

AFEL: Towards Measuring Online Activities Contributions to Self-Directed Learning

Mathieu d'Aquin¹, Alessandro Adamou², Stefan Dietze³, Besnik Fetahu³,
Ujwal Gadiraju³, Ilire Hasani-Mavriqi⁴, Peter Holtz⁵, Joachim Kimmerle⁵,
Dominik Kowald⁴, Elisabeth Lex⁴, Susana López Sola⁶], Ricardo A.
Maturana⁶, Vedran Sabol⁴, Pinelopi Troullinou², and Eduardo Veas⁴

¹ Insight Centre for Data Analytics, National University of Ireland, Galway
mathieu.daquin@insight-centre.org

² Knowledge Media Institute, The Open University, UK
{alessandro.adamou, pinelopi.troullinou}@open.ac.uk

³ L3S Research Center, Leibniz University Hanover, Germany
{dietze, fetahu, gadiraju}@l3s.de

⁴ Know-Center Graz, University of Technology Graz, Austria
{ihasani, dkowald, vsabol, eduveas}@know-center.at, elisabeth.lex@tugraz.at

⁵ The Leibniz-Institut für Wissensmedien, Tübingen, Germany
{p.holtz, j.kimmerle}@iwm-tuebingen.de,

⁶ GNOSS, Spain

{susanalopez, riam}@gnoss.com

Abstract. More and more learning activities take place online in a self-directed manner. Therefore, just as the idea of self-tracking activities for fitness purposes has gained momentum in the past few years, tools and methods for awareness and self-reflection on one's own online learning behavior appear as an emerging need for both formal and informal learners. Addressing this need is one of the key objectives of the AFEL (Analytics for Everyday Learning) project. In this paper, we discuss the different aspects of what needs to be put in place in order to enable awareness and self-reflection in online learning. We start by describing a scenario that guides the work done. We then investigate the theoretical, technical and support aspects that are required to enable this scenario, as well as the current state of the research in each aspect within the AFEL project. We conclude with a discussion of the ongoing plans from the project to develop learner-facing tools that enable awareness and self-reflection for online, self-directed learners. We also elucidate the need to establish further research programs on facets of self-tracking for learning that are necessarily going to emerge in the near future, especially regarding privacy and ethics.

1 Introduction

Much of the research on measuring learners' online activities, and to some extent much of the research work in Technology-Enhanced Learning, focus on the restricted scenario of students formally engaged in learning (e.g. enrolled in

a university program) and where online activities happen through a provided eLearning system. However, whether or not they are formally engaged in learning, more and more learners are now using a large variety of online platforms and resources which are not necessarily connected with their learning environment or with each other. Such use of online resources tends to be self-directed in the sense that learners make their own choices as to which resource to employ and which activity to realize amongst the wide choice offered to them (MOOCs, tutorials, open educational resources, etc). With such practices becoming more common, there is therefore value in researching the way in which to support such choices.

In several other areas than learning where self-directed activities are prominent (e.g. fitness), there has been a trend in recent years following the technological development of tools for self-tracking [15]. Those tools quantify a specific user's activities with respect to a certain goal (e.g. being physically fit) to enable self-awareness and reflection, with the purpose of turning them into behavioral changes. While the actual benefits of self-tracking in those areas are still debatable, our understanding of how such approaches could benefit learning behaviors as they become more self-directed remains very limited.

AFEL⁷ (Analytics for Everyday Learning) is an European Horizon 2020 project which aim is to address both the theoretical and technological challenges arising from applying learning analytics [6] in the context of online, social learning. The pillars of the project are the technologies to capture large scale, heterogeneous data about learner's online activities across multiple platforms (including social media) and the operationalization of theoretical cognitive models of learning to measure and assess those online learning activities. One of the key planned outcomes of the project is therefore a set of tools enabling self-tracking on online learning by a wide range of potential learners to enable them to reflect and ultimately improve the way they focus their learning.

In this paper, we discuss the research and development challenges associated with achieving those goals and describe initial results obtained by the project in three key areas: *theory* (through cognitive models of learning), *technology* (through data capture, processing and enrichment systems) and *support* (through the features provided to users for visualizing, exploring and drawing conclusions from their learning activities). We start by describing a motivating scenario of an online, self-directed learner to clarify our objective.

2 Motivating Scenario

Below is a specific scenario considering a learner not formally engaged in a specific study program, but who is, in a self-directed and explicit way, engaged in online learning. The objective is to describe in a simple way how the envisioned AFEL tools could be used for self-awareness and reflection, but also to explore what the expected benefits of enabling this for users/learners are:

⁷ <http://afel-project.eu>

Jane is 37 and works as an administrative assistant in a local medium-sized company. As hobbies, she enjoys sewing and cycling in the local forests. She is also interested in business management, and is considering either developing in her current job to a more senior level or making a career change. Jane spends a lot of time online at home and at her job. She has friends on Facebook with whom she shares and discusses local places to go cycling, and others with whom she discusses sewing techniques and possible projects, often through sharing YouTube videos. Jane also follows MOOCs and forums related to business management, on different topics. She often uses online resources such as Wikipedia and online magazines. At school, she was not very interested in maths, which is needed if she wants to progress in her job. She is therefore registered on Didactalia⁸, connecting to resources and communities on maths, especially statistics.

Jane has decided to take her learning seriously: She has registered to use the AFEL dashboard through the Didactalia interface. She has also installed the AFEL browser extension to include her browsing history, as well as the Facebook app. She has not included in her dashboard her emails, as they are mostly related to her current job, or Twitter, since she rarely uses it.

Jane looks at the dashboard more or less once a day, as she is prompted by a notification from the AFEL smart phone application or from the Facebook app, to see how she has been doing the previous day in her online social learning. It might for example say “It looks like you progressed well with sewing yesterday! See how you are doing on other topics...” Jane, as she looks at the dashboard, realizes that she has been focusing a lot on her hobbies and procrastinated on the topics she enjoys less, especially statistics. Looking specifically at statistics, she realizes that she almost only works on it on Friday evenings, because she feels guilty of not having done much during the week. She also sees that she is not putting as much effort into her learning of statistics as other learners, and not making as much progress. She therefore makes a conscious decision to put more focus on it. She adds new goals on the dashboard of the form “Work on statistics during my lunch break every week day” or “Have achieved a 10% progress compared to now by the same time next week”. The dashboard will remind her of how she is doing against those goals as she goes about her usual online social learning activities. She also gets recommendations of things to do on Didactalia and Facebook based on the indicators shown on the dashboard and her stated goals.

While this is obviously a fictitious scenario, which is very much simplified, it shows the way tools for awareness and self-reflection can support online self-directed learning, and it provides a basis to investigate the challenges to address in order to enable the development of tools of the kind that are described, as discussed in the rest of this paper.

⁸ <http://didactalia.net>

3 Theoretical Challenge: Measuring Self-Directed Learning

One result of the advent of the Internet as a mass phenomenon was a slight change in our understanding of constructs such as “knowledge” and “learning”. In such contexts as described above, it is by no means a trivial task to identify and to assess learning. Indeed, in order to understand how learning emerges from a collection of disparate online activities, we need to get back to fundamental, cognitive models of learning, as we cannot make the assumption that usual ways to test the results of learning are available.

Traditionally, the acquisition metaphor was frequently used to describe learning processes [19]: From this perspective, learning consists in the accumulation of “basic units of knowledge” within the “container” (p. 5) of the human mind. Already before the digital age, there was also an alternative, more socially oriented understanding of learning, which is endowed in the participation metaphor: Here, knowing is equaled to taking up and getting used to the customs and habits of a community of practice [10], into which a learner is socialized. Over the last two decades however, the knowledge construction metaphor has emerged [17] as a third important metaphor of learning. Building upon a constructivist understanding of learning, the focus lies here on the constant creation and re-creation of knowledge within knowledge construction communities. Knowledge is no longer thought of as a rather static entity in form of a “justified true belief”; instead, knowledge is constantly re-negotiated and evolves in a dynamic way [16]. In this tradition, the co-evolution model of learning and knowledge construction [2] treats learning on the side of individuals and knowledge construction on the side of communities as two structurally coupled processes (see Figure 1). Irritations of a learner’s cognitive system in form of new or unexpected information that has to be integrated into existing cognitive structures can lead to learning processes in the form of changes in the learner’s cognitive schemas, behavioral scripts, and other cognitive structures. In turn, such learning processes may trigger communication acts by learners within knowledge construction communities and stimulate further communication processes that lead to the construction of new knowledge. In this model, shared artifacts, for example in form of digital texts such as contributions to wikis or social media messages, mediate between the two coupled systems of individual minds and communicating communities [8].

When talking about learning in digital environments, we can consequently define learning as the activity of learners encountering at least partly new information in form of digital artifacts. In principle, every single interaction between a learner and an artifact can entail learning processes. Learning can either happen occasionally and accidentally or in the course of planned and at least partly structured learning activities [12]. Planned and structured learning activities can either be self-organized or follow to a certain degree a pre-defined curriculum of learning activities [13]. In both cases, the related activities will constitute a certain learning trajectory [21] which comprises of “the learning goal, the learning activities, and the thinking and learning in which the students might engage”

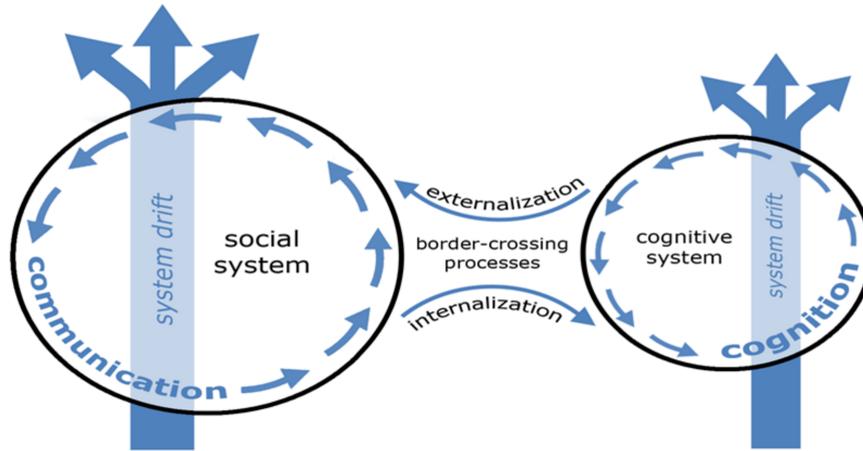


Fig. 1. The dynamic processes of learning and knowledge construction [8] (p. 128).

(p. 133). Successful learning will result in increases in the learner’s abilities and competencies; for example, successful learners will be able to solve increasingly difficult tasks or to process increasingly complex learning materials [23].

Based on these theoretical considerations, the challenge in building tools for self-tracking of online, self-directed learning is to recognize to what extent encountering and processing a certain artifact (a resource) induced learning. In the co-evolution model, we assume that what we can measure is the friction (or irritation) which triggers internalization processes, i.e. what does the artifact bring to the cognitive system that leads to its evolution. At the moment, we distinguish three forms of “frictions”, leading to three categories of indicators of learning:

- New concepts and topics: The simplest way in which we can think about how an artifact could lead to learning is through its introduction of new knowledge unknown to the learner. This is consistent with the traditional acquisition metaphor. In our scenario, this kind of friction happens for example when Jane watches a video about a sewing technique previously unknown to her.
- Increased complexity: While not necessarily introducing new concepts, an artifact might relate to known concepts in a more complex way, where complexity might relate to the granularity, specificity or interrelatedness with which those concepts are treated in the artifact. In a social system, the assumption of the co-evolution model is that the interaction between individuals might enable such increases in understanding of the concepts being considered through iteratively refining them. In our scenario, this kind of friction happens for example when Jane follows a statistics course which is more advanced than the ones she had encountered before.

- New views and opinions: Similarly, known concepts might be introduced “in a different light”, through varying points of views and opinions enabling a refinement of the understanding of the concepts treated. This is consistent with the co-evolution model in the sense that it can be seen either as a widening of the social system in which the learner is involved, or as the integration into different social systems. In our scenario, this kind of friction happens for example when Jane reads a critical review of a business management methodology she has been studying.

What appears evident from confronting the co-evolution model and the types of indicators described above with the scenario of the previous section is that such indicators and models should be considered within distinct “domains” of learning. Indeed, Jane in the scenario would relate to different social systems for example for her interest in sewing, cycling, business management and statistics. The concepts that are relevant, the levels of complexity to consider and the views that can be expressed are also different from each other in those domains.

We call those domains of learning *learning scopes*. In the remainder of this paper, we will therefore consider a learning scope to be an area or theme of interest to a learner (sewing, business, etc.) to which are attached (consciously or not) specific learning goals, as well as a specific set of concepts, topics and activities.

4 Technical Challenge: Making-Sense of Masses of Heterogeneous Activity Data

Considering the conclusions from the previous section, the key challenge at the intersection of theory and technology for self-tracking of online, self-directed learning is to devise ways to compute the kind of indicators that are useful to identify and approximate some quantification of the three types of frictions within (implicit/emerging) learning scopes. Before that, however, we have to face more basic technical challenges to set in place the mechanisms to collect, integrate, enrich and process the data necessary to compute those indicators.

4.1 Data capture, integration and enrichment

The AFEL project aims at identifying the features that characterize learning activities within online contexts across multiple platforms. With that, we contribute to the field of Social Learning Analytics that is based on the idea that new ideas and skills are not only individual achievements, but also the results of interaction and collaboration [20]. With the rise of the Social Web, online social learning has been facilitated due to the participatory and collaborative nature of the Social Web. This has posed several challenges for Learning Analytics: The (online) environments where learning activities and related features are to be detected are largely heterogeneous and tend to generate enormous amounts of data concerning user activities that may or may not relate to learning, and even

when they do, the relation is not guaranteed to be explicit. A key issue is that, even with an emerging theoretical model, there is no established model for representing the data for learning that can span across all the types of activities that might occur in online environments. With respect to data capture, it may be hard to track all relevant learning traces and some indicators such as readership data may be misleading due to switches between the online and offline world [4].

Therefore, AFEL adopted an approach to identify reliable data sources and to structure their capture process, which is based on an effort to classify data sources, rather than the data themselves. Such an exercise in classification is important as it is the result of an effort to understand what dimensions of the activities through the Web should be captured, before setting out to detect specific learning activity factors. The resulting taxonomy revolves around a core of seven types of entities that a candidate data source has a potential for describing; these are further specified into sub-categories that capture a specific set of dimensions, some of which are common to users and communities (e.g. learning statements), or to users (e.g. indicators of expertise) and learning resources (e.g. indicators of popularity). Those categories are at the core of the proposed AFEL Core Data Model⁹, an RDF vocabulary largely based on schema.org and which is, amongst other things, used to aggregate the datasets that AFEL makes publicly available¹⁰.

The following challenge for AFEL is to integrate data from a large number of sources into a shared platform, using the core data model to integrate and make them processable. The approach taken is to create a “data space”, which keeps most of the data sources intact at the time of on-boarding and being integrated at query time through a smart API, following the principles set out in [1]. Using this platform, the project has already created a number of tools, called extractors, which can extract data about user activities from several different platforms, creating a consistent and processable data space for each AFEL user who can choose to enable some of those tools. At the time of writing those extractors include browser extensions for extracting browsing history, applications for Facebook and Twitter, as well as analytics extractors for the Didactalia portal from AFEL partner GNOSS.¹¹ We also integrate resource metadata from several open sources related to learning.

Beyond data storage and integration, the key to enable extracting the features necessary to compute the kind of indicators mentioned in the previous section is to connect those datasets at a semantic level, i.e. to enrich the raw data into a more complete “Knowledge Graph”. In other words, connecting the different entities with each other and extracting from unstructured or semi-structured sources entities of interest that can connect the data from a wide range of places. In AFEL, we use entity linking approaches [5] as well as natural language pro-

⁹ <http://data.afel-project.eu/catalogue/dataset/afel-core-data-model/>

¹⁰ <http://data.afel-project.eu/catalogue/learning-analytics-dataset-v1/>

¹¹ <http://gnoss.com>

cessing [11] and specific feature extraction approaches to turn a user data space into such a semantically enriched knowledge graph. Examples for such feature extraction approaches are computing the complexity of a resource [3], determining the semantic stability of a resource [22], or to assess influencing factors in consensus building processes in online collaboration scenarios [7].

Additionally, AFEL provides a methodology to determine the characteristic features, which allow learning activities to be detected and described, and consequently the attributes that instantiate them, in different data sources identified within the project. This methodology facilitates an initial specification of the features relevant to learning activities by presenting an instantiation of them on some of key data sources. Furthermore, with our methodology, we also outline a top-down perspective of feature engineering indicating that features identified in AFEL are applicable in different use cases, in general online contexts and that they can be extracted from our data basis.

4.2 An example: Learning scopes and topic-based indicator in browsing history

In this section, we present a short pilot experiment in which we implemented an initial version of showing indicators based on topics included in the learning activities of a user (consistently with what described in Section 3). This relies on some of the technical aspects described above, including data capture and enrichment.

The data: We use approximately 6 weeks of browsing history data for a user, obtained through the AFEL browser extension¹², which pushes this information as the user is browsing the web. Each activity is described as an instance of the concept `BrowsingActivity` in the AFEL Core Data Model, with as properties the URL of the page accessed and the time at which it was accessed. In our illustrative example, this corresponds to 42 707 activities, making reference to 12 738 URLs of webpages.

Topic Extraction: The first step to extracting the learning scopes from the activity data is to extract the topics of each resource (webpage). For this, we first use DBpedia Spotlight¹³ to extract the entities referred to in the text in the form of Linked Data entities in the DBpedia dataset¹⁴. DBpedia is a Linked Data version of Wikipedia, where each entity is described according to various properties, including the categories in which the entity has been classified in Wikipedia. We therefore query DBpedia to obtain up to 20 categories from the ones directly connected to the entities, or their broader categories in DBpedia's category taxonomy.

¹² <https://github.com/afel-project/browsing-history-webext>

¹³ <http://spotlight.dbpedia.org>

¹⁴ <http://dbpedia.org>

For example, assume the learner views a YouTube video titled *LMMS Tutorial — Getting VST Instruments*.¹⁵ When mining the extracted text (stripped of HTML markup), DBpedia Spotlight detects that the description of this video mentions entities such as `<http://dbpedia.org/resource/LMMS>` (dbp:LMMS for short - a digital audio software suite) or `dbp:Virtual_Studio_Technology`. Querying DBpedia reveals subject categories for dbp:LMMS, such as `<http://dbpedia.org/resource/Category:Free_audio_editors>` (or short, `dbc:Free_audio_editors`) or `dbc:Software_drum_machines`. The detected category `dbc:Free_audio_editors` is in turn declared in DBpedia to have broader categories such as `dbc:Audio_editors` or `dbc:Free_audio_software`. All of these elements are included in the description of the activity that corresponds to watching the above video, to be used in the next step of clustering activities.

On our browsing history data, running the resources through DBpedia Spotlight extracted 20 876 distinct entities, each being added 20 categories on average. To give an idea of the scale, the final description of the 6 weeks of activities of this one learner takes approximately 1.1GB of space and took between 1 and 15 seconds to compute for each activity (depending on the size of the original text, using a modern laptop with a good internet connection).

Clustering activities: In the next step, we use the description of the activities as produced through the process described above in order to detect candidate learning scopes, i.e. groups of topics and activities that seem to relate to the same broader theme. To do this, we consider the set of entities and categories obtained before similarly to the text of documents and apply a common document clustering process on them (i.e. TFIDF vectorization and k-Means clustering). We obtain from this a set of k clusters (with k being a given) that group activities based on the overlap they have in the topics (entities and categories) they cover. We label each cluster based on the entity or category that best characterizes it in terms of F-Measure (i.e. that covers the maximum number of activities in the cluster, and the minimum number of activities outside the cluster), representing the target of the topic scope.

The clustering technique we applied (k-Means) requires to fix the number of clusters to be obtained in advance. We experimented with numbers between 6 and 100, to see which could best represent the width and breadth of interests of this particular learner. Here, we used 50 as it appeared to lead to good results (as future work, we will integrate ways to automatically discover the ideal number of clusters for a learner). Figure 2 shows the clusters obtained and their size. The gray line describes all activities in the topic scope, i.e. all activities that have been included in the cluster. As can be seen, the clusters are unbalanced between the ones with a large number of activities (Google, Web Programming) with thousands of activities, and the ones representing only a few hundreds of activities.

Topic-based indicator: In the initial scenario we are considering here, we focus on a topic-based indicator which consist in checking whether an activity introduces

¹⁵ <https://www.youtube.com/watch?v=aZKra7rNspU>

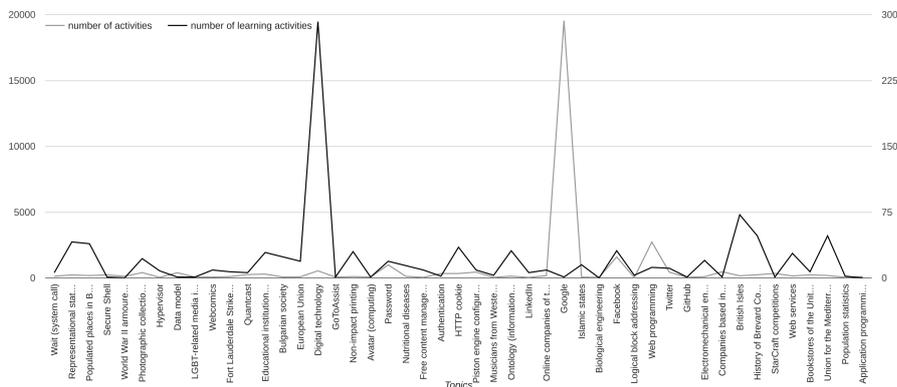


Fig. 2. Topic scopes obtained from the learner’s browsing activities. The gray line and left axis indicate the size of the cluster in total number of activities. The black line and right axis only include activities detected as being learning activities.

new topics (entities or categories) into the learning scope (cluster) in which it is included. We therefore “play back” the sequence of browsing activities from the learner’s history, checking at each time how many new topics are being introduced that were not present in the previous activities of the learner in this scope.

Looking again at Figure 2, it is interesting to look at the difference between the gray line (number of activities in the topic scope) and the black line, representing the number of activities that have integrated new topics in the scope and can therefore be considered learning activities. For example, since the user uses many Google services for basic tasks (such as Gmail for emails), it is not surprising that the Google scope, while being the largest in activities, does not actually include much detected learning activities. What is obvious however is that the balance is much different for other clusters that can be clearly identified for including large amounts of learning activities.

Indeed, we can see the value of the process here by comparing the learning trajectories of the learner according to the definition of contributions to different learning scopes considered. For example, the scope on Digital Technology, representing the largest number of learning activities, can be seen in Figure 3 (top) as a broad topic on which the learner is constantly (almost everyday) learning new things. In contrast, the learning scope on Web Programming, although very related, is one where we can assume the learner already has some familiarity and only makes a significant increment in their learning punctually, as can be seen by the jump around 08 September in Figure 3 (bottom).

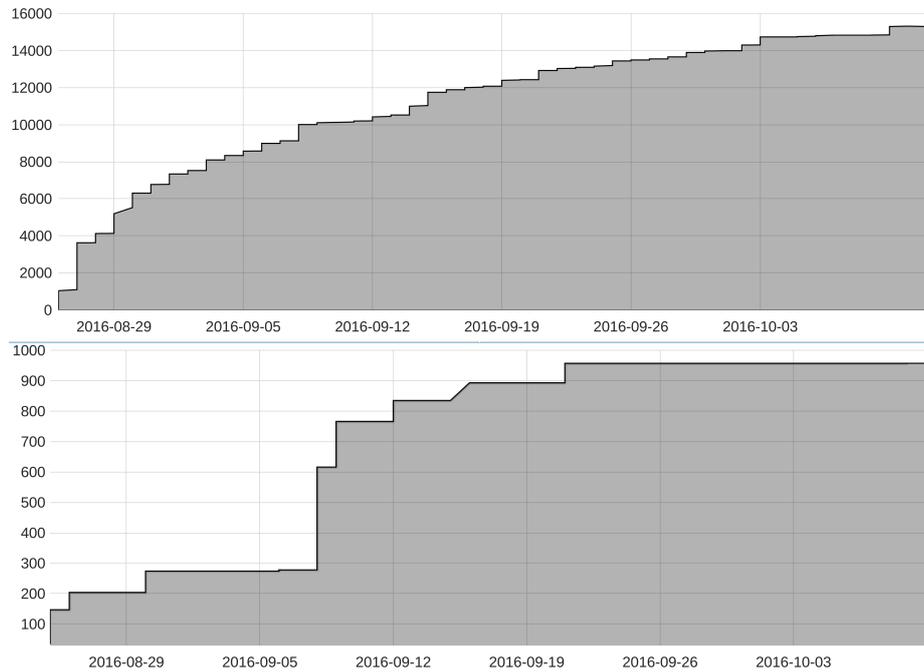


Fig. 3. Trajectory in terms of the contribution (in number of topics) to the learning scope in Digital Technology (top) and Web Programming (bottom).

5 Support Challenge: From Metrics to Actions

The current state in the implementation of the aforementioned aspects takes the form of a prototype learner dashboard, available from the Didactalia platform. The dashboard illustrated in Figure 4 includes initial placeholder indicators for the kind of frictions identified in Section 3 and is implemented on the technologies described above. It is however a preliminary result, showing the ability to technically integrate the different AFEL components into an first product. It will be further evolved in order to truly address the scenario of Section 2, including user feedback and more accurate indicators.

A key aspect to achieve the goal in our everyday learning scenario is that the user should have control over what is being monitored. Indeed, the learner should be able to decide what area of the data should be displayed, according to which indicator and which dimension of the data (e.g. specific topics, times, resources or platforms). Our approach here is to rely on a framework for flexible dashboards based on visualization recommendation, implemented through the VizRec tool [14]. At the root of VizRec lies a visualization engine that extracts the basic features of the data and guiding the user in choosing appropriate ways



Fig. 4. Screenshot of the prototype learner dashboard.

to visualise them. Hereby, a learning expert may design a dashboard with an initial view of set of learning indicators, but VizRec also empowers the user in choosing what area of the data to show. This includes the ability to add new charts to the dashboard that can be selected based on the characteristics of the data (e.g. show a map for geo-graphical data). The tool can learn the user preferences, and therefore show a personalized dashboard which is always consistent with the visualization choices made by the user. Figure 5 shows an example of VizRec displaying multidimensional learning data. A scatterplot correlates the number of previous attempts with studied credits, showing that the number of previous attempts is smaller when studied credits is high. The grouped bar chart displays the number of previous attempts for female (right) and male (left) students, with genders being further subdivided by the highest level of education (encoded by color). It is obvious that education level has a very similar effect for both females and males. Notice that in the VisPicker (shown on right) only some visualizations are enabled, which is a direct consequence of the data dimensions which were chosen by the user: gender, highest education, number of previous attempts (shown on left). The user is free to choose only the enabled, meaningful visualizations, with the optional possibility of the system recommending the optimal representation based on previous user behavior. As the title of this

section calls, it is important to move from metrics to action and consider what the learner should do, having seen her status.

One way to move the learner to action is via recommending learning resources that appear to be relevant considering the current state of the learner [9]. Here, the monitoring of learning activities has a direct benefit in supplying recommendations to the learner. The current implementation of such a recommender system is based on two well-known approaches: (i) Content-based filtering, which recommends similar resources based on the content of a given resource, and (ii) Collaborative filtering, which recommends resources of similar users based on the learning activities of a given user [18].

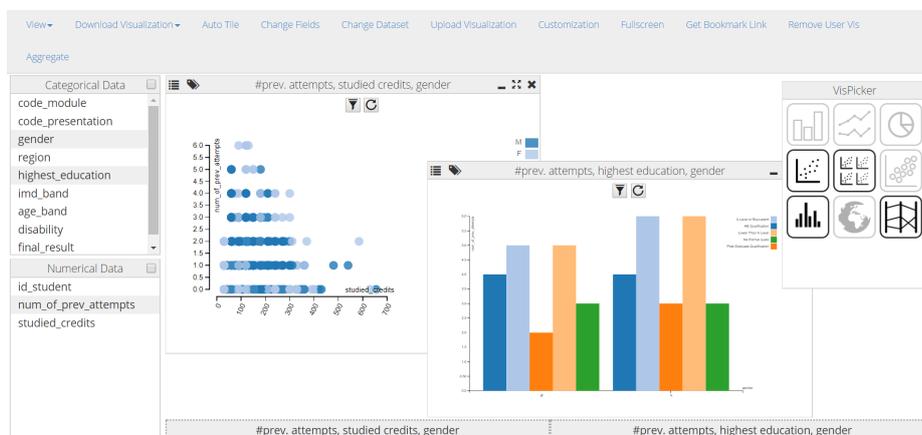


Fig. 5. Example use of the VizRec tool for personalized dashboards.

However, an important aspect, which is still missing, is how such measures of similarity can be based on metrics that are relevant to learning rather than on basic content or profile similarity. Indeed, the objective here would be to recommend learning resources (or even learning resource paths) that have already been helpful for other users with a similar learning goal and a similar learning state (in terms of the concepts, complexities and views already encountered). In other words, the recommendations can be based on a meaningful view of what the suggested resources might contribute to learning.

6 Discussion: Towards Wide-Availability, Ethical Tools for Self-Tracking of Online Learning

In the previous section, we discussed how to theoretically and technically implement tools for self-awareness targeted at self-directed online learning. Those tools are currently at early stages of development. Beyond those aspects however, other challenges will be faced by the AFEL consortium. One of them includes facilitating the adoption of these tools by a wide variety of users. Indeed, the actual usefulness and value of such personal analytics dashboards and learning

assistant technologies have not been formally assessed and the participation of the learner community in their development is necessary in order to ensure that they reach their potential. The approach taken by AFEL here is to start with the community of learners in the Didactalia platform, enabling the dashboard for them and through that, supporting them in integrating data from other platforms. With a large number of users, we will be able to collect enough data to understand how such monitoring can truly support users in reaching awareness of their learning behavior, and how this can help them take decisions with respect to their own learning.

Another aspect which is not discussed in this paper is the ethical implications of realizing such tools and reaching a wide-adoption. As mentioned above, each of the learners is assigned their own data space on the AFEL platform, which is only accessible by them. However, as mentioned in the scenario of Section 2, support to the learner might be better achieved by enabling them to compare their own behavior with others, and we aim to make some aggregated data available to others for research purposes. Proper anonymisation techniques need to be applied in order to ensure that external parties cannot infer information about specific learners from having access to those tools and data.

Beyond privacy however, it is also important to ensure that the effect of the tool might not turn out to be negative. Existing work have shown a number of ethical harms that might come out of enabling self-governance in a number of domains, despite the obvious positive effects [24]. Those include introducing biases towards common learning behaviors or pushing learners towards excessive behaviors for the purpose of improving the values of indicators that are necessarily only approximate representations of learning. Activities within and connected to the AFEL project have for specific objective to tackle those aspects, through establishing contrasting scenarios of the possible effect of self-tracking tools as a basis to engage with users of those tools about the ways to avoid the negative effects while keeping the positive ones.

Acknowledgement

This work has received funding from the European Union's Horizon 2020 research and innovation programme as part of the AFEL (Analytics for Everyday Learning) project under grant agreement No 687916.

References

1. A. Adamou and M. d'Aquin. On requirements for federated data integration as a compilation process. In *Proceedings of the PROFILES 2015 workshop*, 2015.
2. U. Cress and J. Kimmerle. A systemic and cognitive view on collaborative knowledge building with wikis. *International Journal of Computer-Supported Collaborative Learning*, 2(3), 2008.
3. S. A. Crossley, J. Greenfield, and D. S. McNamara. Assessing text readability using cognitively based indices. *Tesol Quarterly*, 42(3):475–493, 2008.
4. M. De Laat and F. R. Prinsen. Social learning analytics: Navigating the changing settings of higher education. *Research & Practice in Assessment*, 9, 2014.

5. S. Dietze, S. Sanchez-Alonso, H. Ebner, H. Qing Yu, D. Giordano, I. Marenzi, and B. Pereira Nunes. Interlinking educational resources and the web of data: A survey of challenges and approaches. *Program*, 47(1):60–91, 2013.
6. R. Ferguson. Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4(5-6):304–317, 2012.
7. I. Hasani-Mavriqi, F. Geigl, S. C. Pujari, E. Lex, and D. Helic. The influence of social status and network structure on consensus building in collaboration networks. *Social Network Analysis and Mining*, 6(1):80, 2016.
8. J. Kimmerle, J. Moskaliuk, A. Oeberst, and U. Cress. Learning and collective knowledge construction with social media: A process-oriented perspective. *Educational Psychologist*, (50), 2015.
9. S. Kopeinik, E. Lex, P. Seitlinger, D. Albert, and T. Ley. Supporting collaborative learning with tag recommendations: a real-world study in an inquiry-based classroom project. In *LAK*, pages 409–418, 2017.
10. J. Lave and E. Wenger. *Situated learning: Legitimate peripheral participation*. Cambridge University Press, 1991.
11. C. D. Manning, M. Surdeanu, J. Bauer, J. R. Finkel, S. Bethard, and D. McClosky. The stanford corenlp natural language processing toolkit. In *ACL (System Demonstrations)*, pages 55–60, 2014.
12. V. Marsick and K. Watkins. Lessons from informal and incidental learning. *Management learning: Integrating perspectives in theory and practice*, 1997.
13. C. McLoughlin and M. Lee. Personalised and self regulated learning in the web 2.0 era: International exemplars of innovative pedagogy using social software. *Australasian Journal of Educational Technology*, 1(26), 2010.
14. B. Mutlu, E. Veas, and C. Trattner. Vizrec: Recommending personalized visualizations. *ACM Trans. Interact. Intell. Syst.*, 6(4):31:1–31:39, Nov. 2016.
15. G. Neff and D. Nafus. *The Self-Tracking*. MIT Press, 2016.
16. A. Oeberst, J. Kimmerle, and U. Cress. What is knowledge? who creates it? who possesses it? the need for novel answers to old questions. *Mass collaboration and education*, 2016.
17. S. Paavola, L. Lipponen, and K. Hakkarainen. Models of innovative knowledge communities and three metaphors of learning. *Review of Educational Research*, 4(74), 2004.
18. P. Seitlinger, D. Kowald, S. Kopeinik, I. Hasani-Mavriqi, T. Ley, and E. Lex. Attention please! a hybrid resource recommender mimicking attention-interpretation dynamics. In *Proceedings of the 24th International Conference on World Wide Web, WWW '15 Companion*, pages 339–345, New York, NY, USA, 2015. ACM.
19. A. Sfard. On two metaphors for learning and the dangers of choosing just one. *Educational Researcher*, 2(27), 1998.
20. S. B. Shum and R. Ferguson. Social learning analytics. *Journal of educational technology & society*, 15(3):3, 2012.
21. M. Simon. Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education*, 2(26), 1995.
22. D. Stanislavljevic, I. Hasani-Mavriqi, E. Lex, M. Strohmaier, and D. Helic. Semantic stability in wikipedia. In *International Workshop on Complex Networks and their Applications*, pages 379–390. Springer, 2016.
23. H. Stubbé and N. Theunissen. Self-directed adult learning in a ubiquitous learning environment: A meta-review. In *Proceedings of the First Workshop on Technology Support for Self-Organized Learners*, 2008.
24. J. R. Whitson. Foucaults fitbit: Governance and gamification. *The Gameful World - Approaches, Issues, Applications*, 2014.