Voglioilruolo: an Italian Success Case about Public Data Reuse

Fabio Dellutri[1], Ilaria Dalmasso[2], Vittorio Ottaviani[3], Gianna Palmarelli[4]

Mitecube S.r.l
Viale G. Mazzini, 69 Frosinone - Italy

[1]fabio.dellutri@mitecube.com
[2]ilaria.dalmasso@mitecube.com
[3]vittorio.ottaviani@mitecube.com
[4]gianna.palmarelli@mitecube.com

Abstract— **www.voglioilruolo.it (i.e. "I want the role" – "role" is the name of permanent contract for Italian public school teachers) is a Web platform that exposes services designed for the Italian temporary teachers and school workers, to facilitate them in understanding data produced by Ministry of Education and to help them to make career choices. [1]**
The platform scrapes and analyzes public data from the Ministry of Education websites, reusing it adding value by combining information. The resulting data are searchable, related each other and, generally speaking, more reliable and comfortable to the user, moreover the community of users provides additional information in a "social" way.
In this paper we will discuss about the Voglioilruolo experience and the sustainability model. The results we obtained in Italy not only show that the chosen sustainability model works, but also that is it possible to do business with the public data reuse, a scenario that will offer more opportunities with the spread of open data.

Keywords— **Open Data, Public Data Reuse, School, Temporary School Workers**

INTRODUCTION

www.voglioilruolo.it is the first public data reuse activity that targets to the temporary teachers of the Italian public school.

Voglioilruolo was born to offer an innovative service and a set of tools, designed to assist teachers and ATA[1]s in addressing career choices. The Italian Ministry of Education (*MIUR*) does not offer any tool aimed to optimize those choices and to increase the chances of being hired with a permanent contract[2]. Moreover, the PA does not facilitate access to their data collection by an Open Data approach. Voglioilruolo make access to data about temporary teacher ranking lists easier. From these lists MIUR draws on for recruiting and hiring; furthermore our service continuously collects and analyses data by scraping them from MIUR's provincial websites (more than an hundred), and then aggregates them. Voglioilruolo hosts several value-added services which allows to browse the ranking lists of teachers and of ATAs, to search by name a person and to find, by specifying a school year, for which rankings he results enrolled in: a ranking list exists, for each province, and for each school matter, so a person could be enrolled in more than one ranking list, e.g. if she teaches more than a matter or if she is willing to work in more than one province. By aggregating and correlating collected data, and by piecing them together to build teacher's profiles, Voglioilruolo is able to calculate individual chances to get hired in a certain province and, generally, it offers hints and suggestions about career opportunities for each worker profile, province and matter.

Finally, Voglioilruolo provides statistics about collected data (in order to give more visibility to the temporary teachers phenomenon), includes a question and answering community, and contains a map of Italian private and public schools.

STATE OF ART

Italian public school teachers are divided in three main classes:

- Permanent teachers;
- Certified temporary teachers;
- Not certified temporary teachers;

---

[1] Technical administrative assistant: janitors, administrative staff, assistants, etc.

[2] In Italy permanent contract is historically assumed as a really important asset, due to a non-flexible labour market.

Permanent teachers are those certified teachers, with a permanent contract, such kind of teachers are certified and have matured an experience that grants their professionalism.

To become certified teacher of a specific matter, the teacher has followed a post-graduation training course lasting at least two years (e.g. SSIS, "Scuola di Specializzazione all'Insegnamento Secondario") or she has held and won a national contest to obtain this qualification after graduation.

Not qualified are those teachers who work temporarily at school even without qualification, e.g. by replacing absent colleagues.

Certified teachers are the only one who can be hired for an indefinite period with a permanent contract (the "role") and they are recruited from ranking lists called "graduatorie ad esaurimento".

The same described recruiting process goes for the ATA school staff.

The extensive and intelligent consultation of ranking lists is the only mean by which teachers can evaluate their career opportunities. The MIUR's internal informative system (SISSI) manages these data, allowing the teacher to read only its own position and ranking's score thru a web interface ("istanze online"), but preventing to browse data among different

teacher profiles.

A. The Public Data Reuse in Italy

In Italy there is still not a well-defined policy about the public data reuse. Public data are published without rights and without rules that define who is able to reuse them. Moreover, in case of personal data (such as those processed by Voglioilruolo) the Italian *Data Protection Authority* ("Garante per la Protezione dei Dati Personali") [2] imposes restrictions on their treatment and processing. In particular, for non-sensitive data, there is a duty to inform who is involved in the treatment. By acquiring data on more than 300.000 individuals there was an objective impossibility to fulfil the duty, so a request to simplify the procedure was requested to the Authority. The Organization recognized the innovation and gave credit to the social utility of the service, by issuing a bill to simplify the duty providing several measures. [3]

Recently in Italy there is an awareness campaign about the reuse of public data [4]. In early 2012 the MIUR gave a positive sign to release school-related open data by starting up a project called "Scuola in chiaro" (i.e. School in clear) [5], that aims to give more transparency over Italian schools, more information to the families who plan to enrol their children at school, and a reduction of work for school's administration. These objectives will be achieved by providing updated (open) data. So far data published still remain related to school's statistics (e.g. for each school they provide numbers about the staff, the percentage of passed students, numbers about the equipment, etc.).

B. Reference Market

Voglioilruolo addresses Italian school temporary workers, who are certified by MIUR. The population consists of about 300.000 teachers and ATAs. The choice to address this market was due to the lack of institutional tools able to shed light on the actual job opportunities.

C. Competitors

In Italy there are several union schools that manually collect data related to teachers hiring, and provide them to members (cardholders) thru their offices located throughout the country. However the union school's role is rather to provide advice and legal protection to the worker.
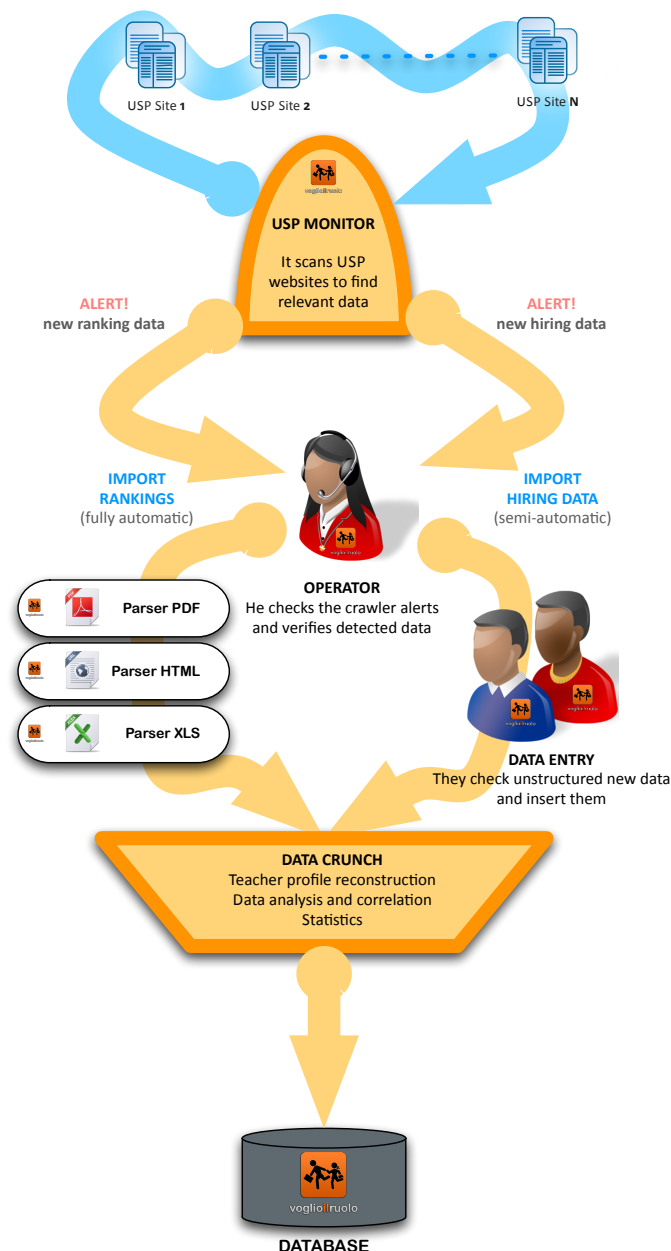
SYSTEM OVERVIEW

A. Data capture and import platform

Voglioilruolo needs to collect and to import several kind of data. Almost always these data are not stored in any Open Data consumable format (e.g. CSV, XML, SPARQL, Linked Data or any other machine-readable format), but they are contained in documents that require ad-hoc import procedures, which must be able to detect relevant data and to cast them to a normalized form. In some cases (such as for rankings and

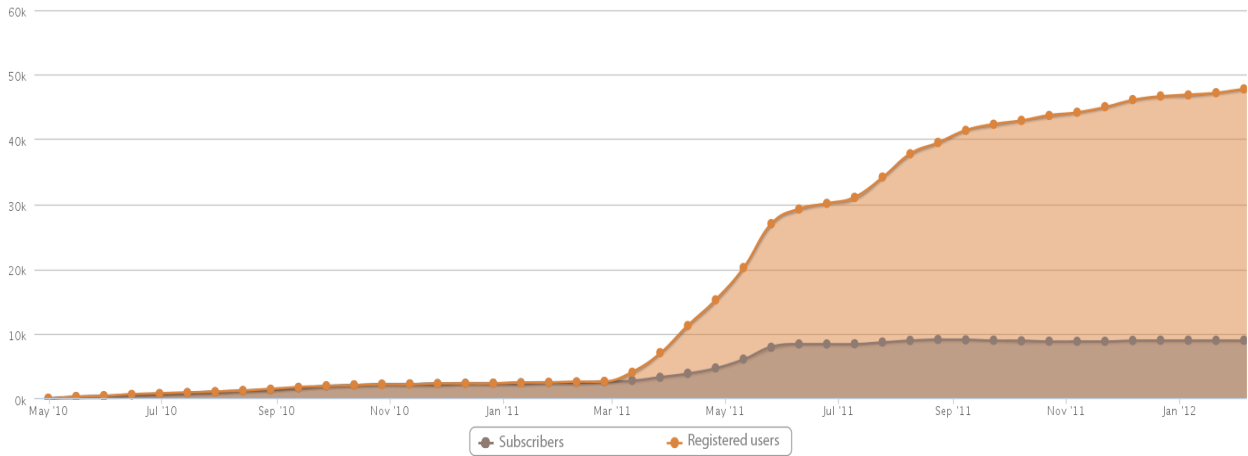Fig. 1 Voglioilruolo data capture and import platform.

Fig. 2 User growth: subscribers and registered users trend.

for retirements) data are stored in a structured (tabular) format: in this case we developed a set of flexible and adaptive parsers which run the entire import process; in other cases (such as for data about hiring) data are implicitly written inside unstructured bulletins that do not leave any chance to be imported with an automatic strategy.

In Table 1 there is a summary of which data Voglioilruolo imports and how.

TABLE I
VOGLIOILRUOLO PUBLIC DATA AND IMPORT COMPLEXITY

Data type	Data format	Import complexity
• Ranking lists of certified teachers • Retirements	Structured (table format) PDF, HTML, XLS	Fully automatic, flexible parser.
• Data about permanent hiring or temporary hiring	Unstructured, they are contained inside bulletin written without any format.	Manual reading and data entry.
• Data about schools	HTML tables before, currently in CSV (thanks to the PA project "Scuola in chiaro")	CSV reading

Periodically Voglioilruolo's web crawler indexes the MIUR provincial websites (called "USP", ~100 websites); at every cycle, it evaluates content's differences towards the previous content indexing, and if those differences match a set of rules, the crawler tags the content as relevant and sends an alert to an operator. Depending on the content found, the operator decides to run an automatic import procedure, or make request to data entry staff to import data manually.

Collected, normalized and stored data are further analysed by the platform in order to find which data matches with which teacher or ATA profile, and to associate them. In this way all data are linked with a profile. Statistical analysis are started after each data import, in order to keep up to date the aggregate data and the graphs.

B. Value added services

Voglioilruolo's users accesses through a web platform that exposes all services for which the user is enabled. The main services are the following:

Ranking lists browsing: facilitates the ranking lists consultation, and provides a linked browsing among the other teacher's or ATA's rankings.

Search by name: allows the user to find all ranking lists where a teacher or ATA is enrolled, and allows her to browse ranking.

Best provinces: currently is the feather in the cap of the platform. Given the matter taught (or the specialization) and the score ranking, the algorithm is able to do a list with the top provinces that would offer the best positioning in their ranking.

Schoolmap: a map-driven, search engine over the Italian schools (public and private). Thanks to open data released recently by the MIUR, data presented in the results are enriched with the number of teachers, the facilities, the offered services, etc.

Statistics: obtained from aggregating collected data. Numbers earned give more visibility on phenomenon of the temporary teachers in Italy.

Q&A: is a question-answering platform. Even if it does not make use of the public data collected and processed, the Q&A offers an invaluable value added to the platform, because it provides a consultative platform driven by the answers of colleagues and other school professionals.

All the services are oriented to the direct participation of users that help each other to find solution to common problems and share information about ranking lists.

In this paragraph we will analyse the first two years of project activity, which is divided in three phases.

1) Phase-1 "Startup/Develop": In the first phase the system architecture was designed and most technical decisions took place. All data collection procedures and data processing features were defined. In this phase the legal aspect of the personal data processing was defined and authorization by the
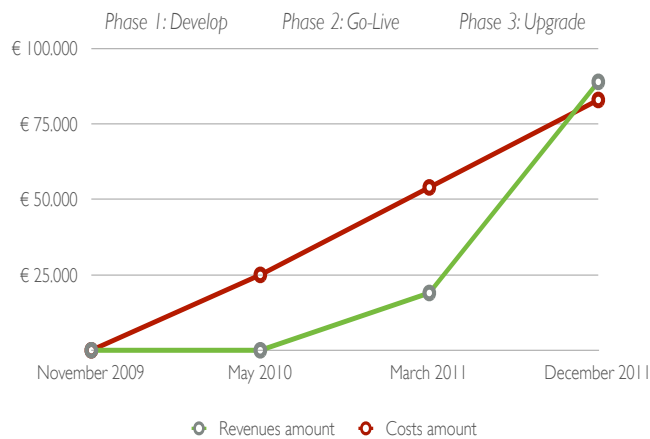

Fig. 3 Business growth.

Data Protection Authority was requested and obtained.

2) Phase-2 "Go Live": The system launch took place at the first may 2010 (on the occasion of the Labor Day). In this phase a subscriber-only business model was chosen: in this model, only paying customers (subscribers) were allowed to access to system features ("Schoolmap" and "Statistics" were delivered as free services instead) and come into contact with the data. From the marketing perspective, the target audience was already well-defined, so we focused on acquiring teachers and school workers by advertising in specialized trade magazines, blogs and online forums. Also article marketing and social media word of mouth were applied.

3) Phase-3 "Upgrade": In the third phase a new business model, a new version of frontend website and new tools were developed (the Q&A community and a new access model to the ranking lists). Such improvements have motivated many users to register them and to use the platform on a daily basis. Therefore the new platform was migrated to a cloud-driven server architecture, to easily scale up (or down) resources following the traffic.

In the phase 2, we noticed that the market had a tepid reaction to the innovation, because people have seen Voglioilruolo as an unclear resale of personal data that, moreover, were already owned by them. An inappropriate business model for such type of service generated this vision. The new business model chosen was the *Freemium model* ("free" and "premium"), in which users can register for free, can see data, can use a set of services, and can pay only for value-added premium services, such as best provinces and advanced teacher profile browsing.

The choices to adopt a new business model (march 2011, clashing with an upgrade of rankings by MIUR), and to introduce the Q&A community have changed drastically the user perception of Voglioilruolo. In Figure 2 is depicted the growth of users.

Such strategy had a positive impact on revenues, which led to a return on the investments within two years, and which shown that the sustainability model chosen is able to maintain the project and to make margins on costs.

CONCLUSIONS

Before Voglioilruolo there was not other platform aimed at easing the access to ranking list's data and to support temporary teachers in their career choices. Voglioilruolo stepped in where there was no clear information, by providing tools to help teachers to calculate their best choice and to take cognizance of their real chances.

In middle 2011 Italian teachers were facing a difficult choice, whether to change province in which they would have worked or not. Using Voglioilruolo a huge quantity of teachers decided to go to the province where they had more chance. Before Voglioilruolo, teachers based their decisions on incomplete and fragmentary information.

Voglioilruolo also demonstrated that doing business by processing data produced by the Public Administration and by developing value-added services over them is possible.

REFERENCES

[1]   Voglioilruolo website, http://www.voglioilruolo.it
[2]   The Italian Data Protection Authority, http://www.garanteprivacy.it/garante/doc.jsp?ID=1669109.
[3]   Garante della Privacy, "Esonero dall'informativa per un sito web che raccoglie e diffonde dati già disponibili online" *http://www.garanteprivacy.it/garante/doc.jsp?ID=1721169*, Prot. 0009374/U, 26 march 2010.
[4]   F. Morando et al. *Libro bianco per il riuso dell'informazione del settore pubblico*, EVPSI, http://www.evpsi.org/librobeta
[5]   "Scuola in chiaro", http://www.dati.gov.it/content/miur-la-'scuola-chiaro'-formato-aperto
[6]   D. J. Teece, *Business models, business strategy and innovation*, Long Range Planning, 2010 - Elsevier