A real time monitoring system to measure the quality of the Italian Public Administration websites

F. Donini¹, M. Martinelli¹ and L. Vasarelli¹
¹Institute of Informatics and Telematics, National Research Council (CNR), Pisa, Italy

Abstract -In 2013, the Institute of Informatics and Telematics of the National Research Council (IIT-CNR) and the Department for Digitalization and Technological Innovation of the Presidency of Council of Ministers, carried out a pilot project, whose main aim was to create a monitoring system in order to verify and analyze the capacity of Italian Public Administrations to activate and manage information and services by means of the Internet. This paper describes the SmartPA (System for the Real Time Monitoring and Analysis of the Italian Public Administrations) application, developed by IIT-CNR, which analyses and monitors Italian PAs, evaluating in real time conformity and coherence of contents with Guidelines regarding PA websites, issued in the years 2010 and 2013. SmartPA, through the publication and presentation of results on a public portal, aims to establish an open and transparent communication between the citizen and Public Administration, thus bringing about a process of continuous quality improvement of PA websites.

Keywords: e-Government, websites quality, case studies, websites monitoring tools, e-Transparency

1 Background

Within the 2012 e-Government plan and in the context of modernization of Public Administration, the 2009/11/26 Directive of the Ministry for the Public Administration and Innovation [1] [2], in order to rationalize Public Administration websites (PA) and to improve the quality of services and on-line information for the citizen, sets out specific Guidelines. These are aimed at providing clear indications regarding general criteria and operating tools for the rationalization of online content, for the reduction of obsolete public websites and improvement of active sites [3].

These Guidelines, which are an essential feature of the initiatives aimed at innovating the PAs, have the objective of establishing an open and transparent communication with the citizen, so bringing about a process of continuous quality improvement of public websites.

In particular, the Guidelines pay special attention to the definition of a map of minimum essential requisites. These features, on the basis of current legislation (e.g. Digital Administration Code, CiVIT Guidelines, Garante della Privacy, etc.), must be included within institutional PA websites.

Within this context, this project foresees the creation of a pilot monitoring system (SmartPA - System for the Real Time Monitoring and Analysis of the Italian Public Administration websites). This system, starting from .it domain names registered by Italian PAs¹, enables verification and analysis of the capacity to activate/manage information and services by means of the Internet and web channels. This evaluation has been carried out through the study and definition of a concise indicator of PA websites, measured in terms of coherence or compliance with regards to a minimum set of compulsory prerequisite contents, indicated in the Guidelines. This includes, merely as an example, the sections: organization chart; external relations office (URP); transparency; evaluation and merit of personnel; administrative procedures; public tenders; job competitions; on-line services; legal notices; certificate electronic mail (PEC).

SmartPA, exploiting the potential offered by ICT technologies, therefore provides real time measuring and evaluation tools for the quality of public services. For some time this has been the focus of PA modernization, with a specific emphasis on closely involving citizens.

2 Architecture

The SmartPA architecture is of a modular type (Figure 1). The system is subdivided into two macro-components, the engine and the portal. The engine is the elaborating part of the system, whereas the portal is the interface that enables the management of the engine and consultation of the results elaborated.

The engine module, which uses the Akka framework [4] for the creation of concurrent elaborations [5] [6], implements all the SmartPA logic, collecting, elaborating and retrieving data. The website analysis has been made by means of the parsing library Beautiful Soup [7].

The portal module, created by means of the Play Framework [8] web application, thanks to the use of the most advanced

¹.it domain names are managed by the .it Registry, that is a service of IIT-CNR
web technologies (HTML5, CSS3, JavaScript, etc.), is compatible with the main Internet browsers and mobile devices.

**Figure 1 - SmartPA architecture**

All the results (metadata, partial and total elaborations regarding analyzed subjects, references to downloaded web pages, etc.) are stored in a dedicated database. The html pages, which make up the monitored website, are stored in a file system, so as to enable future research and comparison. There now follows a detailed analysis of the two macro components that make up SmartPA: the *Engine* and the *Portal*.

### 3 The Engine

The SmartPA engine is subdivided into three main stages (Figure 2):

- **Data Collection** – this identifies the data set of relevant domain names (domain list) on the basis of which the analysis and monitoring of domain names is initiated;

- **Data Processing** – this uses the data of the Data Collection stage and is the key stage of the system. In fact, during this stage all the main activities and elaborations that make up the website analysis and monitoring process are carried out;

- **Data Discovery** – this is based on the Data Processing stage and identifies and associates a Public Administration with its specific institutional website.

**Figure 2 - SmartPA engine schema**

The activity of Data Collection is mainly carried out during the initial stage of database population and in order to add new subjects to the monitoring process. On the other hand, Data Processing and Data Discovery activities can be initiated upon request of the user (real time) or periodically by the system (pre-established time intervals).

#### 3.1 The Data Collection stage

Domain names, by their very nature, can be assigned to different types of registrants (public bodies, natural persons, companies, etc.) identifiable, in the .it Registry, by means of a specific univocal code. Using this code it was possible to retrieve the domain names assigned to Italian Public Administrations [9], to identify the institution they belong to, and for each registrant, identify the category (local institutions, research institutes, school, companies, local health authorities, etc.).

This activity also involved a preliminary process of data cleanup that gets Registry database data prepared to work correctly: in fact, various corrections were made regarding errors that had occurred during the registration of the domain names.

#### 3.2 The Data Processing stage

On the basis of the data set of domain names obtained in the Data Collection stage, websites of each public administration were analyzed and conformity to the Guidelines were assessed. This phase involved a set of elaborations carried out by three ad hoc modules: the *tracker*, the *crawler* and the *parser* modules.

The *tracker* module identifies the URL of the homepage of the website associated with the domain and is able to filter replications of websites hosted by different domains, but belonging to the same PA. Particular attention was paid to the study of false positives such as white pages, splash screens, sites under construction, etc.

The *crawler* module uses as input the URL of the homepage obtained by the tracker and downloads the associated website. Given that the minimum contents foreseen by the Guidelines can be found also in second level subsections, for each website a download of at least three levels was carried out.
The parser module has the task of verifying the presence and compliance of the contents foreseen by the Guidelines. For each content item the parser gives three different results (compliant, partially compliant, not compliant), in this way identifying a final score for each website.

3.3 The Data Discovery stage

Each PA can have multiple websites, but only one of these should correspond to the so-called Institutional website, the one that identifies the PA institutionally and highlights the structure, organization, staff, salaries, services offered, etc.

The purpose of the system at this stage is to identify the most likely institutional website of each PA. This relies on a scoring system (ranking) that takes into account, as weighted average, the results obtained by the parser (number of compliance to the Guidelines), the domain name pattern and the website URL pattern (Table 1).

Table 1 - Identification of the Institutional website according to the scoring system

<table>
<thead>
<tr>
<th>Rank</th>
<th>URL</th>
<th>Comp.</th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><a href="http://www.comune.pisa.it">http://www.comune.pisa.it</a></td>
<td>66</td>
<td>Institutional website (identified)</td>
</tr>
<tr>
<td>2</td>
<td><a href="http://comune.pisa.it">http://comune.pisa.it</a></td>
<td>66</td>
<td>Missing “www”</td>
</tr>
<tr>
<td>3</td>
<td><a href="http://www.comune.pi.it">http://www.comune.pi.it</a></td>
<td>66</td>
<td>Not the best pattern “pi” less clear than “pisa”</td>
</tr>
<tr>
<td>4</td>
<td><a href="http://cittapisa.it">http://cittapisa.it</a></td>
<td>3</td>
<td>Clearly not institutional for the numbers of compliances</td>
</tr>
</tbody>
</table>

4 Portal

The Portal consists of two modules and two separate sections, the public section dedicated to the presentation of data and statistical results (Front-end) and the system administration section (Back-end). The portal was designed to be intuitive, easily usable for the general public, accessible and equipped with a clear response layout for mobile devices (Figure 3).

Access to the portal is therefore in two circumstances: in order to take advantage of the contents and results made available (as an anonymous visitor without credentials or as a user registered in the portal), or for its management (as administrator, operator and maintainer).

4.1 Front-end

The front-end module implements the portal section dedicated to the display of results. Among its main features this section includes the search (exact and partial) of a PA, and visualization by means of reports, charts and graphs of its descriptive information and compliance of the institutional website with the Guidelines (Figure 4).

This section also enables the user to search and, consequently display statistical aggregate data according to selectable parameters, such as the territory (on a national, regional and provincial basis), the type of PA or a specific Guideline (Figure 5).

4.2 Back-end

The back-end module implements the section of the portal dedicated to system administration and interaction with the procedures of data collection. This section (Figure 6) offers an extensive set of features that enable the user to manage and monitor the main features of the system (information processing, domain names, organizations, users, system logs, general settings, etc.).
Through the back-end it is possible, for example, to launch Data Processing (3.2) and Data Discovery (3.3) processes and visualize the elaborations in real time, thanks to dedicated visual media. In the case of multiple instances of same time elaborations, they are managed by means of an ad hoc queue.

Other important features that are implemented in the back-end module are: visualization of the archive of the operations performed, consultation of the systems log, modifications of the profiles of the subjects undergoing verification (personal details, domain names associated with them, Institutional website URL, etc.) (Figure 7) and management of the system users.

5 Main results

The main statistical results obtained after the entry into operation of the analysis and monitoring system of the websites of the Italian PAs, are indicated below.

In January 2014 around 2,635,000 .it domain names were active, 34,115 (1.29%) of which were registered by Italian Public Administrations. 45% of these domain names (15,541) have been associated with subjects belonging to categories of interest of the present study (Councils, Provinces, Regions and Ministries). Within these subjects, over 8,232 PAs, 314 do not have a .it domain name. Some sample surveys showed that the main reasons for this situation are, for example, the availability of a domain name other than a .it (.es , .eu, .net, .org), or the absence of a specific web site for the institution, etc.

Figure 8, shows the ratio between registered domain names and PublicAdministationon a regional basis. Ideally, ratio should tend as much as possible to 1 as for the regions of Molise, Val d’Aosta, and Abruzzo. On the contrary, for what concerns regions as Tuscany, Emilia Romagna, Lazio, Marche the ratio turns out to be considerably higher.

Taking into account the analysis of the compliance of PA websites with the Guidelines, results show that on average, the degree of compliance is 47.7%. Therefore, around 33 Guidelines out of 69 are totally or partially respected in the
websites of the PAs analyzed.

The highest average compliance at the regional level is achieved by the PAs of Sardegna (70.7%) region. On the other hand, the lowest average compliance can be found in Trentino Alto-Adige, with only 6.4% of compliant or partially compliant websites (Figure 9, Figure 11).

At a national level, results show (Figure 10) that most of the bodies have almost fully compliant websites or fully not compliant websites:

- 154 PAs have a website 100% compliant with the Guidelines (including 90 PAs that have a website partially compliant with the Guidelines);
- 1248 PAs have the institutional website that does not comply with any Guideline

Taking into account the situation at a Macro Area level (North, Centre and South), we can see that (Figure 11):

- PAs of the North have a degree of compliance with the Guidelines of 50.56% (35 guidelines out of 69 are respected);
- PAs of the Centre have a degree of compliance of 44.35% (31 out of 69 Guidelines are respected);
- PAs from the South and the Islands have a degree of compliance with the Guidelines of 48.19% (33 out of 69 Guidelines are respected).
6 Conclusions

The aim of SmartPA is to suggest criteria and tools for the rationalization of online contents, the reduction of obsolete public web sites and improvement of active sites.

Results show that in Italy there are still many Public Administrations that do not fulfill the requirements stated in the Legislative Decree dated 14 March 2013, n. 33.

This study also highlights the difficulties that a citizen may encounter while accessing a PA website, considering multiple domain names associated to it. In order to solve this problem, it is suggested the use of an electronic stamp that indicates the institutionalism of the reference web site of the body.

At a geographical level, there are no major differences for what concerns the quality of results. Only for Trentino Alto-Adige results differ from the national average; a possible explanation could be the presence of contents written in another language (German), considering the presence of bilingualism.

7 Future works

In order to contribute to the improvement of PA websites, SmartPA is constantly evolving.

A set of tools and advanced features are foreseen in order to increase awareness and, at the same time facilitate PAs during the phase of adjustment to this Decree. These features include an extension of the analysis and monitoring to all Italian PAs, an user-reserved dashboard that enable the PAs to manage their own profile and independently analyze the “quality” of their institutional website.

There will be also a tool for validating the accessibility of Institutional PA websites, able to verify adherence and compliance with the Guidelines for web content accessibility foreseen by the Stanca Law [10], the historical record of PA websites, able to show the trend over time of the quality of websites and services and the possibility to export of information and results in Open Data format [11].

Farther is foreseen a creation of a quality stamp to be assigned to the PAs with high levels of adherence and compliance with the Guidelines.

SmartPA is easily adaptable to other international e-Government contexts; due to its modularity, it allows the creation of custom rules to evaluate the quality of Public Administration websites.

8 References


[9] Codes of the councils, provinces and regions according to ISTAT: http://www.istat.it/it/archivio/6789

