Towards Analytics for Educational Interactive e-Books

The case of the Reflective Designer Analytics Platform (RDAP)

Sokratis Karkalas
London Knowledge Lab
UCL Institute of Education
London WC1N 3QS, UK
s.karkalas@ucl.ac.uk

Manolis Mavrikis
London Knowledge Lab
UCL Institute of Education
London WC1N 3QS, UK
m.mavrikis@ucl.ac.uk

Oliver Labs
University of Potsdam
Am Neuen Palais 10, House 9
14469 Potsdam, Germany
labs@cermat.org

ABSTRACT

This paper presents an analytics dashboard that has been developed for designers of interactive e-books. This is part of the EU-funded MC Squared project that is developing a platform for authoring interactive educational e-books. The primary objective is to develop technologies and resources that enhance creative thinking for both designers (authors) and learners. The learning material is expected to offer learners opportunities to engage creatively with mathematical problems and develop creative mathematical thinking. The analytics dashboard is designed to increase authors' awareness so that they can make informed decisions on how to redesign and improve the e-books. This paper presents architectural and design decisions on key features of the dashboard and discusses future steps with respect to the potential for exploratory data analysis.

Categories and Subject Descriptors

J.1 [**Administrative Data Processing**]: Education; K.3 [**Computers and Education**]: Miscellaneous

1. INTRODUCTION

The growing interest on learning analytics dashboards is partly due to their potential of providing, both in real time and retrospectively, an opportunity for awareness and decision-making that is otherwise difficult if not impossible. As educational applications are being adopted and used at scale, understanding their usage and their impact on learning is important. We are interested in the particular genre of digital, interactive 'books' (or e-books) that are beginning to be established as a possible alternative to static textbooks offering several advantages both practical (such as portability or low cost) and pedagogical (such as interactivity and potential for formative feedback) [4, 9, 6]. We see carefully designed dashboards as having a lot to offer in the design cycle of educational resources. While the emergence of authoring software for e-books is making it easier to create or modify e-books based on their preferences, there is very little work to make this process more evidence-based. Similar to the growing interest in the possible synergies between learning analytics, learning design and teacher inquiry [3] we are observing a need for informing the design and re-design of resources based on empirical data from their usage. While with the advent of data science and analytics in general, there are several analytical tools that are making their appearance, most are not targeted to educational e-books. Publishers and authors are interested in (and to some extent only have access to) high level information such the number of pages read, average reading times and other details that reveal reading patterns that can correlate with sales figures. However, from an educational point of view teachers, designers require a more in-depth analysis of learners' interaction with the e-books. The ease of data collection offers unprecedented opportunities in enabling e-book authors to make informed decisions as to how these books can be improved and serve better their original design goals. The latter is one of the goals of the EU-funded M C Squared project that is investigating the design and use of digital, interactive, creative, mathematics e-books. A key characteristic of these e-books is the inclusion of dynamic, interactive widgets that target creative mathematical thinking and problem solving rather than procedural knowledge [6]. The project engages the members of several so-called 'Communities of Interest' (CoI) across EU in the creation of e-books and parallel requirements elicitation and analysis.

This paper presents our approach towards a dashboard and associated visualisations particularly targeted to assist designers to reflect on the use of their e-books. First, in Section 2, we present related work. Section 3 describes our overall design methodology and initial requirements after a series of workshops, sustained online interaction and engagement of key members of the CoIs. Section 4 and 5 present our underlying architecture and the dashboard itself. Section 6 concludes the paper.

LAK '16, April 25 - 29, 2016, Edinburgh, United Kingdom

ISBN 978-1-4503-4190-5/16/04…$15.00
DOI: http://dx.doi.org/10.1145/2883851.2883943

2. RELATED WORK

There is a growing trend of learning analytics dashboards for online, face-to-face, and blended learning settings, largely targeted to teachers. As we cannot review them all here, we refer the reader to relevant reviews in the field (e.g. [15]). The closest area of work is that in collaborative or open-ended digital learning environments that demonstrated the potential of tools for increasing awareness, supporting reflection and facilitating decision-making and intervention [10, 5]. GLASS [7] is a web-based modular system that is based on data that follows the CAM schema [17]. It is a versatile solution that can be used with any CAM datasource. It focuses mainly on user activity and the detection of the most common events. A system that is more closely related to our project is eLAT [2]. This is a framework and an implementation of a Learning Analytics Platform that is designed exclusively for teachers. The aim is to offer teachers opportunities for exploratory data analysis and the ability to evaluate and reflect on teaching practices and interventions.

3. METHODOLOGY AND REQUIREMENTS

As mentioned in the Introduction, this work is set in the context of the EU-funded M C Squared project that engages several designers across EU in Communities of Interest that design mathematical e-books. For the development of the dashboard, we are following a methodology the largely resembles the Learning Awareness Tools User eXperience (LATUX) workflow [10]. Although that paper focuses mostly in the design and deployment of awareness tools in the classroom, the workflow applies in the case of targeting designers as it consists of an iterative process of five stages commonly found in software engineering and user experience approaches — problem definition, low- and higher-fidelity prototyping, pilot studies and validation in-the-wild.

Despite the fact that our focus of attention is e-book designers e.g. rather that teachers, through our iterative design process it has become clear that we need to also pay attention to the pedagogical requirements behind the design of the tools as they have potential for classroom use by teachers. Nevertheless, in this paper, we focus on the design, prototyping and early pilot studies targeting our main stakeholders — tech savvy designers (but not necessarily developers) and authors of e-books. in some cases these are teachers. The starting point in this part is the identification of usage scenarios for the e-books as specified by the CoIs in the project. We first need to take into account that digital resources like e-books are being used either directly in the classroom or in 'blended' learning scenarios (e.g. for practice exercises at home) or in a 'flipped' learning model where students read and interact with the e-book content online (e.g. at home) and complete other parts of the e-book in the classroom with the help of other students or the teacher. So neither context can be excluded. Based on the above usage scenario, in early stages of our design cycle we identified the process of analysing e-book interaction as having similar characteristics as exploratory data analysis [14].

We also identified the following high-level requirements. Designers should be able to:

1. utilise the service at any time and from anywhere without any restrictions and dependencies on technologies and platforms.
2. retrieve, process and analyse data about any chosen period of an e-book's lifetime, which can be changed as one is working the dashboard.
3. perform different types of analysis of the same data at different times throughout a session.
4. analyse an e-book at different levels of granularity (book, page, widget, user).
5. go back in time and inspect past data (flashback operations).

It is also worth clarifying that while, in principle, a dashboard like the one we are presenting here could be used by teachers to support their work in the classroom, our focus here is primarily authors (who could, of course, be teachers). For previous related work on dashboard for teachers we refer the reader to related work (e.g. [11, 5]). So an additional requirement is being able to monitor the usage of an e-book unit in real-time. Early iterations and feedback on low-fidelity prototypes from members of the CoIs helps us convert these requirements to a specification and eventually to a higher-fidelity prototype that is the version we present in this paper.

4. ARCHITECTURE

The analytics platform is designed as an external pluggable application that can provide its full functionality in a totally service oriented manner through standardised interfaces. It comprises two main parts: The analytics data repository and the dashboard. These two parts are not physically or logically interdependent. The data repository comprises a dedicated DBMS instance and a set of RESTful web services that can receive, validate and process xAPI messages [1]. The data services are optimised to handle different types of requests and decompose incoming data in case it is sent as a batch. That implies that the learning platform can optionally implement its own optimisers and take advantage of these optimisations. A simple scenario would be to perform temporary caching whenever possible and send user actions cumulatively as a batch. This is much more efficient than sending each individual action event separately. The following figure (fig. 1) depicts the architecture for that part of the application:

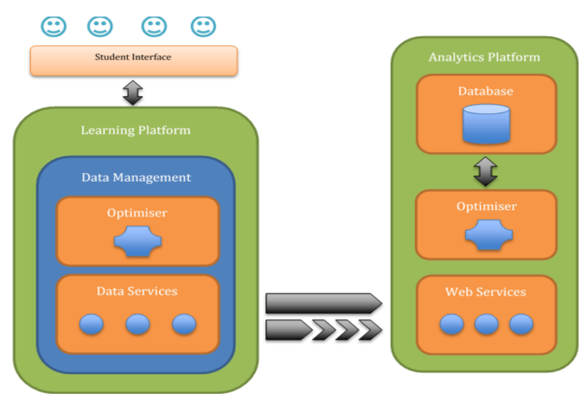

Figure 1: Data Repository

The dashboard communicates directly with the data repository. It constantly checks to see if there is new data recorded and discretely informs the user. It limits the number of data requests through caching and analysis services can process local data and construct visualisations (fig.2).

Integration with the learning platform requires nothing more than a url to be passed along with a set of launch parameters for the dashboard. These parameters are needed so that the dashboard can construct dynamically the structure that represents the learning environment. The assumption here is that we always expect to have entities like users

[1]http://www.adlnet.gov/tla/experience-api

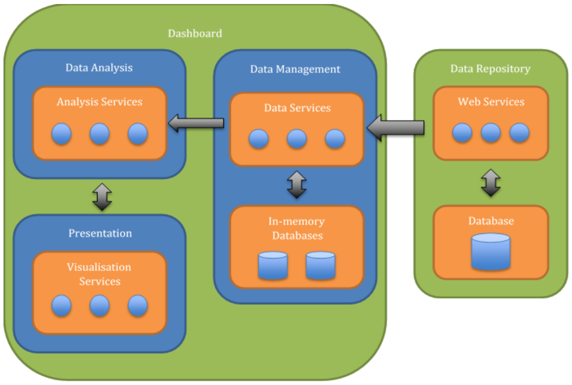

Figure 2: The Dashboard

(students), widgets (activities) and webpages (containers for these activities). This structure has a dual purpose in the system. It can be used to dynamically create the necessary visualisations at the correct level and it can also be used by the authors to navigate themselves through these visualisations during the analysis. The information passed in these parameters is expected to match the identifiers in the user data that will be received from the repository.

4.1 Data Management

As mentioned above this is a web-based application designed to process large volumes of data in real-time and deliver configurable analytics to authors and teachers. This is a service that may need to be utilised both in a synchronous and in an asynchronous manner. In any case it is not known in advance what will be requested by the user. That implies that data preprocessing in the server is not a viable option. Large datasets might have to be transferred, processed and delivered in real-time to the client-end of the application. When requirements like these apply, it is obvious that data management becomes a matter of crucial importance and therefore requires special attention. In order to have the complete picture of what influenced the design decisions, we must also consider the constraints. The constraints follow:

1. The service must be delivered in a distributed manner over the web. That satisfies the first requirement but imposes a problem of potential bandwidth limitations that may affect the ability to transfer large quantities of data in a timely fashion between tiers.
2. The service must be communicated through a web browser without any dependencies on components that are not inherently supported. This is a consequence of 1 that satisfies the second part of the first requirement but imposes an additional problem. That is the potential memory limitations of the browser and its subsequent inability to store large quantities of data.
3. Another problem related to 2 is the fact that JavaScript engines in browsers follow a single-threaded model. That means that concurrency and its respective performance gains are typically not possible. Fortunately, HTML5 offers the ability to distribute processing through web workers.
4. The data processing cannot be performed in the server. All the requirements apart from 1 (especially 7) converge on that.

Considering all the above, the decision was to create a web-based platform with a sophisticated data management sub-component that offers the following:

1. It provides data caching capabilities. It maintains data in local JavaScript databases. Data is synchronised with the source in an asynchronous mode using Ajax. That guarantees that this operation is a non-blocking process in case it takes a noticeable time to complete.
2. It provides the ability for incremental updates. If more data is needed (or less), it is not required to download the entire dataset again but only synchronise the missing parts.
3. It is discreet enough to inform the author about the availability of new data without interrupting what is currently being processed.
4. It offers a clear distinction between synchronisation, analysis and presentation. That helps the user operate the system in a more efficient way.
5. It offers the option to process different time ranges within the range the cached data covers. That means that if all the required data has been downloaded, the rest of the session can be completed in a purely disconnected mode.

It is obvious that the main objectives here were to keep the amount of data to transfer and the number of round-trips to the server to the minimum so that we can utilise in the most efficient way the available bandwidth. The user is able to perform as many operations as needed on the local data without incurring additional network traffic and workload to the data services. This connection-less approach makes the application more scalable, since the data services are able to process more requests, and more responsive, since all the processing takes place at the client side. The authors are given full control over what is synchronised and processed. All that is needed is sensible decisions and careful handling. The application provides all the information about the amount of data that is available. It also provides the ability to select a time range that corresponds to the period that needs to be analysed. The authors must have a certain degree of IT literacy so that they can understand the limitations of the system and use it responsibly.

4.2 Distinguishing Features

RDAP is loosely coupled with the learning environment it relates to. Communication takes place through the standardised xAPI specification that looks a lot like ActivityStreams [2] but allows more flexibility in the structure and the definition of verbs. It also allows the inclusion of widget-specific data that may follow totally different data models. That provides flexibility without compromising diversity. Another difference is in RDAP the primary focus is to enable the author of the material to revisit the initial design and use the feedback from the dashboard to verify the extend to which the objectives have been met. In this process the author is expected to identify flaws in the design that had as an effect the appearance of unexpected patterns in students' behaviour. Finally, the tree-like dynamic structure of the learning platform in RDAP provides the ability

[2] http://www.w3.org/TR/activitystreams-core/

to easily navigate in a random manner between different levels of specificity and thus perform exploratory data analysis with minimal cognitive overhead.

5. THE DASHBOARD

The dashboard is initially empty. There is no data that can be used for analysis and visualisation. The only information that is available is the structure of the e-book and the timestamps that define the start and the end of the time period recorded in the analytics data repository. The available controls that can be used for parameterisation and execution of commands are organised in areas called ribbons. There are currently three ribbons available in the application (Toolbar, Configuration and Event Log). The figure below displays the configuration ribbon. This ribbon hosts controls that can be used primarily for data-related settings and operations. The green area in the data range part is a special slider control that is equipped with two handles. The entire area covered by the control corresponds to the available data in the server. The two text fields above the slider display the starting and ending dates of this period and the text in the green area displays the duration. If at the same time the e-book under investigation is being used by students, the tool gets automatically updated with the changes. If the author wants to analyse a smaller period than that the handles can be used to adjust the starting and ending dates.

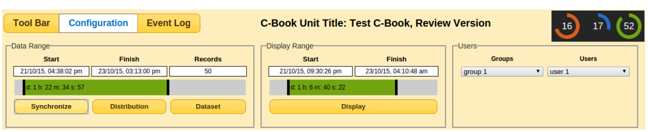

Figure 3: Data Configuration

As the user adjusts the data range period, the display range period gets automatically updated. The display range corresponds to the data that will eventually be analysed and display results. After the adjustments the author can press the 'synchronize' button to start the data synchronisation process. After the completion of this operation the new data is stored in the local databases and becomes available. The next thing to do is to select which part of this data needs to be analysed using the display range slider and press 'display'. After the analysis is completed the visualisations are displayed in the dashboard. The tool bar ribbon can be used for further analysis of existing data. The inspection part of the ribbon hosts another slider that can be used for flashback operations. This slider is initially empty and inactive. The first time it gets activated is when the display button is pressed and a successful analysis completes. When that happens it takes the time period of the currently selected display range. As the slider moves back and forth the author can see immediate changes in the visualisations. The changes are so fast that appear like animations to the human eye. If step-by-step flashback is needed then the dropdown list and the respective buttons in the inspection part must be used.

5.1 The Visualisations

The available visualisations are categorised and presented in three tabs: Widgets, Users and Usage. The first tab focuses on the structure of the learning environment. On the left we can see the structure of the unit (fig. 4). In this case the e-book consists of three pages each one of which contains two widgets. The nodes in the tree are selectable. The author can use them to navigate to different levels of the e-book and display the respective visualisations.

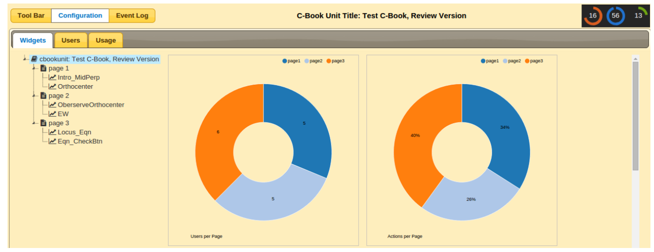

Figure 4: Multi-level Analysis. The structure of the e-book is shown on the left and the user can select nodes to see the corresponding data visualisations.

The third tab focuses on widget usage but from a different perspective. It shows in the same graph how the users relate to widgets in terms of intensity of usage. A cross-tabulation table is used to present this information. The intersection of a row and a column shows information about how intensely a particular widget is used by a particular user. The intensity of the colour in the box corresponds to the proportion of indicators generated by that user for this particular widget in relation to the total number of indicators for this widget. Widgets may be given different colors depending on how heavily they are used. If the total number of indicators for a widget is greater than the average activity per widget then the respective column is displayed in hues of green. If it is more than 80% of the average it is displayed in hues of blue and if it is less than that it is displayed in hues of red. If there are no indicators at all the column is displayed in white color.

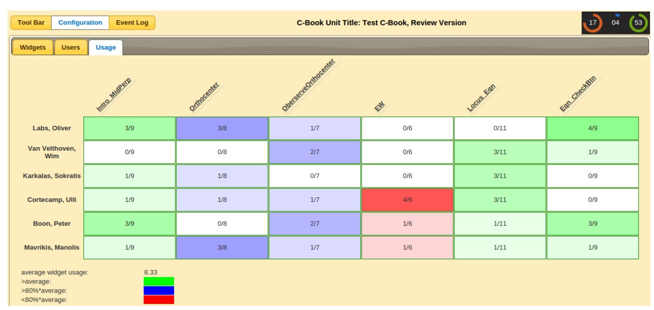

Figure 5: Widget Usage

These three tabs present data from different perspectives. There is often an overlap between them and two or more visualisations can complement each other and provide a view that is more representative of what really happens in the classroom. That implies that if there are ties between them wherever possible, that may support a more exploratory type of data analysis, which is desirable. An example of that approach can be seen on the way the 'Usage' tab is linked to the 'Widgets'. If the author identifies something in the former that requires further investigation, she can click on the header of the table column (widget) and move directly to the page that displays the set of more detailed visualisations that correspond to this particular widget.

6. CONCLUSION

This paper presented the Reflective Designer Analytics Platform (RDAP) that helps learning material authors reflect on their designs and how they meet their original objectives. A prototype was designed and tested with the MC Squared learning platform that utilises interactive e-books that include dynamic widgets aimed at enhancing students' mathematical problem solving and creativity. RDAP is designed to operate as a standalone, platform-independent application that communicates with partner systems through standardised interfaces and data formats. It features a highly efficient data management mechanism that enables incremental synchronisation of data and disconnected operation at the client side. This eliminates server bottlenecks, prevents excessive network load, increases the analytical capacity of the tool and delivers the results through a highly responsive user interface. From an analytical viewpoint the strong points of the system are the ability to analyse data from diverse and dissimilar widgets (learning activities) and the ability to switch between different levels of specificity with ease (shallow, deep analysis). The latter is especially advantageous in exploratory data analysis.

The system has been thoroughly tested for usability and analytic potential. Preliminary evaluation results provided us with positive feedback that feeds into discussions of new features and modifications. One key part we need to pay attention to is cognitive overload and therefore allowing the configuration of the dashboard according to users' preference seems to be important. As mentioned, our focus here has been on designers conducting retrospective data analysis but, technically, the system can be used for real time monitoring and therefore we plan to extend it with appropriate visualisations to support teachers as well. We are, therefore, at the time of this writing testing and improving scaleability issues. Finally, because of the extensibility of the system we can easily include new visualisations and plan to include more advanced algorithms to help analysts identify important events or patterns that are worth exploring further.

ACKNOWLEDGEMENTS

The 'M C Squared' project has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement N°610467.

7. REFERENCES

[1] P. Bradford, M. Porciello, N. Balkon, and D. Backus. The blackboard learning system: The be all and end all in educational instruction? *Journal of Educational Technology Systems*, 35(3):301–314, 2007.

[2] A. L. Dyckhoff, D. Zielke, M. Bültmann, M. A. Chatti, and U. Schroeder. Design and implementation of a learning analytics toolkit for teachers. *Journal of Educational Technology & Society*, 15(3):58–76, 2012.

[3] V. Emin-Martinez, H. Cecilie, R. Triana, M. Jesus, W. Barbara, M. Yishay;, D. Mihai, F. Rebecca, and J.-P. Pernin. *Towards teacher-led design inquiry of learning*. 2014.

[4] E. Gardiner and R. Musto. *The Oxford companion to the book*, chapter The electronic book. Oxford University Press, Oxford, 2010.

[5] S. Gutierrez-Santos, E. Geraniou, D. Pearce-Lazard, and A. Poulovassilis. Design of Teacher Assistance Tools in an Exploratory Learning Environment for Algebraic Generalization. *IEEE Transactions on Learning Technologies*, 5(4):366–376, 2012.

[6] C. Kynigos. Designing constructionist e-books: New mediations for creative mathematical thinking? *Constructivist Foundations*, 10(3):305–313, 2015.

[7] D. Leony, A. Pardo, L. de la Fuente Valentín, D. S. de Castro, and C. D. Kloos. Glass: a learning analytics visualization tool. In *Proceedings of the 2nd international conference on learning analytics and knowledge*, pages 162–163. ACM, 2012.

[8] J. Lewis. Usability testing. In *Handbook of Human Factors and Ergonomics*, pages 1275–1316. John Wiley, New York, NY, 2006.

[9] R. Martin. The road ahead: ebooks, etextbooks and publishers' electronic resources. In M. Brown, M. Hartnett, and T. Stewart, editors, *Future Challenges, Sustainable futures: Proceedings ASCILITE 2012*.

[10] R. Martinez-Maldonado, A. Pardo, N. Mirriahi, K. Yacef, J. Kay, and A. Clayphan. The latux workflow: designing and deploying awareness tools in technology-enabled learning settings. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*, pages 1–10. ACM, 2015.

[11] Noss et al. The design of a system to support exploratory learning of algebraic generalisation. *Computers & Education*, 59(1):63–81, 2012.

[12] J. A. Ruipérez-Valiente, P. J. Muñoz-Merino, D. Leony, and C. D. Kloos. Alas-ka: A learning analytics extension for better understanding the learning process in the khan academy platform. *Computers in Human Behavior*, 47:139–148, 2015.

[13] M. Scheffel, H. Drachsler, and M. Specht. Developing an evaluation framework of quality indicators for learning analytics. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*, pages 16–20. ACM, 2015.

[14] J. W. Tukey. *Exploratory Data Analysis*. Addison-Wesley, 1977.

[15] K. Verbert, S. Govaerts, E. Duval, J. L. Santos, F. Van Assche, G. Parra, and J. Klerkx. Learning dashboards: an overview and future research opportunities. *Personal and Ubiquitous Computing*, 18(6):1499–1514, 2014.

[16] A. Vozniuk, S. Govaerts, and D. Gillet. Towards portable learning analytics dashboards. In *Advanced Learning Technologies (ICALT), 2013 IEEE 13th International Conference on*, pages 412–416. IEEE, 2013.

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

[18] H. Zhang, K. Almeroth, A. Knight, M. Bulger, and R. Mayer. Moodog: Tracking students' online learning activities. In *World Conference on Educational Multimedia, Hypermedia and Telecommunications*, volume 2007, pages 4415–4422, 2007.