### The Mandela Portal – an assessment of global visitors using web analytics

<table>
<thead>
<tr>
<th>Journal</th>
<th>Digital Library Perspectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>DLP-01-2016-0002.R1</td>
</tr>
<tr>
<td>Manuscript Type</td>
<td>Case Study</td>
</tr>
<tr>
<td>Keywords</td>
<td>Big Data, Data Trustworthiness, Google Analytics, Nelson Mandela, Mandela Archive, Portal</td>
</tr>
</tbody>
</table>
1.0 Introduction

The Nelson Mandela Foundation (NMF) was established in 1999 as the post-presidential office for Mr. Nelson Mandela upon his retirement as South Africa’s first democratically elected president. Mr. Mandela had promised to serve only one five-year term, and upon retirement felt that he needed to continue working to promote social justice (Harris, 2012). He was actively involved in the work of the NMF for the first five years, but in 2004 announced he was “retiring from retirement” (Mandela, 2004). His stepping away from public life, and from direct involvement in NMF operations, sparked a review of the organisation’s mandate, vision and mission. The review was completed in 2006, and adoption of the review report by the NMF Board of Trustees introduced a five-year transition process designed to transform the institution into the Nelson Mandela Centre of Memory (Josias, 2006).

In early 2010, Mr. Mandela stopped going to the premises of the NMF as he went into full retirement and the institution embarked on its transition beyond being his post-presidential office. In the last five years the transition into a streamlined non-governmental organization, promoting social justice through memory and dialogue work has continued in earnest with various elements including changes in the building infrastructure and organizational activities. At the core of the mandate of the NMF is to document and facilitate access to the Mandela Archive which is infinite, fragmented and scattered both geographically and institutionally. Since the NMF does not envisage bringing the archive into a physical location, the objective is to use the Mandela Portal as the key avenue to provide access the Archive.

1.1 Conceptualising the Mandela Portal

The NMF conceptualised a multi-layered virtual archive or portal that would be accessible through its website. The website was initially developed in the early 2000s and had already undergone several phases of development by the late 2000s (Katuu et al., 2011). The drafting of the Mandela Portal began in 2008 with an architecture that had four design elements:

- databases providing a dense description of resource materials;
- linkages to actual materials, to other websites and to different layers within the website;
- digitised materials including paper, sound and moving images; and
• a surface layer of stories and information (Nelson Mandela Foundation, 2009a, p. 22-23).

These elements were further integrated with social media platforms that relate to each of these elements were launched in April 2011 as an integrated Mandela Portal (Nelson Mandela Foundation, 2012a, p. 14). By 2014 the Portal had “established itself as the most trusted and widely used internet resource for research on the life and times of Nelson Mandela” (Nelson Mandela Foundation, 2014, p. 11). In this regard the Mandela Portal that was initially developed to provide web presence for the nascent Centre of Memory had grown to becoming its premier communication platform. Over the years the Portal has offered a rich resource of content including databases on Mr. Mandela’s speeches, archival material on the Rivonia Trial, Speeches, and a bibliography of books. In addition, it has a tributes databases that are an inventory of thousands of civic honours and awards given to Mr. Mandela from educational, arts and cultural as well as sports institutions (Nelson Mandela Foundation, 2016). These databases have opened the resources of the Portal to new audiences and "provided traditional users a way to access, use and repurpose materials" without the mediation of archivists and other information professionals at institutions such as the Centre of Memory (Ndebele, 2013, p. 1, Prom, 2011, p. 160). Therefore, web analytics tools facilitate monitoring and gathering of information about visitors’ online interactions taking place around the clock on a global scale.

1.2 Assessing visitors to the Mandela Portal
The formal publications as well as the scholarly analysis provided valuable insight as a basis for
the continuous improvement of the Portal in its early years. For instance, the analysis in 2009
led to the development of a mobile version of the Portal with statistics of traffic published in the
included design elements that allowed responsive features for mobile devices to the Portal
therefore eliminating the need for a mobile platform.

Another example of the important of analysis emanated from the 2009 report that demonstrated
that 80% of the visitors to the Portal had English as the default language setting on their web
browsers (Katuu, 2009). This prompted the inclusion of language translation in the 2012
upgrade of the Portal with options of nine languages: Afrikaans, Dutch, Finish, French, German,
Italian, Portuguese, Russian, Spanish and offering the possibility to use Google Translate for
other languages (Nelson Mandela Foundation, 2012b).

 Nonetheless these early efforts at analysis as well as the annual reports have not yet offered a
composite view of the longer term trends. This article is a longitudinal assessment of the visitors
to the Mandela Portal using web analytics over a period of seven years, between 2009 and
2015. This is a period that witnessed many events including the long-term hospitalisation and
eventual passing away of Mr. Mandela. In the seven year period, the Portal has endeavored to
establish itself as the most trusted resource on the life and times of Mr. Mandela. This article
describes the methodology, findings as well as a discussion on trends and issues arising from
the assessment activities.

2.0 Methodology
Archivists, library and other information professionals often lack a “systematic understanding” of
how people interact with descriptive information and digital objects that are accessible through
their institutions’ website portals (Arendt and Wagner, 2010, Fang, 2007). Web analytics is seen
as a method they could use to “measure user actions, to understand some aspects of user
behavior” and to initiate ways of improving their online offerings (Prom, 2011, p. 158).
Web analytics is defined as “the measurement, collection, analysis and reporting of Internet data for the purposes of understanding and optimizing web usage” (Digital Analytics Association, 2008, p. 3). Web analytics constitutes four steps: collection of data, processing, developing performance indicators and formulating an online strategy to meet institutional goals (Jansen, 2009). The four steps are illustrated in the diagram below (Wikipedia, 2015).

**Figure 1: Basic steps of the web analytics process**

Web analytics has mostly been used by commercial entities to enhance online marketing strategies (Prom, 2011, p. 162). In a few cases libraries and cultural institutions in the not-for-profit sector have used web analytics methodology to help understand their visitors’ needs and behaviour (Fang, 2015, Marek, 2011, Woody, 2011). This article uses the first two steps of web analytics to assess the global visitors to the Portal in order to identify longitudinal trends, discuss their implications and demonstrate their impact on development work on the Portal.

There are several web analytics tools in the market but the most influential and freely available one is Google Analytics which the NMF adopted from November 2007. The first year of its use was spent integrating it to the website platform as well as understanding the reporting.
mechanisms. Therefore, this analysis draws data on the web traffic to the Portal between January 2009 and December 2015. The seven-year period provides a rich source of information on, among other things, the number of unique visitors and countries they accessed the Portal from, average time spent on the Portal as well as favourite pages and the different settings on their web browsers.

3.0 Findings

There is considerable debate on the reliability of data captured by web analytics tools (Clifton, 2012, p. 26-29, Thomas, 2013, p. 17). For instance, the means to identify a unique visitor are still undergoing refinement (Menken, 2012, p. 20). As well, whether a page view consists of one hit to the page or different parts of the page is also debatable. Therefore, for the purposes of this study the base statistics are drawn from what Google Analytics terms as sessions defining each as “a group of interactions that take place on the website within a given time frame” (Google, 2016a). A session combines screen views, events social interactions and transactions and therefore, the statistical figure will be much less than the number of page or screen views but represents a more organic engagement with a website.

This section describes the findings of the study covering the period between 1st January 2009 and 31st December 2015 that had a total of 13,055,218 sessions on the Mandela Portal. The findings begin with general trends of global figures and explore more specific trends using an analysis of sub-continents, top three and top ten countries as well as language preferences.

3.1 Cumulative global statistics between 2009 and 2015

The seven year period shows a growth in the number of visitor doubling from 2009 to 2010, dipping in 2011 and going up again in 2012. Understandably in 2013, due to the dramatic events of the long illness and eventual passing on of Mr. Mandela, the statistics are more than three times all the previous years combined. 2014 shows a reduction of visitors by almost 60% and 2015 another 60% reduction. Nonetheless, even the 2015 figures are quite high compared to the pre-2013 figures. The figure below illustrates the trend of visitors over the seven year period.
3.2 Cumulative statistics of sub-continents between 2009 and 2015

Section 3.1 provides a very general picture and this section examines the statistics at the sub-continental level. Google Analytics offers statistical rendering using geographical mapping. The diagram below is a graphical representation of the statistics by sub-continent, where the darker colour are a higher number of sessions than the lighter colour.
According to the data, Google Analytics identified a total number of 24 sub-continents around the world with the ranking showing the lowest statistics as 12 and the highest at 4,465,522 visitor sessions. The table below shows the ranking of the top ten sub-continents and the percentages of web traffic statistics.

<table>
<thead>
<tr>
<th>Sub-continent</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northern America</td>
<td>34.02%</td>
</tr>
<tr>
<td>Southern Africa</td>
<td>13.49%</td>
</tr>
<tr>
<td>Northern Europe</td>
<td>11.89%</td>
</tr>
<tr>
<td>Western Europe</td>
<td>8.46%</td>
</tr>
<tr>
<td>Sub-continent</td>
<td>Percentage</td>
</tr>
<tr>
<td>-----------------</td>
<td>------------</td>
</tr>
<tr>
<td>(not set)</td>
<td>4.81%</td>
</tr>
<tr>
<td>Southern Asia</td>
<td>4.72%</td>
</tr>
<tr>
<td>Australasia</td>
<td>4.08%</td>
</tr>
<tr>
<td>Southern Europe</td>
<td>3.75%</td>
</tr>
<tr>
<td>South America</td>
<td>3.51%</td>
</tr>
<tr>
<td>Eastern Africa</td>
<td>2.79%</td>
</tr>
</tbody>
</table>

Table 1: Cumulative statistics of sub-continents between 2009 and 2015

The table illustrates the vast geographical spread of the visitors to the Portal as well as the clear domination of Northern America that carries one third of the statistics, which is more than the second, third and fourth sub-continents combined. In addition, there is the curious categorization of “not set”. Google (2016c) acknowledges that this is the result of using third-party vendors who may not have an accurate record of the visitor location and a more extensive discussion on the issues is in Section 4. Nonetheless, it is worth noting the “not set” category has large enough statistics to be considered the fifth position in the sub-continental ranking.

3.3 Annual statistics of the top three countries between 2009 and 2015
The sub-continent discussion is Section 3.2 offered a first level of specificity and this section provides addition insight on the top three countries out of a total of 242 ranked countries. The illustration below shows the four countries that have held the top three positions between 2009 and 2015.
Figure 4: Annual statistics of the top three countries between 2009 and 2015

As shown above the top three countries between 2009 and 2015 were United States, South Africa and United Kingdom. The United States has always held the first position, South Africa second place except in 2013 where the United Kingdom was second. The United Kingdom always held the third position except in two years, 2013 when it was second and 2015 when it was fourth. The most dramatic change is the sudden rise of Kenya into third position in 2015 and this will be discussed in more detail in Section 4.
3.4 Cumulative and annual statistics of the top ten countries between 2009 and 2015

Section 3.3 offered specific insight on the top three positions and this section casts the net wider to the top ten countries out of a total of 242 ranked countries. This section looks at the statistics at two levels, first based on the overall figures in the seven year period and then individually year-by-year. The table below shows the cumulative figures from 2009 to 2015 for the top ten ranking countries.

<table>
<thead>
<tr>
<th>Country</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>29.79%</td>
</tr>
<tr>
<td>South Africa</td>
<td>13.21%</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>9.06%</td>
</tr>
<tr>
<td>“not set”</td>
<td>4.81%</td>
</tr>
<tr>
<td>Canada</td>
<td>4.41%</td>
</tr>
<tr>
<td>India</td>
<td>3.89%</td>
</tr>
<tr>
<td>Australia</td>
<td>3.49%</td>
</tr>
<tr>
<td>Germany</td>
<td>2.82%</td>
</tr>
<tr>
<td>France</td>
<td>2.63%</td>
</tr>
<tr>
<td>Brazil</td>
<td>2.30%</td>
</tr>
</tbody>
</table>

Table 2: Cumulative statistics of the top ten countries between 2009 and 2015

The table above offered a number of observations. First, “not set” is a category that ranks fourth constituting 4.81% of the statistics. Secondly, while Section 3.3 shows that Kenya was ranked third in 2015, its cumulative statistics were not enough for the country to appear in the top ten ranking over the period of seven years. Third, when these results are compared to those in Section 3.2, there are curious similarities in the top four sub-continents. Northern America has the first as well as fifth ranking countries and Southern Africa has the second ranking country. Northern Europe had the third ranking country while Western Europe had the eighth and ninth ranked countries. The illustration below has the 14 countries that ranked in the top ten listing between 2009 and 2015.
**Figure 5: Annual statistics of the top ten countries between 2009 and 2015**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>100,086</td>
<td>193,317</td>
<td>185,348</td>
<td>189,040</td>
<td>185,784</td>
<td>960,080</td>
<td>911,766</td>
</tr>
<tr>
<td>South Africa</td>
<td>73,050</td>
<td>114,451</td>
<td>132,011</td>
<td>146,964</td>
<td>148,932</td>
<td>293,080</td>
<td>373,195</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>35,463</td>
<td>74,904</td>
<td>54,057</td>
<td>64,382</td>
<td>77,113</td>
<td>235,203</td>
<td>210,318</td>
</tr>
<tr>
<td>New Zealand</td>
<td>17,845</td>
<td>17,499</td>
<td>33,390</td>
<td>61,559</td>
<td>110,590</td>
<td>108,936</td>
<td>103,315</td>
</tr>
<tr>
<td>Canada</td>
<td>5,423</td>
<td>50,050</td>
<td>51,632</td>
<td>72,086</td>
<td>234,274</td>
<td>130,076</td>
<td>108,936</td>
</tr>
<tr>
<td>India</td>
<td>15,326</td>
<td>94,740</td>
<td>31,005</td>
<td>24,441</td>
<td>189,379</td>
<td>162,541</td>
<td>108,934</td>
</tr>
<tr>
<td>Australia</td>
<td>10,180</td>
<td>35,048</td>
<td>16,556</td>
<td>23,905</td>
<td>100,736</td>
<td>134,627</td>
<td>60,709</td>
</tr>
<tr>
<td>Germany</td>
<td>4,882</td>
<td>13,896</td>
<td>8,367</td>
<td>27,030</td>
<td>101,387</td>
<td>70,536</td>
<td>54,781</td>
</tr>
<tr>
<td>France</td>
<td>4,680</td>
<td>10,507</td>
<td>6,367</td>
<td>25,801</td>
<td>152,806</td>
<td>70,533</td>
<td>54,781</td>
</tr>
<tr>
<td>Brazil</td>
<td>4,571</td>
<td>11,971</td>
<td>8,087</td>
<td>17,070</td>
<td>116,433</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norway</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Netherlands</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kenya</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The table demonstrates the dominance of the three top countries over the others throughout the seven year period. In addition, the illustration exemplifies that “not set” was most prominent in 2010, 2012 and 2013 when it made the top 10 countries. In particular the figures were very high in 2013 which was the peak year for the Portal. This suggests that the identification of visitors was worst when statistics were higher.

When compared with the statistics in Section 3.2, the dominance of Northern America, Southern Africa as well as Northern and Western Europe are more apparent due to the contribution of one or two countries in the particular regions. The presence of Brazil, India and Kenya demonstrates that though certain regions dominate there was increasing traffic from other regions of the world.

3.5 Cumulative statistics of the language settings to web browsers between 2009 and 2015
Section 3.1 to 3.4 relied on one set of data to demonstrate regional diversity by looking at sessions from global, sub-continent and country levels and offering multiple perspectives on the visitors to the Portal. This section uses a different data set, that of the language settings of the web browsers to explore linguistic diversity. Google Analytics identified a total of 615 language settings with the top 20 shown in the illustration below.
Two main observations can be drawn from the illustration above. First, the listing is dominated by languages that originate from Europe but are used in various parts of the world including Australia, Brazil, Canada, South Africa and United States. When compared with statistics from Section 3.4 the linguistic dominance is evident despite the fact that there are countries like South Africa, Kenya and India that have other non-European languages.
Second, the table also shows that there are different variations in the language settings on the browsers. English clearly dominates with six versions of language settings that amount to almost 80% of all the statistics. French is second with two versions of language settings and just above 3%. The other languages in the top 20 list are German, Portuguese, Spanish, Italian, Danish and Dutch with each having two versions and all below 3%. An analysis of more languages in the full list shows that for all the 615 identified there were a large number of varieties demonstrating very granular web browser language settings.

4.0 Discussion
Fang (2007, p. 15) argued that Google Analytics is a great tool for constructing user-centered websites because it offers “a user-friendly interface and information reports that allow for quick identification of problems”. The findings in this article have revealed a number of aspects related to trends observed in the data as well as the analytics tools. This section is a discussion of the issues. Section 4.1 examines the web traffic trends, Section 4.2 explores weaknesses observed in Google Analytics that would be common to web analytics tools, and Section 4.3 outlines the impact of the web traffic trends on the development of the Mandela Portal.

4.1 Web traffic trends
The findings demonstrated a number of consistencies as well as contradictions. First, the statistics from the individual countries showed the clear dominance of three countries throughout the seven year period of the study i.e. United States, South Africa and United Kingdom. The findings also showed the sub-continents of the dominant countries also ranking highly.

Second, the statistics showed the web traffic peaked in 2013 consistent with the global attention on the long-term hospitalization and eventual passing on of Mr. Mandela. After 2013 there was a gradual decline from the peak numbers. Upon further examination of the figures in 2014 and 2015 there is a decline in specific countries such as the United Kingdom, France and Canada in 2015. A few countries like South Africa generally maintained the same level but Kenya increased drastically in 2014 and 2015. The gradual decline post-2013 could be attributed to a number of factors such as the once-in-a-lifetime unique events of 2013 and possibly, to a
smaller extent, the increasing dispersion of visitors through social media platforms used by the NMF.

Third, the issue of Kenya’s dramatic change to being amongst the top three countries in 2015 deserves further exploration. The dramatic ascendance could be attributed to a number of factors. First, there is a generally held view that there is an increase in access to internet within developing countries and particularly in sub-Saharan Africa. While there are studies that show increasing numbers of internet users, no studies show such dramatic increases within a period of one or two years (Gillwald et al., 2013, Hansson and Jobe, 2015, West, 2015). Second, is the notion that the Portal may have been particularly attractive to a Kenyan web audience. There has been Kenyan related content on the Portal but this predates the 2014 and 2015 period because it includes Mr. Mandela hosting President Uhuru Kenyatta in 2009 at the time when he was Deputy Prime Minister (Nelson Mandela Foundation, 2009b). A third and more likely reason is that Google Analytics may have increased the accuracy of its data on Kenya between 2014 and 2015. This is supported by various reports that demonstrate Google’s office in Kenya being involved in a number of mapping activities including building of small and medium size businesses beginning in 2012 culminating in the current Kenyan Business Online portal (Google, 2016b, Halliday, 2012). In addition, Google has supported mapping of households in health related activities (Centers for Disease Control and Prevention [United States], 2015) as well as the transit system (Mindock, 2015, Walker, 2014). Additional research is needed to explore whether other African countries with similar mapping projects would also see a rise in their web traffic figures.

Fourth, the findings demonstrate that English has dominated the web browser language settings which is consistent with the observations made in the 2009 study where the figure was almost 80% (Katuu, 2009). This shows that during the course of the seven year period visitors may come from different regions but still primarily engage in English language. This would suggest that either nothing has changed or if there has been a change then it cannot be detected by the analytics tool. One thing that has not changed is that most of the content on the Portal is in English. Gäde (2014) noted that, in general, English dominates most language settings in web browsers suggesting that this is a global phenomenon. Nonetheless, the NMF made efforts as early as the first quarter of 2012 to offer different language options on the Portal through the
Google Translate feature (Nelson Mandela Foundation, 2012b). The lack of change in the figures may suggest that additional research is needed to explore the number of those who may access the Portal in English but use the translation feature to convert the content to their own language.

4.2 Weakness in web analytics tools

Fundamental to any web analytics tool is the ability to provide trustworthy statistics but such tools struggle to conclusively determine visitor statistics. Menken (2012, p. 20) described it in the sense that “the unique visitors for each day in a month do not add up to the same total as unique visitors for that month”. This is commonly referred to as the “hotel problem” illustrated through the analogy of a hotel with two rooms and three visitors John, Mark and Jane as shown in the table below.

<table>
<thead>
<tr>
<th></th>
<th>Day 1</th>
<th>Day 2</th>
<th>Day 3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Room A</td>
<td>John</td>
<td>John</td>
<td>Mark</td>
<td>2 unique users</td>
</tr>
<tr>
<td>Room B</td>
<td>Mark</td>
<td>Jane</td>
<td>Jane</td>
<td>2 unique users</td>
</tr>
<tr>
<td>Total</td>
<td>2 total visitors</td>
<td>2 total visitors</td>
<td>2 total visitors</td>
<td>?</td>
</tr>
</tbody>
</table>

Table 3: Determining a web visitor using the hotel problem

The table shows that the hotel had two visitors each of the three days and therefore the sum should be six visitors in total. However, during each period each room had two unique visitors and therefore the sum of each of the rooms should be four unique visitors. This creates problems because the numbers do not add up considering there were actually only three individuals over the three day period in the hotel’s two rooms.

The inability to determine unique as well as total number of visitors is probably at the core of the “not set” phenomenon observed in the findings. If visitors could not be determined precisely then it would be difficult to connect them to particular geographic locations. Google (2016c) admitted that it collects geographic location information from third party vendors and relies on them for accuracy stating...
“Google Analytics uses a third-party data source to determine your visitors’ geographical locations. If our third-party vendor does not have an accurate record of the visitor location, Google Analytics displays a (not set) entry”

Since geographic locations are determined using visitors IP addresses complications arise with visitors that use virtual connections or access the web via mobile devices that could be moved easily across terrain as demonstrated in the hotel analogy (Barnes, 2015). This may explain why Google Analytics identified 242 countries yet the officially recognised country figures are less than 200 countries (United Nations, 2016a, United Nations, 2016b). There are other reasons for inconsistent data including the need to filter out data from one’s own IP address since it may skew the totals and the ability of web users to visit a site anonymously which may not be accurately identified by the analytics tool (Hines, 2011). Therefore, it is not surprising there was a recurring “not set” theme in sections 3.2 and 3.4. In cumulative statistics “not set” was the fifth largest sub-continent and the fourth largest country. It is also apparent that the bulk of the “not set” data was collected in 2013 consisting of just over 8% of statistics that year. Hence the impact of the anomaly cannot be understated.

As demonstrated above, since there are obvious weaknesses in the analytical tool statistics. Additional questions would be asked about other statistical information such as number page views, bounce rates, pages per visit and average time per visit particularly when such information could also be generated by web robots that perform automated tasks over the internet. This is the likely reason Google Analytics chooses session statistics as the default or preferred data grouping. A session is defined as “a group of interactions that take place on the website within a given time frame” (Google, 2016a). It combines screen views, events social interactions and transactions. Therefore, the figure will be much less than the number of page or screen views and represents an organic engagement with a website and is more likely to be the result of human interactions. When measurable concepts such as the number of unique visitors, length of time on a site, page views, bounce rate and frequency are combined they can be related to engagement concepts such as recognition, reciprocity, expectations and duration (Padua, 2012, p. 170, Prom, 2011, p. 163). While web analytics concepts are measurable, engagement concepts are not, one uses the measurable aspects to determine the
unmeasurable ones in order to determine whether, over time, visitors demonstrate trust in a website or portal.

Web analytics is seen as the precursor to the emerging discipline of big data that, unlike web analytics, has wider applications beyond the virtual web environment (Sun et al., 2015, p. 206-207, Thomas, 2013, p. 155, Verhoef et al., 2016, p. 194). Thomas (2013, p. 155) argued that the discussions on big data focus “primarily or exclusively on the technology” thereby crowding out discussions about “the real purpose of implementing big data systems, which is to enable people across organizations to rely on data” in order to inform and support decision making processes. The fundamental discussion on the trustworthiness of data generated by web analytics tools is one that should concern big data experts who also use tools to capture, analyze and curate information. As demonstrated in this article, the categorization “not set” is a legitimate concern and research is needed to validate the trustworthiness of the data collected in the years it was most prominent. This may mean revisiting of the data in order to revalidate the observations but this presumes that it would be possible to audit what was captured and how it was curated at that moment in time. The aspect of trustworthiness is necessary in order to guarantee “documentary truth” i.e. whether the moment of action and moment of documentation are close enough to trust the data (Duranti, 1990, p. 4). Lemieux (2014, p. 138) discusses big data and argued

To address missing data, organizations can seek to actively create trustworthy sources of data, which includes capturing suitable metadata and information about data schemas to enable interpretation and analysis of data. Active steps must also be taken to preserve data, which begins with assigning appropriate responsibilities and accountabilities for the management of data over time and putting in place the necessary resource description infrastructure (i.e. metadata) to preserve data.

This article didn’t explore other statistical aspects of the Mandela Portal data such as bounce rates, pages per visit, average time per visit all of which would deserve in-depth analysis in order to contribute to the debates about the trustworthiness of visitor statistics. The discussion demonstrates the web analytics could offer incredible insight but this requires sustained efforts in exploring avenues to guarantee the data being analysed can be trusted.
4.3 Impact of Web Analytics on the Mandela Portal

Fang (2007, p. 10) argued that while Google Analytics “can report facts about a monitored website” it is “unable to make suggestions” on how to improve such a website. Section 4.1 of this discussion demonstrated that trends may sometimes show consistencies or contradictions. Section 4.2 demonstrated that there are inherent weaknesses in web analytics tools. Therefore, interpreting trends drawn from data requires understanding a much larger context than the analysis results and may also challenge the trustworthiness of the data itself. Nonetheless, web analytics tools are still useful, evidenced in the number of institutions that have used web analytics in similar ways to the Mandela Portal (Arendt and Wagner, 2010, Fang, 2015, Prom, 2011, p. 176).

Prom (2011) described how the University of Illinois Archives implemented a pilot project in 2007 to test “the utility of Google Analytics for measuring and analyzing” use of their website. A direct consequence of the project was the redesign of a common landing page of the website and the showcasing of descriptions of their archival material (Prom, 2011, p. 176). Similarly, between 2009 and 2011, the NMF made efforts to support the developments of the Mandela Portal by incorporating design elements based on early analysis made from Google Analytics data (Katuu et al., 2011, Katuu and Hatang, 2010). Based on this analysis the Portal upgrade implemented in 2012 included design elements that allowed responsive features for mobile devices to the Portal. In addition, between 2012 and 2015, as the influence of social media grew globally the NMF increased its use of social media platforms to supplement its communication avenue (Nelson Mandela Foundation, 2013, p. 11, Nelson Mandela Foundation, 2014, p. 29). The use of Google Analytics has made it possible to monitor the number of referrals made from social media back to the Portal. For instance, in 2015 the NMF (2015, p. 57) reported that 2.74% online visitors landed on the Portal through referrals from Facebook. While this figure is still quite small compared to the 77.33% of the visitors that used search engines and 13.35% of the visitors that landed on the Portal directly, an awareness of trends over time allows for informed decision making on how best to leverage opportunities brought about from social media platforms in further fulfilling the Mandela Portal mandate. Additionally, based on recent analysis of the Google Analytics data, the Portal has implemented a Universal Search tool which references all deep databases as well as surface layer content (Davies, 2016).
5.0 Conclusion

Web analytics is evolving and its methodology as well as terminologies are consistently being refined (LaCugna, 2013, Sostre and LeClaire, 2007, p. 309). This article used data drawn from Google Analytics tool to analyse web visitors to the Mandela Portal and discuss some of the trends and patterns as well as their implications on the Portal’s development. The Mandela Portal is at the core of facilitating global access to the Mandela Archive and is hosted by the NMF. This article outlined the conceptualisation of the Mandela Portal and the need to do an assessment of the visitors to the Portal. The period between 2009 and 2015 witnessed just over 13 million sessions of web visitors to the Portal. The ranking of countries showed a dominance of three nations: United States, South Africa and United Kingdom. However, an inconsistency is the dramatic rise of visitors from Kenya in 2014 resulted in a top three ranking in 2015. An exploration of reasons suggests this is more likely the result of weaknesses in accuracy of data collected by the tool. Finally the rankings also showed the dominance of English language in the web browser settings which is consistent with an earlier study (Katuu, 2009).

While web analytics has predominantly been used to formulate online strategies that are often in corporate monetary terms, this study demonstrates the value created to a not-for-profit institution that used it to better understand its web audience and make appropriate portal development decisions. For instance, this was initially critical in the decision to create separate mobile site between 2009 and 2010 and informed the decision to explore web design aspects that would be responsive to different kinds of devices (Katuu, 2009, Nelson Mandela Foundation, 2011, p. 50, Nelson Mandela Foundation, 2012a, p. 73).

This study explored a sliver of the vast amount of data available on the visitor statistics on the Mandela Portal. However, this initial exploration reveals that while web analytics tools are useful there are questions related to trustworthiness of the data that need further exploration.

Acknowledgement: The author was a staff member at the Nelson Mandela Foundation for a period of five years from 2005 and contributed to the development of the Mandela Portal. The
author gratefully acknowledges the thoughtful readings of previous iterations of the article by Lee Davies, Verne Harris and Sello Hatang. The views expressed herein are solely those of the author and should neither be attributed to his current employer, the International Atomic Energy Agency nor any of his previous employers including the Nelson Mandela Foundation.

References


Clifton, B. (2012), *Advanced web metrics with Google Analytics*, John Wiley & Sons, Indianapolis, IN.

Davies, L. (2016), "Personal email communication", Johannesburg, South Africa.


Menken, I. (2012), Web Analytics Complete Certification Kit - Core Series for IT, Emereo Pty Ltd, Brisbane, Queensland.


Sostre, P. and LeClaire, J. (2007), Web Analytics for Dummies, John Wiley & Sons, Indianapolis, IN.


