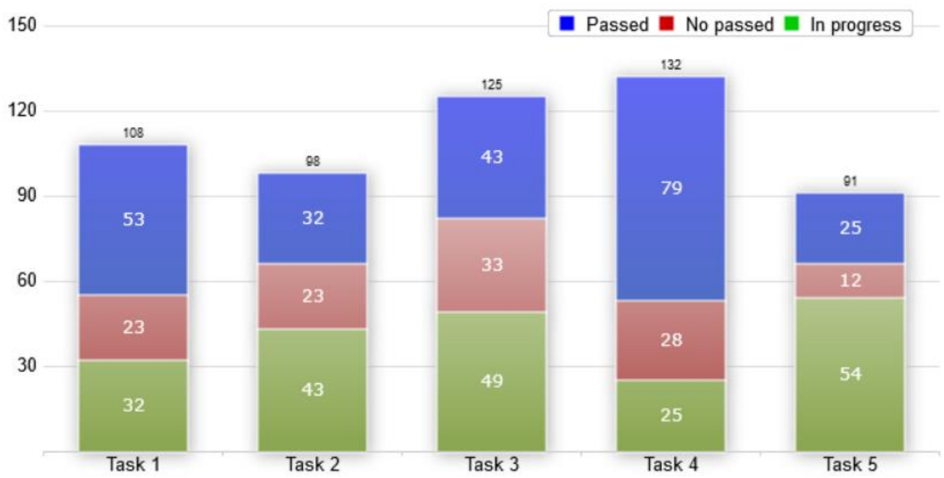
As we say, charts and graphics should give teachers the possibility to define Learner Activity values. Because of that, Learner Activity box should have a little box in order to be activated or deactivated, as it is described at Figure 10:

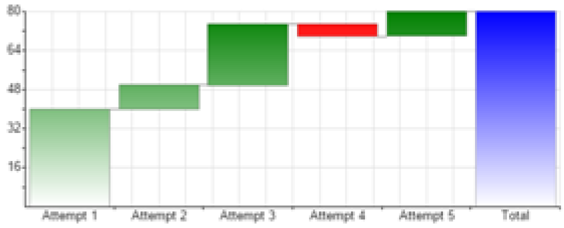
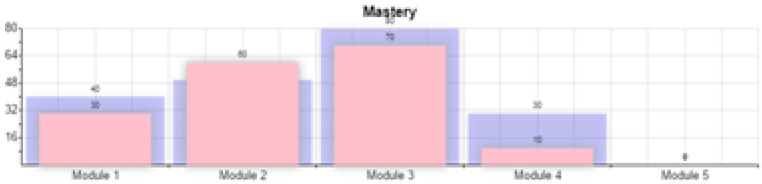
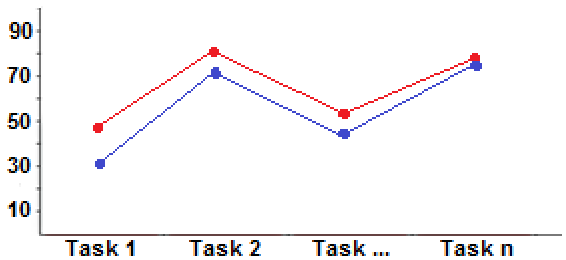
- If it is activated: graphics change, relying exclusively on Learner Activity values. Consequently, it will be necessary that the data would be divided in three different graphics (one per each group: "inactive", "active", "very active"), either only one graphic ensembles with three different coloured data per each indicator.
- If it is deactivated: graphics show whole course figures, without relying on Learner Activity values

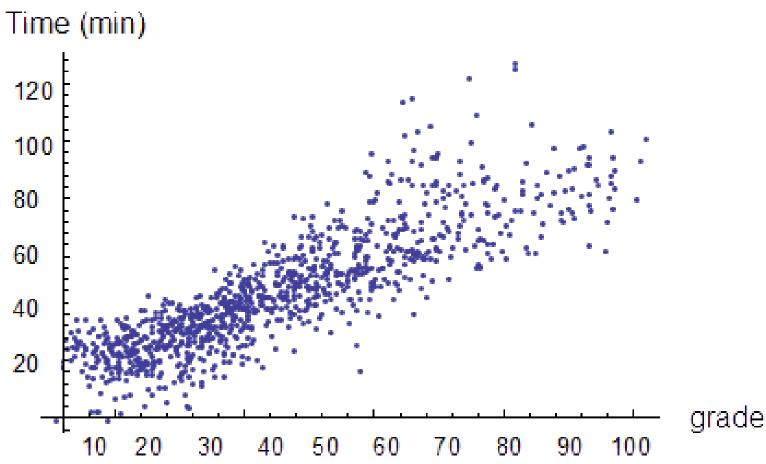
It is important to underline that the "Definition of Student Activity" is an extra, a complement to the participants' classification available at chapter 2 ("completing learners", "auditing learners", "exploring learners").

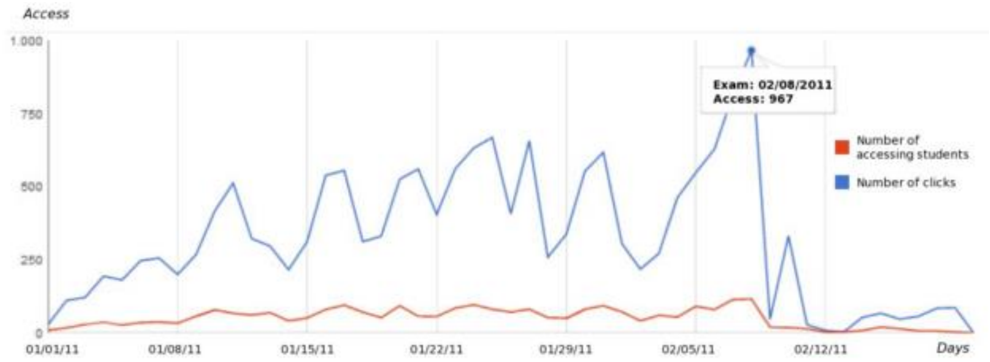
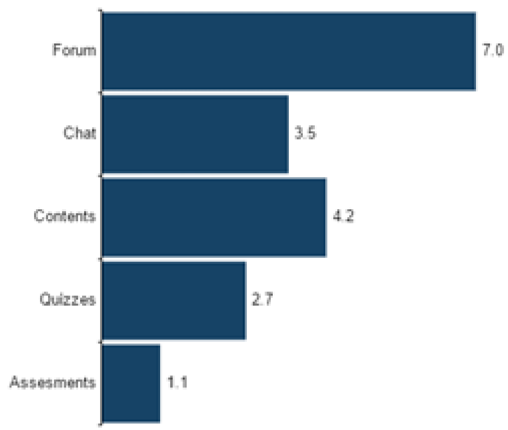
The difference between these two categorizations is that the "Definition of Student Activity" relies exclusively on the number of days a student logs into the platform and that, on the contrary, the second categorization relies on logins, readings, assessments, and other profile matters, just as it is explained at Chapter 2 of this deliverable. Moreover, the numbers which determine how "Definition of Student Activity" works are configurable by the teacher, as opposed of the other categorization, which relies on indicators previously defined, as it is indicated at Chapter 2.

Table 31: Types of control for indicators

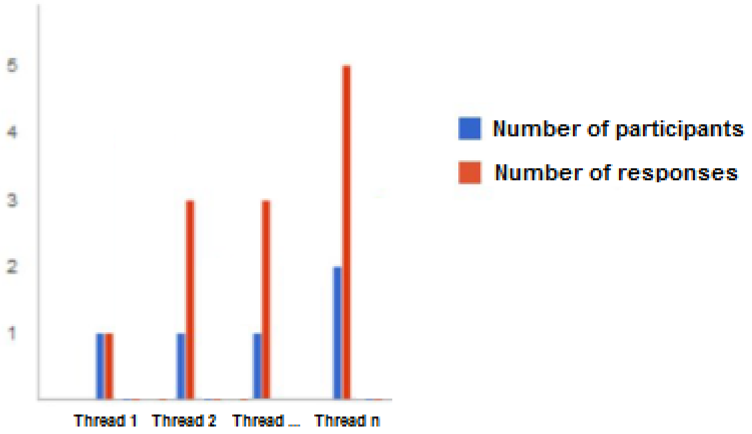
Indicators	Type of control or chart
Performance by mandatory tasks	We can use text boxes: ● Number of passes ● Number of fails ● In progress  The y-axis represents the grading scale. With floating numbers we represent the % participants performed the task.
Mastery: knowledge gain	We can use the following multiple text boxes. ·Number of passes and Number of fails and Number of participants who have needed or decided doing the task more than once (three bars per task). Floating numbers will indicate the % participants performed the task A waterfall chart can be used to show the improvement or decline in the different attempts. This figure shows an example of this type of graph:

	 In order to obtain a visual representation of the global mastery along the course, a bar chart with bars-in-bars can be useful.  (Not with modules, but with tasks)
Effort: average time spent among passes and average time spent among fails	We can use a line chart. Blue line represents time invested in passing a task. Red one, time invested among people who have failed. We always take about average figures. With floating numbers, above coloured points, we represent the average number of attempts and % participants performed the task.  The y-axis represents minutes. Another way to represent it is as follows.

	 The average time spent in an activity can be represented using two gauge charts. The first one represents average time spent by learners that passed, and the second one average time spent by learners who failed. Other interesting graphs can be show the correlation between effort (time spent) and performance (grades). The following is an example of correlation graph: 
Engagement: accesses to the platform	A basic timeline chart can show Number of accessing participants and Number of clicks, allowing teacher to zoom in and out over a day, over a week and global.

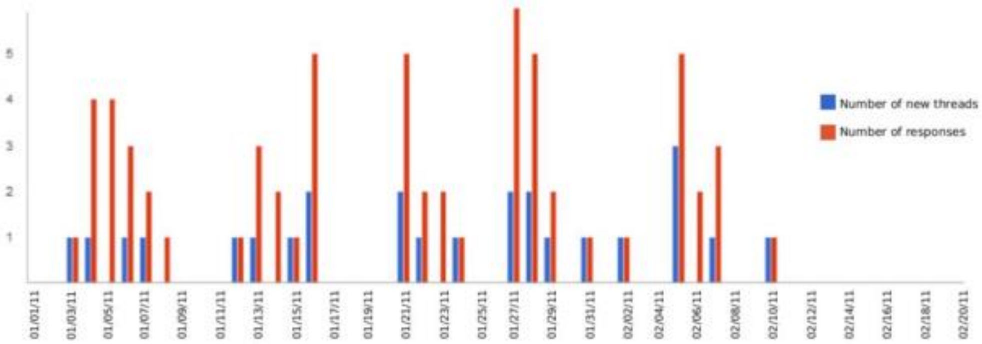
	 Another alternative is to graph a two-line chart to compare the individual accesses with the groups with similar goals. Another basic timeline chart, differentiating participants who pass and participants who fail with blue and red lines, will include the "Average of accesses" and "Average of clicks". It will be possible to differentiate groups defined at chapter 2.
Engagement: Forums, social networks, wikis and blogs participation	Horizontal bar charts can show the average of days per week which participants accessed to each type of resource in the course: 

About forums



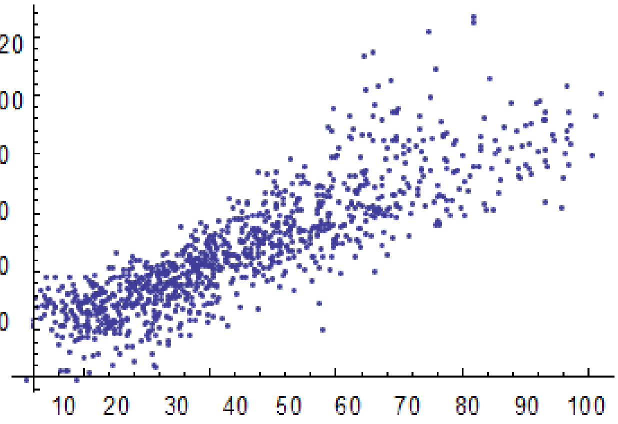
With floating numbers we represent also "Average of messages among participants who participated"

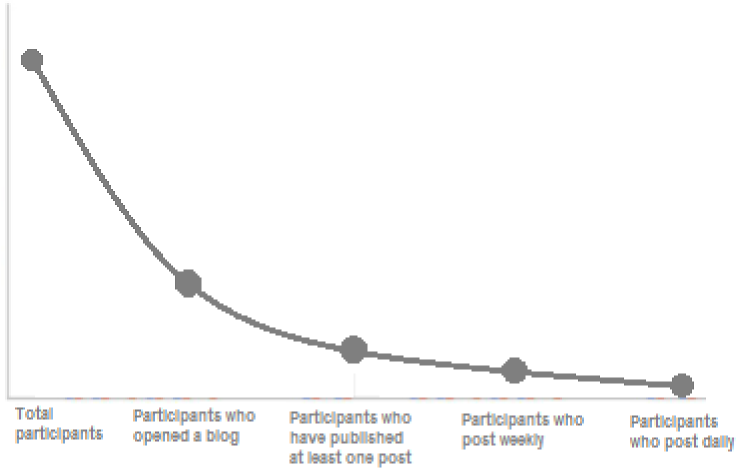
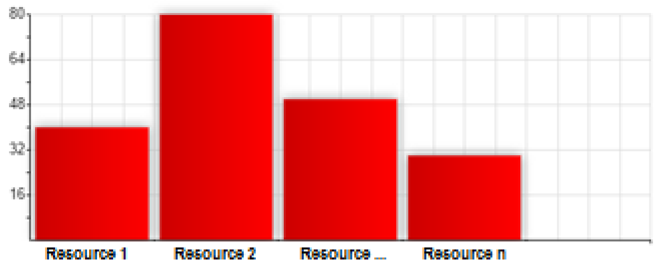
About forums, it will be also useful a timeline with opened threads and responses.

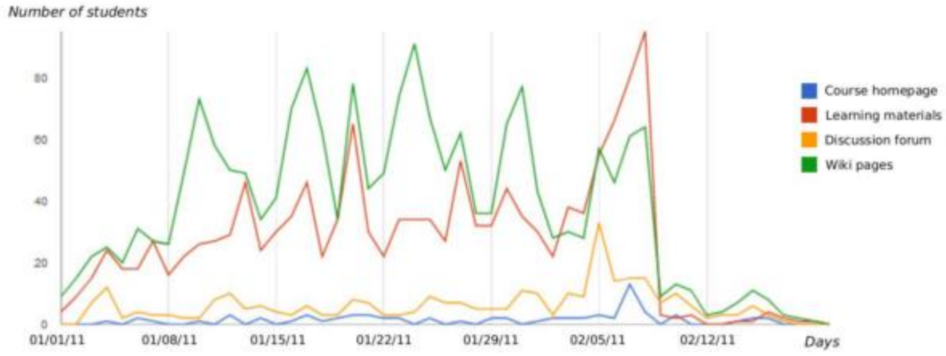
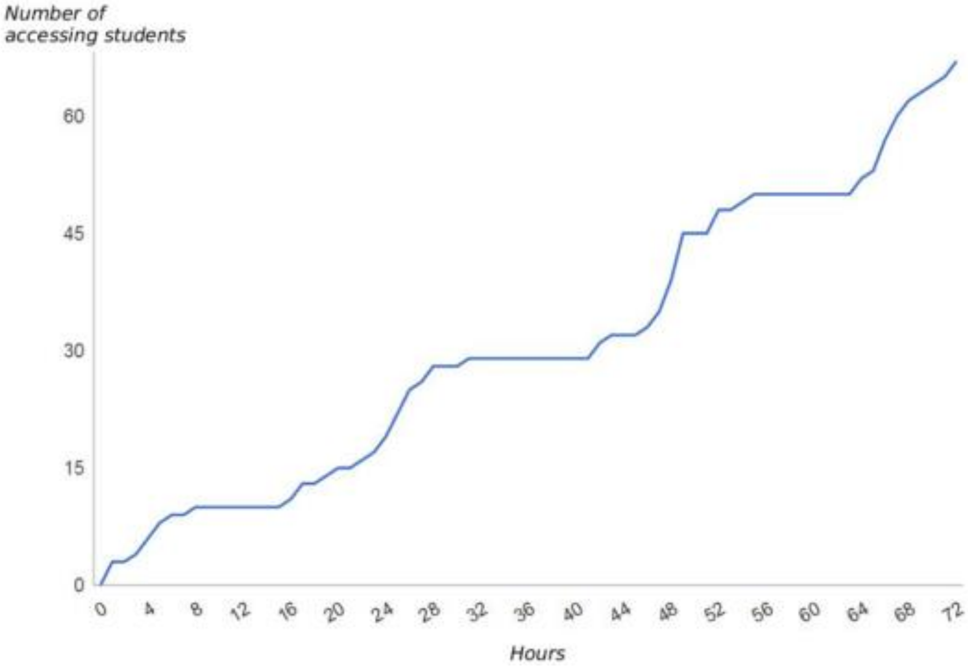


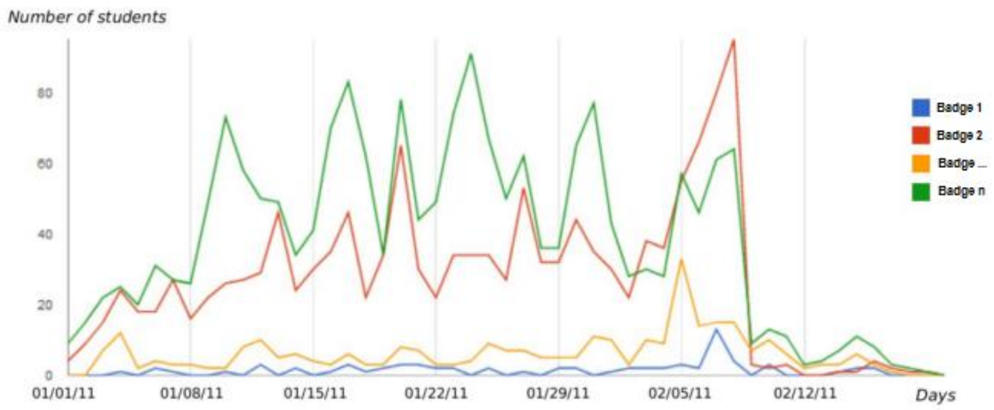
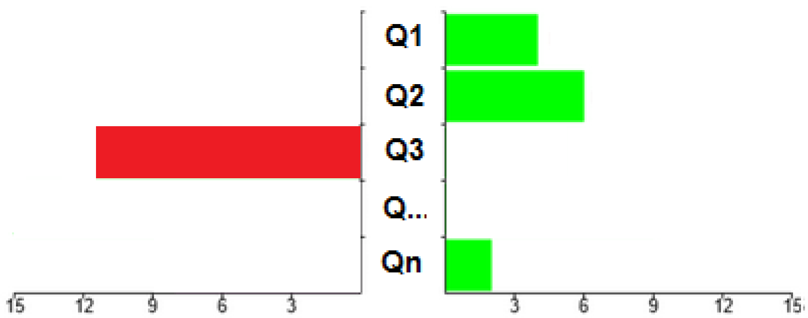
About wikis, we can use the same chart as forums, but exchanging "Number of responses" for "Number of characters". With floating numbers we will represent the "Average of characters written by each participant".

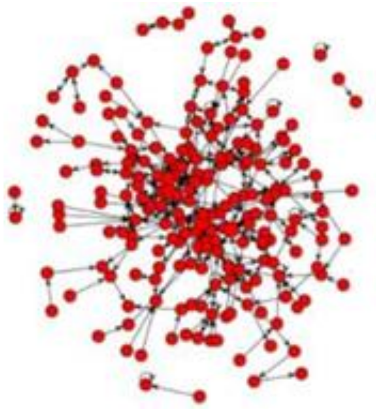
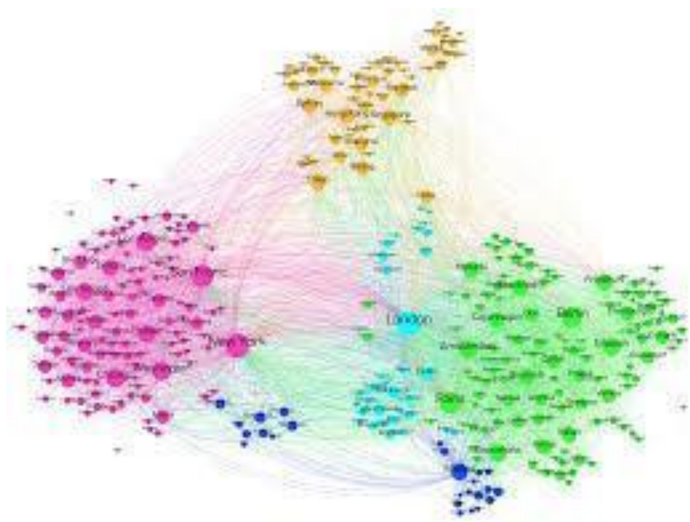
About forums and wikis, we can also use graphs, as the following one, in order to represent the characters written by each participant and the number of

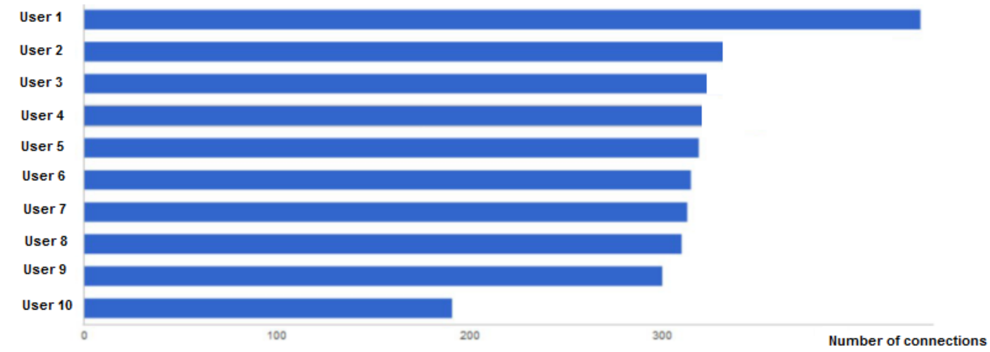
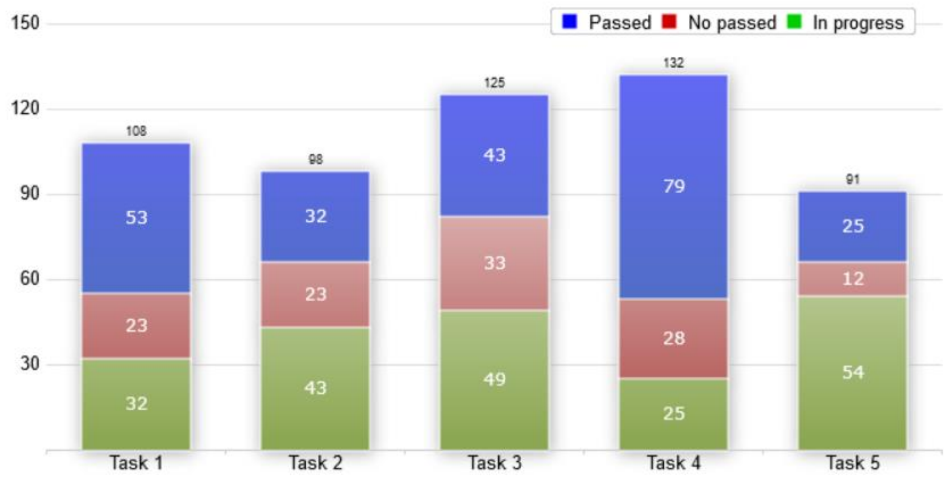
	contributions (for wiki, we assume a contribution as every time participant logs into a wiki and contributes; for forums, we assume a contribution as a post, either a new thread or a response).  The y-axis represents characters written by a participant. The x-asis represents numbers of contributions by a participant. About blogs, we use a line chart, as follows. With floating numbers we will represent the exact figures.

	
Resources consumption	We can use text boxes, with a bar per each resource.  The y-axis represents the number of downloads.
Courseware: activity area	A timeline can be useful to represent Activity area, which shows how participants move on in the platform

	
Courseware: adoption rate	A timeline is necessary because adoption rate deals with the time period from uploading a selected learning material to the time of access by students. In case of MOOCs, all learning materials are uploaded from the beginning of course. So, we will know which materials are first consumed.  We will need one chart per resource.
Badges	A timeline, with floating numbers which will represent the exact figures of achievements. It will be necessary to allow teacher to select only the badges he or she want to visualise.

	 In addition, it will be recommendable to use badges logos in the legend.
Satisfaction	Satisfaction measures the average of participants' opinion. We use a double bar chart. We assume that participants' opinion is positive (green) when it is higher than the half and negative when it is lower than the half. Consequently, it is impossible to have a green bar and a red bar in the same question. If the average is just in the middle, we don't need to draw any bar. 
Social affiliation	In order to visualize interactions resulting from discussion forum post and replies, we propose the generation of graphs using tools as SNAPP. The network visualisations allow teachers to evaluate and identify user behavioural pattern learners and any stage of course progression. The following figure shows an

	example of interaction graph, each node represents a learner.  SNA (Social Network Analysis) techniques can also be used to community detection in the network. 
Social recognition and leadership	A horizontal bar chart will be useful to know the 10 most connected participants.

	 The same with: ● Top 10 contributors with favourite marks ● Top 10 most well marked participants ● Top 10 participants with more contributions
Social responsibility and plurality: forums, social networks and wikis.	A stacked radar chart will be necessary, one per each thread in forum, in order to visualise better the data. The same mechanism will be use regarding forums, social networks and wikis.
Social responsibility and plurality: peer evaluation	A stacked bar chart can display the number of: ● Participants who have passed the task ● Participants who have failed the task ● Participants who are waiting for being evaluated  About "Distributions mark" and "Average time between uploading the task and the complete evaluation by peers", we will use boxplots.

<table>
<tr><td>Logging: way of access</td><td>We can use a circle chart.

With floating numbers we will represent the number and % of participants.</td></tr>
<tr><td>Logging: device of access</td><td>As well as “Logging: way of access”, we can use a circle chart.</td></tr>
</table>

7.5 Other visualisations

Many of the visualisation and LA indicators are valuable to the course designer as well, because it provide information on how to improve the course.

State transition diagrams can be used to analyse the learners' engagement with the course material. An example of use of these visual elements can be found in (Coffrin, et al 2014). They use state transition diagrams to represent how a system moves from one state to another state over a sequence of events. These diagrams are typically visualized as a graph where nodes represent the states and the lines connecting nodes reflect probabilistic or weighted transitions between states. For instance, the next figures show state transition diagrams associated to videos grouped by conceptual themes or topics. It is typical in MOOCs to split lectures into short micro-lectures and each one can represent a state in the diagram.

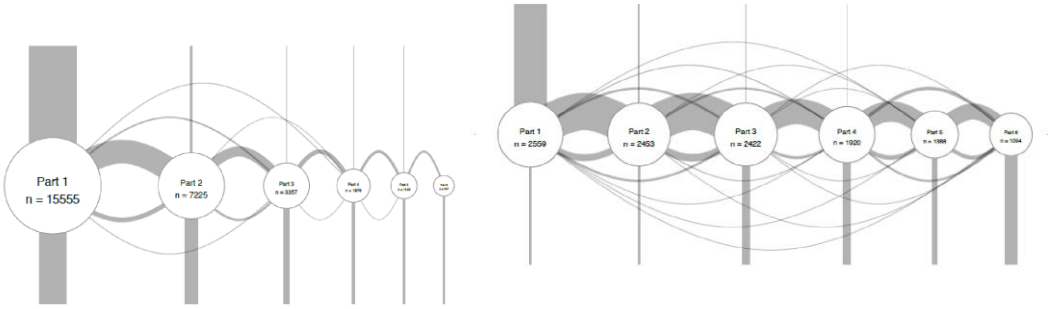


Figure 11: State transition diagrams showing different patterns of engagement for learners that passed or failed

The diagrams above (Coffrin, et al 2014) show learner video viewing transitions broken by two subgroups: non-qualified and qualified learners. This visualization clearly shows how different types of users show different patterns of transition between key MOOC resources and assessment activities.

7.6 Tools

Several resources and tools can be used in the construction of these graphical elements to be included in learning analytic dashboards. *Rgraph* is a library for generating interactive charts using JavaScript and HTML (http://www.rgraph.net).

Google offers a set of chart tools that are powerful, simple to use, and free. A gallery of interactive charts and data tools is accessible in the webhttps://developers.google.com/chart/. Using the APIs from Google all the charts that are part of a dashboard can be integrated.

Another example is *wxWidgets*, a powerful open source framework for developing native cross-platform GUI applications (http://docs.wxwidgets.org/trunk/index.html)

7.7 References

Cobo, A.; Rocha, R.; Rodríguez-Hoyos, C. (2014). Evaluation of the interactivity of students in virtual learning environments using a multicriteria approach and data mining. Behaviour & Information Technology, 33(10), pp 1000-1012. DOI: 10.1080/0144929X.2013.853838

Coffrin, C.; Corrin, L.; de Barba, P.; Kennedy, G. (2014). Visualizing patterns of student engagement and performance in MOOCs. In Proceedings of the Fourth International

Verbert, K.; Govaerts, S.; Duval, E.; Santos, J.L.; Van Assche, F.; Parra, G.; Klerkx, J. (2014). Learning dashboards: an overview and future research opportunities. Personal and Ubiquitous Computing. Volume 18, Issue 6, pp 1499-1514.

8 Data for learning analytics indicators

Learning analytics of course relies on data. In order to increasing the learner and teacher success and performance the learning environment has to track, record and log data that can be used as input for calculations and visualisations. The results are then shown to learners, possibly groups or types of learners, and teachers in the learning analytics dashboards. The indicators in the dashboard should give learners insight into their progress and performance and assist them to get the most out of the MOOC according to their personal learning objectives. Teachers can use the indicators to monitor progress and performance in general, discover whether there are learners at risk of dropping out and decide whether they need to take corrective actions. Other analytics are used to assist in evaluation of course design, and its effectiveness and efficiency. This data is not only relevant to the course designers and teachers, but also to management and institution.

There are several sources of data and several types of data that can be traced and recorded. That largely depends on the learning design of the course, the features the learning environment offers and the way the course is designed to make use of these features. Furthermore there are several ways to persist these tracking and logging data, but to be able to make use of these data it pays to standardize data formats. Several metadata and paradata formats are available, in addition to specifications developed to record activities and events. A brief description will be given below.

Although learning environments usually log some access and behaviour data, it is imperative to ensure that the data that we require for meaningful learning analytics are tracked, traced, logged and recorded. Most of these data are coming from the behaviour and actions of the learner while participating in the MOOC and making use of the features of the platform. To prevent a cold-start problem, some data are gathered through a questionnaire put to learners at registration and enrolment, prior to starting the course, mainly to gather insights into intentions and learning goals. Some qualitative data are gathered as well both in the entry questionnaire as through an evaluation questionnaire upon completion or end date of the course.

Although theoretically learner data could be gathered from applications external to the ECO platforms, there are legal and privacy issues to consider, because learners might not be aware that ECO would be gathering these data. Moreover, it is very likely that users generate a lot of data on external applications that are not related to their participation in the MOOC. Until these issues are resolved, ECO will refrain from including this kind of external data.

In general ECO has to treat any data as confidential and handle data with due diligence. The learning analytics service has to ensure that no identifiable data are made available to anybody other than the owner and anonymise data for indicators. Some ethical concerns and issues related to privacy of data are mentioned in the next chapter.

8.1 Gathering additional information through questionnaires

It is important to provide meaningful learning analytics that suit the intentions and learning goals of the learner. Because type of learner is not yet known at the start of the course, the best approximation would be to ask the learner about it. Consequently, the following types of questions should be put to the learner.

8.1.1 At registration

At the time of registration in the ECO portal demographic data should be asked. These should be stored into the user profile where the learner always can update these again.

- demographics: age, gender, employment status, highest level of education achieved, years of working experience, geographic location
- previous experience with MOOCs
    - first time registering for MOOCs: yes/no
    - how many MOOCs registered
    - how many MOOCs completed

8.1.2 At enrolment

Whenever a learner enrols for a particular MOOC, an entry questionnaire should enquire about:

- intentions for enrolling (allowing multiple answers)
    - -enhance my resume for career or college advancement
    - -it is free
    - -it is fun and challenges
    - -obtain credits
    - -interested in the topic
    - -experiencing what a MOOC is about
    - -to collect information and resources
    - -to have contact with people with the same interests
    - -to learn a specific topic
- intentions for use (only one option allowed)
    - -view all lectures, do all activities, submit all assignments and assessment, including self-assessment
    - -going for certification
    - -view lectures, do some activities, do not submit assignments/assessments
    - -select several topics to read or download
- learning goals:
    - -selection of learning goals as set by course designer
- main intention for registering previous MOOCs
    - enhance my resume for career or college advancement

- it is free
- it is fun and challenges
- obtain credits
- interested in the topic
- experiencing what a MOOC is about

-main reason for not completing:

- topic not interesting
- topic not relevant
- could not combine with other activities
- course too difficult
- course too easy

8.1.3 At end of MOOC or upon completion

The post questionnaire should be developed further in conjunction with WP4 because it might contain further questions that are relevant to evaluation. The questionnaire should be sent to all participants, not just those that complete the MOOC.

- Overall experience with the course: 5-point likert scale from very poor to excellent
- Did you complete the course
- Reasons for quitting /not completing a MOOC (conditional question only to be shown to those that did not complete)
  - -Topic did not meet expectations
  - -Materials did not meet expectations
  - -Perception of low quality delivered
  - -Wasn't proficient enough in the MOOCs language(s)
  - -Course was too easy/didn't add to my knowledge
  - -Course was too difficult/didn't have enough background knowledge
  - -Too much time required
  - -Content did not match presentation
  - -Never intended to complete it/when found what I needed, I quit
  - -Never intended to complete it/did it just out of curiosity
- verify intentions for use
  - -view all lectures, do all activities, submit all assignments and assessment, including self-assessment
  - -going for certification
  - -view lectures, do some activities, do not submit assignments/assessments
  - -select several topics to read or download
- actual intentions reached
  - -view all lectures, do all activities, submit all assignments and assessment, including self-assessment
  - -going for certification

-view lectures, do some activities, do not submit assignments/assessments
-select several topics to read or download
- learning goals actually achieved
  -selection of learning goals as set by course designer
- The effort required measured in hours per week.
- Workload for learner measured as 5-point likert scale from low to very high
- Level of difficulty to follow the course measured as 5-point likert scale from very low to very high.
- Evaluation of pedagogical aspects measured as 5-point likert scale from very poor to excellent.
- Adequacy of the activities and resources for learning measured as 5-point likert scale from very poor to excellent.
- Course usefulness measured as 5-point likert scale from very poor to excellent.
- Evaluation of results in accordance with the personal goals of the learner measured as 5-point likert scale from very poor to excellent.
- Degree of overall satisfaction with the course measured as 5-point likert scale from very poor to excellent.
- Aspects to be improved in the course (free text)

8.2 Gathering data for indicators

In order to provide the learning analytics for the indicators as presented in this deliverable, data need to be gathered to determine answers to a set of questions. Data gathered from individual learners need to be aggregated to arrive at corresponding metrics for the indicators that are described in chapters 3 and 4.

This section lists the raw data and the questions that are needed to calculate the metrics.

8.2.1 Whole population, all MOOCs

From the whole population, over all MOOCs, data which must be gathered to answer the following questions is:

- No of participants registered
  - Userid and date-time of registration at the portal
- No of learners actually logged in, once, and more than once.
  - Userid, MOOCid, date-time of first login
  - Userid, MOOCId, date-time of last login
- No of learners enrolled in all MOOCs
  - Userid, date-time of enrolment, MOOCid
  - MOOCid
  - Number of MOOCs in portal
  - Number of MOOCs for which at least one learner enrolled
- No of learners enrolled in each MOOC
  - Userid, date-time of enrolment, MOOCid

- No of learners actually started by accessing at least one learning activity.
  - Userid, date-time first access, MOOCid, pageid/objectid (page can be MOOC, lesson, module, unit, resource)
  - Userid, MOOCid, date-time access, pageid/objectid (page can be MOOC, lesson, module, unit, resource)
- Average number of enrolments per MOOC
  - Userid, date-time of enrolment, MOOCid
  - Number of learners enrolment in each MOOC divided by total number of MOOCs in portal
  - Number of learners enrolment in each MOOC divided by number of MOOCs for which at least one learner has enrolled.
- Average number of learners per MOOC who logged-in once.
  - Userid, MOOCid, date-time of first login
  - Number of MOOCs for which at least one learner has enrolled
- Average number of learners per MOOC who logged-in more than once.
  - Userid, MOOCid, date-time of first login
  - Userid, MOOCid, date-time of last login
  - Number of MOOCs for which at least one learner has enrolled
- Average number of active learner per MOOC. Active defined as actually having accessed at least 5% of the learning material.
  - Userid, MOOCid, date-time access, pageid/objectid (page can be MOOC, lesson, module, unit, resource)
  - Number of MOOCs in portal
  - Number of MOOCs for which at least one learner has enrolled
- Average number of learners who completed MOOC
  - Userid, MOOCid, date-time completion, score completion
  - Number of MOOCs for which at least one learner has enrolled
  - Userid, MOOCid, date-time certification
- Percentage of registered participants who become active learners, over all MOOCs and per MOOC
  - Userid, date of registration
  - Userid, date of enrolment, MOOCid
  - Userid, date of access, MOOCid, pageid/objectid
  - Userid, MOOCid,access at least 5% of total learning material
  - Total number of registered participants
  - Total number of participants in each MOOC
  - Total number of active learners across all MOOCs
  - Total number of active learner in each MOOC
  - Total number of MOOCs in the portal
  - Total number of MOOCs for which learners have registered
  - Total number of MOOCs for which learners actually started
- Percentage of logged in learners who become active learners
  - Userid, date of registration
  - Userid, date of enrolment, MOOCid

  - Userid, date of first login, MOOCid
  - Userid, date of access, MOOCid, pageid/objectid
  - Userid, MOOCid, access at least 5% of total learning material
  - Total number of registered participants
  - Total number of participants in each MOOC
  - Total number of active learners across all MOOCs
  - Total number of active learner in each MOOC
  - Total number of MOOCs in the portal
  - Total number of MOOCs for which learners have registered
  - Total number of MOOCs for which learners actually started
- Percentage of logged in learners who completed for all participants and for category 'completing learners'.
  - Userid, date of registration
  - Userid, date of enrolment, MOOCid
  - Userid, date of first login, MOOCid
  - Userid, MOOCid, date-time completion, score completion
  - Number of MOOCs for which at least one learner has enrolled
  - Userid, MOOCid, date-time certification
  - Total number of registered participants
  - Total number of participants in each MOOC
  - Total number of active learners across all MOOCs
  - Total number of active learner in each MOOC
  - Total number of MOOCs in the portal
  - Total number of MOOCs for which learners have registered
  - Total number of MOOCs for which learners actually started
- Number of learners participating in forums (also per type of learner):

  Merely accessing at least 1 thread
  number of threads accessed
  number of replies to thread
  number of posts (post is first message in thread)
  number of messages in thread, thread-length
  - Userid, MOOCid, forumId, threadid, date-time, postid, content
  - Userid, MOOCid, forumid, threadid, date-time, type (post/reply), position in thread (for replies also postid to which it refers)
  - MOOCid, forumid, threadid, postid
  - Number of forums
  - Number of threads in each forum and over all forums
  - Number of messages in thread in each forum and over all forums

    MOOCid, forumid, threadid, thread length: number of messages in thread

8.2.2 Whole population, individual MOOCs

From the whole population, for individual MOOCs. Several of these queries require the same data as mentioned above. These data allow several questions to be answered, for example to analyse collaborative actions and discussion.

- No of learners registered
- No of learners enrolled in the individual MOOC
- No of learners actually logged in, once, and more than once.
- No of active learners actually started by accessing at least one learning activity.
- No of learning activities accessed
- Timeline accessing learning activities and resources
    - Userid, date-time access, pageid/objectid (page can be MOOC, lesson, module, unit, resource)
- No of quizzes submitted
    - Userid, MOOCid, quizid, grading-scale, mandatory/optional
- No of quiz attempts per quiz
    - Userid, MOOCid, quizid, attemptid
- No of quizzes submitted successfully
    - Userid, MOOCid, quizid, attemptid, grade/mark, score > cut-off set by designer
- No of peer assessments
    - Userid, MOOCid, peerassessmentid, peerid, grade/mark, grading scale, score > cut-off set by designer

- Number of learners participating in forum
- Number of threads, average number of messages per thread.
- Number of learners, and percentage of active learners, participating in threads
- Number of learners, and percentage of active learners, initiating threads.
- Number of learners, and percentage of active learners, replying to threads.
- Number of threads learners contribute to.
- Number of threads posters initiate
- Number of threads replies respond to.
- Who is initiating?
- Who is replying?
- How many threads (and percentage) get replies? How many replies?
- How many threads (and percentage) get answered?
- How many threads (and percentage) do not get answered, despite replies.
- How many threads (and percentage) do not receive a reply.
- Number of published messages (per forum over all threads, over all forums in the MOOC)
- Number of replies (per forum over all threads, over all forums in the MOOC)
- Number of initiated conversations (per forum over all threads, over all forums in the MOOC)

- Number of initiated conversations without replies (per forum over all threads, over all forums in the MOOC)
- Number of conversations where the learner has posted (initial or reply) a message (per forum over all threads, over all forums in the MOOC)
- Number of forums where the learner has posted a message

- Number of contacts added to contact list
    - Userid, Contactid, type of contact
- Number of followers
    - Userid, Followid, type of follower (following, being followed)

8.2.3 Individual learners

For individual learners the following data should be stored. Some of these data will be gathered through the questionnaires.

- Intentions for use
- Individual learning goals
- Registration date-time
- First logged in date-time
- Date-time of logging in (every time)
- Last logged in date-time

- Page accessed, date-time, pageid, type of page, objectid, type of object. Page can be course, module, learning activity, learning resource, task, assignment, assessment, social media, etc. A page can contain multiple objects, e.g. displaying activity instruction, resources plus assessment.

- Most recent page accessed, date-time, pageid, type of page, objectid, type of object

- Date-time accessing a particular learning activity
- Date-time completing a particular learning activity
- For each quiz/assignment/assessment

Date-time assessing a quiz
Date-time saving a quiz
Date-time submitting a quiz

Number of quiz attempts, recorded by date.
Time between quiz attempts
Time between quizzes
Score obtained/Grade achieved, grading-scale

- Whether contributing in forum

  Date-time first accessing a forum thread
  How often and which forum threads are accessed
  Date-time of accessing forum threads
  Date-time of reply to a particular forum thread
  Date-time of posting a new thread
  Number of threads participated in
  Date-time and rating/liking score for contribution

- Posting a blog

  Personal blog
  Course blog
  Date-time of post, content, learnerid, blogid, postid, type (personal/course)

- Adding to a wiki/Editing a wiki page

  Personal wiki
  Course wiki
  Date-time of edit, content, learner_id, wiki_id, page_id, type (personal/course)

- Social activities

  Viewing/liking/rating/following profile of other participant (learner, teacher, tutor, other)
    Date-time of viewing/liking/rating/following
    Id of another participant
    Role of other participant
    Score given for rating/liking
  Adding other participant to contact list (learner, teacher, tutor, other)
    Date-time of adding
    Id of another participant
    Type of contact
  Number of contacts

  Number of followers

  Viewing/liking/rating/following blog of other participant
    Date-time of viewing/liking/rating/following
    Id of another participant
    id of blog_input

like or rate score

Viewing/liking/rating/following wiki of other participant

Date-time of viewing/liking/rating/following
Id of another participant
id of wiki_page
like or rate score

Commenting on blog/profile/comments/wiki

Karma/reputation score

- Badge received

Id of activity
Date-time
Type of badge

- Downloading of resources

date-time, pageid, type of page, objectid, type of object

8.3 xAPI statements

Even when some of the ECO platforms are already tracking and logging required data - and they can continue doing so -, for the LA service to work properly, the data needs to be tracked and recorded as xAPI statements. A brief explanation of xAPI is given in chapter 5. To recap, the xAPI specification is used to track and log relevant actions the learner takes in the MOOC platform. It registers who performs what activity with what object at which time and in what context. xAPI statements take the form *noun verb object. Result, context and timestamp* are also part of the statement.

In this section a start is made with the mapping from actions by the learner in the ECO platform to xAPI events. It is not the intention to provide the full event statement, but to highlight the mapping from actions to verbs and objects/activities. All xAPI statements should have a timestamp, and most will have a context and result. The table is not yet complete on context and results, but only lists the most important once.

Verbs and objects are taken from the Tin Can Registry[6] and are case sensitive; some terms appear twice, either being defined by Activity stream, or by ADL. xAPI and Activity stream do not provide all verbs and objects that are required for ECO. New verbs and objects need to be defined in conjunction with WP3 and be in accordance with the features offered by each of the ECO MOOC platforms. Particular attention has to be paid to capture all events related to networked learning and social media. These are not yet fully covered by ADL or Activity stream. Whenever events of groups of learners or groups of activities have to be

[6] https://registry.tincanapi.com/

captured, the statements have to be expanded with paradata syntax as explained in chapter 5. It pays to collaborate with other learning analytics and MOOC projects that are also using xAPI to synchronise and standardise events and arrive at a set of events that is suitable for MOOC platforms.

A user identifier is required. Personalised and identifiable data is only used to report to specific learners, all other analytics are performed on anonymized data. Nevertheless data are tracked from identifiable users. An anonymised unique identifier would be sufficient, and is referred to as userId. It is common to use the email address as user identifier, but that is not anonymous.

The xAPI specification requires that verbs and objects are identified through a namespace URL. For briefly sake, this URL is not given in the table. Therefore the table does not indicate whether the verb or object is defined by activity stream or by ADL. An exception is made object question (defined by activity stream) and Question (defined by ADL). Moreover, the various types of learning resources have not been fully identified. Separate xAPI statements have to be generated for each of the relevant learning resource types.

Table 32: Mapping learner actions to xAPI verbs and objects

[*]to be modified from xAPI, [#] new, question[1:] defined by activity stream, Question[2]: defined by ADL

Event	xAPI verb	xAPI object/activity	context	result
User registers at ECO portal	registered[*]	service		
Learner logs in	Log in	service		
Learner enrolls in MOOC	registered	Course		enrolled
Learner unrolls from MOOC		Course		not enrolled
Learner accesses MOOC	accessed	Course		
Learner accesses entry questionnaire	accessed	page/Assessment	entry questionnaire_id	
Learner selects/specifies intentions	preferred	question[1]	entry_questionnaire, question	selected intentions
Learner selects/specifies learning goals	preferred	question[1]	entry_questionnaire, question	selected learning goals
Learner access a page	accessed	page	page type	
Learner access a module	accessed	module	module type	
Learners access syllabus	accessed			
Learner access a learning activity	accessed	activity[#]	type of activity	
Learner completes learning activity	completed	activity[#]	type of activity	ticked/completed by
Learner access a task	accessed	task	module, task	

Event	xAPI verb	xAPI object/activity	context	result
Learner submits task	submitted	task	module, task	submitted or score
Learner completes task	completed	task	module, task	ticked, completed by or score
Learner access assignment	accessed	assignment#	type of assignment	
Learner submits assignment	submitted	assignment#		submitted or score
Learner uploads assignment product	attached	file	assignment id	
Learner completes assignment	completed	assignment#		ticked, completed by or score
Learner access assessment	accessed	Assessment		
Learner attempts assessment	attempted	Assessment	attempt id	saved, intermediate results
Learner answers question	responded	Question[2]		response
Learner submits assessment	submitted	Assessment	attempt id	submitted or score
Learner access peer assessment	accessed	peer assessment#		
Learner access peer product	accessed	peer assessment#		
Learner submits peer feedback	submitted	review	peer assessment Agent	
Learner indicates activity/task/assignment completed	completed	activity#/task/ assignment#		ticked, completed by
Learner accesses learning resource	accessed	learning resource# page Media …	resource type	
Learner downloads resource		page Media	resource type	

Event	xAPI verb	xAPI object/activity …	context	result
Learner watches video	Played	video		
Learner listens to audio	listened	audio		
Learner listens to podcast	listened	podcast#		
Learner accesses forum	accessed	discussion	forumid	
Learner reads forum message	read	discussion	forumid, threadid	
Learner post new forum thread	authored	comment		
Learner replies to forum message	authored	comment	inReplyTo forumid, threadid	
Learner accesses blog	accessed			
Learner accesses blog page	accessed			
Learner accesses individual blogpost	accessed			
Learner posts blogpost	authored	blogpost#	blog	
Learner comments on blogpost	authored	comment	blog, blogpost	
Learners access wiki	accessed			
Learner access wiki page	accessed			
Learner creates wiki page	authored			
Learner edits wiki page	updated			
Learner access game	accessed	game		
Learner submits game	submitted	game		
Learner updates personal profile	updated	page	profile	
Learner views another user's profile	viewed	page	profile Agent	
Learner adds contact to list	appended			
Learner removes contact from list	removed			
Learner requested friend	requested friend			
	contacted			

Event	xAPI verb	xAPI object/activity	context	result
Learner added friend	made friend			
Learner follows user	followed			
Learner removes friend	removed friend			
Learner unfollows user	unfollowed			
Learner bookmarks activity	bookmarked		activity#	
Learner bookmarks resource	bookmarked			
Learner rates resource	rated			rating
Learner rates blog post	rated			rating
Learner rates forum message	rated			rating
Learner liked resource	liked			
Learner liked blog post	liked			
Learner liked forum message	liked			
Learner liked tweet	liked	note		
Learner unliked	unliked			
Learner shares resource	shared		user/medium	
Learner shares blog post	shared		user/medium	
Learner unshared	unshared			
Learner access tweet feed				
Learner tweets	tweeted	note		
Learner retweeted	retweeted	note		
Learner replied to tweet	replied to tweet	comment	note inReplyTo	
Learner follows twitter	followed	note		
Learner access activity stream	accessed			
Learner clicks (and views) activity	views		user/activity	
Learner follows activity stream	followed			
Learner follows RSS feed	followed			
Learner joins a group	joined	group		
Learner creates a group	created	group		
learner leaves a group	left	group		
Learner receives a badge	received	badge	activity/task/assig	

Event	xAPI verb	xAPI object/activity	context	result
			nment/course	
	scored			
	passed			
	mastered			
	preferred			
	favored			
	completed			
		audio		
		article		
		badge		
		binary		
		comment		
		file		
		game		
		image		
		note		
		page		
		Media		

9 Possible scenarios for implementation

In this chapter we give some scenarios to illustrate how the major metrics for progress, performance and the group of learners at risk of dropping out could be implemented and visualised.

We start by describing how the operational versions of OpenMOOC and Logi Assist currently are tracking learner progress and performance. OpenMOOC is chosen because most hubs are using this platform. Logi Assist is taken because in addition to a web platform it also offers mobile access.

Then we take one of the MOOCs that have been offered in the first pilot to illustrate how the approach suggested in this deliverable could be applied. The course is first described according to the course design model. Next a mapping of data to indicators is suggested.

Next there is a section about progress and performance monitoring. This section is intended to illustrate the complexity ECO faces when having to deliver learning analytics that meet the learner's objectives, as that means that many variations are possible. It describes various ways in visualising progress and goes into some background information for performance and drop-out metrics.

9.1 Current progress tracking in OpenMOOC

9.1.1 Course design

OpenMOOC have three levels for contents: course, units and pills.

A **course** consists of:

- General information on the course, teachers.
- The units that form.

A **unit** consist of pills and there are three types of units:

- Normal, to expose the contents and usually published weekly.
- Homework, to promote the individual and group work, usually published weekly and related with the contents exposed in the normal unit of that week.
- Exam, to measure the assimilation of content by students.

Homework and exam are limited in time, with a start and end time.

A **pill** (or **nugget**) consist of:

- One online video from Youtube or Vimeo (or prezi or scribd content) like central content.
- Comments by the teacher to clarify something exposed in the video.
- Additional material to extend the content exposed in the video.

- Some file attached.

You can add forms or peer review tasks to a pill.

The optimal use is adding some questions to normal units (not in all the pills) and not use peer review. It's better the use of peer review in homework units.

9.1.2 Calculation of final score

This is an example about how OpenMOOC compute the final score.

Let a course with three units

- Unit 1, with 20% as weight in the final grade. This unit has two nuggets each with a p2p task. Each nugget has the same internal weight in the unit. In that case if 0% weight for each nugget is given, it will mean "equal weight".
- Unit 2, with 80% as weight in the final grade. This unit has three nuggets, only one has a quiz and all the nuggets have the same weight (we've given 0% as weight).
- Unit 3, with 0% as weight in the final grade. This unit has three nuggets.

Let a student has the following results

- Unit 1
  - Nugget 1, he/she done the p2p task and obtained an average score of N peer reviews as 2.5 on 5, is 5 on 10.
  - Nugget 2, he/she done the p2p task and obtained an average score of N peer reviews as 3 on 5, is 6 on 10.
- Unit 2
  - Nugget 1, he/she viewed the video => correct . Involves only correct if you finish watching the video, if not, it's wrong, I mean 0 points. 10 points.
  - Nugget 2, he/she fail in quiz => incorrect. 0 points.
  - Nugget 3, he/she viewed the video => correct. 10 points.
- Unit 3
  - Nugget 1, he/she viewed the video => correct. 10 points.
  - Nugget 2, he/she viewed the video => correct. 10 points.
  - Nugget 3, he/she viewed the video => correct. 10 points.

Then, the final grade for this student in this course is as follows

Final grade = 0.2 x (6 + 5) / 2 + 0.8 x (10 + 0 + 10) / 3 + 0 x (10 + 10 + 10) / 3 = 6.43.

9.1.3 Course progress monitor for teachers

The admin->stats menu entry provides teachers with progress information.

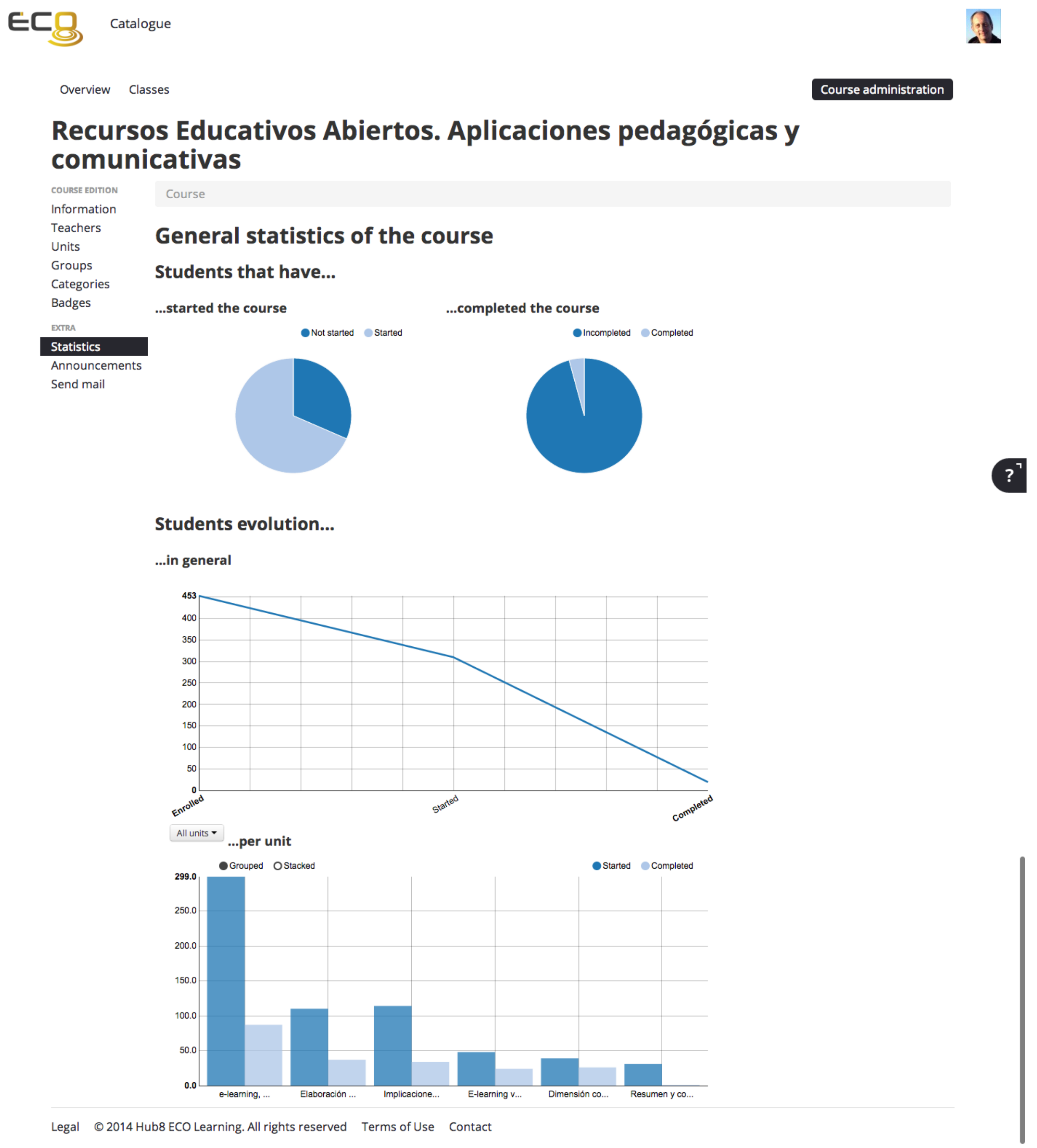


Figure 12: OpenMOOC teacher dashboard progress monitor

If you want to see the data on unit 2, you can choose this unit from the dropdown menu on the bar graph.

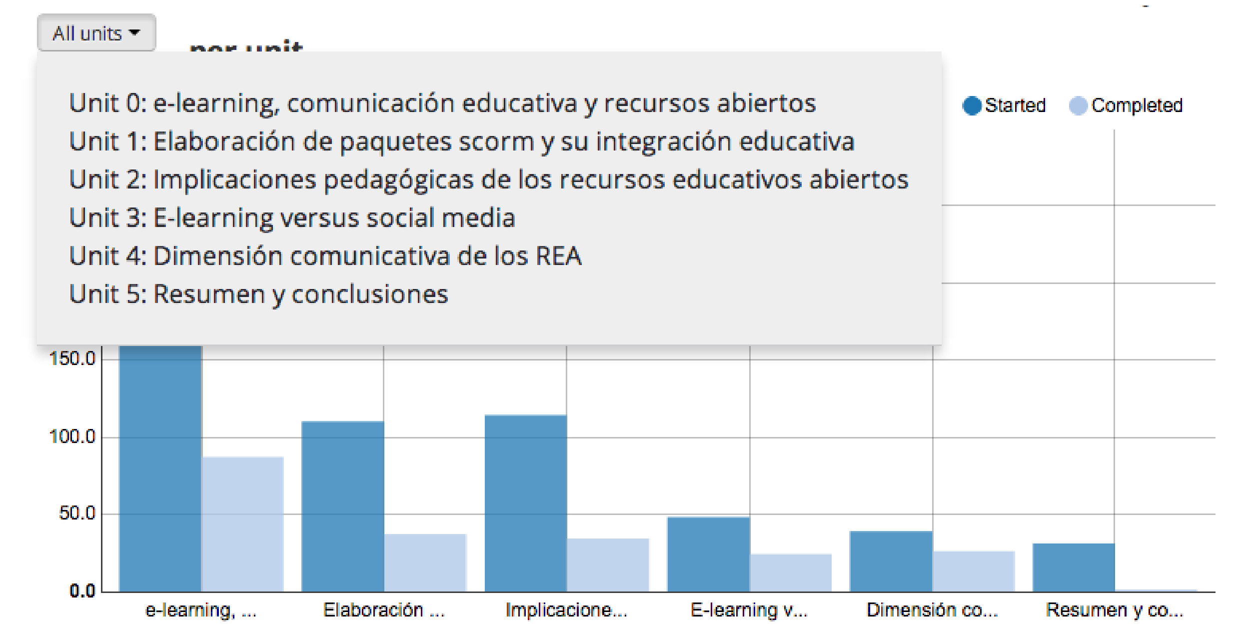


Figure 13: OpenMOOC teacher dashboard progress on tasks

Then you can see the data on nugget 3, which is a peer review task

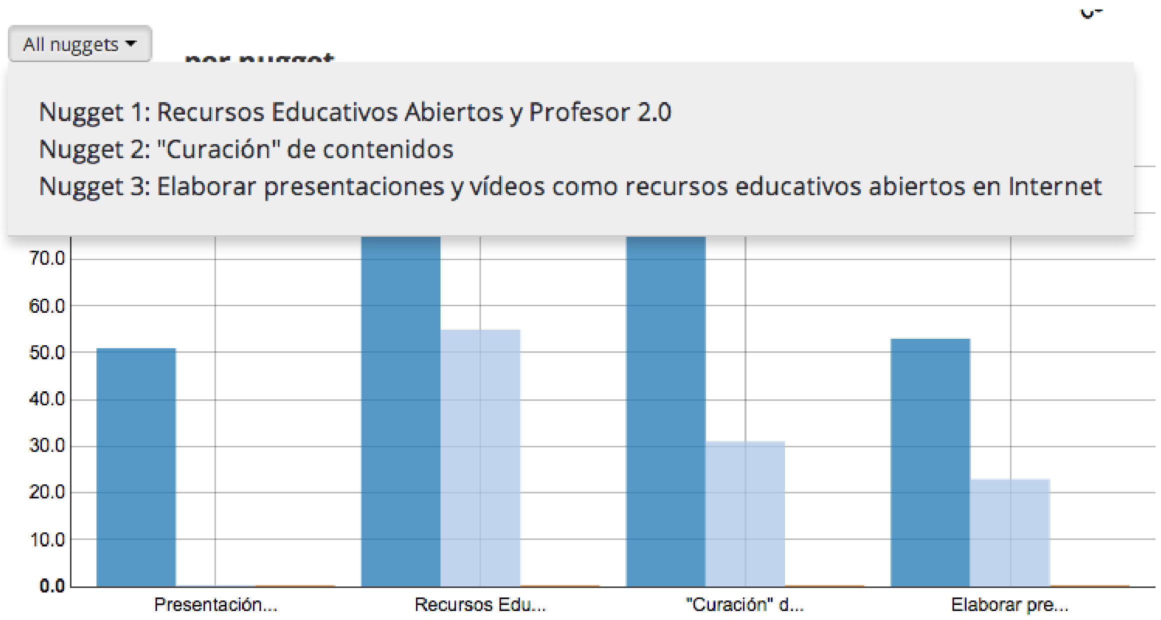


Figure 14: OpenMOOC teacher dashboard progress for a peer feedback task

This is the data about tasks sent and reviewed by peers on nugget 3 in unit 2.

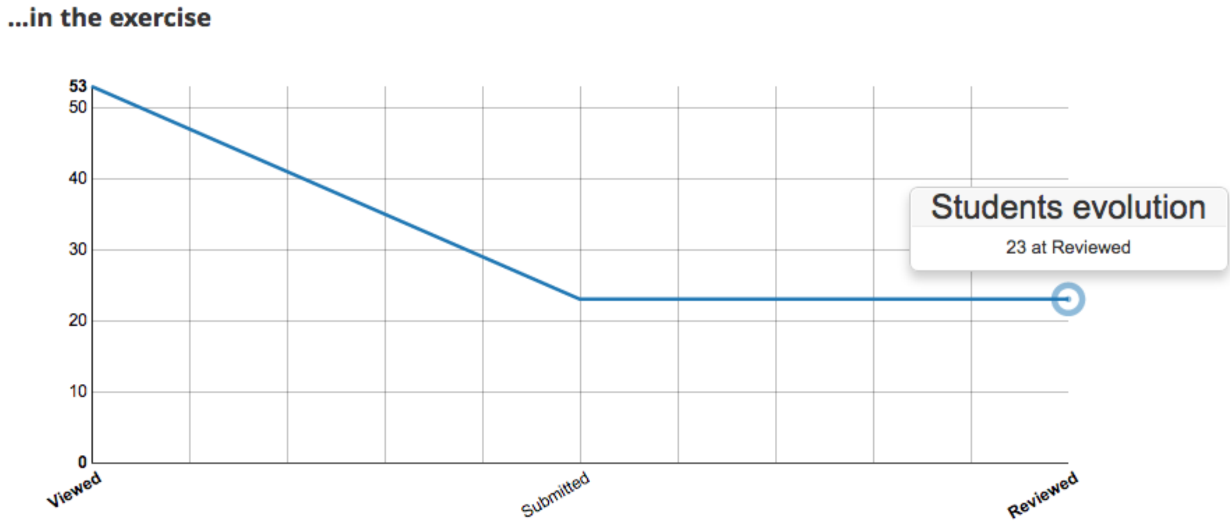


Figure 15: OpenMOOC teacher dashboard status peer feedback

Finally, you can pass the mouse on any part of the draw to see some extra info.

9.1.4 Other data

Additionally, you can extract all the information from students of the course as surnames and emails, ratings and other data each.

The OpenMOOC team is working hard now in access and display data from outside for exploitation in learning analytics. Especially important is the relationship between different information, links between people and geolocation data.

9.2 Progress and user tracking in Logi Assist

In this section progress tracking methods and progress visualizations are detailed for the ECO MOOC platform Logi Assist. The educational system of Logi Assist has several ways to track learner's progress. First of all, courses in Logi Assist can have different unlockable chapters and contents. A course administrator can set up conditions that have to be fulfilled to unlock new contents of a course. In this way Logi Assist provides dynamic courses, which can evolve depending on a learner’s knowledge. Secondly progress of all course contents are tracked and saved in the database and sum up to an overall course progress. In conclusion, additional information about learner’s behaviour can be determined while the learner is interacting with the course.

9.2.1 Course progress

A course in Logi Assist contains arbitrary number of chapters and contents. The latest can be PDF files, texts, audio files, video files, exams or surveys. Each type of content has a different way to track learner's progress.

- *Progress of PDF files using browser plugin*

  In Logi Assist, there are two ways to display pdf files in a course. One is using the native PDF plug-in of the learner's browser. This is an efficient way to display large PDF files in a course regarding loading and compatibility issues. The limitation of using this technique is that there is no way to track the learner's interaction within the PDF. Thus, if PDF files are displayed using the native browser PDF plug-in, the progress will be either completed (equals 100%) or not completed (equals 0%) depending if the file is opened or not.

- *Progress of PDF files using the custom PDF viewer of Logi Assist*

  Progress can be tracked more finely, when the PDF file is rendered in the browser using Logi Assist's own PDF processor plug-in. This is a JavaScript based PDF viewer based on pdf4js. Processing and rendering time is slower than using the browser's native plugin, but grants the opportunity to track the user interaction within the PDF. Meaning the platform can track which pages of the PDF file are opened. In this way more detailed progress data are available.

- *Progress of video and audio files*

  The progress of video and audio files is tracked depending on how long the learner is watching or listening to the corresponding content. The progress of such multimedia contents is updated when defined boundary values are reached. Assuming that the full length of the video or audio is representing 100%, the progress is updated if a threshold is once reached by the learner (25%, 50%, 75% and 100% of the audio or video file).

- *Progress of text contents*

  A course can contain two different types of text contents in the platform Logi Assist. The first one grants the administrator access to a simple text editor, where a customized text can be created, edited and updated. The other possibility is using a special text editor allowing the administrator to create a text with headlines, images, tables or listings all following the style of the portal when displayed in the learner's browser. Progress tracking for both options is the same and is tracked by observing the scrolling behavior of the learner. Similar to the progress tracking of video and audio files, the progress of text contents is updated if a threshold is once reached by the learner (25%, 50%, 75% and 100% of the text content).

- *Progress of exams and surveys*

  In addition to all the described content types, exams or surveys can be added to a course. Both contain an arbitrary number of questions. There are three available types of question, which can be used:

  - Single Choice questions are allowing only one option to be chosen by the learner.
  - Multiple Choice questions are allowing multiple options to be chosen by the learner.

- User Input questions are forcing the learner to enter the correct answer into a corresponding field.

Both surveys and exams are tracking the given answer values of a learner.

In an exam, given answers are evaluated and, depending on the result, points are given. When an exam is finished, the overall progress of the run-through (a percentage value representing correct answers and the achieved points) represents the progress of the exam. The best result of all run-throughs defines the current progress of the learner's corresponding exam.

In a survey, given answers are not evaluated. Thus a survey's progress is set to 100% if the survey was completed.

In both surveys and exam, the time spent by the learner is tracked and an average duration for a learner's run-through is calculated.

These are the progress tracking methods of course contents in Logi Assist. All those content progresses are summed up in an overall course progress, representing the current completion (as percentage) of a course.

9.2.2 Tracking of user behaviour

In addition to the progress tracking described above, the Logi Assist platform does track the learner's behaviour while interacting within a course. For each content inside the course, the time spent by the learner while working on a certain content is tracked and saved. The number of viewings of a particular content and of a course is observed as well. This additional user behaviour evaluation allows the calculation of an average time spent on content level as well as on course level. From those data, additional statistics can be derived like e.g. complexity of a course, which course is in general the most viewed course, etc.

9.2.3 Progress rules

A course in the Logi Assist platform can be configured using so called "progress rules". An administrator has the possibility to design a course depending on the success rates of a learner. For each specified content inside a course, individual progress rules can be defined, allowing a dynamic design of a course. Meaning conditions are defined, which specify when contents of the course should be unlocked (or made visible) depending on the achieved progress of the learner. For example, if a learner has reached 40% in an exam, Content A will be unlocked. By reaching 90% of the same exam, the learner will be able to get an additional content B.

Progress rules can be specified by comparing the achieved progress by an operator and a value. Available operators are "*less than*", "*less or equal than*", "*equals*", "*greater or equal than*" and "*greater than*". As a matter of fact the value can be any value between the range of 0% and 100%. The observed content of a progress rule can unlock an arbitrary number of additional contents. In conclusion, this concept allows the design of dynamic courses.

Moreover, by coupling the progress and behaviour tracking methods described above, additional statistics of user interactions can be derived like e.g. how many users did unlock a specific content, etc.

9.2.4 Progress visualization for the learner

A learner in Logi Assist has different ways to see the current achieved progresses. There are two main pages on the platform where a learner can inspect his individual educational statistics: on the course list or on his education dashboard.

- *Course list*

  In this page, a list of all available courses is displayed as a tile grid. Every course tile holds a progress bar representing the current overall progress. Additional information and overall statistics of the course can be seen, when a course is selected. On the right hand side panel, the learner can see how much time was spent on this course and how often the course was viewed. Information like the date of assignment and the date of completion (date when the learner has reached 100%) are also available.

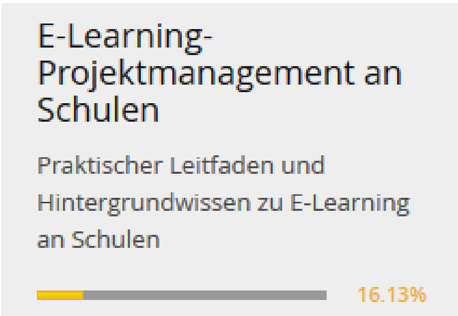


Figure 16: Logi Assist visualising course progress for learners with course title

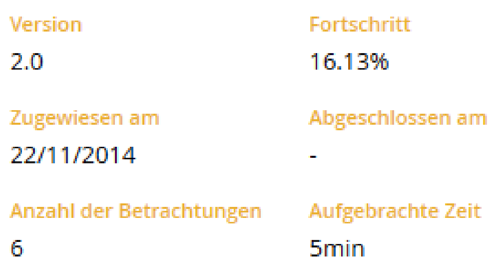


Figure 17: Logi Assist showing additional information to the learner in the detail pane of the course indicating when started, how much time was spent, current progress and indicating that the course is not yet completed

- *Education dashboard*

  The education dashboard shows the learner's overall educational statistics in the whole Logi Assist platform. Accordingly the number of assigned courses and completed courses as well as the number of unlocked course content and completed course contents are displayed in an overview. This allows the learner to evaluate his current progress. Furthermore the overall average spent time (based on all spent times of all assigned courses) can be visualized as well the total amount of course interactions.

  In the dashboard, the learner also sees the last three unlocked course contents as well as the last three completed course contents. That information allows the learner to see directly, where he lastly completed certain content. Thus a positive feedback can be stimulated for the learner. Additionally newly unlocked course contents are shown, motivating the learner to solve new challenges.

  The progress bar at the bottom of the dashboard indicates the overall achieved progress, regarding all assigned courses of the learner. If this progress bar is filled completely the learner has successfully completed all assigned courses.

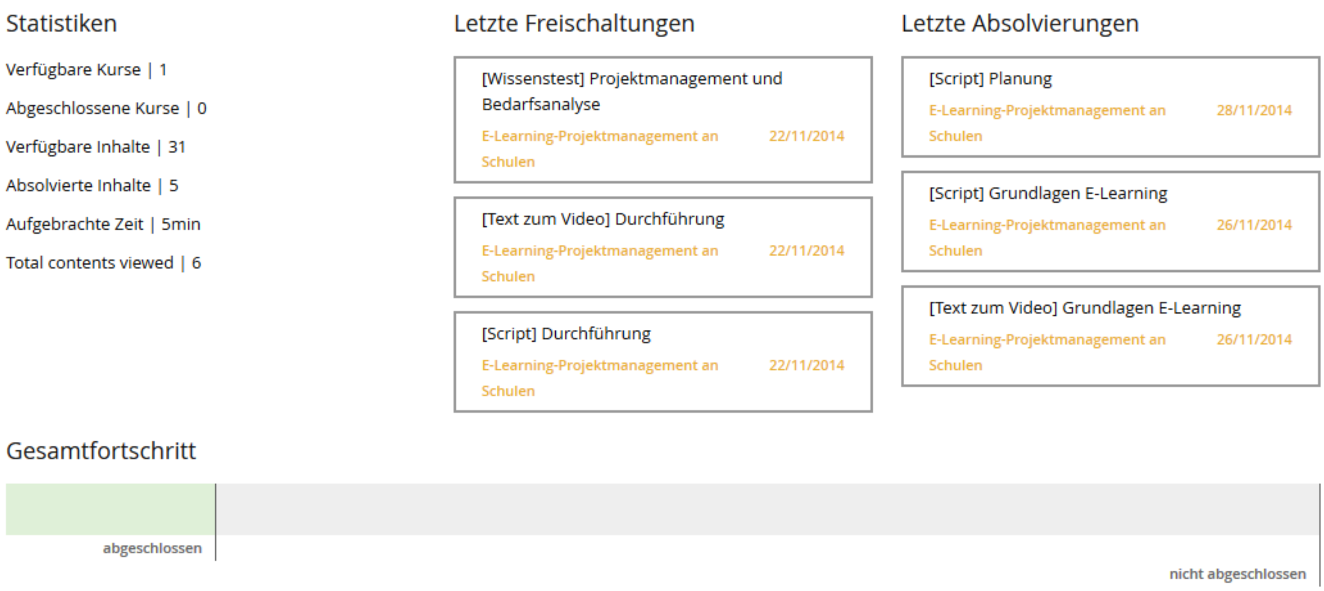


Figure 18: Logi Assist learner dashboard

The figure depicts the dashboard with at the top the statistics, in the middle and right new and completed content, and the progress bar at the bottom.

9.2.5 Summary

As a conclusion, the Logi Assist platform offers a lot of possibilities to track and to visualize the progress of a learner. With the recorded data of progress and behaviour, many useful statistics can be generated. Moreover the opportunity to visualize the learner's own progress allows the learner to get an overview of his progress and overall performance. With those techniques, the learner is efficiently informed, motivated and supported to reach his educational goals.

9.3 How to apply ECO learning analytics

The main goal of this scenario is to show how ECO learners' indicators can be calculated from an ECO course hosted in an e-learning platform such as OpenMooc or Logic Assist. Likewise we want to highlight which data must be provided by teachers and platforms to accomplish learning analytics task.

Next we describe the organisation of one of the ECO MOOC courses launched in November 2014 entitled "Educational Innovation and Professional Development. Possibilities and limits of ICT". The following table shows the units designed as well as the learning goals and the optional and compulsory tasks established.

Table 33: Mapping course design

Unit	Objectives	Optional Tasks	Compulsory Tasks
Unit 1. Stirring ideas about innovation **(1 week)**	Use ECO environment in which the course is developed. Use of the different communication tools available.	1) Share information in the forum about their background, experience, etc. 2) Answer a quiz about their educative experience, MOOC experiences, etc. 3) Watch different videos and discuss some questions in the forum.	
Unit 2. Lighting the change: key concepts, facilitators and barriers to innovation	Students will be able to differentiate the main concepts that have been used to refer to changes in educational contexts and	1) Watch a video interview and discuss about it in the forum. 2) Read news and	1) Study of the basic materials (videos, pdfs, etc.). 2) Assessment: Quiz

(1 week)	identify factors that facilitate the innovation process and those that hinder (barriers).	participate in discussions with other participants through Facebook. 3) Share on Twitter the problems you found when you use ICT in educational contexts. 4) Review optional materials (academic paper, etc.).	(15 questions).
Unit 3. Innovation and teacher development **(1 week)**	Understand some strategies to teachers' professional development through the use of ICT. Identify key concepts such as professional development or long life learning.	1) Watch a video and relate the concepts learned in this unit with those proposed in the video. 2) Read an academic article and establish relationships between the grounds of public policy for the integration of ICT and models of teacher training. 3) Share in Twitter information about models of training with ICT in which you have participated as student. 4) Review optional materials (academic paper, etc.).	1) Study of the basic materials (videos, pdfs, etc.). 2) Assessment: Quiz (15 questions).
Unit 4. Basic dimensions of innovation with	Define the main dimensions of innovation with ICT (curricular,	1) Analyse how ICTs are being integrated in the professional or	1) Study of basic materials (videos, pdfs,

technology: curriculum, training and organizational **(2 weeks)**	organizational and training). Use indicators to identify the best practices.	training context and share it in Twitter (e.g. with a picture). 2) Read an academic article and answer different questions in the forum. 3) Share information about projects in which has participated and establish relations with the contents of the unit. 4) Review optional materials (academic paper, etc.).	etc.). 2) Assessment: Peer to peer activity. Case analysis.
Unit 5. The realization of innovation projects with ICT **(2 weeks)**	Design of contextualized educational innovation projects with ICT. Systematise data for the evaluation of projects.	1) Read an academic article about practice communities and answer different questions in the forum. 2) Watch a video and answer different questions in the forum. 3) Share information on Facebook about communities of innovative practices in which you are participating or you know it. 4) Review optional materials (academic paper, etc.).	1) Study of basic materials (videos, pdfs, etc.). 2) Assessment: Peer to peer activity. Design an innovation project with ICT.
Unit 6. Rebuilding the	Present the main	1) Collective creation	

itinerary **(1 week)**	conclusions drawn in this MOOC course development. Make those mandatory and optional activities that he could not have made.	of the main conclusions of the course. 2) Invite other students to perform the MOOC with video, audio, etc. messages. 3) Make a quiz to evaluate the experience 4) Share information on Twitter about the type of innovation project you would like to launch.	

The following table summarises the tools that platforms must provide in order to teachers can fully specify the tasks as well as the data that these platforms must write in their log to perform the calculation of ECO learners' indicators.

Table 34: Required tools for teacher and required data for learner indicators

Category: Performance PI1: Personal grade in module PI2: Personal grade in the course		
Course Activities	Platform must provide teachers with tools to specify	Data
Applicable to compulsory assessable tasks from units 2 to 5.	- Weight of each assessable task in the module -Weight of each module in the course -Scale of grade -Minimum grade to consider the task passed	- Weight of each assessable task - Grade achieved in each assessable task

	-A mechanism which allows participants to establish a mark in peer evaluation	
Category: Mastery		
Course Activities	Platform must provide	Data
1) Study of the basic materials (videos, pdfs, etc.) 2) Review optional materials (academic paper, etc.). 3) Answer different questions on the forum. 4) Make quizzes 5) Peer to peer activities	An action button which student pushes when he or she considers that has completed the activities categorised in 1) and 2) Regarding activities from 3), category the list of messages sent to forum with their publication date. Regarding activities from 4) and 5) category, platforms must register the number of attempts of each quiz with its grade and date in which was performed. A mechanism which allows participants to establish a mark in peer evaluation must be enabled	-Date of task completion -Number of attempts -Grade achieved in each attempt
Category: Progress PrI1: Progress in each module for compulsory tasks PrI2: Progress in each module for optional tasks		
Course Activities	Platform must provide	Data

Applicable to all the activities/resources over any established time period.	A mechanism to mark each task as completed.	-Date of task completion
Category: Engagement (EI1) Access frequency (EI2) Resource use (EI3) Percentage of participation in communicative/collaborative tasks (EI4) Ranking of valuation of participation in forum (EI5) Ranking of valuation of participation in forum (EI6) Post duration		
Course Activities	Platform must provide	Data
Applicable to all the activities/resources over any established time period.	A mechanism to mark each task as completed. A mechanism to mark forum messages as favourite or give a valuation (e.g. 1 to 5). Teacher should establish deadlines for each activity to be done	-Time/date of log-in. -Time/Date of access to each different resource and/or activity -Deadline for each activity to be done -number of threads accessed -number of replies to thread -number of messages marked as favourite

Category: Effort (EfI1) Time per resource (EfI2) Time vs performance		
Course Activities	Platform must provide	Data
Applicable to all the activities/resources over any established time period.	-Weight of each assessable task in the module -Weight of each module in the course -Scale of grade -Minimum grade to consider the task passed -A mechanism which allows participants to accept a peer evaluation task -A mechanism which allows participants to establish a mark in peer evaluation	ime spent in quizzes ime spent in forum ime spent in reading or watching activities or a mechanism to estimate it rade achieved in each unit and in the course
Category: Satisfaction (SI) Set of indicators which summarises the learners' opinion		
Course Activities	Platform must provide	Data
Make the quiz in unit 6	A general ECO post-questionnaire should be sent when the course finishes, an example is proposed in section 5, chapter 7 in order to unify	The value given to each question
Category: Social affiliation (SAI) Social graph		

Course Activities	Platform must provide	Data
Answer different questions on the forum. Share information on Twitter, Facebook, etc.	A mechanism to follow participants A mechanism to value the contribution of each participant in forums A mechanism to ask permission to measure their contribution in external social network A mechanism to read message network and valuation of contribution in external social network A mechanism to collect learner's profile	- Number of contacts added to contact list - Number of followers - Viewing/linking/ rating/following profile of other participant - Viewing/linking/ rating/following messages in the forum - Learner's profile: age, gender, karma reputation and so on
Category: Social Recognition (SRI1) Followers (SRI2) Contributors		
Course Activities	Platform must provide	Data
-Answer different questions on the forum. -Share information on Twitter, Facebook, etc.	A mechanism to follow participants A mechanism to value the contribution of each participant in forums A mechanism to ask permission to measure their contribution in external social network A mechanism to read message network and valuation of contribution in external social network	-Number of contacts added to contact list -Viewing/liking/rating/ following profile of other participant -Viewing/liking/rating/ following messages of other participant

Category: Social Responsibility (SRI1) Average time to answer a question addressed to this learner (SRI2) Number of mandatory peer-evaluation tasks carried out on time out of total (SRI3) Position in relation to the number of voluntary evaluations of peer-to-peer tasks performed. (SRI4) Valuation given by his peers		
Course Activities	Platform must provide	Data
Answer different questions on the forum. Assessment: Peer to peer activity	-A mechanism which allows participants to accept a peer evaluation task -A mechanism which allows participants to establish a mark in peer evaluation -A mechanism which allows participants to assess your peers	-Date of accessing internal mail -Date of reply to a particular mail - Date of evaluation of P2P task - Deadline of P2P tasks -Number of P2P tasks evaluated voluntarily

9.4 Variations on progress and performance scenarios

In the previous chapters we defined important concepts and described relevant indicators and metrics for learners and teachers. We stressed the importance of providing learning analytics that fit the learner's objectives and the need for indicators according to the various types of learners that participate in MOOCs.

In this section we describe several alternative scenarios. Alternative, because they might need to be applied when the ECO learning analytics requirements can't be met. Alternative, because they informed the decisions on which the ECO learning analytics service is based.

9.4.1 Progress metrics

The Cambridge Dictionaries Online give various definitions of progress that all are applicable in a learning context.

- movement to an improved or more developed state, or to a forward position
- to improve or develop in skills, knowledge, etc.
- to continue gradually

In its simplest form 'to continue gradually' progress is only an indication of how far somebody has moved along, while the first two definitions entail some kind of assessment. Both perspectives, a) how far along am I and b) how well am I doing are of interest to a learner.

As explained before, progress needs to be measured against the intentions and learning goals of the individual learner. Nevertheless, a simple scenario of 'how far along am I' might suit all categories of participants.

How far along am I

Progress is depicted for individual learners in a progress bar.

The image below is a progress bar, indicating the start and end dates of the course and every blue rectangle indicating a week. No progress is indicated here.

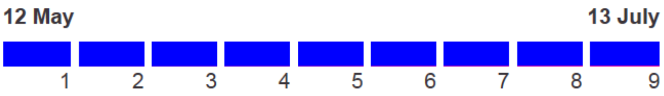


Figure 19: Progress bar; no progress as yet

The next image shows a progress bar with partial progress. It depicts start and end dates, and indicates relative progress in each week. It also indicates through the small pink triangle what the current week is. Mouse over can indicate the percentage completed. The preferred duration of a sMOOC is six to eight weeks, but there might not always be a set starting and end date. At least during the duration of the project ECO sMOOCs are launched at specific date and thus have a starting date and duration. The figure above is intended as example.

Figure 20: Progress bar with relative progress indication in every week, plus indication of current week

When a MOOC provider makes all content available from the start of the MOOC and does not deliver content in weekly intervals, progress still can be depicted in intervals. A weekly interval seems suitable. Alternatively, the same progress can be depicted by replacing weeks with learning-activity, module, topic, or whatever the main method is to structure the course. It then turns into a table of content, with nodes and leaves.

More detailed progress information can be given per week/learning activity/node, by indicating for each of the activities whether the learner has accessed the activity page at least once. Compulsory assignments as set by the course designer that matches the learner's learning goal should receive emphasis.

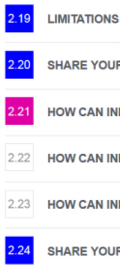


Figure 21: Detailed progress indication within a week/topic/module

Of course, there needs to be a measure of progress. The simplest form would be only considering a visit to the activity page, a more advanced form would need learners to indicate that they consider the activity to be completed, or can be determined by the system through the course design, e.g. when instructor has set a minimum time or qualification to achieve.

Scenario 1: full progress, no learner completion

In scenario we assume that a learner follows what the teacher has designed. ECO sMOOCs should last 6 - 8 weeks and every week should contain a set of activities. In its simplest form, progress here means an indication whether the learner has visited at least one activity in a week. A better indicator takes into account the number of visited activities relative to the total number of activities that week. This is a very simplistic scenario and should only be used when no other option is available. Better scenarios to use are those that rely on measures of completion. Those measures should be suitable for the type of activity. Reading a forum message is not the same as reading a lecture or submitting a quiz.

Assumptions:

- fixed start and end date
- fixed number of weeks
- teacher has designed x activities
- there are no optional or extra activities (no bank of challenges)
- learner can't indicate activity completion
- system tracks userid
- system tracks which activity learner has accessed on which date

Assuming, learners can't indicate completion, the formula for calculations would be:

> Progress per week = number of activities accessed at least once (accessed page) divided by total number of activities defined for that week.

Visualisations as shown in Figure 20 and Figure 21 are based on accessing activities.

Scenario 2: full progress, learner indicates completion

In a more refined progress measure, the learner can indicate that an activity has been completed. In this scenario we assume that a learner follows what the teacher has designed. ECO sMOOCs should last 6 - 8 weeks and every week should contain a set of activities. In its simplest form, progress here means an indication whether the learner has visited at least one activity in a week. A better indicator takes into account the number of visited activities relative to the total number of activities that week.

Assumptions:

- fixed start and end date
- fixed number of weeks
- teacher has designed x activities
- there are no optional or extra activities (no bank of challenges)
- learner can indicate activity completion
- system tracks userid
- system tracks which activity learner x has accessed on which date
- system tracks which activity learner x has completed on which date

> Progress per week = number of activities completed divided by total number of activities defined for that week.

Visualisation is similar to Figure 20 and Figure 21, only based on different data. For Figure 21 a distinction is made between activities completed, activities accessed but not yet completed and activities that have not yet been accessed.

Assuming that learner can indicate completion, the calculations take into account merely accessing, plus completion status, resulting in three progress indicators that can be marked by use of colour.

- No fill: not accessed, progress = 0
- Light colour: accessed at least once
- Dark colour: indicated by learner as completed

Scenario 3: individual progress based on learner type

An entry questionnaire can be used to ask learners about their intentions and learning goals. For learners with the intention to access every activity and complete the MOOC, progress can be shown according to scenario 2 (completion status of activities). For learners who indicate that they intend to access most activities, but do not intend to submit assignments and/or quizzes (auditing learners) and for those learners that indicate just to sample some of the activities or resources (sampling learners) progress can be visualised according to the first scenario (based on access data).

Scenario 4: progress in relation to optional activities

Whenever a MOOC designer includes a bank of challenges, i.e. additional, optional activities, or provides alternative learning paths, any progress indication is more complicated without knowing learner intentions. Progress against optional activities should be indicated separately from the progress as shown in Figure 20 and Figure 21. The visualisation needs to indicate the number of optional activities and show progress either as completed or accessed depending on type of learner.

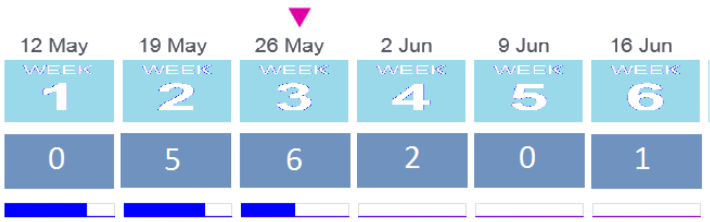


Figure 22: Progress for optional activities in bank of challenges

How well am I doing

More importantly than ‘How far along am I’ for learners to know is ‘How well am I doing?’. Of course this entails progress but includes some form of assessment. First of all, it is measured against the set of chosen learning objectives. These learning objectives can be those designed by the teacher or the ones chosen by the learner. Next it entails a comparison. A major comparison is against the ‘golden standard’ set by the course designer, other comparisons are against the whole learner population, against learners with similar learning objective, and maybe against previous cohorts. In this section about progress, this scenario is accomplished without resorting to assessed performance and is purely based on comparing progress of individuals against others.

This means that the progress scenarios 1 and 2 as described above are appended with a comparison against the average for the whole population, scenario 3 with a comparison against the average of the whole population plus a comparison against the average of similar users. While scenario 4 can be compared against the average of those participants who also have chosen to perform optional activities.

9.4.2 Performance metrics

Like progress, performance has to be measured against learner’s objectives and is therefore even harder to assess than progress. Moreover, performance can only be measured when learners engage in activities that can be assessed. From a course designer perspective this could entail (graded) assignments, (graded) quizzes, peer assessments, etc., but also more diffuse activities set by the course designer as having to take part actively, e.g. by submitting self-assessments, engaging in discussions, providing content such as blogposts or wiki contributions. The course designer defines the maximum attainable grade or score and has to specify a scoring rubric. The scoring rubric can be based on combinations of (weighted) scores for assignments, assessments and self-assessments. Individual learner performance is calculated according to this scoring rubric.

When performance is not assessed by graded assignments, homework and assessments, another way of measuring learning need to be applied. Learning can be seen as knowledge, skills and competences being gained and put to use. In conjunction with the constructivist and connectivist approach taken in ECO, evidence of learning can be gotten from learner's actions and behaviour, and collaboration and interactions with others (Stahl, 2006, Suthers, Dwyers, Medina, & Vatrapu, 2010). Uptake of knowledge can be evidenced by content and discussions created by the learner. Ideally natural language processing and text processing techniques in conjunction with speech acts and/or discourse analysis are applied to codify relevant constructs and speech acts or turns in the artefact that indicate uptake of knowledge and knowledge creation or find indications of successful collaboration (Arvaija et al., 2011). In a simplified form Apache Lucene could be used to determine relevant index terms in artefacts. By determining topics and identified frequency and timelines of these topics, several types of discussions can be identified. In the context of knowledge uptake and creation, sustained discussions are most relevant (Tobarra et al., 2014). Other types of discussion originate around events, and although relevant for learners, are less suitable as performance measure. Tobarra et al. (2014) defined several algorithms that can be automatically applied to determine topics and type of discussions. Wang et al (2001) applied discourse analysis and found that total number of responses, frequency of response and number of times that the students was the first to respond were positively correlated with conventional performance measures.

However, NLP is a highly specialised domain and discourse analysis, speech act analysis usually relies on manual coding of artefacts. These conditions might preclude the use of these approached in ECO.

Another approach would be to analyse whether a learner is active and contributes in forum, wiki and blogs. Here it is important to differentiate between posters, i.e. learners who create new messages and posts, and repliers, learners who do not initiate but reply and comment. This needs to be set off against the total number of messages and posts and the total number of learners that are involved in these activities. Posters are more likely to show a higher performance. As some indication of quality, reputation and karma scores can be involved.

This entails calculating:

- Number of published messages
- Number of replies
- Number of initiated conversations
- Number of initiated conversations without replies
- Number of conversations where the participant has posted a message
- Number of forums where the participant has posted a message

Some indication of topics dealt with can be gained from using Apache Lucene or OpinionFinder to capture sentiment of the contribution (Wilson et al., 2005).

9.4.3 Drop-out metrics

It will be really difficult to determine at what stage a learner can be considered to be 'dropped' out, in particular when we want to assess that for each of the distinguished categories of learners. Dropped-out in regular education is conventionally taken to mean a student who never finished a course or never took the exam course. Contrary to conventional education, most MOOCs do not finish with an exam, although there might be compulsory assignments and assessments.

In the context of ECO we define a participant at risk of drop-out as a participant with a remarkable low participation, progress and performance with relation to those who have got the same goals.Therefore, we propose the definition of at-risk-of-dropout. The limit to be at risk of dropout could be established as a 20% or 25% of the activity performed in the course by their peers, but only those with the same goals.

To measure drop-out in this context is very difficult since participants can adapt their pace of learning to their needs, can change their initial goals extending or reducing the goals to be reached.

Creating predictions like this is quite challenging as it would require textual analysis of learner's contributions in addition to analysis of the network and design. Ramesh et al. (2014) found that there are several latent indicators that can predict learner engagement. Relevant indicators are the viewing of content, interacting with others and topic and tone of interactions. They differentiated between active and passive engagement and were able to create predictions at an early stage. At the start of the MOOC, the latent predictors already give a good prediction of participants who will complete, but obviously it is most important to analyse the mid section of the MOOC. Viewing lectures and viewing forum posts turned out to be the most valuable predictors. Although this analysis was based on xMOOC model, there should be some use of this in ECO sMOOCs.

Some information can be gained from time of registration as well. Anderson et al. (2014) found that only 60% of participants registered before the course started, while another 18% registered only after the course has ended. The remainder registered during the course. Learners who register late are more likely to be auditing or sampling learners or learners who only download content.

Most studies on drop-out consider drop-out to occur in the first week or so, while a large proportion of registered participants never engage with MOOC contents or drop out after the first week. However, there is a large proportion of participants that struggle to keep up, and drop out along the way. These are the learners that are worthwhile to try to keep in by supporting them. Moreover these studies focused on providing summative information about type of participants instead of looking into how to promote that learners keep engaged with the course.

It seems that learners who keep up with the MOOC starting date are more likely to complete than learners who start at a later time. They tend to start with the contents of week they joined instead of from the start.

They also seem to have problems getting integrated into community discussion (Yang et al., 2013, Sinha, 2014). Learners who start at later weeks of the MOOC tend to post less frequently, do not catch up with earlier contributions and do not return.

Yang et al. (2013) applied a survival model. a proportional log odds logistic model, to determine what factors affect drop-out over time and found that cohort membership, post-duration (time difference between first and last post in a particular period of measurement) and authority score (learners who engage others in discussion) were indicative of drop out. Meaning that learners who started in the first week of the MOOC,or learners that have a higher than average post-duration, or a higher than average authority score are much less likely to drop out than others.

Balakrishnan (2013) applied hidden markov chains to predict drop-out on a weekly basis and used cumulative percentage of available video lectures watched, daily-unique forum threads viewed, number of posts in the forum (be it starting discussions or replying), and the number of times the course progress page was checked. Viewing the progress page was very indicative. Watching videos is important, but it becomes less important to view them in their entirety towards the end. Learners who watch at least 50% have a low risk of dropping out. Viewing forum threads becomes indicative of drop-out at later weeks. And learners who do not post are more likely to drop-out, although not as much as with the other factors.

Some information on risk of dropping out can be gathered from how active a learner participates in the forums, combined with social network analysis. There are some reports that mention that active forum contributors. i.e. those that belong to the top 5% of forum participants with respect to number of posts, are more engaged, are enrolled in multiple MOOCs and have a higher performance (Huang et al, 2014). However, it seems that learners who join the MOOC at a later stage do not get integrated into ongoing active discussions. Moreover, only a few learners persistently engage in the discussions (Huang et al, 2014; Sinha, 2014). Social network analysis can support in determining those active learners that act as bridges to others. Authority scores and hub scores can be used to get some indication of quality. Learners with high authority scores are those who engage others into discussion, while learners with hub scores get engaged into discussions.

Above illustrates that drop-out is not a static metric but depends on MOOC design, MOOC content, learning objectives and intentions and learning goals of the learners, as well as their behaviour. It can change during the duration of the MOOC, and has to be dynamically determined.

In addition some easy metrics can be used. A learner can leave the course by unenrolling, or can be considered to have left the course when being absent for more than three weeks in a row (based on the assumption that MOOCs last between 6 and 8 weeks). A learner can be considered to be inactive or dropped-out when at any given moment in time less than 20% of available activities, learning resources, assignments etc. have been accessed or performed.

9.5 References

Anderson, A., Huttenlocher, D., Kleinberg., J., & Leskovec, J. (2014). Engaging with massive online courses. In proceedings of the 23rd International Conference on World Wide Web, WWW'14, pp. 687-698. Republic and Canton of Geneva, Switzerland.

Arvaja, M., & Pöysä-Tarhonen, J. (2011). Tracing discursive processes of shared knowledge construction in a technology-enhanced higher education setting. Interactive Learning Environments, 1-17. doi: 10.1080/10494820.2011.559171

Huang, J., Dasgupta, A., Ghosh, A., Manning, J., & Sanders, M. (2014). Superposter behavior in MOOC forums. Paper presented at the Proceedings of the first ACM conference on Learning @ scale conference, Atlanta, Georgia, USA.

Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. Paper presented at the Proceedings of the Third International Conference on Learning Analytics and Knowledge, Leuven, Belgium.

Milligan, C., Littlejohn, A., & Margaryan, A. (2013). Patterns of Engagement in Connectivist MOOCs. MERLOT Journal of Online Learning and Teaching 9(2), pp. 149-159

Ramesh, A., Goldwasser, D., Huang, B., Daumé III, H., Getoor, L. (2014) Learning latent engagement patterns of students in online courses. Proceedings of the Twenty-Eight AAAI Conference on Artificial Intelligence

Sinha, T. (2014). Supporting {MOOC} Instruction with Social Network Analysis. CoRR, abs/1401.5175 http://arxiv.org/abs/1401.5175

Tabaa, Y., Medouri, A. (2013). LASyM: A Learning Analytics System for MOOCs (IJACSA) International Journal of Advanced Computer Science and Applications, 4 (5), 1-11.

Tobarra, L., Robles-Gómez, A., Ros, S., Hernández, R., & Caminero, A. C. (2014). Analyzing the students' behavior and relevant topics in virtual learning communities. Computers in Human Behavior, 31(0), 659-669. doi: http://dx.doi.org/10.1016/j.chb.2013.10.001

Wang, A. Y., Newlin, M. H., & Tucker, T. L. (2001). A Discourse Analysis of Online Classroom Chats: Predictors of Cyber-Student Performance. Teaching of Psychology, 28(3), 222-226. doi: 10.1207/s15328023top2803_09

Wilson, T., Hoffmann, P., Somasundaran, S., Kessler, J., Wiebe, J., Choi, Y., Cardie, C., Riloff, E., & Patwardhan, S. (2005). OpinionFinder: A system for subjectivity analysis. In Proceedings of HLT/EMNLP on Interactive Demonstrations.

Yang, D., Sinha, T., Adamson, D., & Rosé, C. P. (2013). Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses. Proceedings of the 2013 NIPS Data-Driven Education Workshop.

10 Legal, privacy and ethical concerns

Anytime an application deals with personal data care needs to be taken to guarantee privacy of those data and ensure that data are dealt with in ethical and transparent manner. There are large misconceptions on what constitutes personal data and what can be done with it. The scope of “personal data” is much broader than most people think, and includes absolutely everything which makes data personally identifiable.

In the learning analytics framework developed by Greller & Drachsler (2012) the constraints dimension defines those aspects that impact use of learning analytics and need careful consideration

10.1 Legal requirements

Legal restrictions in the area of Learning Analytics are for example data protection and copyright laws. Data that identifies a person is strictly protected in most Western countries, and access to such data needs to be specifically sought from the person in question. Copyright and Intellectual Property Rights (IPR) on the other hand limit the sharing of datasets collected by proprietary tools. In fact, the data sharing community is investigating possible ways to apply sharing licences to datasets to make them more widely reusable.

10.2 Privacy

According Pardo and Siemens, privacy is defined as the regulation of how personal digital information is being observed by the self or distributed to other observers.

Learning Analytics and other data services raise serious questions about people's privacy. Although not as expressly protected as personal data, the right to privacy is a fundamental principle for a democratic society. Privacy regulations include among other things the right to be left alone, or the right to form your own opinion. As more and more technologies collect data about people's whereabouts, behaviour, and feelings, this is increasingly perceived as intrusion into one's private life. As the data economy evolves, we need to ask the question: who owns the data of a person's actions and thoughts?

Palen and Dourish (2003) proposed characterizing privacy in digital environments on tree boundaries:

1. Discourse, comprises the simultaneous need to maintain some information private, and at the same time, some other information publicly available and known.
2. Identity, is used to symbolize the tension between the private and public information from the self to others.
3. Temporality, refers to the temporal nature of information. The information collected in the past can be used to infer future behaviour.

Some studies suggest that the concerns of users about privacy vary significantly depending on what is being observed, the context and the perceived value when granting access to personal information (Klasnja, Consolvo & Choudhury, 2009). This variation partially derives from the vague definition of what is considered personal information. Although personal data can be easily identified, the definition tends to also include the processes applied to these data and the resulting inferences. It is with the inclusion of such ample terms that the definition becomes much broader (Narayanan & Shmatikov, 2010).

10.3 Ethics

In the digital context, ethics is defined as the systematization of correct and incorrect behaviour in virtual spaces according to all stakeholders.

Applying Learning Analytics to the benefit of the learner and other stakeholders, may be the original objective of system designers, but there is no guarantee that this is actually taking place in the way intended. As data analysis uncovers information about people or groups of people with similar attributes, this may lead to confirmed prejudices and discrimination instead of help and support. There is no real way to protect against information about a person being used as a mechanism to exercise pressure in order to manipulate them into certain behaviour. Therefore we need to be aware that personal freedom and creativity might suffer.

The EU Data Protection Directive provides nine themes that need to be covered for any use of personal data including learning analytics applications.

1. Legitimate grounds – why you should have the data in the first place
2. Purpose of the data – what you want to do with it
3. Data quality – minimisation, deletion, etc.
4. Transparency – informing the students
5. Inventory – knowing what data you have and what you do with it already
6. Access – the right of the data subject to access their data, when can you have access to it and what can you see
7. Outsourcing – and the responsibilities of your institution as data controller and the third party as data processor
8. Transport of data – particularly problematic if outside the EU
9. Data security

In particular to note is that access to data about learners' activities is limited to learners themselves and certified access for teachers, researchers etc, based on the purpose of the data gathered. This implies it is not always possible to use data unless the purpose has been made clear upfront. Moreover it is imperative that all users are informed about the fact that data is being collected, for what purpose and to what goal and that the persons of whom data is being gathered, provide their consent. In case of minors, consent needs also to be given by parents or legal guardians.

Whenever data is being collected, **all** data on an individual should be provided at **any** time they request it. Moreover, individuals have the right to request that all their data be removed at any time in retrograde. An exception can be made for data an institution is obliged to keep for other reasons (e.g. names, date of birth, final grades). In the context of learning analytics applications that transgress the boundaries of the institutional learning or MOOC environment consent must be obtained for collection of all data from all non-institutional sources (e.g. Twitter).

Whether data can be used is determined by contextual integrity and depends on whether learners are aware about the processes, whether it is clear what the possible consequences are for learners, and whether safeguards are in place. Circumstances in which it is not allowed or not advisable to collect and process data are those situations

- when data are collected without explicit purpose;
- when data are gathered outside the learning context
- when data are gathered of which the learner/participant is not aware
- when data are gathered that pose a risk to the learner/participant
- when data are not properly protected.

10.4 Confidentiality of data

Although some data as generated by participants in MOOCs will be publicly available to those who enrolled and can access the MOOC that does not necessarily mean that these data can be used for other purposes, including learning analytics. Even when consent has been obtained to track, log and record personal data, some data never can be made public. This applies to all logging data and the statements recording participant actions and events in the platform, such as the Tin Can statements. Proper care has to be taken to ensure that these LA services are private and secured.

When users register at the ECO platforms they should accept the general terms of use. These should regulate the permission to user data explicitly (e.g. showing names in forum posts) or anonymously in aggregated form (e.g. ratings, tags). In addition, purpose of tracking usage data and social data should be explained as well as the distribution of these data for providing personalised recommendations and advice.

Aggregated data should be used anonymously and therefore all identifying attributes have to be removed.

Users, i.e. learners and staff, remain owner of the data and should be able to determine privacy and access of their data and products.

10.5 Actions for ECO project

Although some general guidelines on Intellectual Property Rights (IPR), security and privacy are described in Deliverable 1.2 IPR, security and privacy report, the ECO project has to ensure that all platforms and services properly manage all aspects related to this.

The ECO platforms and learning analytics services and dashboards make extensive use of personal data of learners and staff like teachers and tutors, including tracking their behaviour and actions in the platforms. Consequently, these platforms and services have to be designed and operated in a legally compliant manner. Ownership of data should be passed to the user. Users should be properly informed; informed consent should be obtained for acceptance and agreement of the use of the personal data. Moreover, services should take into account that users have the right to have their data withdrawn at any time and be designed accordingly.

This entails that the ECO platforms have to present a Terms of use and a disclaimer to users of the platform. It is advisable to make this part of the registration process, but Terms of use and disclaimer should be easily accessible at any time. The terms of use should address rights and obligations both of the service provider and the user. Sections to consider are registration, right to use, social data collection and protection, general aspects as to validity of the terms, and legal aspects. A distinction should be made between the data that are generated because the participants make use of the platform and any additional information that is gathered through questionnaires and surveys. Participants can’t be forced to submit the survey but should be advised that ECO can provide them with better support when they fill out these questionnaires and surveys.

ECO has to ensure that only those data are logged and recorded that are needed to calculate the indicators as described in this deliverable, and to the purpose thereof.

Currently, multiple platforms are in use and hosted by individual consortium partners. The servers running those platforms could in theory be located anywhere. This can have repercussions as to what legislation is applicable. The ECO project needs to decide whether to adhere to a national or EU framework.

10.6 Terms of use

An example of terms of use according to the EU framework would be the following. The example pertains to questionnaires, but should equally be provided for use of the platform and participation in the MOOC.

> You need to agree with the following before you answer our questions:
>
> I agree to answer the questions collected via the ECO platforms, for the purpose of statistical records and analysis. I understand that the information I provide will only be used in processing the research data and presenting the results in publications of the research project named “ECO (Elearning, Communication and Open-data: Massive Mobile, Ubiquitous and Open Learning)”.
>
> I also understand that I have the right to refuse to participate in this questionnaire, that I have at any time the right to access, rectify and to request the deletion of any of my personal data that I have provided during this interview, and that I also have at any time the right to have recourse to the European Data Protection Supervisor.

Right to access data, Article 13 Regulation (EC) No. 45/2001 of the European Parliament and of the Council of 18 December 2000 on the protection of individuals with regard to the processing of personal data by the Community institutions and bodies and the free movement of such data. Hereafter referred to as Regulation 45/2001.

Right to rectify, Article 14 Regulation 45/2001.

Right to erase, Article 16 Regulation 45/2001.

Right to have recourse at any time to the European Data Protection Supervisor (EDPS). Article 41 Regulation 45/2001.)

The information will be kept in a secure environment by the ECO Consortium according to data protection guidelines (Article 5 Regulation 45/2001)

Contact details need to be added to allow the user to put in any requests.

10.7 References

Drachsler, H. (October 2014). Ethics & privacy issues in the application of learning analytics. Presentation at NSF Meeting - Big data in education. Washington, USA, 9 - 11 October 2014. Retrieved from http://de.slideshare.net/Drachsler/ethics-and-privacy-in-the-application-of-learning-analytics-ep4la

Greller, W., & Drachsler, H. (2012). Translating Learning into Numbers: A Generic Framework for Learning Analytics. Educational Technology & Society, 15 (3), 42–57.

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. British Journal of Educational Technology, 45(3), 438-450.

11 Conclusions

This deliverable describes how a learning analytics service should be implemented that takes the unique circumstances of MOOCs into account. MOOCs provide learning opportunities and as such learning analytics can be applied to provide learners with additional learner support to optimise their learning experience and performance, to provide teachers with indications of when additional actions are required and inform course designers about the efficiency and effectiveness of their courses. However, in contrast to conventional educational contexts, MOOCs – being massive open online course – pose additional challenges. The sheer number of participants is one, but a real challenge is the heterogeneity of the population of participants who are attracted to MOOCs. Although the course designer has designed the course with certain learning objectives, not all participants intend to follow those. They participate in the MOOC to satisfy their own learning objectives. The open character of MOOCs means that no prerequisites can be set and consequently that participants enter with a variety of educational background and knowledge level. That means that any support given should be aligned with their personal objectives. Moreover, the online character asks for a different educational approach.

There are further requirements that the learning analytics services need to deal with in the context of the ECO project because ECO wants to support social and seamless MOOCs, sMOOCs. ECO argues that MOOCs need to follow a pedagogical model that is based on social, networked learning and provide authentic situated learning experiences. Therefore the learning analytics should not only take academic goals into account, but also has to support social and networked learning aspects as well as those arising from ubiquitous learning.

Metrics and indicators have to deal with the fact that learners enter MOOCs with their own particular learning goals that might be different from the learning objective as set by the course designer. Moreover, the learning analytics need to adapt to any changes that arise in participants' intentions. Only a minority of MOOC participants tend to follow the course from start to finish as designed by the course designer and complete the course. Many participants are only interested in parts or particular subjects, or simply can't keep up. Learning analytics can be applied to assist those learners to get the best out of their participation in the MOOC, but needs to be able to determine the type of participants.

Visual analytics provide a powerful mechanism to display indicators to participants in an intuitive and meaningful way in the form of different graphs or charts as gauge controls, bars, progress bars, donut charts, stacked radars and other types of graphs. These visualisations are commonly done in dashboards. Dashboards capture and visualize traces of learning activities, in order to promote awareness, reflection, and sense-making, and to enable learners to define goals and track progress toward these goals. For learners the dashboard displays individual indicators for a particular learner, but also allows the learner to compare with others. While dashboards for teachers provide an aggregated overview over the various types of participants and groups of learners.

The choice of the xAPI specification to log the learner's behaviour in the platforms guarantees interoperability both between the various ECO platforms and with external platforms, because actions are registered in a standardised format.

Finally, learning analytics makes heavy use of personal data of individuals and therefore imposes several ethical concerns that need to be dealt with accordingly. Privacy and confidentiality of data need to be guaranteed and ECO needs to ensure that participants are informed properly.

The next step is to work with work package 3 to implement the learning analytics taking into account the features of each of the MOOC platforms.