YOUR NUMBERS ARE WRONG
Ensuring High Quality Web Analytics Data
Introduction

“Your numbers are wrong.”

This is a phrase that strikes fear in the hearts of web analysts worldwide. You’ve heard these dreaded words before, haven’t you? If not, then you must be new – welcome to the world of web analytics.

Uncertain data quality is a reality of web analytics. Even with only a single web page tracking a single metric – visits – a range of challenges can bring the accuracy of your Visits Report into question. Some are technical. Are you using log files or javascript? Some are caused by your visitors’ behavior. Are they using an anonymous browser, or do they have javascript disabled? Some are a matter of interpretation. How exactly are you defining a visit? What if a customer reloads the page? How long between page loads before you would add an additional visit? And these are the obvious considerations.

The potential for bad data, or misinterpreting good data, increases exponentially as we add site complexity.

There is a full spectrum in the severity of data issues, from inconsequential to embarrassingly bad. In many ways, we are fortunate when faced with the worst type. These are the kind of errors that anyone can spot on a trendline – represented by a sudden and unexplained shift from historical norms. These are embarrassing, especially if caught by a Director or VP instead of the web analytics team. However, the data is so obviously wrong that it is easy to identify.

After the fear and embarrassment of finding this error subsides, a more insidious thought emerges. “If this data was not so obviously wrong, would I have even noticed? And, how many other data points in my reports are questionable?”

Indeed, it is the lesser errors that can be more dangerous to the health of an organization. As they are less likely to be spotted, the potential increases that the less extreme errors will be used in support of business decisions that could negatively impact your company’s performance.

Data quality is a serious issue but a given in this line of work. A healthy fear and understanding of data quality must exist in all web analytics organizations. This paper will examine the components of what drives data quality problems and how firms can work towards effectively managing and minimizing the risks they present.
A QUICK CLARIFICATION:
DATA PRECISION VERSUS DATA ACCURACY

The primary concern regarding web analytics data quality is the precision of data. Many industry experts have discussed the relative importance of data precision versus data accuracy. Accuracy answers the question “how close to another reference point is your data?”. This other reference point is usually other internal data. Precision answers the question “can the data be relied upon to detect changes in performance?”.

The current state of data quality

All organizations have a problem with data quality. It is only the volume and severity of data issues that varies from one company to the next. Even if an organization accomplishes a perfect implementation of its web analytics platform, data quality problems will begin to appear as soon as any website changes are made: new products, new pages, new features, new tools, etc. Ironically, it’s often web analytics that drives that data quality drift, as the very purpose of web analytics is to provide compelling evidence to make website changes and drive continuous optimization. As a result, the more effective your web analytics team is, the higher the likelihood of more serious data quality problems occurring. Thus, if your goal is to create a world class, consistently effective web analytics organization, it is not possible to get there without a keen focus on data quality.

The dynamic nature of today’s websites and the rapid change in internet technologies ensure that there is a never-ending data quality “to do” list. Some of the more common events that should always trigger data quality checks include:

- Any new page creation, including geographic or product expansion, new site sections, landing pages for new marketing campaigns, etc.
- New content and/or site restructuring, resulting from strategic changes, web analytics or A/B testing on the site
- Launch of new website features and technologies
- Business-driven changes to classification tables or filters (especially product classifications for retailers)
- Changes to VISTA rules (or other data processing rules) or tags

The key piece of the puzzle for successful, high quality websites that is regularly overlooked is good quality data and the processes, skills and organizational rigor that drive it. In the following sections, we will review key components of a robust data quality regime and how firms can leverage strategic third party partners to cost effectively execute these programs.
Three Organizational Pillars for High Quality Data

There are three prerequisites in any web analytics organization that are critical for ensuring high quality data. The absence of or deficiency in any of these will ensure substandard data quality and ongoing deterioration.

1. Do you have the right number of resources to ensure high quality data, relative to the size and complexity of the web analytics implementation?

2. Do you have the correct skill sets, in the correct internal organizations, properly coordinated to identify and address data quality problems?

3. Do you have a commitment to data quality at every level of the organization?

DO YOU HAVE ENOUGH RESOURCES TO ENSURE HIGH QUALITY?

It is widely agreed that most organizations are underinvested in web analytics human resources, and this is one of the key contributors to low ROI on many web analytics programs. Avinash Kaushik’s oft-quoted construct that organizations should spend $9 on web analysts (people) for every $1 that they spend on analytics tools is met with applause by every overwhelmed web analyst, but met with cringes by those who manage budgets.

Kaushik’s justification for this high headcount investment stems from the inherent complexity of deriving actionable insights from the complicated sets of web data. It takes intelligence and insight to understand and anticipate the 2nd, 3rd, and 4th degree impacts that ripple through web analytics data for every primary event. Clearly, it is a poor investment to gain access to such a robust and complex set of data without investing in a sufficient number and quality of web analysts to interpret it and unlock its true value. Given the ever increasing volume and complexity of data, this staffing need continues to grow.

Implicit in Kaushik’s staffing ratio is that the human capital investment is not all focused on one skill set. As we discuss in the next section, strong analytical and insight skills don’t always include strong data quality and management capabilities. Consequently, the need is not only for right sizing the staff but also the skill set composition of the team and coordination between them all. Good analysis is wasted without good data.
DO YOU HAVE THE RIGHT SKILLS TO ENSURE HIGH QUALITY DATA?

Even if an organization has the budget to hire a well staffed team, it will still prove difficult to find the right mix of skills, in your location, necessary for good data stewardship. The challenge lays in the specialized skill set required to detect, troubleshoot, and fix the causes of inaccurate data.

Bill Gassman of Gartner articulated the problem in recent Web Analytics Demystified blog post comment titled "Does your data quality still suck":

"Data quality is important, and there is a mature discipline around achieving it. Unfortunately, the knowledge and skills for data quality disciplines generally lie in the IT department, under those responsible for data integration or business intelligence services. For too many organizations, this is the ‘don’t go there’ zone."

The individuals involved in the implementation of a web analytics platform and ensuring the ongoing accuracy of the data are often referred to as “technologists” in the web analytics community and are usually in a separate organization (often IT) from those using the analytics tool. Leaving data quality to IT can be problematic as many IT organizations won’t accept full responsibility for data quality of applications and data hosted off-premise, as most web analytics platforms are. Even a cooperative IT organization will struggle with data quality as they often lack the context of the issue and the needs of the business.
The skills required for a successful implementation and ongoing quality control are difficult to find in one person, or even one organization. The table below lists the skills essential for maintaining high quality data, along with a general assessment of the skill level of web analysts and technologists in these areas.

<table>
<thead>
<tr>
<th>SKILL REQUIREMENT</th>
<th>WEB ANALYST</th>
<th>TECHNOLOGIST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Understanding of business objectives and tracking requirements</td>
<td><img src="image" alt="Skill Level" /></td>
<td><img src="image" alt="Skill Level" /></td>
</tr>
<tr>
<td>Ability to translate business needs into technical requirements for coding</td>
<td><img src="image" alt="Skill Level" /></td>
<td><img src="image" alt="Skill Level" /></td>
</tr>
<tr>
<td>Ability to understand how tagging methodology impacts interpretation of end data</td>
<td><img src="image" alt="Skill Level" /></td>
<td><img src="image" alt="Skill Level" /></td>
</tr>
<tr>
<td>Ability to identify questionable data in reports</td>
<td><img src="image" alt="Skill Level" /></td>
<td><img src="image" alt="Skill Level" /></td>
</tr>
<tr>
<td>Ability to identify root cause of questionable data</td>
<td><img src="image" alt="Skill Level" /></td>
<td><img src="image" alt="Skill Level" /></td>
</tr>
<tr>
<td>Ability to code the tags with correct logic to achieve desired business objectives</td>
<td><img src="image" alt="Skill Level" /></td>
<td><img src="image" alt="Skill Level" /></td>
</tr>
<tr>
<td>Knowledge of tag validation tools to identify missing or erroneous tags</td>
<td><img src="image" alt="Skill Level" /></td>
<td><img src="image" alt="Skill Level" /></td>
</tr>
<tr>
<td>Thorough QA process for identifying questionable data</td>
<td><img src="image" alt="Skill Level" /></td>
<td><img src="image" alt="Skill Level" /></td>
</tr>
<tr>
<td>Thorough QA process for creating and modifying tags</td>
<td><img src="image" alt="Skill Level" /></td>
<td><img src="image" alt="Skill Level" /></td>
</tr>
</tbody>
</table>

The skill level estimates in the table above are obviously generalizations as there are definitely professionals in IT and web analytics teams with varying combinations of the above skills. However, the fact remains that the skills required to maintain data quality are so diverse that it is unlikely that any one person, or even organization within a company, possess all of them. Quite apart from whether the full complement of skills exist in an individual or organization is the fact that most often incentives and compensation (and therefore interest) for ensuring good data quality are not properly structured. So the challenge becomes two-fold: i) ensuring the skills are present; and ii) effectively coordinating and corralling them.
IS YOUR ORGANIZATION COMMITTED TO DATA QUALITY?

The best web analysts have the ability to learn the skills required to close gaps in the data quality process. However, if the organizational culture, and the analyst’s performance plan, do not place value on the analyst’s contribution to solving data quality issues, he or she will invest time and effort on other activities with a higher perceived value.

Web analysts, from a career development, compensation and personal interest perspective are motivated to deliver great insight. Rarely is this the case for ensuring and delivering great data quality. How can your organization overcome the natural tendency of analysts to deprioritize data quality issues in favor of delivering more analysis (even with suspect data)?

A critical first step is obviously senior management buy-in on the importance of data quality and a commitment to building a coordinated infrastructure to bring it to bear. Taking into account the three pillars of data quality, they must then evaluate how to most effectively implement a sound data quality regime, in terms of cost as well as execution.

A solution some more advanced web analytics organizations have successfully pursued is to integrate an experienced third party web analytics / data management expert, such as eClerx, into their web analytics organizations to undertake the less glamorous but highly critical data quality components. As they are bound by SLAs, the external partner is correctly motivated to identify and drive strong QA, leaving their in-house analysts to focus on delivering high value insights and their IT teams to focus on valuable development and key implementations.
Building Robust Data QA Processes

After an initial implementation of a web analytics solution, there are a number of ongoing activities that must be performed to ensure high quality data. Engaging a reputable third party like eClerx can provide the skill, focus and rigor necessary to build a strong foundation of quality data. There are three general areas that logically are outsourceable to such partners that will drive data quality and value:

1. Ongoing tag audits
2. New tag implementation QA
3. Troubleshooting data issues flagged by end users

ONGOING TAG AUDITS
The only constant with today’s websites is change. New site sections, new pages, new features, new content. These items are often launched and managed by a wide range of stakeholders, often geographically distributed, making tagging governance all the more difficult.

Engaging with an experienced third party to establish a cadence and discipline of ongoing tag audits, using a combination of automated site scanning and debugging tools and focused manual checks, can identify missing tags, uncover javascript errors, and create a close-to-real-time map of custom variable tags throughout a site. In building a continuous audit culture, the firm gains a higher level of confidence in their tagging accuracy and the quality of their web analytics implementation.

NEW TAG IMPLEMENTATION QA
The ongoing tag audits described above help uncover tagging problems on pages and features already live on your site. However, ideally a company will execute a QA process that ensures appropriate site tagging before new pages or functionality go live.

Sophisticated organizations often have development environments for their websites. Ideally, testing and validating web analytics tagging is part of the functionality of these "dev environments". An even more advanced approach is to test different tagging logic and optimize data capture, reporting and insight pre-go live.

As more complex tagging logic is added, the risk of unintended consequences rises and the need to perform QA becomes more critical. The typical QA process is not particularly complicated, but it can be tedious and time consuming, ultimately interfering with your team focusing on the higher value added analysis.

For a variety of reasons, most organizations do not have a separate team performing tagging QA. Instead, this QA is often performed by the same coders and javascript experts that are placing the tags on the site. Clearly, this is not ideal. First, it is poor QA process to use the same person (or team) to perform both the execution and the quality assurance – unintended biases may lead to less critical evaluation. Second, it is a misuse of highly skilled developers’ time and capabilities – with finite hours in the day, QA will always take a back seat. By layering in a structured, third party solution, the developers can focus on development, implementation and other key initiatives. Their only demands from a QA perspective are reduced to rectifying issues identified by the audit team, rather than trying to identify any implementation issues themselves.
TROUBLESHOOTING DATA ISSUES
FLAGGED BY END USERS
Despite all efforts to ensure accurate tagging pre-go live, it’s inevitable that your web analytics tools will sometimes produce unexpected data. As web analysts and other end users perform analysis and study dashboards, they will periodically uncover questionable data. When suspect data is found, the fire drills, deep dives and troubleshooting begins. While the exact steps depend on the nature of the problem, they often include:

• Trend analysis to identify the day the unexpected change occurred, to determine if a known site change, IT release, or tagging change could have caused the issue
• Segmentation or filtering analysis to drill down on the specific problem
• Deploying site scanning tools to identify which locations on the site are contributing to the issue. For example, on which pages is a custom variable tag firing?
• Using a debugging tool to see what data is being passed to the web analytics server for the variable(s) in question

It can take hours, or even days, to thoroughly troubleshoot, document, and fix erroneous tags. While important to resolve, is this the best use of a web analytics team’s time to stop or slow their analytics activities to resolve data issues? Companies often struggle with commitment to a QA process and it always feels like a fire drill. This third area is also well suited for external partners to undertake, ensuring the in house teams aren’t pulled from their analytics responsibilities while still driving resolution of identified problems.

CONCLUSION
There are never enough hours in the day or dollars in the budget to undertake all that’s possible in web analytics. And even with what we do focus on, we are haunted by the specter of bad data undoing good analysis. However, the power of analytics is gaining traction in the executive suite and forward thinking firms are realizing that data quality and analytics must go hand in hand. As executives structure their analytics organizations, opportunity now exists to cost effectively layer in expert third party partners like eClerx to undertake the critical but time consuming heavy lifting of data management and QA. Such partnerships raise everyone’s potential and focuses their efforts on what they want to do:

• Web analysts can focus on the high value add of delivering valuable, business changing insights
• Technology teams can focus on developing and implementing high value applications and systems
• The firm benefits from having a more motivated, productive workforce delivering functionality and insight that drives competitive advantage

CITATIONS & REFERENCES
AUTHOR BIOGRAPHY

DAVID NELSON is the Practice Lead for Web Analytics and Competitive Intelligence at eClerx, where he works with clients to build teams and customized processes for organizations involved in web analytics, sales and marketing analytics, pricing, and competitive intelligence. He has over 13 years of experience supporting and managing analytics teams at Fortune 50 companies. During his time at Intel and Dell, David led analytics efforts across a wide range of corporate activities, including web analytics, online testing and targeting, multi-tiered distribution channel analysis, pricing and promotion optimization, competitive intelligence and benchmarking, mergers and acquisitions due diligence, and capital and operating expense analysis. David holds an MBA from the McCombs School of Business at the University of Texas and a Bachelors of Science from Indiana University (Bloomington).

About eClerx

eClerx is the leading knowledge process outsourcing company providing middle and back office operational support, data management and analytics solutions to sales and marketing organizations of over 50 Global Fortune 500 and Internet Retailer 500 scale companies. With a particular focus on the burgeoning digital marketing universe and the industry's largest offshore third party web analytics team, our clients are among the world's leading Online Retail & Distribution, Software, High Tech and Industrial Manufacturing, Interactive Media and Entertainment and Travel & Leisure companies.

Incorporated in 2000, we provide our clients with expertise in consulting, rigorous process work-flows, automation and improvement and robust outsourcing solutions to help reduce costs, drive actionable insights to lead fast-paced markets and correct long-standing process inefficiencies. With over 4,500 professionals across five delivery centers worldwide, no job is too large or complicated. Our corporate offices are based in New York, London, Silicon Valley, Austin, Chicago, Dublin, Singapore and Dalian (China).

For more information, please visit www.eclerx.com.

Contact eClerx
Americas
Scott McCartney
+1 646-368-6161
scott.mccartney@eclerx.com

EMEA
Joseph Sursock
+44 (0) 20-7529-6008
joseph.sursock@eclerx.com

Asia Pacific
LiChien Koh
+65 9625-5078
lichien.koh@eclerx.com