Developing a Framework for Web Analytics

Verena Hausmann  
*Universität Koblenz-Landau, Germany*, verena.hausmann@uni-koblenz.de

Susan P. Williams  
*Universität Koblenz-Landau, Germany*, susan.williams@uni-koblenz.de

Petra Schubert  
*Universität Koblenz-Landau, Germany*, petra.schubert@uni-koblenz.de

Follow this and additional works at: [http://aisel.aisnet.org/bled2012](http://aisel.aisnet.org/bled2012)

**Recommended Citation**

[http://aisel.aisnet.org/bled2012/11](http://aisel.aisnet.org/bled2012/11)
Developing a Framework for Web Analytics

Verena Hausmann
Universität Koblenz-Landau, Germany
verena.hausmann@uni-koblenz.de

Susan P. Williams
Universität Koblenz-Landau, Germany
susan.williams@uni-koblenz.de

Petra Schubert
Universität Koblenz-Landau, Germany
petra.schubert@uni-koblenz.de

Abstract
Gathering and analysing data about the usage of websites can assist in understanding customer behaviour, improving services and potentially increasing profit. However, changing technologies and evolving web analytics methods and processes present a challenge to organisations embarking on web analytics projects. In this paper we examine current theory and practice in the field of Web Analytics (WA) to identify methods and techniques, issues and challenges. Our analysis identifies the need to move from a data-centric view of WA towards a more user- and business-focused view. We propose a framework for establishing sustainable WA programmes, which widens the focus of the WA process to fully account for business requirements and to focus attention on achieving actionable outcomes.

Keywords: Web Usage Analysis, Web Analytics Framework, Process Cycle

1 Background and Motivation
Identifying who uses a website and how it is used has been of interest to website owners since Tim Berners-Lee developed the first web browser in 1990 (Berners-Lee 1989; Berners-Lee & Fischetti 2000; Connolly 2000; CERN 2008). The influence of the online channel now eclipses all other media (Lovett et al. 2009) and interest in monitoring website usage and user activities has intensified. This has led to significant developments in the field of Web Analytics (WA), defined as the:
“objective tracking, collection, measurement, reporting, and analysis of quantitative Internet data to optimize websites and web marketing initiatives” (http://www.webanalyticsassociation.org/ 2006).

As online channels have become more important, understanding users and their needs is essential for ensuring effective website design and management, efficient mass customisation and effective marketing (Kosala & Blockeel 2000). In turn, companies are increasingly interested in converting the usage data generated from their websites into Business Intelligence (BI).

However, Internet technologies and the field of WA are constantly evolving and the scope and range of WA tools currently available on the market is enormous (Hassler 2010). Organisations are faced with the significant challenge of identifying which tool to use, understanding the differences between tools and metrics, and deciding what must be considered when undertaking WA programmes. They now have access to many tools for Web Analytics; however “unfortunately, many organisations are not getting the most value out of their initiatives, and some are questioning the value of doing anything at all” (Gassman 2009). In particular, small and medium-sized enterprises (SMEs) and non-profit organisations that operate informational websites struggle; often lacking the financial and staff resources to operate a WA programme in an effective manner. A recent Gartner report estimates that organisations need a budget of at least $250,000 per annum to properly fund a WA program and some organisations spend millions of dollars (Gassman, 2009).

The academic literature on WA focuses largely on the advantages and limitations of specific algorithms for analysis and the development of improved methods for web mining; little attention is directed towards the user and organisational aspects of WA. The aim of this paper is to broaden the focus of academic research by addressing the wider organisational and user issues. The work is part of a long-term research programme to assist organisations to improve their WA strategies and approaches to deliver greater value. Our goal is to develop a framework and methodology to guide the development of a reliable, usable and sustainable WA programme. The topic is consistent with this year’s Bled conference theme; we look at WA from the point of view of ensuring dependability and reliability in the evaluation and design of web resources and e-services. We place specific emphasis on understanding the needs of small- and medium-sized enterprises, where the resources and skills required to run a successful WA programme are more limited and thus hinder the WA usage possibilities.

The paper is structured as follows. We begin with an overview of the research aims, methodology and research design. This is followed by an analysis of the key issues and challenges for WA. These provide the basis for the development of a design for a WA framework. Finally we discuss our findings and directions for future work.

2 Research Approach

The aim of this research is to develop insights into the organisational and user aspects of Web Analytics in order to extend current theory and provide practical guidance for organisations in the development of their WA programmes.
Specifically our research objectives (RO) are to:

RO1: identify key trends in WA methods, metrics and processes

RO2: provide an overview of the WA tool market with a focus on SME/non-profit capabilities

RO3: investigate the issues, challenges and stages involved in the development of a WA programme.

These objectives provide an understanding of the field of WA in general and key aspects of both theory and practice, and form the basis for the development of a WA framework, which can be used to guide organisations in setting up and maintaining a sustainable WA programme. This framework forms part of a wider research project; the framework will be further tested and refined in future work.

2.1 Research Design & Data Collection

The research process for the overarching project is divided into four phases as shown in Figure 1. To meet the research objectives outlined above, this paper addresses elements of three of these phases. Phase 4 is not addressed here as work is still underway. Due to space limitations, phase 2 (tool analysis) is not covered in this paper; our focus here is on the development of the framework.

Figure 1: Research Steps

2.1.1 Phase 1: Identification of key WA methods, techniques and WA market

Phase 1 takes the form of a literature review and analysis including an analysis of the WA market. It provides the foundation for the study and addresses RO1: identification of key trends in WA methods, metrics and processes and RO2: providing an overview of the WA tool market with a focus on SME/non-profit capabilities. The aim of the literature analysis is to investigate current trends in WA and to identify key WA issues and challenges. The market analysis provides an overview of the major players and WA tools.

2.1.2 Phase 2: Analysis of findings and WA methodology requirements

Phase 2 of the research design draws on the findings of Phase 1 and addresses RO3 issues, challenges and stages involved in the development of a WA programme. The insights and findings of the literature review provide the starting point for the interpretive case study, which follows a small, not-for-profit organisation embarking on developing a new WA programme.
2.1.3 Phase 3: Development of WA framework

The outcomes of the literature and tool analysis and the case study provide input to Phase 3. Here we consolidate the issues, challenges and processes to identify the key phases in the development of the WA programme. We present and discuss the framework to guide the development of a WA programme.

3 Web Analytics Literature Analysis

The aim of the literature analysis is to identify the key methods, techniques, issues and challenges associated with Web Analytics. Due to space limitations we provide only a summary of the key literature, an extensive review and detailed discussion can be found in (Hausmann, 2012).

The area of WA has its roots in the field of Data Mining, more specifically in Web Mining. In early attempts to categorise Web Mining two different approaches were identified: Web Content Mining and Web Usage Mining. Later, a third category was added: Web Structure Mining (Kosala & Blockeel 2000; Madria et al. 1999). Web Content Mining uses information such as text and images available on the Web to discover useful information from these data sources and to improve information finding and filtering (Kosala & Blockeel 2000; Pierrakos et al. 2003). Web Structure Mining is concerned with the underlying link structure of the website itself and is used for the categorisation of web pages (Kosala & Blockeel 2000). Finally, the discovery of user browsing and access patterns is achieved through Web Usage Mining (Cooley et al. 1997).

The overarching term Web Analytics (WA) is used to describe the whole process from data gathering to recommendations for website redesign. WA draws together all three areas of Web Content Mining, Web Structure Mining and Web Usage Mining.

3.1 Type of website and WA requirements

Websites are a key communication channel for organisations and “improving Web communication is essential to better satisfy the objectives of both the Web site and its target audience” (Norguet et al. 2006, p.430). However, to gain relevant insights it is important to understand the purposes served by the website. In terms of WA, different types of websites have different intentions and goals and require different indicators to measure their success. For example, WA for an e-commerce website is likely to examine increased revenue or gross online margin whereas this key performance indicator would not make sense for a government information website. What constitutes success may be very different according to the website type.

 ISSUE 1: A key issue for organisations is to clearly identify the purpose and goals of the website and the corresponding WA requirements, and to understand what they need to measure and why.
3.2 Data collection and data analysis methods

A wide range of web usage data types and data collection methods exist. These can be distinguished into quantitative data/collection methods and qualitative data/collection methods (Table 1).

WA tools work with the quantitative data. Log files and JavaScript tagging are the two most widely used data collection methods. Log files of web servers were the first method for WA data collection (Kaushik 2007; Hasan et al. 2009). They log server-side information such as IP-address, time stamps, client requests, referrer, etc. A few years ago the majority of web metrics were generated with the use of server log files. However, page tagging is now more commonly used and overcomes the limitations of web server logs. It can also be used to analyse more types of data.

JavaScript tagging requires the inclusion of a short JavaScript codelet in each web page. This code captures client-side data. The variety of data that can be captured is enormous and ranges from clicks and position of the cursor, to mouse moves and keystrokes, to the window size of the browsers and installed plug-ins. Any information that can be captured by log files can also be captured with tagging (Kaushik 2007, p.54). JavaScript tagging is currently the most used data gathering technique (Kaushik 2007, p.30ff; Hassler 2010, p.29). However, both methods have challenges and limitations, for example, dealing with data from non-human users e.g. spiders and bots, problems with the deactivation of cookies or measuring time spent on the last page.

**ISSUE 2: Identifying the appropriate data types to be collected and methods for data collection that are commensurate with the organisations WA requirements.**

Understanding the limitations of WA data collection methods.

3.3 Data analysis process and methods

Web usage data itself is not useful until it is analysed. To reach the point of interpretation of metrics and to identify user browsing trends and patterns several steps are required. Figure 2 gives an overview of the WA process as commonly described in the literature.

The process can be split into four steps. It begins with the data collection phase which includes the extraction of usage data from various sources (logs, tags, packet sniffer, etc.) and the identification of their content and structure (Pierrakos et al. 2003, p.320). Next is the pre-processing phase which includes the identification of which of the collected data is useful for the mining process and the preparation of this data for further analysis and pattern discovery (Jaideep Srivastava et al. 2000). Session and user identification are key tasks as these form the basic building block for pattern discovery (Pierrakos et al. 2003, p.327).
Figure 2: Web Analytic Process (adapted from Cooley et al. 1997; Jaideep Srivastava et al. 2000; Pierrakos et al. 2003; Grace et al. 2011)

Pattern discovery encompasses multiple approaches including: statistical analysis, path analysis, discovery of association rules and sequential patterns, clustering and classification (Cooley et al. 1997; Grace et al. 2011). In order to find the interesting patterns, rules and to generate reports and statistics the step of pattern analysis is required. Within pattern analysis the results of the pattern discovery are validated, interpreted and visualised (Etminani et al. 2009; Grace et al. 2011).

**ISSUE 3: Understanding and establishing an effective process for web usage analysis.**

### 3.4 Market overview

The use of WA tools is relatively new (Nakatani & Chuang 2011), however the number of tools available on the market is extensive and diverse (Hassler 2010, p.39). In 2004 the worldwide market for WA tools was estimated at $292 million (Sen et al. 2006, p.86). A more recent study estimates that US businesses alone will spend around $953 million on WA software in 2014. The average spending for WA technologies is around $15,000 per year (Lovett et al. 2009). The range and complexity of tools presents another challenge, the functionality and scope of tools is extensive and for a simple informational website much of this functionality may not even be required. We performed a review of reports and statistics on WA tool usage and identified the following key WA tools currently on the market (shown in Table 2).

<table>
<thead>
<tr>
<th>Adobe (Omniture)*</th>
<th>SiteCatalyst</th>
<th>Coremetrics (IBM)*</th>
<th>Netmind GO</th>
<th>SiteCensus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analyser NX (AT Internet)*</td>
<td>Etracker</td>
<td>Piwik</td>
<td>Webtrekk Q3</td>
<td></td>
</tr>
<tr>
<td>Click Tale</td>
<td>Google Analytics*</td>
<td>Sitestats</td>
<td>Webtrends Analytics*</td>
<td></td>
</tr>
<tr>
<td>comScore Analytix*</td>
<td>Digital NetInsight</td>
<td>StatCounter</td>
<td>Yahoo! Web Analytics*</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2: Key Web Analytic Tools on the market**
Tools marked with an asterisk (*) are those tools most often used by enterprise-class clients with revenue of more than US$10 million. Currently Google Analytics is the most popular WA tool globally (Stanhope et al. 2011).

ISSUE 4: Organisations must understand the functionality and capability of different WA tools in order to select a tool that meets their specific requirements.

3.5 Summary of WA issues

An analysis of the current research literature identified a number of broad challenges outlined above. These range from macro-level issues such as understanding the business requirements for WA, selection of appropriate WA tools and compliance with legal requirements to more micro-level issues such as selection of data types, data analysis and data interpretation methods. Further, whilst there is a well-understood process for handling the micro-level issues (see Figure 2) there is no overarching framework for guiding the development of a complete and sustainable WA programme. This is the motivation for this research; to develop a WA framework to assist companies to deal with the challenges identified above.

We build on these findings from the literature through the use of a case study (presented below) to connect and elaborate on them with practical field study insights.

4 Case study: WA in practice

To provide insights into the real-world challenges of developing a WA programme, an interpretive case study was conducted alongside the literature analysis. In a first phase the literature analysis built the basis for the course of action. However, the process described in the literature does not address all required steps to setting up a WA programme as the case study reveals.

The case study organisation is a not-for-profit group that developed and maintains an informational website on the topic of doing business in European Union countries. The multi-language website was developed within an EU project between 2005 and 2007 and contains information about cultural, business and work variations between EU countries. Currently the site is undergoing a complete revision to improve the information architecture and usability, add extensive new content, new language translations and implementation of social media. The authors’ role in this broader project was concerned with analysing old web usage data to improve the site’s information architecture and to develop a WA methodology to ensure the sustainability of the future web site. In the following sections we highlight key moments during this work and provide additional insights into the requirements of the WA process.

4.1 Business requirements

In the case study, whilst log data was collected from the old website it was not gathered in a systematic manner and in the four years that the site had been live no web usage analysis had been conducted. A first major problem arose with the availability of the log files. Due to technical problems arising from the captured log files the website itself as well as the log files were no longer readily available. A sample of two months of
interactions of the old log files (comprising 33,250 lines stored in a spreadsheet) could be restored, however this data was less than optimal. A key challenge was to understand the nature and structure of the old log files, which had not been captured in a standard log file format. Further, there was no formal record of why specific data types had been collected and for what purpose. Despite intensive investigations no WA tool could be found that was capable of analysing the log files.

Thus, the first insight from the case study was the need to clearly define business requirements in order to undertake sustainable actions. What are the desired outcomes and which data should be kept?

4.2 Integrated and iterative analysis

To understand the website as a whole and to conduct the process of Web Analytics it was necessary to conduct content, structural and visual design analyses before any of the existing log files could be analysed to get insights into user behaviour. The resulting forensic information design assisted us in identifying the various information categories, content types, linkages between content and the overall information architecture. Beside the quantitative analysis (WA) this qualitative analysis was also conducted in the form of an heuristic usability study. This background knowledge about the design of the website was necessary to later interpret and understand the results of WA. Had a sustainable WA programme and an assigned analyst already been in place these activities would not have been required.

In terms of a WA framework this revealed the second insight that it requires an integrated analysis that brings together all the usability and analytics data to identify and understand the form of the current site and its limitations. Further the process must be iterative, reviewing both the outcomes of the WA process and evaluating the WA methods themselves.

4.3 Data format and tool selection

Even though several tools could be identified which are capable of analysing log files, none were able to handle the log data from the old website as it had been saved in a non-standard log file format. Since no analytic tool could be used the log file analysis was undertaken by hand. It was clear that this would not bring the same effectiveness as with a WA tool, however it was undertaken to provide at least some insights into web usage.

In addition, this procedure gave us the opportunity to think about the Web Analytics process as described in the literature in more depth. It became clear, that when using a WA tool, that although some steps would be completed by the tool itself (e.g. session and user identification) additional steps involving the analyst him/herself would be required (e.g. customizing of the WA tool).

The main tasks within the pre-processing phase were to clean the data and identify which of the collected data was useful for the mining process. Further it was necessary to find users and to build sessions by hand and entries such as images, error logs and bots needed to be removed. 18,847 bot entries were removed leaving 14,403 remaining log file entries. Due to capacity limitations it was decided to examine two weeks of log
files (708 logs) for further manual analysis. Single users and sessions had to be identified; this was done on the basis of IP addresses. A session end was defined after 30 minutes of inactivity. In this process, 76 different users and 105 sessions could be identified. On this basis, a few usage statistics could be discovered. For instance, 15% of the 76 visitors became returning visitors within the analysed two weeks. This statistic could have different meanings, e.g. that users could not find the information they were looking for and therefore did not come back, or that they found exactly what they searched for and got the information first time. The most visited web pages were the 6 top-level topic pages of the website. 24 referrer URLs were captured which showed searched key phrases and external pages linking to the one selected by the user. English was the most used language and the United Kingdom the most viewed country. However, these findings are not surprising. English is the default language of the site and the UK the default country for the homepage. The data as it had been collected allowed little more than this very basic analysis and showed how essential specifying data formats and tools is to the future success of a WA programme.

The case study is a good example of the shift that has taken place in recent years in both websites and WA. The way this website was originally built and usage data collected is no longer appropriate, making it difficult to analyse the site and its usage in any depth. The case provided the possibility to think about critical aspects involved in setting up a WA programme. The insights we gained served to reinforce the need for a clear business requirements analysis (to identify relevant usage criteria) and to engage in a detailed data type analysis to ensure that the appropriate data are captured and in a format suitable for analysis using the selected metrics.

4.4 Clearly defined methodology and process

As identified in the literature, there is no well-established methodology for designing and maintaining a successful WA programme. Where WA methods are currently strong is at the data analysis stages (see Figure 2). However, evidence from the case study findings and the literature show that organisations are failing to apply WA effectively. The case study served to emphasise the need for a broader methodology that considers WA as a business analysis and design activity and provides guidance on the entire process in a sustainable manner taking into account business goals, business requirements, the necessary data types and metrics, the affordances of WA tools and the cycle of analysis.

5 Web Analytics Framework

Building on our findings from the literature and the case study, the outcome of this research is a WA analytics framework. As described above we need to provide a broader framework than that shown in Figure 2 in order to address all tasks that should be undertaken in preparing and coordinating an effective WA programme. Araya et al. (2004) made some progress in this direction by extending the existing process by adding 2 steps: “identification of objectives and available data” at the beginning and “business integration” at the end. We have built further by consolidating all our findings and structuring them into a WA process framework, consisting of two parts. Figure 3 shows the high level WA process cycle with 5 steps. These 5 steps are then elaborated in more
Hausmann, Williams, Schubert

detail in Figure 4. The colours of the different steps within the process cycle are according to the corresponding steps within the framework.

The 5 phases are described below.

The process begins with Phase 1, Business Requirements. The need for a clear understanding of the business requirements for Web Analytics is addressed by identifying the information needs and the information architecture of the website.

Phase 2, Planning for Web Analytics builds on the business requirements by conducting detailed analysis of possible metrics and analysis outcomes and selecting those that most usefully address the business requirements. When metrics and outcomes have been decided then a suitable tool that can meet these requirements should be selected.

Phase 3, Developing a Data Collection Capability involves setting up the selected tool and customising it so that it captures and processes relevant data types and the required criteria. This phase is the one that is most fully developed in the current research literature.

Phase 4, Achieving Useable and Actionable Results requires the organisation to have a clear pathway for action taking based on the insights developed through usage analysis. This requires a clear understanding of what benefits will be achieved from the action taking and developing metrics to further measure this, e.g. increased downloads, fewer abrupt exits from the site, etc.

Finally, Phase 5, Evaluation of Actions questions if the WA process and actions for improvement have resulted in positive outcomes.

Each of these steps is further outlined and divided into subtasks within the Web Analytics process framework.
Developing a Framework for Web Analytics

Figure 4: Web Analytics Framework
The gear cogs within the framework are used to indicate cycles of action. It is necessary to establish the different action cycles and make decisions about how frequently analysis takes place. The data processing for example needs to be done continuously whereas the pattern analysis and identification should be conducted on a weekly basis. Here it needs to be monitored and checked that the right data is being collected. If new items are required then the tool must be further adjusted to capture these new data types. Monthly action taking involves a more detailed usage analysis and leads to, for example, fine tuning of the website, adjusting marketing strategies or fixing structural errors. Beyond this, the two-yearly action taking involves a major review of the website and its usage and may lead to, for example, major strategic re-positioning of the website (e.g. to move from an informational site to a transactional site), or a major re-design of the site to take advantage of new technologies or social/business trends (e.g. the addition of social media tools, taking advantage of the benefits of HTML5 etc.). The exact action cycle frequencies will be determined according to the organisation's specific requirements.

6 Concluding Remarks and Future Work

In this paper we examine the current state of Web Analytics (WA) through the use of a literature analysis and a case study. Each approach identified a range of challenges and issues that need to be addressed when conducting WA. We placed specific focus on the needs for informational websites of SMEs and non-profit organisations in order to understand the challenges they are facing concerning resource limitations. While the academic literature describes the analytic process from a data-centric view, the tasks that need to be undertaken to make effective use of a WA tool are more user- and business-focused.

We propose a framework for developing a sustainable WA programme that widens the focus of the WA process to fully account for business requirements and to focus attention on achieving actionable outcomes. The WA framework is currently being tested in the case study organisation to assist them to develop a sustainable WA programme. The next cycle of this research project involves an assessment of the effectiveness and if necessary, further refinement of the WA framework.

References


