Attention Please!
Learning Analytics for Visualization and Recommendation

Erik Duval
Dept. Computer Science
Katholieke Universiteit Leuven
Celestijnenlaan 200A
B3000 Leuven, Belgium
erik.duval@cs.kuleuven.be

ABSTRACT
This paper will present the general goal of and inspiration for our work on learning analytics, that relies on attention metadata for visualization and recommendation. Through information visualization techniques, we can provide a dashboard for learners and teachers, so that they no longer need to "drive blind". Moreover, recommendation can help to deal with the "paradox of choice" and turn abundance from a problem into an asset for learning.

1. INTRODUCTION
Attention is a core concern in learning: as learning resources become available in more and more abundant ways, attention becomes the scarce factor, both on the side of learners as well as on the side of teachers. (This is a wider concern, as we evolve towards an 'attention economy' [10].)

Learners and teachers leave many traces of their attention: some are immediately obvious to others, for instance in the form of posts and comments on blogs, or as twitter messages. These explicit traces are human readable, but can be difficult to cope with in a world of abundance [29]. Although some refer to 'information overload', we prefer Shirky's "filter failure" as a way to think about the problem of dealing with this abundance [30]. In any case, human attention traces are extremely valuable, but do not scale very well.

In this paper, we will explain how machine readable traces of attention can be used to filter and suggest, provide awareness and support social links.

This paper is structured as follows: section 2 provides a brief background on the field of analytics in general. The section thereafter focuses on applications from the world of jogging, as these provide a particularly rich source of inspiration for our work. The general concept of goal oriented visualizations is at the core of learning dashboard applications: that is why it is the topic of section 4. In order to scale up the current work on learning analytics and achieve broad adoption, it is imperative to establish a global open infrastructure, as we briefly explain in section 5. The two next sections briefly present the two approaches we've explored so far to leverage learning analytics: learning dashboard (section 6) and learning recommenders (section 7). Before concluding the paper, we briefly mention exciting opportunities that learning analytics provides for data based research on learning in section 8.

This text is a PREPRINT. Please cite as: E. Duval Attention Please! Learning Analytics for Visualization and Recommendation. To appear in: Proceedings of LAK11: 1st International Conference on Learning Analytics and Knowledge 2011.

2. BACKGROUND ON ANALYTICS
The new field of learning analytics is quite related to similar evolutions in other domains, such as Big Data [1], e-science [14, 9, 26], web analytics [7], educational data mining [25]. All of these have in common that they rely on large collections of quite detailed data in order to detect patterns. This detection of patterns can be based on data mining techniques, so that for instance recommendations can be made for resources, activities, people, etc. that are likely to be relevant. Alternatively, the data can be processed so that they can be visualized in a way that enables the teacher or learner rather than the software to make sense of them.

In fact, the research in my team gets much of its inspiration from tools like wakoopa (`http://social.wakoopa.com`) or rescuetime (`http://www.rescuetime.com`), that install tracker tools on the machine of a user and then automatically record all activities (applications launched, documents accessed, web sites visited, music played, etc.) by that user.

A typical illustration of the visualizations that such tools provide is figure 1, where a simple overview is presented of the software applications I used last week and last month and how their usage is distributed over time. (Tuesday-Thursday were travel days...) In this way, such an application can help a user to be more aware of her activities.

Moreover, based on these tracking data, the wakoopa tool can compare the activity of a user with that of other users and recommend software or contacts - see figure 2. It doesn't require much imagination to see how similar visualizations could be useful in the world of learning, for instance to chart learning activities, tools used or recommended, and peer learners or suitable teachers in one's social network.

It is important to note that the tracking occurs without any manual effort by the user - although it is of course important

Figure 1: My wakoopa dashboard

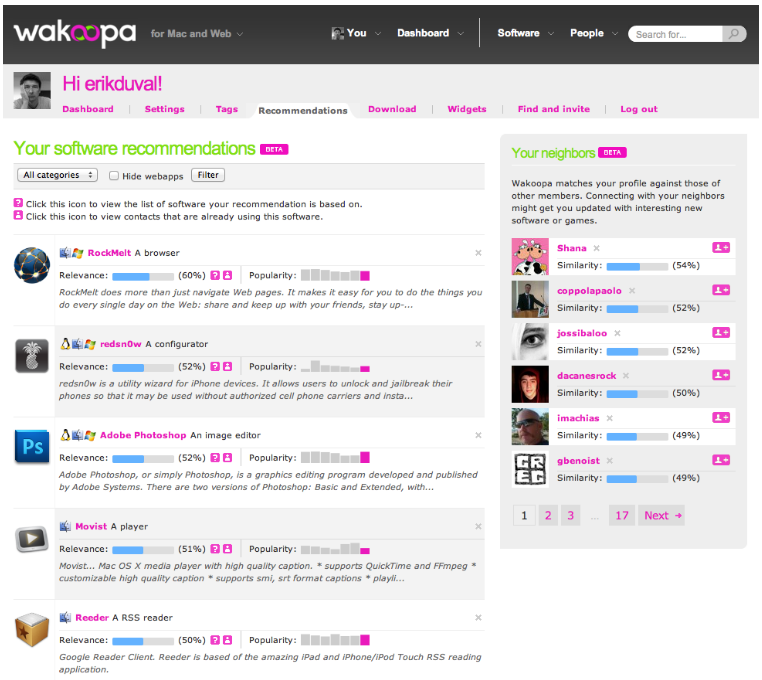

Figure 2: My Wakoopa recommendations

that the user is aware that her activities are being tracked. Actually, such tools typically also make it possible to pause tracking. Some applications allow users to set goals ("spend less than 1 hour per day on email" or "play computer games for less than 1,5 hour per day" or "write more than 3 hours per day") and will notify them when they are in danger of not meeting their goal, when they get close to the self-imposed limit - or signal them that they did reach their goal. Moreover, they provide quite detailed visualizations of all the activities of a user, so that she can analyze where most of her on-line activity takes place and make better informed decisions on how to manage these activities.

A similar tool is tripit (`http://www.tripit.com`): noteworthy about this tool is that when a user forwards flight or hotel reservations to a tripit email address, all the structured data is extracted and a calendar is created with all the relevant information. This is an excellent example of automatic metadata generation [6] or information extraction [5], an essential technology if we want to collect metadata of resources, activities and people at scale. Note also that, if other people from the user's network are near, tripit will mention that - see figure 3.

Of course, more mainstream tools like google offer similar functionality, such as for instance "google history" that provides an overview of every search that a user ever submitted (when logged in) or that indicates who from a user's social circle tweeted about an item included in a search result, etc.

3. INSPIRATION FOR OUR WORK

A particularly inspiring set of applications comes from the domain of jogging, and sports in general, where applications like nikeplus (`http://nikerunning.nike.com/`, see figure 4) or runkeeper (`http://runkeeper.com/`) provide detailed statistics on how fast, far, often, etc. one runs.

What is particularly relevant in a learning context is that many running applications also help runners to set goals ("run a marathon in november"), develop a plan to achieve that goal, find running routes in a foreign town, locate other runners with a similar profiles, challenge them so as to maintain motivation, etc. Sometimes, such tools even take a more pro-active role and send messages to users to enquire why they have stopped uploading activities, whether they need to re-define goals and plans, or want to be connected to other users that can help, etc.

Although there are few studies that show whether these special purpose social networks actually change user behavior, [16] did find that "users' weight changes correlated positively with the number of their friends and their friends' weight-change performance" and "users' weight changes have rippling effects in the Online Social Networks due to the social influence. The strength of such online influence and its propagation distance appear to be greater than those in the real-world social network.". An early overview of how the combination of tracking and social network services can lead to a more patient-driven approach to medicine is provided in [32].

One assumption underlying our work is that similar applications can be built to track learner progress, to assist in developing and maintaining motivation, to help define realistic goals and develop plans to achieve them, as well as connect learners or teachers with other learning actors, etc. In that way, they can help to realize a more learner-driven approach to education, training and learning in general.

4. GOAL ORIENTED VISUALIZATIONS

Many of these inspiring applications take a visual approach. Yet, if visualizations are to have any effect beyond the initial "wow" factor, it would be useful to have more clarity on what the intended goal is and how to assess whether that goal is achieved. Many visualizations look good - and some are actually beautiful. But how we can connect visualization not only with meaning or truth, but with taking actions? This is very much a "quantified self" approach (see `http://quantifiedself.com/`) [31], where for instance a visualization of eating habits can help to lead a healthier life, or where a visualization of mobility patterns can help to explore alternative modes of transport, etc. Such visualizations are successful if they trigger the intended behaviour (change). That can be measured, as in "people smoke less when they use this visualization" or "people discover new publications based on this visualization" (we are actually evaluating such an application) or "people run more using this visualization" etc.

It would be really useful if we could draw up some guidelines to design effective goal oriented visualizations. As an example, it is probably kind of useful to be able to visualise progress towards a goal - or lack thereof. If you want to run further, a visualization can help you to assess whether you're making progress. Or if you want to spend less time doing email, a simple visualization can help. Another guideline could relate to social support, that enables you to compare your progress with that of others.

5. TECHNICAL INFRASTRUCTURE FOR LEARNING ANALYTICS

If we want to apply learning analytics at a broader scale, then it is imperative that we realize an infrastructure that can support the development of tools and services. Such an infrastructure will need basic technical agreement on common standards and protocols [8].

A first question is how to model the relevant data. Our early work on Contextualized Attention Metadata (CAM) [19] [36] defines a simple model to structure attention metadata, i.e. the interactions that people have with objects. The ontology-based user interaction context model (UICO) [24] focuses more on the tasks that people carry out while interacting with resources. Either we need to better understand how to map and translate automatically between different such models, or we need to find a way to achieve broad consensus on and adoption of a common schema or a small set of schemas, as in the case of learning resources where nearly everyone has now adopted Learning Object Metadata (LOM) or Dublin Core (DC) [8].

Similar to the way we manage learning objects and their metadata [33], we will need a service architecture that can power a plethora of tools and applications. One interesting approach is to rely on technologies like widgets that

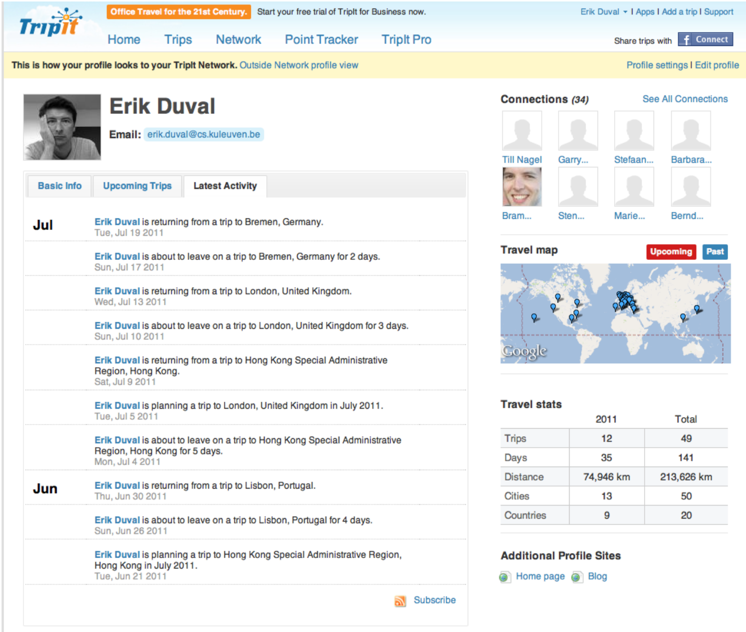

Figure 3: My tripit dahsboard

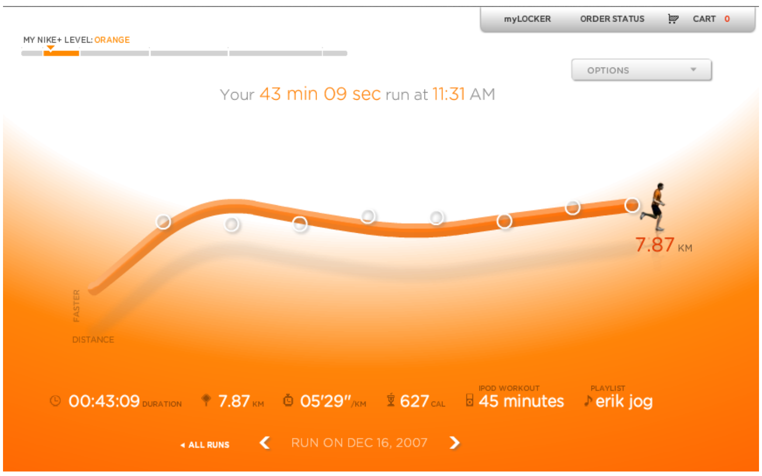

Figure 4: My Nike Plus dashboard

enable the dynamic embedding of small application components - an approach at the core of Personal Learning Environments (PLE's), researched in the ROLE project on Responsive Open Learning Environments (see `http://www.role-project.eu/`) [15] [12]. Another approach is the Learning Registry architecture that makes "user data trails" available through a network of nodes that provide services to publish, access, distribute, broker or administer paradata (see `http://www.learningregistry.org/`).

6. LEARNING DASHBOARDS

For learners and teachers alike, it can be extremely useful to have a visual overview of their activities and how they relate to those of their peers or other actors in the learning experience.

In fact, such visualizations can also be quite useful for other stakeholders, like for instance system administrators. Figure 5 provides an early example of such a visualization that displays the number of events in different widgets deployed in the ROLE context [27]. From the visualization, it is rather obvious that users were most active in the May-July period (towards the left of the diagram), that they enter chat rooms (top of area 4 on Figure 5) much more often than they post messages (third row of area 4 on Figure 5), etc. Such information can help a teacher to re-organize the activities or even to retract or add widgets that learners can deploy in their PLE.

Similarly, [28] describes a tool that includes a "zeitgeist" of action types (opening a document, sending a message, etc.) and specific user actions. By selecting a time period and the relevant action types, the user can control the visualization of relevant data (see also `http://www.role-showcase.eu/role-tool/cam-zeitgeist`).

Following a similar visual approach, the Student Activity Monitor (SAM) supports self-monitoring for learners and awareness for teachers [13]: In area A on figure 6, every line represents a student in a course. The horizontal axis represents calendar time and the vertical axis total time spent. If the line ascends fast, then the student worked intensely during that period. If the line stays flat, the student did not work much on the course. For example, student s1 started late and worked very hard for a very short time. Student s2 started early and then worked harder in about the same period as student s1. At the bottom, a smaller version of the visualization is shown with a slider on top to select a part of the period for analysis of data dense areas. Area 2 displays global course statistics on time spent and document use. The colored dots represent minimum, maximum and average time spent per student and the time spent for the currently logged in user and for a user selected in one of the visualizations. The recommendation area in Box 3 enables exploration of document recommendations (see also section 7). The parallel coordinates in area B display

1. the total time spent on the course,
2. the average time spent on a document,
3. the number of documents used and
4. the average time of the day that the students work.

For example, the green line (the logged in user) works on average in the early evening and is spending an average time in line with the majority. He does not use so many different documents and on average looks at these for a short time. He scores the worst here. The average student of the class (in yellow) is also presented. This is a somewhat complex visualization, but our evaluation studies show that students considered the visualizations clear [13]. They rated the tools as usable, useful, understandable and organized.

A much more simple such visualization is edufeedr [21], where a matrix includes a row for every student that displays his progress along a series of assignments. A nice feature is that such progress can take place on the individual blog of the student, outside of the institutional Learning Management System (LMS), Virtual Learning Environment (VLE) or even institution provided PLE widgets. Rather, the coherence of the course is maintained through the trackback mechanism between the teacher blog and those of the students.

What these visualizations have in common is that they enable a learner or teacher to obtain an overview of their own efforts and of those of the (other) learners. This is the essence of our "dashboard" approach to visualizations for learning that remedy the "blind driving" that often occurs on the side of teachers and learners alike. Similar approaches have proven to be beneficial in for instance software engineering [3] and social data analysis [18].

7. LEARNING RECOMMENDERS

By collecting data about user behavior, learning analytics can also be mined for recommendations, of resources, activities or people [17]. In this way, we can turn the abundance of learning resources into an asset, by addressing "l'embarras du choix" that is at the core of "the paradox of choice" [29]. Of course, similar approaches have been deployed for books, music, entertainment, etc. Yet, only by basing recommenders on detailed learning attention metadata can they take into account the learning specific characteristics and requirements of our activities.

In one particular tool, we applied this approach to filter and rank search results when a learner searches for material in YouTube (`http://www.youtube.com/`), SlideShare (`http://www.slideshare.net/`) and Globe (`http://www.globe-info.org/`): as figure 7 illustrates, every search activity in our tool is tracked in the form of attention metadata that are stored in a repository. The user can indicate whether search results are relevant or not and that feedback is also stored in the attention metadata repository. Search results are filtered and ranked based on earlier interaction by the user and by other users in her social network, as made available through OpenSocial. Although we need to do more user evaluations, the first results are very encouraging [11, 20].

8. DATA BASED RESEARCH ON LEARNING

On a meta-level, learning analytics provides exciting opportunities to ground research on learning in data and to transform it from what is currently all too often a collection of opinions and impossible-to-falsify conceptualizations and

Figure 5: The CAM dashboard [27]

Figure 6: The Student Activity Monitor (SAM) [13]

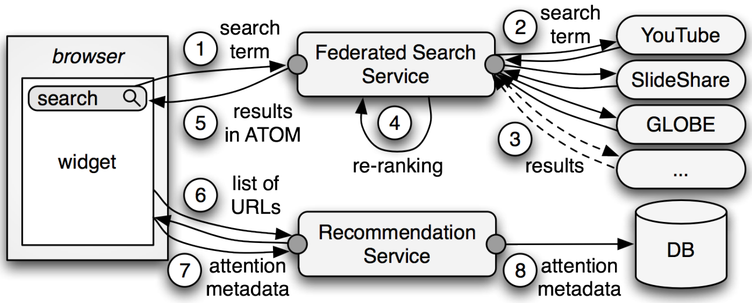

Figure 7: Storing attention metadata in federated search [11]

theories [23].

As a precursor to making that happen, it is important that we agree on ways to share data sets, in an "open science" approach [26, 9, 14]. That is why a group of interested researchers has started an initiative around "dataTEL" (`http://www.teleurope.eu/pg/groups/9405/datatel/` [34]. The main objective is to promote exchange and interoperability of educational data sets.

9. CONCLUSION

Of course, one of the big problems around learning analytics is the lack of clarity about what exactly should be measured to get a deeper understanding of how learning is taking place: typical measurements include time spent, number of logins, number of mouse clicks, number of accessed resources, number of artifacts produced, number of finished assignments, etc. But is this really getting to the heart of the matter?

Moreover, there are serious issues about privacy when detailed data of learner interactions are tracked [4]. An interesting early approach to deal with these issues was the proposal of the no longer active not for profit "Attention-Trust" [35]: their guiding principles included

- property: the data about a person's attention remain the property of that person;
- mobility: it should be possible to move data about a person out of one system and into another system - see also google's recent data liberation initiative (see `http://www.dataliberation.org/`);
- economy: a person should be able to sell data about his attention;
- transparency: it should always be clear to a person that she is being tracked.

Especially that last principle seems key: tools like ghostery (`http://www.ghostery.com/`) enable a user to know when she is being tracked on a web site. As we evolve towards a world where not only learning activities, but virtually everything will be tracked [2], this issue is likely to become even more important.

Some people are quite concerned about the "filter bubble" that personalization and recommendation engines may create [22]: we agree that there is a certain danger there, but we also believe that more advanced algorithms and ethical reflection can help us to address these issues.

In any case, we believe that learning analytics can be used to put the user in control, not to take control away in an Intelligent Tutoring Systems kind of way, by using attention to filter and suggest, provide awareness and support social links.

10. ACKNOWLEDGMENTS

This research has received funding from the European Community Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 231396 (ROLE) and no. 231913 (STELLAR). Much more importantly, the support, comments and feedback from my team and students have thought me much more than I will ever be able to teach them.

11. REFERENCES

[1] C. Anderson. The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired Magazine*, 16(7), 2008.

[2] G. Bell and J. Gemmel. *Your life, uploaded.* Plume, 2010.

[3] J. Biehl, M. Czerwinski, G. Smith, and G. Robertson. FASTDash: a visual dashboard for fostering awareness in software teams. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 1313–1322. ACM, 2007.

[4] D. Boyd. Facebook's Privacy Trainwreck: Exposure, Invasion, and Social Convergence. *Convergence: The International Journal of Research into New Media Technologies*, 14(1):13–20, 2008.

[5] C.-H. Chang, M. Kayed, M. R. Girgis, and K. F. Shaalan. A Survey of Web Information Extraction Systems. *IEEE Transactions on Knowledge and Data Engineering*, 18:1411–1428, 2006.

[6] N. Corthaut, S. Lippens, S. Govaerts, E. Duval, and J.-P. Martens. The integration of a metadata generation framework in a music annotation workflow. Oct. 2009.

[7] A. Croll and S. Power. *Complete Web Monitoring.* O'Reilly Media, Inc., 2009.

[8] E. Duval and K. Verbert. On the role of technical standards for learning technologies. *IEEE*

Transactions on Learning Technologies, 1(4):229–234, Oct. 2008.

[9] J. Fry, R. Schroeder, and M. den Besten. Open science in e-science: contingency or policy? *JOURNAL OF DOCUMENTATION*, 65(1):6–32, 2009.

[10] M. H. Goldhaber. The Attention Economy and the Net. *First Monday*, 2(4), Apr. 1997.

[11] S. Govaerts, S. E. Helou, E. Duval, and D. Gillet. A Federated Search and Social Recommendation Widget. In *Proceedings of the 2nd International Workshop on Social Recommender Systems (SRS 2011) in conjunction with the 2011 ACM Conference on Computer Supported Cooperative Work (CSCW 2011)*, pages 1–8, 2011.

[12] S. Govaerts, K. Verbert, D. Dahrendorf, C. Ullrich, S. Manuel, M. Werkle, A. Chatterjee, A. Nussbaumer, D. Renzel, M. Scheffel, M. Friedrich, J. L. Santos, E. Duval, and E. L.-c. Law. Towards Responsive Open Learning Environments : the ROLE Interoperability Framework. In *ECTEL11: European Conference on Technology Enhanced Learning*, Lecture Notes in Computer Science, 2011.

[13] S. Govaerts, K. Verbert, J. Klerkx, and E. Duval. Visualizing Activities for Self-reflection and Awareness. In *Proceedings of the 9th international conference on Web-based Learning*, pages 91–100. Springer, 2010.

[14] T. Hey and A. E. Trefethen. Cyberinfrastructure for e-Science. *Science*, 308(5723):817–821, 2005.

[15] U. Kirschenmann, M. Scheffel, M. Friedrich, K. Niemann, and M. Wolpers. Demands of Modern PLEs and the ROLE Approach. In M. Wolpers, P. Kirschner, M. Scheffel, S. Lindstaedt, and V. Dimitrova, editors, *Sustaining TEL: From Innovation to Learning and Practice*, volume 6383 of *Lecture Notes in Computer Science*, pages 167–182. Springer, 2010.

[16] X. Ma, G. Chen, and J. Xiao. Analysis of An Online Health Social Network. In *Proceedings of the 1st ACM International Health Informatics Symposium*, pages 297–306. ACM, 2010.

[17] N. Manouselis, H. Drachsler, R. Vuorikari, H. Hummel, and R. Koper. Recommender Systems in Technology Enhanced Learning. In F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, editors, *Recommender Systems Handbook*, pages 387–415. Springer US, 2011.

[18] M. McKeon. Harnessing the web information ecosystem with wiki-based visualization dashboards. *IEEE transactions on visualization and computer graphics*, 15(6):1081–8, 2009.

[19] J. Najjar, M. Wolpers, and E. Duval. Contextualized attention metadata. *D-Lib Magazine*, 13(9/10), Sept. 2007.

[20] X. Ochoa and E. Duval. Use of contextualized attention metadata for ranking and recommending learning objects. In *CAMA06: Proceedings of the 1st international workshop on Contextualized attention metadata: collecting, managing and exploiting of rich usage information*, pages 9–16, 2006.

[21] H. Põldoja. EduFeedr-following and supporting learners in open blog-based courses. In *Proceedings of OpenEd2010*, number 2010. Universitat Oberta de Catalunya, 2010.

[22] E. Pariser. *The Filter Bubble: What the Internet Is Hiding from You.* Penguin Press, 2011.

[23] K. Popper. *The Logic of Scientific Discovery.* Routledge, 1959.

[24] A. S. Rath, D. Devaurs, and S. Lindstaedt. UICO: an ontology-based user interaction context model for automatic task detection on the computer desktop. In *Proceedings of the 1st Workshop on Context, Information and Ontologies*, CIAO '09, pages 8:1—-8:10, New York, NY, USA, 2009. ACM.

[25] C. Romero and S. Ventura. Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, 33(1):135–146, July 2007.

[26] S. S. Sahoo, A. Sheth, and C. Henson. Semantic provenance for eScience - Managing the deluge of scientific data. *IEEE INTERNET COMPUTING*, 12(4):46–54, 2008.

[27] J. L. Santos, K. Verbert, S. Govaerts, and E. Duval. Visualizing PLE Usage. In *Proceedings of EFEPLE11: 1st Workshop on Exploring the Fitness and Evolvability of Personal Learning Environments.* CEUR workshop proceedings, 2011.

[28] H. Schmitz, M. Scheffel, M. Friedrich, M. Jahn, K. Niemann, and M. Wolpers. CAMera for PLE. In U. Cress, V. Dimitrova, and M. Specht, editors, *Learning in the Synergy of Multiple Disciplines*, volume 5794 of *Lecture Notes in Computer Science*, pages 507–520. Springer, 2009.

[29] B. Schwartz. *The paradox of choice - Why more is less.* HarperCollins, 2007.

[30] C. Shirky. *Here Comes Everybody: The Power of Organizing Without Organizations.* Penguin Press, 2008.

[31] E. Singer. The Measured Life. *Technology Review*, 2011.

[32] M. Swan. Emerging patient-driven health care models: an examination of health social networks, consumer personalized medicine and quantified self-tracking. *International journal of environmental research and public health*, 6(2):492–525, Feb. 2009.

[33] S. Ternier, K. Verbert, G. Parra, B. Vandeputte, J. Klerkx, E. Duval, V. Ordonez, and X. Ochoa. The Ariadne Infrastructure for Managing and Storing Metadata. *IEEE Internet Computing*, 13(4):18–25, July 2009.

[34] K. Verbert, E. Duval, H. Drachsler, N. Manouselis, M. Wolpers, R. Vuorikari, and G. Beham. Dataset-driven Research for Improving TEL Recommender Systems. In *1st International Conference on Learning Analytics and Knowledge*, Banff, Canada, 2011.

[35] M. Wolpers, J. Najjar, and E. Duval. Workshop report on the international {ACM} workshop on contextualized attention metadata: collecting, managing and exploiting rich usage information (cama 2006), June 2007.

[36] M. Wolpers, J. Najjar, K. Verbert, and E. Duval. Tracking actual usage: the attention metadata approach. *Educational Technology and Society*, 10(3):106–121, 2007.