Improving the user experience of a digital content viewer through advanced analytics

Ralph Paul Marschall
IT Department, National Library of Luxembourg (BnL), Luxembourg, Luxembourg.
E-mail address: ralph.marschall@bnl.etat.lu

Abstract:

Every company, institution or any other entity that has a web presence faces the challenges of discovering and understanding who their real visitors and users are, what they are doing on their platform(s) and what changes or additions could be made to a particular service that would lead to an improved user experience. Since libraries and other cultural heritage institutions must ensure access to their digital collections to the widest audience, they have the additional challenge of trying to make a "service for everyone".

In such cases, one cannot exclusively rely on general web analytics or on small user groups to test and measure user interfaces. Advanced analytics and user insights are a crucial part of a continuous development process. Such tools allow for fine-grained configurations and event tracking and can directly be applied on end users in real time in order to discover facts that were previously ignored and help to develop services or even prioritize future content.

The National Library of Luxembourg (BnL) has adopted and implemented a self-hosted, open-source, analytics platform which has been integrated into the existing digital content viewer for digitized newspapers and books at www.eluxemburgensia.lu. The generated data has been used by the BnL to drive the development of a new digital content viewer. In this paper, we share our experience of implementing custom analytics into a digital content viewer. We also present how the data has helped us create a new digital content viewer and describe how other libraries could use such technology to gain more insights from their services and take better decisions.

Keywords: digital content viewer, user analytics, user experience, digitization

1. Introduction

In this Internet and mobile age, the number of Internet users will soon reach four billion [1]. More and more people use the Internet in their daily lives and depend on it. Every day, new applications, services and social platforms are being created and hope to attract and retain the greatest number of users. In front of every online service lies an interface that allows the user to interact with that service. Among the first things that the user sees, feels and experiences is the presented user interface. This includes, for example, the layout, content organization, colors, fonts, sounds, animations as well as the loading times and responsiveness. In a fraction of a second, the user will be able to judge if the interface feels right or wrong and decide

whether to stay or look somewhere else. With so much noise on the Internet in general, it is difficult to keep the attention of the user. Companies, institutions, individuals or any other entities having a web presence are confronted to the challenge of finding, understanding, attracting and keeping their real users. Ultimately, web services are made for users and, therefore, they must be tailored to them in order to be able to deliver the best value possible.

General web analytics has allowed many to monitor the traffic coming to a website. However, more advanced analytical tools are necessary in order to get deeper insight into the behavior of the users and the real usage of the user interface. Getting such detailed metrics is an important part of a continuous development and improvement process. By combining general metrics with advanced and specific analytics, one can discover new insight of how a certain services are used and thus support the decisions for adding or changing services and/or features. The main goal of using advanced web analytical tools is to measure very specific events that are happening on a specific user interface, which would not be possible by monitoring the traffic only. Those events can then be analyzed and evaluated in order to gain a deeper understand of the true usage of client-side features.

This paper is specifically focused on web services for digital content viewers, which is very common for libraries and other cultural institutions, and presents the approach that the National Library of Luxembourg (BnL) has used to implement advanced and custom analytics. The method has allowed the BnL to gather user data that was previously ignored and later apply the knowledge to improve an existing digital content viewer, the BnL Viewer [18] and develop a new user interface prototype.

This paper is organized as follows. First the current state of the art regarding web analytics and digital content viewers is covered. Then the approach of the BnL along with the implementation details is presented. At last, a 2-month period of gathered data is being used to showcase the insights and results, followed by the conclusion.

2. State of the Art

This section will cover the state of the art related to currently available open source technologies in the category of user analytics and in the category of digital content viewers.

Starting with the task of tracking, storing and analyzing user activity on any device across web, mobile or even desktop applications, there exist several solutions, some requiring more technical involvement than others. A short mention should be made to Google Analytics [2], which is the leading freemium web analytics service. It is a very powerful and complete service, however not open source and thus does not provide full and private access to the data collected.

At the moment, the closest open source competitor of Google Analytics is Piwik [3]. Piwik is a full featured PHP and MySQL software that is installed on your own servers. Then, similar to other solutions, a JavaScript code snippet has to be copied and pasted on your website and the tracking already begins.

Among the most important features are customizable dashboards with real time data updates, e-commerce support, tracking of events (clicks or any other interactions), searches, campaigns, goal conversions and, of course, user specific data such as geolocation or even device information. At last, Piwik is fully extensible, customizable and ensures that the user privacy

is protected and that the operator stays in complete control and ownership of the data. With more than 7.000 stars, 165 contributors and 24.000 commits on GitHub [4] at the time of writing this paper, Piwik is, despite its relatively old technology stack, the best fully open source solution for web analytics.

The next alternative to Piwik is Open Web Analytics [5]. It shares several features of Piwik and is interesting due to its feature to generate heat maps of web pages, which allows displaying on which part of the page the user engages the most and the least. However, the last minor release has been on 2012 and their presented feature comparison matrix is against features of Google Analytics on 2011 and has not been updated since. Finally, with only 433 stars, 1.650 commits, 5 contributors on GitHub [6], this project has clearly less activity in the open source community than Piwik.

Unfortunately, there are not more open source and full featured user analytics software. Many others do exist but fall in the category of business intelligence and require, most of the time, an already available and populated database of user action tracking items.

The last option that we will cover comprises the well-known solution called Elastic Stack [7]. It consists of several open source products working together to reliably and securely collect data from any source and to provide search, visualization and analysis capabilities. More precisely, the technology stack use Logstash and/or Beats to collect data from log files. This data is then sent to Elasticsearch, which is an analytics and search engine. Finally, Kibana is the software able to query Elasticsearch and generate any kind of plots. This solution requires more work and setup on servers, but it has been proven to give many more insights on big data that was previously ignored [8, 9].

Since this paper not only covers the technologies and usage of web analytics, but also how an existing digital content viewer can be improved, the next few paragraphs will focus on the state of the art for digital content viewers.

Nowadays, there exist numerous open source content viewers [10, 11, 12]. Each have their own specificities and are implemented using different web frameworks, but all do share a set of core technologies and standards. The International Image Interoperability Framework (IIIF) aims at providing access to high quality image resources. For this, they defined community driven Application Programming Interface (API) standards, such as the IIIF Image, Presentation, Authentication and Search API [13]. The IIIF Image API has been implemented by all major open source image servers that are specifically designed to serve high quality images [14, 15, 16].

On the client side, services supporting the IIIF Image API can provide a unified viewer, capable of zooming deeply into images while keeping a high display resolution and able to connect to any image server following the same standard. The most prominent client-side library is OpenSeadragon [17]. Having an image server that implements the IIIF Image API, means that the front end and/or the back-end itself can easily be swapped with another implementation as long as it supports the same interfaces.

3. Implementation of Advanced and Custom Analytics using Piwik

At the national library of Luxembourg, we have built and deployed a custom, web-based digital content viewer [18] back in 2009 in order to make available to the public our digitalization campaign results through the website www.eluxemburgensia.lu. Later, a mobile iOS application has also been developed [19].

Figure 1 shows a representation of the current viewer. Since the first release, a few minor features has been added, but the interface stayed unchanged for the most part. The viewer implements many features that you would expect from a digital content viewer, mainly zoom, pagination, table of content navigation, search and word highlighting. Additionally, the BnL viewer also supports article-level selection and highlight.

Figure 1: Two screenshots of the BnL Viewer. (Left: Page View, Right: Article View)

From a technical point of view, the digital content viewer itself consists of a JavaEE back-end that parses the METS/ALTO [18] files and delivers the data, along with the converted images, to the front-end viewer, which has been developed using ExtJs.

With the ever-changing, improving technologies and disruptive trends, we should constantly reevaluate our solutions and services, so that we lead or follow the state of the art. Therefore, with the increasing need to implement and support high resolution image serving along with the need to have fully responsive user interfaces, the BnL took the opportunity to work on a completely new user interface for the digital content viewer.

Following the still growing trends of mobile devices [21], a mobile-first should be taken. Not necessarily by building native mobile applications, but by making sure that the web-based viewer will scale and be perfectly responsive on any size of screen and that the user experience is unified and improved across all devices. Then, since the ExtJs framework has a "window-like" look and feel, this was also an occasion to redesign the interface and try to deliver a mobile friendly experience.

While evaluating existing and possible new features, the use of advanced analytics shines, because it lets the real users share what they actually do in an automated and scalable way. It is important to always keep in mind who the end-user is when taking decisions that affects the services they use. Asking users directly through targeted emails, user group interviews or surveys can be one way to achieve this. Measuring and observing users directly, in real-time is, however, much more efficient and allows to gather much more insight.

As a first step toward improving user experience on a service, the key features and actions that the user could perform have been listed. Those actions could later be tracked as a time stamped event by the analytics platform, allowing for detailed analysis of when certain events occur.

Those actions have then been grouped into categories. One category was created per major interface component. In this case, there are five categories: top toolbar, left panel, right panel, full page view and collection view. As an example, making a search or selecting an element in the table of content, are actions / events that happen in the left panel of the BnL Viewer. That way, an overall usage of the different components and areas on a page can easily be investigated. From that list, only a subset is retained to be implemented. The criterion is to keep actions that help to answer specific questions such as "How much is the toolbar used?", "How much is button X used?", "How many users copy parts of the full text?".

Once the desired events have been defined, the focus goes on the technical side of the analysis, which mainly consists of evaluating the technical feasibility of implementing those events. At this point, the BnL has opted for the open source analytics platform, called Piwik [3]. As already mentioned in the state of the art, Piwik allows to track and manage numerous web analytics and it could easily support all current and possible future requirements.

The software has been installed on a dedicated virtual machine at the BnL. Piwik being well documented, both, installing and setting up the project has been straightforward and allowed to move into production in a matter of days. It is worth noting that since Luxembourg is a small country, scalability and traffic issues are less challenging.

Out of the box, by inserting a JavaScript script snippet into the web platforms enables Piwik to already track common analytical data related to, but not limited to, visitors, referrers, browser and device information. The next step is to go deeper into the client-side code and call a one-line Piwik command where the specific events should be recorded. As an example, the code snippet below is inserted in the table of content controller and tracks a click event on the table of content:

```
// _paq is the "Piwik Analytics Queue" object
// Format: The event type, followed by the category, then the action/event.
_paq.push(['trackEvent', 'LeftSidebar', 'TOCElement_Click']);
```

After implementing the tracking code into the client, we are ready to deploy it into production and start gathering data, which can be analyzed as soon as they are available.

4. Result Analysis

This section is organized into subsections, where each focuses specifically on a single or a group of related metrics and covers the data collected along with the different possible interpretations for improving and assisting the development of a new viewer. The last section presents the new digital content viewer prototype of the BnL.

The presented results cover the period from the beginning of February 2017 until the end of March 2017 and were generated by the BnL Viewer web service. During the analysis it turned out that, for the traffic of the BnL Viewer, analyzing one or more months at a time were yielding more insightful numbers than exploring day-to-day or even week-to-week data.

All the analytics has been gathered by the Piwik platform. Data privacy is a concern and fortunately Piwik supports a mode where it automatically anonymizes data as soon as it is collected by the platform. This means that no data or behavior can be associated to a particular user.

Visitors

First, looking at the basic visitor data, Figure 2Figure 2 shows a clear pattern. The BnL Viewer had most of the time between 60 and 120 visitors per day. Out of those visitors, 40 to 100 are unique visitors. Visitors performed around 20 to 40 actions per visits. From Figure 3, it can be seen that this represents around 400 page views per day, with February 15/16 showing a record activity for this month with 2.000 page views, due to an announcement on the official BnL website and on the Facebook page at that period. Taking the difference between visitors and unique visitors, we can observe that a relatively small amount of users come again.
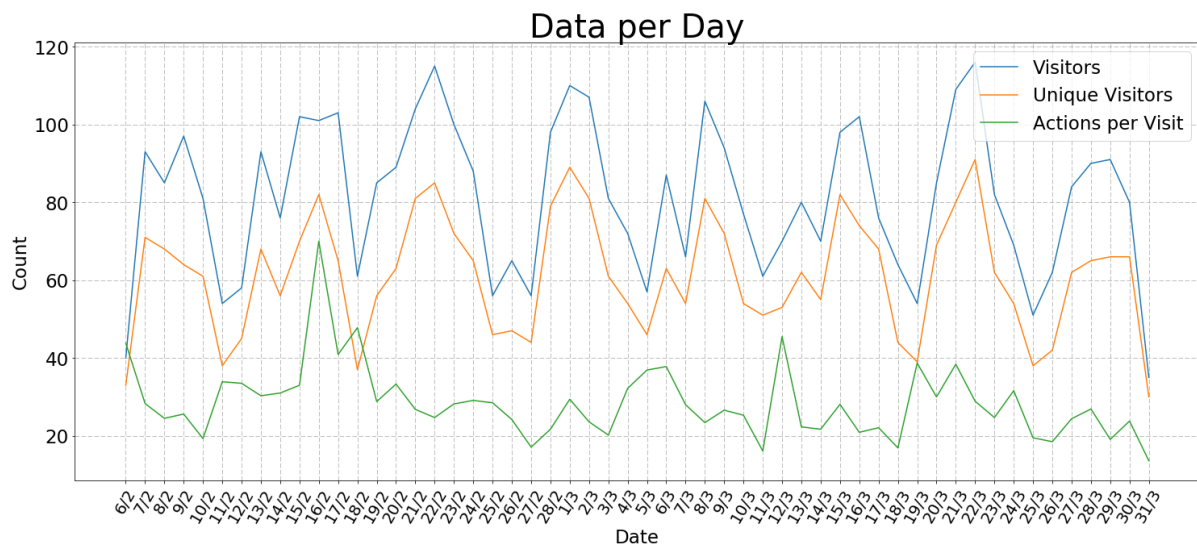
Figure 2: Number of visitors, unique visitors and average actions per visits
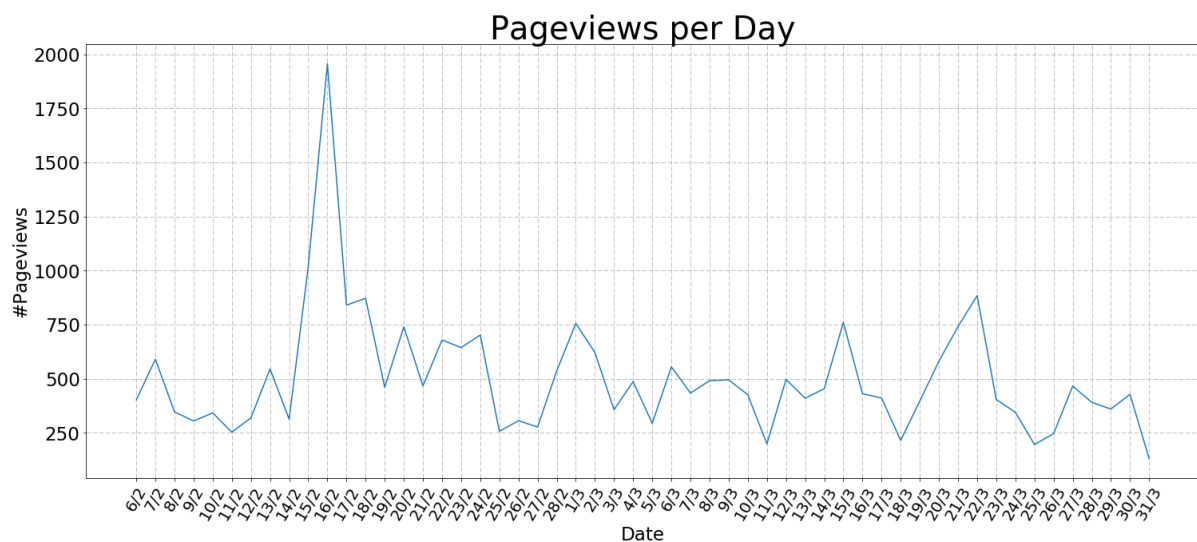

Figure 3: Number of pageviews per day

Next, Figure 4 shows the number of visits per time of the day, as recorded by the server. This means that two visitors from different time zones will be tracked by the local server time and not by their specific time zones. With no surprise, we can observe again a common pattern, where the peak activity happens first in the morning between 10h00 - 11h00 and a second peak between 15h00 and 17h00.
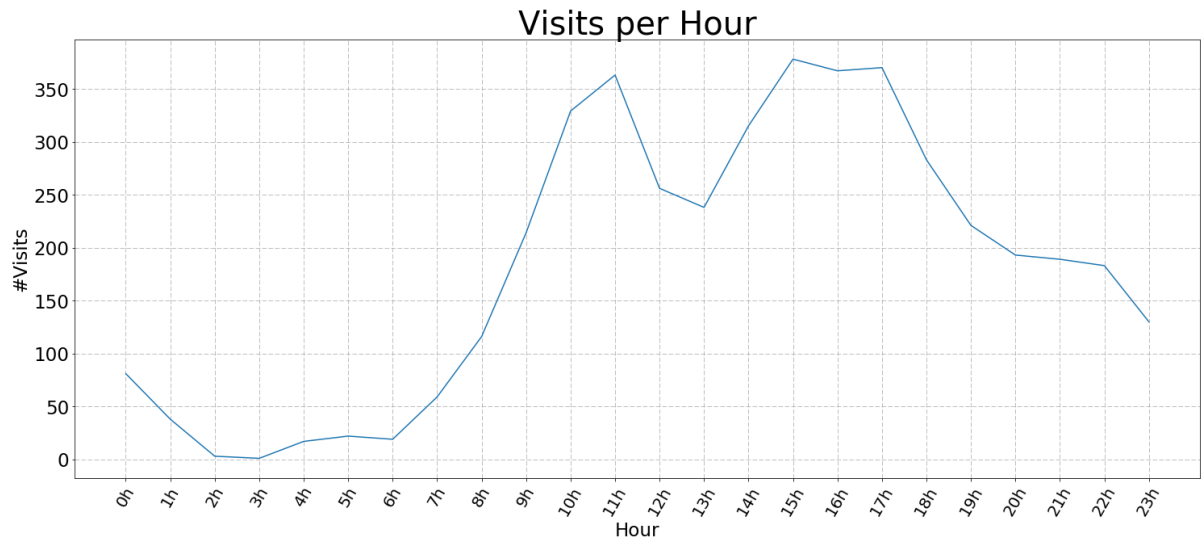
Figure 4: Number of visits per server hour

Figure 5, further reinforces this fact by showing that a large amount of users arriving on the BnL Viewer stay no more than 10 seconds. However, many users seem to have long sessions on the viewer as well. This metric is important and can later be used to measure if a change in the user interface translated into actual improvements, such as increasing the time that users spent on the site.
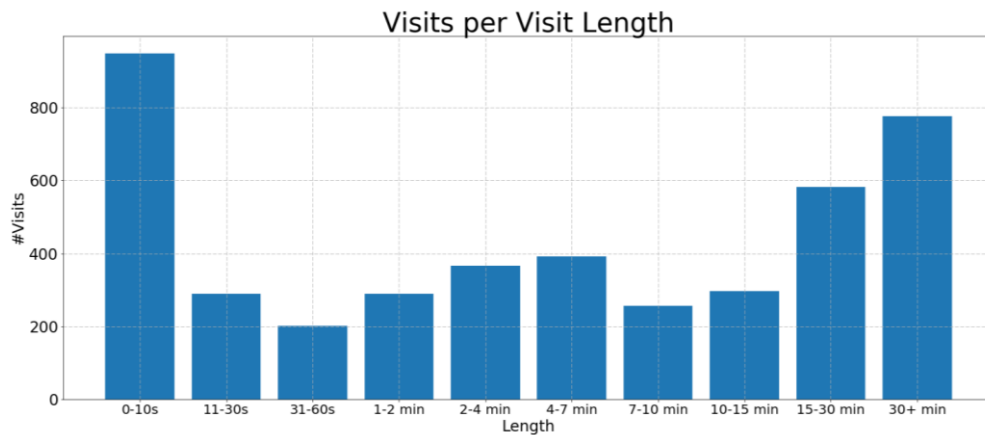

Figure 5: Number of visits per visit length

Finally, Piwik can also track the location of the user. In this case, most visits came from Luxembourg, followed by Germany, then France. This insight, along with the browser language data, which is covered in the next section, can guide several decisions related to the user interface and the content itself. For example, if the analytics shows that the language of the users is completely different than the language of the content itself, then a translation feature might be worth implementing. That way, the same content becomes more attractive to the existing user base.

User devices, browsers and configurations
Figure 6 shows the different device types and operating systems. It is clear that the user base of the BnL Viewer mostly consists of Desktop users on a Windows operating system. Mobile

and tablet users are very low and this data can lead to several interpretations. A possible reason could just be that users prefer to use larger screens for using a digital content viewer. Another possibility could be that the current service is not much used by younger audiences, who are heavy mobile users [20]. And the fact that the current content viewer is not responsive and not well suited to mobile devices also plays a strong role here.
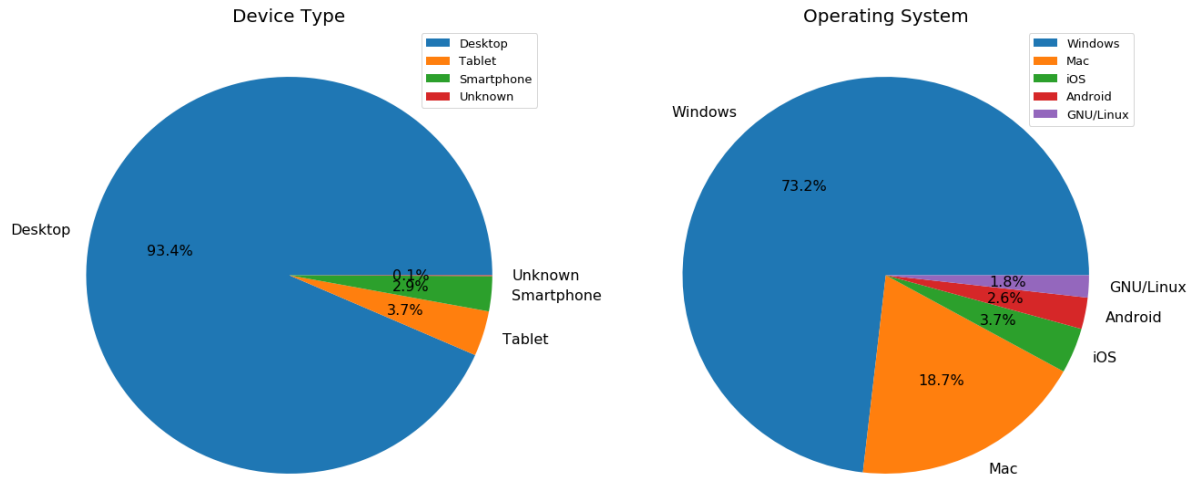
Figure 6: Device types and operating systems

During the 2-month period, 150 different screen resolutions have been tracked, which shows the importance of having a fully responsive web site that can scale and present itself optimally on any screen. Unfortunately, the current web-based BnL Viewer is not optimized for mobile devices. Thus, this is a clear point of improvement. Figure 7 shows the top screen resolutions. When investigating in more detail, 285 different configurations (device-screen-browser combination) have been recorded in those 2 months, insisting on the fact that responsiveness and cross browser/platform support is highly important.
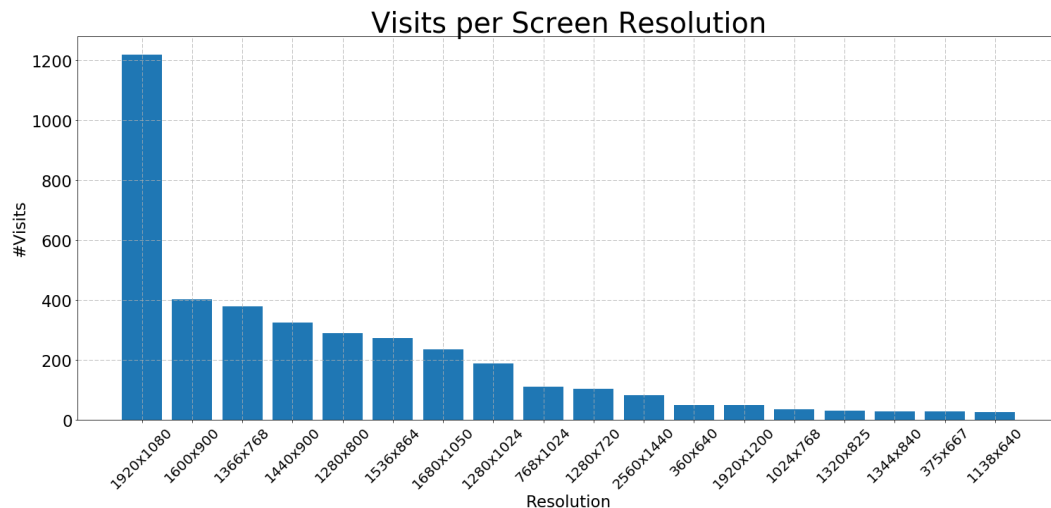
Figure 7: Number of visits per screen resolution

Regarding the language, users have their browser mostly in French, German, English and even Luxemburgish. Since most people in Luxembourg speak 4 or even more languages, this result is expected. Thus, it is also a critical requirement to support internalization of all web services. Finally, we can see in Figure 8 that the majority of visitors are using Firefox and Chrome, but

also IE and Safari. Additionally, all users are not necessarily on the latest version of each browser. This reinforces the need to test any web service on every major browser and across multiple versions.
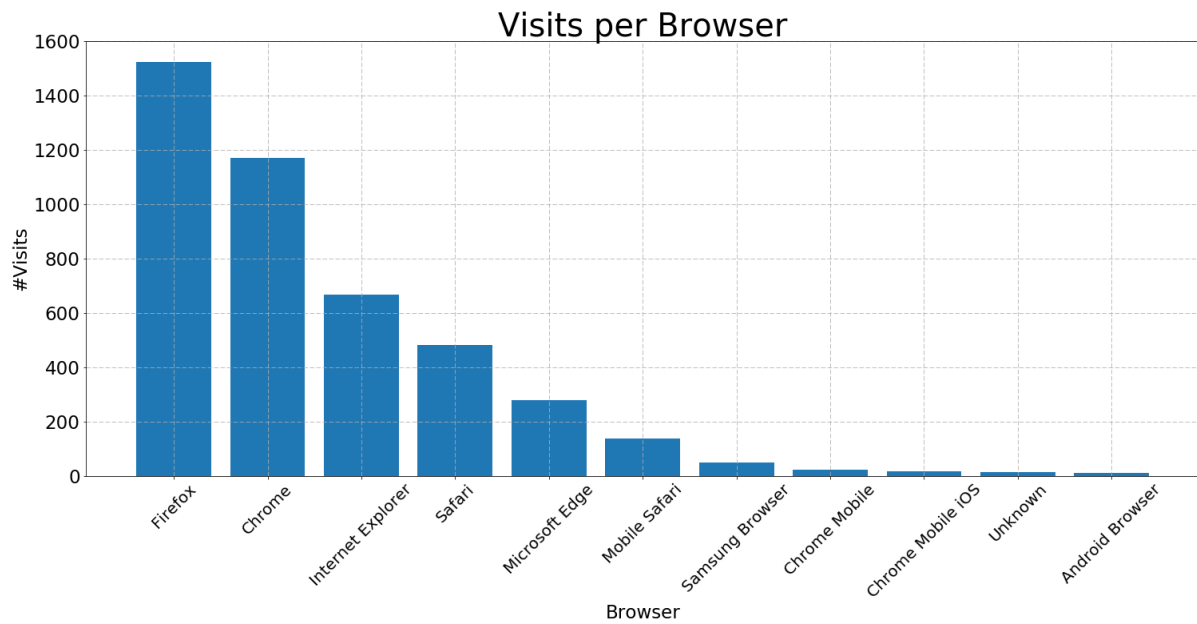

Figure 8: Number of visitors per browser

Cutom BnL Events

As previously mentioned, custom events are tracked in order to analyze and understand which features are being used and which ones are not. Custom events are the most helpful kind of metrics that directly measure a user interface and evaluate the usage of features. Here, the primary aim was to know which features were not used and would be better removed or implemented in a different way. Features that are not used are not automatically useless. They could highlight poor design choices where the user does not find those features, and therefore does not use them. In such cases, the best solution would be to perform A/B testing of the user interface. This consists of deploying two different UI versions and deliver version A to the main user base and version B to a subset of that user base. Then it is measured which one converts more people to a defined goal.
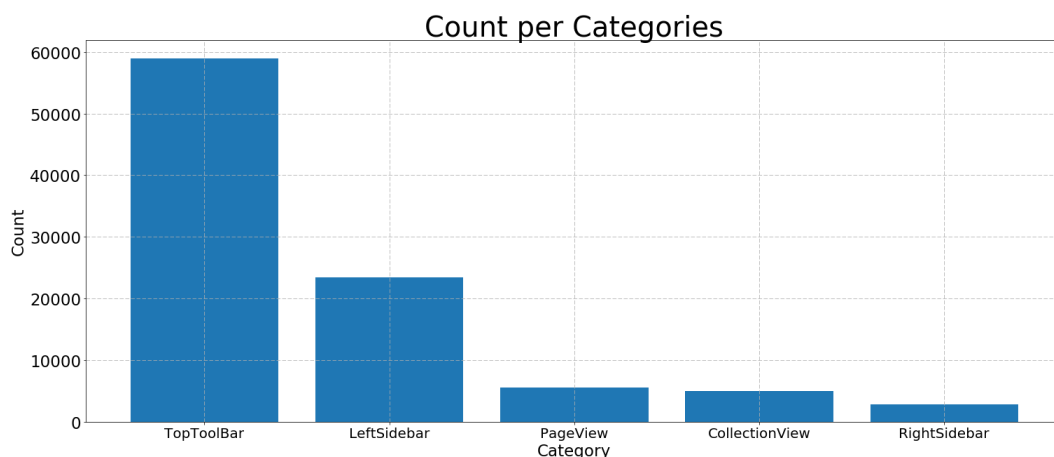

Figure 9: Number of custom events per category

Each event belongs to one category which relates to one main user interface component. Figure 9 shows the number of events per category. At the moment, there are 5 categories.

Regarding events, there are 34 individual events tracked in total. Figure 10 presents the top 20 events across all categories. Analyzing events inside each category is also possible but less recommended, since it can lead to value features that seem to be used a lot inside a single category, but then used rarely compared to all events on that view. This is especially true here, since all categories are part of the same main view.
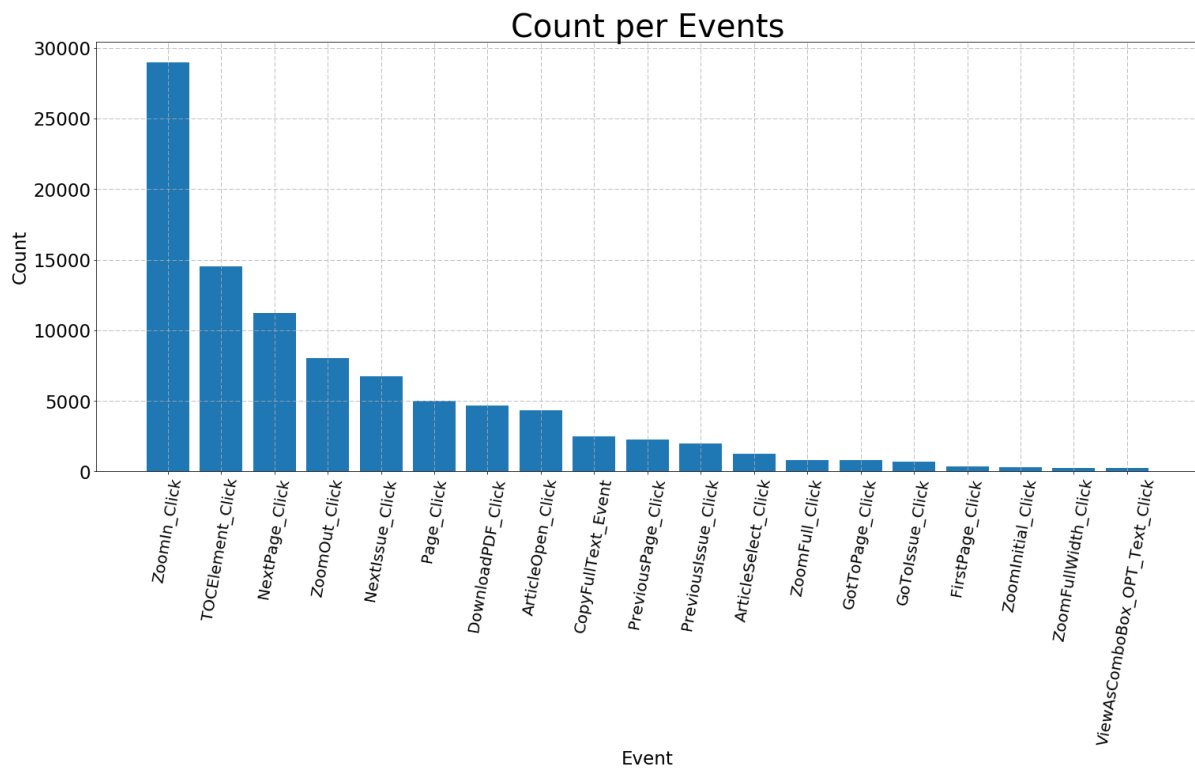

Figure 10: Number of occurances per custom event (top 20)

We can see from the data that the user zooms in a lot. Looking back the design of the current BnL Viewer (Figure 1), we see that the user actually has to zoom in to look into the details of a page and this is mostly done using the zoom button on the toolbar. Those are clicks or taps that could be eliminated by using another image viewing technology, such as OpenSeadragon [17], which allows more intuitive zooming using the mouse wheel or a pinch gesture.

Form the next top events, we can see that the table of content and features to go to the next page or issue are used a lot and this shows that a content viewer not only has the goal of displaying content but also to allow efficient navigation to other related or linked digital elements. Another important point is that there have been quite a lot of PDF downloads as well as copy events on the displayed full text, which hints that users do use features that allow to get all, or parts, of a digital item for usage elsewhere or offline.

It is worth noting that despite the fact that the most important part of the user interface is the presentation of the digital content, most of the clicking actions are happening in the top toolbar or in the left sidebar. Thus, it is crucial for such a digital content viewer to focus on keeping the attention of the user on the digital content and to not add too much distractions around. Features can be numerous, but the user interface should not be cluttered.

Therefore, the aim for the next major iteration of BnL Viewer will be to enable richer interactions with the actual digital content. Everything around the digital content would be better designed in a minimalistic manner, only presenting the user the core functionalities for efficient navigation and search.

Many, less used, functionalities are directly visible on a toolbar or panel. Since, users do not use them, even though they are right there, those features can effectively be relocated to a discrete menu for example or removed altogether.

Search inside Viewer

The BnL Viewer supports article-level selection (highlight) as well as full text search with word highlighting. Thus, tracking what the user searches for is very valuable and allows tracking the usage of the search function as well as what search terms are used. Concerning the data, around 3000 searches have been recorded during for the two-month period. This metric can later be extended by also tracking the search results. As an example, analyzing when a search term gives no results is a very important information that can be used to improve the search engine.

From Data to Prototype

Finally, the BnL took all those results into consideration in order to develop an improved user interface for the BnL Viewer. Figure 11 showcases how the new version would most likely look like on a desktop and on a mobile device. The main user interface components such as the toolbar and the left panel have been kept but their design have been reworked. First, in order to support small mobile screens, it is crucial that the left panel can be hidden and the top bar can be minified. The changes focus on prioritizing the viewing and manipulation of the main digital content. Some unused features have been removed, while others have been organized into non-distracting, yet accessible, setting menus. And, obviously, analytical custom events need to be implemented again into this new user interface, in order to have a continuous understanding of how the platform is being used and take the best decisions for future iterations.
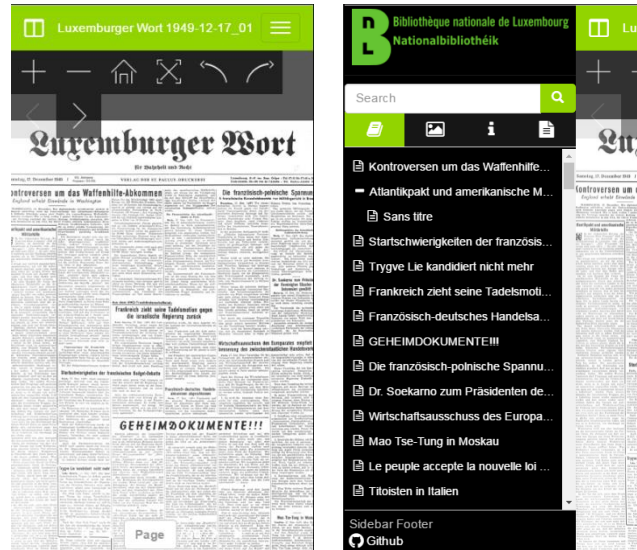
Figure 11: Screenshots of the new BnL Viewer prototype interface

5. Conclusion and Future Work

This work first discussed the state of the art related to user analytic platform solutions and digital content viewer technology, with a focus on open source software only. And then the use case of the National Library of Luxembourg has been covered. There are many solutions available and a few standards, such as the IIIF API, do stand out and allow the different implementations to, ideally, share core functionalities.

Taking into account all those different analytical data results was a tremendous help and guided the BnL into creating a prototype version for a new major iteration of the current content viewer. Focusing on supporting mobile screens from day one and improving the support for high resolution image viewing are natural improvements over the current version. The analytics has proved most beneficial in tracking custom events, which allowed to gain a true insight of how the interface is being used and afterward translate this knowledge into an actual user interface.

Looking at the technology side, using an open source analytics platform, such as Piwik, has proved to be the right solution, enabling the most flexibility and, most importantly, giving the most trust and control over the gathered data. The data can be easily exported in a common format (e.g. JSON, CSV, …) in order to create other plots or perform additional analysis.

Every day, user expectations are evolving as new products are created and trends are adopted. It is the goal of anyone building or maintaining an online service and striving to always deliver a better service to follow and understand those expectations. Users expect great services and will compare your service with the current state of the art that they know. To only cite a few examples; users expect fast response times, clear and easy to understand interfaces and when looking for something, they expect to have the search results that will answer their questions.

Improving user interfaces, and thus, user experiences, is an ongoing effort and, of course, there are many more ways to build and design better services. Custom user analytics are one efficient way to track the behavior of the users and use this data to support the continuous development of a service and take better decisions based on the actual usage.

References

[1]   ITU, ICT Facts and Figures 2016. (2016, June) .Retrieved April 4, 2017, from
        http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2016.pdf

[2]   Google Analytics official website. (2017). Retrieved March 10, 2017, from
        https://analytics.google.com/analytics/web/

[3]   Piwik official website. (2017). Retrieved March 10, 2017, from https://piwik.org

[4]   Piwik GitHub repository. (2017). Retrieved March 10, 2017, from
        https://github.com/piwik/piwik

[5]   Open Web Analytics official website. (2014). Retrieved March 10, 2017,
        from http://www.openwebanalytics.com

[6]   Open Web Analytics GitHub repository. (2017). Retrieved March 10, 2017,
        from https://github.com/padams/Open-Web-Analytics

[7]   Elastic Stack, open source products. (2017). Retrieved March 11, 2017, from
        https://www.elastic.co/products

[8]   Pragya G. & Sreeja N. (2016, January). Survey Paper on Elastic Search. Volume 5 Issue
        1 of International Journal of Science and Research.

[9]   Kononenko, O., Baysal, O., Holmes, R., & Godfrey, M. W. (2014, May). Mining
        modern repositories with elasticsearch. In Proceedings of the 11th Working
        Conference on Mining Software Repositories (pp. 328-331). ACM.

[10]   Mirador, web.based multi-window image viewer. (n.d.). Retrieved March 12, 2017,
        from https://github.com/ProjectMirador/mirador

[11]   Universal Viewer. (2017). Retrieved March 12, 2017, from
        https://github.com/universalviewer/universalviewer

[12]   Diva.js web-based document viewer. (2016). Retrieved March 12, 2017, from
        https://github.com/ddmal/diva.js

[13]   IIIF API specification details. (n.d.). Retrieved March 14, 2017, from
        http://iiif.io/technical-details/

[14]   Cantaloupe, open source image server. (n.d.). Retrieved March 15, 2017, from
        https://medusa-project.github.io/cantaloupe/

[15]   IIPImage, open source image server. (n.d.). Retrieved March 15, 2017, from
        http://iipimage.sourceforge.net/documentation/server/

[16]   Loris, open source image server. (2017, February 3). Retrieved March 15, 2017, from
        https://github.com/loris-imageserver/loris

[17]   OpenSeadragon open source web-based high resolution image viewer. (2017).
        Retrieved March 16, 2017 from https://openseadragon.github.io

[18]   BnL Viewer, METS/ALTO viewer. (2016). Retrieved March 12, 2017, from
        https://sourceforge.net/projects/bnlviewer/

[19]   BnL eLux, iOS application for tablet. (2014, March 22). Retrieved March 12, 2017,
        from https://itunes.apple.com/lu/app/bnl-elux./id461402921?mt=8

[20]   BMascheroni, G. & Cuman, A. (2014). Net Children Go Mobile: Final report.
        Deliverables D6.4 & D5.2. Milano: Educatt.

[21]   Digital in 2017 – Global Overview. (2017, January 24). Retrieved March 25, 2017,
        from https://wearesocial.com/blog/2017/01/digital-in-2017-global-overview