Mozambican Government Portal
Case Study: Visitor Analysis

Valter CUMBI¹, Lourino CHEMANE¹,²
¹UTICT, Guerra Popular Avenue, No. 20, 2nd Floor, Room 201, Maputo, Mozambique
Tel: +258 21 309398, Fax: + 258 21 302289, Email: valter.cumbi@infopol.gov.mz
²University of Stockholm, DSV, Forum 100, 164 40 Kista, Stockholm, Sweden
Tel: +46 8 161175, Fax: + 46 8 7039025, Email: si-lac@dsv.su.se

Abstract: The e-Government initiative in Mozambique is an ambitious and far-reaching movement with immense potential for positive impact on the Mozambican government and economy. Central to this initiative is the Government Portal—a portal dedicated to intuitively and helpfully disseminating government news, services, and information to citizens, the public sector, and the private sector. Such an important project of the e-Government initiative requires frequent evaluation of the project performance and ways in which it can improve. This document attempts to measure the performance of the Government Portal through visitor analysis. It discusses the limitations of analysis through this approach, what can be deduced from the visitor analysis, and possible ideas to improve the project in an attempt to help it reach its objectives.

Keywords: hits, logs, visits, Webalizer, IP address, hostname, e-Governance.

1. Background

In recent years Mozambique has embarked on an ambitious endeavour to re-engineer how the Government operates through the Public Sector Reform Strategy [15] in an attempt to expedite the development of the country as a whole and of its citizens, and to reduce poverty [14]. This endeavour seeks to, amongst other things, foster an environment conducive to a more informed citizen (empowering him to participate in overall governance), and improved government operation and service delivery. Information and Communication Technology (ICT) is an important and central component of this endeavour and it is being strongly leveraged through the adoption of policy and strategic documents in this field such as the ICT Policy [10], ICT Policy Implementation Strategy [1], Universal Access Strategy [11], Telecommunications Policy, and the e-Government Strategy [2].

1.1 Mozambican E-Government

Within the last decade the Mozambican government has been pursuing innovative ideas and projects that fundamentally revolutionize the way in which it operates [1]. This movement is best exemplified by the Public Sector Reform (PSR) program which strives to [15]:
1. Improve service delivery by decentralization and institutional restructuring;
2. Strengthen the policy formulation and monitoring process;
3. Enhance professionalism in the public sector;
4. Improve financial management and accountability;
5. Promote good governance and combat corruption.

To ensure ICT support to the second phase of the PSR, scheduled for 2006-2011, the government is leveraging two critical programs: the ICT Policy and Implementation
Strategy and the Mozambican e-Government Strategy. These programs aim to “increase efficacy and effectiveness of services availed by government agencies and other public entities to the citizen and private sector as well as other government and development partners through the use of ICTs [1].”

While pushing for the use of ICTs, the Mozambican government draws upon the vast knowledge and experience gained from other countries that have successfully integrated ICTs into their government. These countries, ranging from developed to developing, have used ICTs and the reengineering of government processes to provide citizen-centric services [3]. Despite the immense public sector restructuring efforts necessary to provide citizen-centric services, these countries have proved that if correctly implemented the results are “interoperable information systems working in a seamless and coherent way across the public sector to provide better and more efficient services tailored to the needs of citizens and businesses at lower costs [4].” In other words, not only do these programs benefit citizens, but they also catalyze the private sector and create an atmosphere more conducive to business ventures. Furthermore, these programs contribute to more economic utilization of public funds by promoting the adoption of common technological platform for common services, and the reuse of software modules and solutions [16] [17].

In pursuing the abovementioned goals, the ICT Policy Implementation Technical Unit (UTICT), through the Electronic Government Network Project (GovNet), developed the Mozambican Government Portal (www.portaldogoverno.gov.mz). This portal is categorized as a Foundation Project as it is one of the core components of the e-Government Strategy, and therefore the PSR, and facilitates e-public services delivery [2] and the interface between Government and the citizen.

1.2 Portal Content Overview

The Government Portal was designed and developed with four main stakeholders in mind: the citizen, the private sector, civil society, and public servants [8]. Of all the stakeholders, however, the citizen is the primary target of the Portal as its main objective is to “provide a unified point of access and infrastructure to support the Government of Mozambique strategic direction of bringing government public information and service to the citizen [8].”

Factoring in the needs of citizens and other stakeholders before implementing the Portal was an important activity in ensuring that the Portal was designed and developed to meet the needs and objectives for which it was envisioned. The needs of citizens and other stakeholders were considered and then services deemed useful to them were designed into the Portal. These services include [9]:

- Intranet, including: document repository, e-mail, instant messaging, and utilities.
- Forums;
- Search engine.

As well as catering the services to the stakeholders, the Portal was also designed to allow stakeholders to perform tasks and access information that they would most likely need and want on a daily basis. Below is an example of these tasks [9]:

- Access the documents of the different services offered by the government;
- Obtain information about the government structure and operations;
- Access Webmail (email through the internet) [17];
- Read government news;
- Read and download the “Boletins da República” (the Mozambican Government Gazette) and other official Government Publications;
- Check government policies and strategies (as they are developed and approved);
- Read press releases from the Council of Ministers;
• Provide an environment to voice their opinions on different issues concerning the government and the development of the country in general. Below is an illustration of the Government Portal’s homepage.


This Portal was officially launched in July 2006 and has been in continuous operation since then. Like many other websites and portals, the Mozambican Government Portal has built-in technological means to register and monitor the visits and accesses to the information and services it avails. Though the current trend illustrates a general growth in the number of visits, there is no clear indication as to who is accessing what in this Portal.

2. Objectives

This paper analyzes the Government Portal’s progress in reaching its original objectives of anticipating and meeting the needs of its four stakeholders, particularly the citizen. This analysis produces results that can be used as feedback and can contribute in enacting control actions (administrative and technological) to ensure that the growth of the Portal is in the direction that guarantees that it reaches its original goals.

At the heart of this paper is the identification of the typical Portal user, of which their most important characteristics are, which stakeholder group they belong to (if any) and how they use the Portal. The answers to these questions could then enable us to:

• Assess whether the Portal is reaching its intended user, and therefore is reaching its goals;
• Formulate ideas for new services to be offered through the Portal, most likely to be of use to the intended users of the Portal;
• Identify the services that are most probably of value to the user and possibly channel more resources to those services;
• Shed insight into what to expect when undertaking similar projects.

3. Methodology

This paper presents deductions obtained through purely empirical methodology—visitor pattern analysis. Visitor pattern analysis is performed using web server log data (a log contains records of every request the web server receives and its subsequent response). These records are stored in ASCII text format in which each line represents a different request or response, and displays the operation requested, the IP addresses of the requester,
the time and date, and other data. Though useful, in this form the information is raw and extremely difficult to analyse, necessitating the use of a specialized program that reads the log files and creates a human-friendly representation of its content. Webalizer is one such program and it is used on the Government Portal to “produce highly detailed, easily configurable usage reports in Hypertext Markup Language (HTML) format, for viewing with a standard web browser” [5] [16].

In this paper the Webalizer reports undergo a time analysis and page view analysis. The time analysis looks at the different times of day, days of week, and days of month that the Government Portal receives the most visitors. This information helps paint an initial picture of the Portal visitor. The page view analysis adds another dimension to the time analysis by illustrating which of the Portal pages are most viewed by the visitors. The combination of these statistics helped us to infer the characteristics of the general Portal user and draw a clearer picture of their stakeholder group.

4. Limitations of Visitor Pattern Analysis

There are, unfortunately, fundamental problems in the use of logs as a means to analyze visitor patterns on any website/portal. These problems are intrinsic to the Hyper Text Transfer Protocol (HTTP) 1.0, the protocol that computers use to communicate on the World Wide Web (WWW), and as such are unavoidable. HTTP has two properties that contribute to complicating visitor analysis: IP addresses and statelessness [6].

When IP addresses were originally created, each one represented a single computer on the Internet. As the popularity of the internet increased, however, this one-to-one relationship between IP addresses and computers was lost and one IP address can now represent many computers using tools such as Dynamic Host Configuration (DHCP), on a Local Area Network (LAN) or Wide Area Network (WAN) [6] [16]. For instance, it is now common practice for companies to use internal networks and firewalls so as to give the impression from the Internet that every computer within that company uses the same IP address. At home computers are frequently assigned different IP addresses from a pool of addresses owned by the Internet Service Provider (ISP). This lack of one-to-one relationship between IP addresses and computers means that it is now impossible to guarantee that requests made by a single IP address are from the same computer, or that requests from different IP addresses are from different computers. Judging the habits of a particular user is exacerbated by the fact that a user can access the Government Portal from different computers, and hence, different IP addresses.

Statelessness is seeing each request discretely and individually instead of tying them in to future and past requests. In stateless protocols, such as HTTP, it is impossible to guarantee when someone first started accessing a website/portal and when they left. For instance, if there is a 20 minute break between when a particular computer accessed the Government Portal one cannot indubitably ascertain whether it took the user that long to read through the first page or if the original user left and a new one arrived and also decided to navigate on the Portal.

While solutions are continually being devised and implemented to address the above issues (HTTP version 1.1, cookies, etc) [7], these advances are not yet widely implemented. Consequently, this document proceeds taking into account the limitations of HTTP 1.0 and the effect they have on the Webalizer reports. The most important fact for us to keep in mind is that none of the recorded terms are an accurate representation of the actual number of visitors to the Portal, though they can provide an approximation that can help in deriving the information needed to answer our research questions. The closest approximation to the number of Portal visitors is a mixture of the “sites” and “hits” statistic, as described below.
5. Technology Description

Webalizer uses pie charts and bar graphs to illustrate the usage patterns of websites. These charts and graphs display hits, files, sites, visits, and pages. Below is a description of these terms in the context of this program, [5]:

- **Hits** - the total number of requests made to the server during the given time period (month, day, hour, etc.);
- **Files** - the total number of hits (requests) that actually resulted in something being sent back to the user. Some hits do not send data, such as 404-Not Found requests and requests for pages that are already in the browsers cache (i.e. the temporary store for internet files on the user’s computer);
- **Sites** - the number of unique IP addresses/hostnames that made requests to the server. Care should be taken when using this metric. Many users can appear to come from a single site, and they can also appear to come from many IP addresses so it should be used simply as a rough gauge as to the number of visitors to the server;
- **Visits** - occur when some remote site makes a request for a page on the server for the first time. As long as the same site keeps making requests within a given timeout period, they will all be considered part of the same visit. If the site makes a request to the server, and the length of time since the last request is greater than the specified timeout period (default is 30 minutes), a new visit is started and counted, and the sequence repeats. Since only pages will trigger a visit, remotes sites that link to graphic and other non-page Uniform Resource Locators (URLs) will not be counted in the visit totals, reducing the number of false visits;
- **Pages** - those URLs that would be considered the actual page being requested, and not all of the individual items that make it up (such as graphics and audio clips). Some people call this metric page views or page impressions, and defaults to any URL that has an extension of .htm, .html or .cgi.

6. Time Analysis of the Portal Visitors

In [12] are presented different criteria for evaluating and selecting web sources that can also be used to evaluate the Government Portal. These include: web source stability, web data quality, and application specific or contextual issues. Initial determination of the Portal visitors is done purely using contextual issues [12] by analyzing the times in which they visit the Portal. This information, taken in the long-term and short-term, can provide a basic understanding of who accesses the Portal.

6.1 Overall Usage

Since the launch of the Mozambican Government Portal on July 24th, 2006, the Portal has seen a notable increase in its visits. While in June the Portal received 646,154 hits on 16,860 visits, by November those numbers doubled to 1,423,541 hits on 24,132 visits, and by February they reached 61,137 hits and 61,137 visits. This is illustrated in Chart 1 below.

From this increase in visits to the Government Portal, we can deduce that two factors in particular have played a significant role: Portal promotion and user satisfaction. Portal promotion is essentially increasing the number of people that know of the existence of the Portal. Since the Portal launch, UTICT has taken every opportunity to put the Portal in the minds and words of the press through press releases and interviews, while also advertising to smaller groups through activities such as presentations to institutions. Informal methods of advertising have most likely also contributed to the growth of the Portal user base. This includes word-of-mouth, and displaying the Portal address on various places such as folders and email signatures.
User satisfaction is another important factor to the growth of the Government Portal visits. Usually this is the factor that retains the users once they enter the Portal for the first time. This is an extremely important factor to consider as it speaks directly to the utility, and therefore, relevance, of the Portal. Though no concrete numbers are shown, it is reasonable for one to assume that the sustained growth shown above would be improbable were it not for the retention of new users.

Despite the overall increase shown above, it is also necessary to account for the decrease in hits in September, October, and December, and the decrease in visits in September. These can be justified by public holidays. In September there were two public holidays (on the 7th and 25th), as were there in October (4th and 19th) and December (25th, 26th, 29th). This piece of information not only sheds light on the slight drop in traffic on those particular months, but also starts to paint a picture of the type of person that accesses the Portal and from where they access it: a worker (in the public or private sector) or student, from the office or school.

6.2 Daily Usage

Within each month, one can also see the daily usage chart—an illustration of the visits received per day. Chart 2, displayed below, shows that even though the Portal is accessed every day, it receives far more visits from Monday to Friday.

The increased visits received by the Portal on Monday to Friday, reinforces the earlier finding that this Portal is mainly used by workers and/or students at the office or school. Recreational users of the Portal, with Internet access at home, would likely visit it more on the weekends. Furthermore, the slight dip in the visits on the 10th of November, a public holiday in Maputo city, adds to the argument of worker and/or student visitors.
6.3 Hourly Usage

The hourly usage chart illustrates the number of hits and pages, on average, at every hour of the day in November. This chart displays a drastic increase in the hits starting from 8am. The increase then continues steeply until 11am, though not as steeply from 7am to 8am. Between noon and 1pm, there is a dip in the usage, but by 2pm there is a peak usage of the Portal. From this time until late at night there is a gradual decline in the usage. This information is presented below:

The high usage of the Portal between 8am and 5pm is indicative of worker visitors, and is reaffirmed by the dip at noon and 1pm, lunchtime. This hourly usage also provides information on what type of worker visits the Portal. Since the high usage starts at 8am, and the overwhelming majority of government offices open at 7:30am while some private firms open at 7am and commence at 9am, it is logical to conclude that there are a large number of workers from the government departments accessing the Portal.

This does not, however, mean that the site is only viewed by government workers. Actually, the chart above speaks to the contrary. For instance, the continued use of the Portal well after the closing work hour of 3:30pm is unlikely to represent government workers. A more likely explanation is that non-government workers, students, and users...
from other countries also use the Portal, albeit not as early in the day as government workers, and maybe not in as many numbers.

7. Page View Analysis of the Portal Visitor

The analysis of specific pages viewed adds another dimension into ascertaining the identity of the typical Portal visitor. Here we attempt to discern the visitor by the type of services that they seek on the Portal, rather than purely by what time they access the Portal. This analysis is also useful in the formulation of new services to offer on the Portal.

7.1 Top Entry Pages

An entry page is the first page requested in a visit [5]. This page, triggered by the visit statistic, is where people start navigating on the Portal. One should bear in mind, however, that the visit statistics are not very accurate, as explained above.

The top entry pages table contains a list of many pages and documents on the Portal. The top entry page is “/” which, unsurprisingly, signifies the initial page. The other pages are most likely bookmarked (i.e. their links are stored by the browser for easy access).

![Table 1: Top Entry Pages of the Government Portal in November 2006](image)

<table>
<thead>
<tr>
<th>#</th>
<th>Hits</th>
<th>Visits</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15822</td>
<td>1.11%</td>
<td>/</td>
</tr>
<tr>
<td>2</td>
<td>575</td>
<td>0.46%</td>
<td>/info/informatica/</td>
</tr>
<tr>
<td>3</td>
<td>2956</td>
<td>0.69%</td>
<td>/img/logo.png</td>
</tr>
<tr>
<td>4</td>
<td>1622</td>
<td>0.11%</td>
<td>/news</td>
</tr>
<tr>
<td>5</td>
<td>189</td>
<td>0.03%</td>
<td>/legislacao/index.html</td>
</tr>
<tr>
<td>6</td>
<td>945</td>
<td>0.09%</td>
<td>/gov/compos_conselhos_ministros/compos_ministerios/</td>
</tr>
<tr>
<td>7</td>
<td>1035</td>
<td>0.01%</td>
<td>/Mozambique/</td>
</tr>
<tr>
<td>8</td>
<td>255</td>
<td>0.02%</td>
<td>/Mozambique/realizados</td>
</tr>
<tr>
<td>9</td>
<td>122</td>
<td>0.01%</td>
<td>/legislacao/index.html</td>
</tr>
<tr>
<td>10</td>
<td>346</td>
<td>0.02%</td>
<td>/Servicos/empreg</td>
</tr>
</tbody>
</table>

This table illustrates that the most viewed pages, excluding the initial page, are news (“/news”), contact information about the ministries (“/gov/compos_conselhos_ministros/compos_ministerios”), information about Mozambique (“/Mozambique”), and job information (“/Servicos/empreg”). While the Mozambique information and news page shed little light on the type of visitor, ministry contact information and job information points to a worker and/or student.

This data on entry pages can also be applied to the amelioration of the user experience of the Portal users. Knowing where users first access the Portal means services can be placed on these pages that encourage the user to navigate to other pages and even to login. It also means that specific care should be taken to ensure that these pages load quickly and have little chance of containing errors.

7.2 Top Uniform Resource Locators (URLs)

The top URLs table displays the most visited pages. This differs from the above table in that it includes the pages that a user visits after seeing the entry page. This table is the most useful in the study of where users go in the Portal and why.

When looking at the table below, the first thing that one notices is the dominance of “.css”, “.js”, and some other files as the most requested URLs. The reason for this is that these files are used by the visitors’ browsers to present the Portal web page. While the visitor never sees these pages, every computer that accesses the Portal downloads them. So, for the purpose of this document, these files are irrelevant. Of the relevant pages for our
study, the top URLs are the initial page (“/”), and the webmail page (“/webmail/src/read_body.php”), top 9th and 19th, respectively.

Table 1: Top URLs of the Government Portal in November 2006

<table>
<thead>
<tr>
<th>#</th>
<th>Hits</th>
<th>KBytes</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>42466</td>
<td>2.98%</td>
<td>/phone.css</td>
</tr>
<tr>
<td>2</td>
<td>41650</td>
<td>2.83%</td>
<td>/phoneColumns.css</td>
</tr>
<tr>
<td>3</td>
<td>41465</td>
<td>2.01%</td>
<td>/phoneNS4.css</td>
</tr>
<tr>
<td>4</td>
<td>40371</td>
<td>2.04%</td>
<td>/phonePrint.css</td>
</tr>
<tr>
<td>5</td>
<td>40260</td>
<td>2.03%</td>
<td>/phone_javascript_variables.js</td>
</tr>
<tr>
<td>6</td>
<td>40222</td>
<td>2.03%</td>
<td>/phoneCustom.css</td>
</tr>
<tr>
<td>7</td>
<td>40203</td>
<td>2.02%</td>
<td>/phonePresentation.css</td>
</tr>
<tr>
<td>8</td>
<td>34196</td>
<td>2.40%</td>
<td>/phone/EPKes.css</td>
</tr>
<tr>
<td>9</td>
<td>15822</td>
<td>1.11%</td>
<td>/mailal21841</td>
</tr>
<tr>
<td>10</td>
<td>14316</td>
<td>1.01%</td>
<td>/phoneTextSmall.css</td>
</tr>
<tr>
<td>11</td>
<td>14016</td>
<td>0.99%</td>
<td>/phoneTextLarge.css</td>
</tr>
<tr>
<td>12</td>
<td>12824</td>
<td>0.97%</td>
<td>/phone_javascripts.js</td>
</tr>
<tr>
<td>13</td>
<td>11765</td>
<td>0.83%</td>
<td>/phonefonturi18060</td>
</tr>
<tr>
<td>14</td>
<td>11632</td>
<td>0.82%</td>
<td>/logoteans.png</td>
</tr>
<tr>
<td>15</td>
<td>11615</td>
<td>0.82%</td>
<td>/GovNetImages/logoWebMail.png</td>
</tr>
<tr>
<td>16</td>
<td>10396</td>
<td>0.73%</td>
<td>/webmail/src/sight_main.php</td>
</tr>
<tr>
<td>17</td>
<td>9588</td>
<td>0.67%</td>
<td>/img/clamps/logo_down.png</td>
</tr>
<tr>
<td>18</td>
<td>9436</td>
<td>0.66%</td>
<td>/phonesqueezebox_utla.js</td>
</tr>
<tr>
<td>19</td>
<td>6703</td>
<td>0.43%</td>
<td>/webmail/src/read_body.php</td>
</tr>
<tr>
<td>20</td>
<td>5657</td>
<td>0.43%</td>
<td>/ms:image/coords2.swf</td>
</tr>
</tbody>
</table>

The two most requested, and relevant, URLs communicate a lot about the Portal user. The most important information is that the webmail URL is one of the most requested. Webmail is a service provided through the Government Portal and is part of the Centralized e-Mail System provided only for public servants. That this is one of the most requested URLs is a fact in strong support of our argument and initial empirical finding that the typical user of the Government Portal is a public servant.

8. Conclusion and Recommendations

The Government Portal was conceived with the ambitious goal of reaching multiple stakeholders and sectors, namely the public sector, private sector, citizens, and civil society. Though in development phases the Portal strived to cater to the needs of the average citizen, we think the above data indicates that, in fact, the employed citizen makes most use of the Portal. Furthermore, for us, the above data suggests that public servants make up the vast majority of the Portal visitors. The underlying reason for this, it seems, is due to the utility of the webmail service.

In our opinion, with this conclusion in mind, we were left with two new questions. First, how should we use this new information to redistribute resources into the webmail service? This includes electronic resources like increasing mailbox sizes, improving the user interface, and so on. This also includes, however, non-electronic resources like continued training of government institutions to fully capitalize on the utility of webmail.

Another important question is how should the Portal evolve to attract its initially intended audience, the average citizen? One obvious solution is offering webmail services
to the citizen. Another solution is to devise new services available over the Portal targeting
the average citizen, such as electronic forms that can be filled in online and then submitted
immediately to the chosen institution. A further solution could be to promote initiatives that
increase the average citizen’s awareness of the Portal and access to the internet.

Regardless of which path the Portal developers take, it is very clear that analysing the
Portal visitors is a valuable exercise. At the very least we believe this paper has opened
debate on whether or not the Mozambican Government Portal is reaching its intended goals.
The next step should be to confirm or deny the findings of this paper through detailed and
in depth studies such as research, surveys, etc. Armed with these studies we can more
confidently ascertain the Portal user stakeholder group, which will help us to make strides
towards a successful Government Portal.

References

Citizen and Business”, white paper
July 2006
December 2006
Mozambique, 2006
Data Warehouse”, Darmstadt, Germany, 2002
Proceedings, Paul Cunningham and Miriam Cunningham (Eds), IIMC International Information