# An e-government service as PaaS application to serve Switzerland's municipalities

Daniel Müller, Funk Te, Irena Pletikosa Cvijikj
Department of Management, Technology and Economics
ETH Zurich
Zurich, Switzerland
{danielmueller, fte, ipletikosa}@ethz.ch

***Abstract*—Municipalities seek a Service Oriented Architecture (SOA) that allows the integration and interoperability of e-governmental services with their own IT systems. Municipalities benefit from computational insights such as the quantification of the municipalities' business landscape. This enables them to calculate the number of businesses they need to attract in a sector and in which ones they are positioned at or above average. We refer to this variable as *market saturation*. To the best of our knowledge nobody has yet tried to evaluate or build a platform independent service which allows municipalities to answer this question. We provide a solution to this problem and suggest a Platform-as-a-Service (PaaS) architecture, which can assess the number of suitable businesses for a municipality. We calculate these numbers by applying a LASSO regression and are able to predict the amount of businesses with a high degree of precision. Based on our previous findings about the non-linear distribution of businesses in municipalities, we believe we can thus model the number of businesses in Switzerland's municipalities and found a way to distribute this knowledge through a hybrid IT architecture.**

***Keywords – Software Architecture, Portal Interoperability, Service Oriented Architecture (SOA), municipality service, e-government, PaaS, market saturation prediction***

## I. INTRODUCTION

A versatile range of local business offerings in rural municipalities greatly enhances the attractiveness of local communities [1]. The businesses which contribute strongly to the attractiveness of municipalities are generally small and midsized enterprises (SME), representing >99% of all business and about 68% of all jobs in Switzerland [1]. Many of these SMEs and the municipalities they operate in are structurally similar however, to the best of our knowledge, currently there is no platform in Switzerland that provides stakeholders with feedback about the composition of municipalities' business landscape.

With the wide range of interested stakeholders such as governmental and business associations, political parties and financial service providers that primarily serve local demand, the systematic distribution of this knowledge can greatly reduce the failure risk of the SMEs [2]. This insight also helps municipalities to maintain and develop a versatile range of business offerings, which subsequently benefit from reflux of taxes.

Due to the fragmented information systems landscape of municipalities and related stakeholders, the integration of such a business application is a challenging task. Bolstered by the rising demand for a Service Oriented Architecture (SOA), a hybrid configuration of IT systems including *Platform as a Service* (PaaS) components, offers potential to achieve this task [3]. Reflecting the increasing popularity of PaaS systems among industry companies, the suggested approach is promising since all associated stakeholders can develop, run, and manage their Web applications on top of a PaaS system [3], without the need to understand and maintain the machine learning components the systems is build upon. Therefore, a PaaS system is suitable to distribute knowledge about the composition of businesses sectors.

Further, the architectural integration of an e-government service within their own systems enables the stakeholders to comply with the separation of data legislatives [4]. No data of the stakeholders using the service are leaving e.g. the governmental premises, which is required by local law for some applications. This way, developers working for governmental bodies can also benefit from the service without the need to implement and train the base computation logic in their systems.

Our paper makes the following contributions: 1) we elaborate on the challenge of integration and interoperability of governmental services; 2) we propose a hybrid IT architecture using a PaaS systems to accomplish this task; 3) and we provide a preliminary evaluation through a concreate example of an e-government service for municipalities such as a business saturation prediction by calculating a suitable number of businesses within Switzerland's municipalities.

The rest of this paper is organized as follows. Section II reflects on the development and the needs of e-government services, highlights the characteristics of PaaS systems and reflects on the computational advances which allow to model business market saturation. Section III elaborates on a motivating example for our research. The IV. section explains how this service benefits the general public and governmental agencies and section V states the performance of our early model. Section VI states the models performance and the last sections conclude, address the limitation of our preliminary results and give an outlook of the continuation of our research.

## II. Related Work

This research in progress is based upon previous findings in three research areas. It builds upon recent findings in e-governmental initiatives. We further elaborate the understanding of PaaS as part of a hybrid IT architecture. Finally, we reflect on data science methodologies which were previously used in services computing [5] to predict a future customer state of a local market.

### A. E-government initiatives architecture requirements

Within the last few years, new information and communication technologies (ICTs) innovations have been introduced to the public sector at a high speed [1]. E-government services were introduced by many administrative agencies. Further, newly created open data initiatives allowed the internal sharing of information among other government agencies and externally for the general public [4].

These services and initiatives play a key role in the modernization and reform of administrative structures, as governments face the continuing pressure of increasing their performance and evolving to adapt to the new information society [5], while ever increasing their attractiveness for new businesses. However, most e-governmental services such as portals lack integration and interoperability capacity. This fragmentation problem, which causes each government agency to develop and operate its own services [3].

The authors of [7] have identified the need to overcome this constraint and suggest federal, state and municipalities provide their services to consumers in various functionalities in a language independent, platform independent, and protocol independent manner. A suitable architecture needs to enable a seamless integration and interoperability of e-government agencies, e.g. by using a hybrid distribution e-government architecture [3] allowing to solve this collective problem [6].

### B. PaaS systems

*Platform-as-a-Service (PaaS)* is as commonly referred as a category of cloud computing services, which allows consumers to build particular classes of applications and services without the complexity of building and maintaining the infrastructure typically associated with developing and launching a high-level application [3]. PaaS systems offer the feasibility and extensibility of business service demands for versatile industry and governmental stakeholders. They promise great flexibility in offering advanced ICT functionality, while maintaining interoperability between heterogeneous IT architectures of different service consumers and their higher level IT service. This fosters higher-level programming with dramatically reduced complexity and makes the development of applications more effective [3].

### C. Computational business insight

Data science methodologies have proven their value by accurately predicting a local market's state in the future [6]. However as [7] identified, current understanding of market capacity is still in its infancy, wherefore more research needs to be performed in this domain. Advances in the scope of data sources to model market saturation have been brought by [6-9].

## III. A motivating example

Newly created independent SMEs (non-chain) serving local markets face market risk just like any other startup, however they do not have the capacity to analyze demand and competition in structured ways, as in comparison chain stores can. In contrast to the companies which are generally referred to as (technology-) startups, these SMEs rely on proven business models such as general practitioning, plumbing or accounting services. To estimate the local demand for their services, these SMEs have to rely on their subjective understanding of the local market structure. Faced with this challenge, they can seek support from business associations or financial service providers. Also political parties can offer help and advice to enable business activities. However, to the best of our knowledge, none of these stakeholders have access to quantitative insights about the structure of their respective markets. We were approached by a cooperatively owned financial service provider in Switzerland to establish knowledge about the architectural requirements such a service would rely on, how to integrate the service with the municipality IT infrastructure and how such a service serves the needs of local business communities.

## IV. Market saturation estimation in a PaaS system

In this paper, we propose a novel approach to predict market saturation as a PaaS component to service the needs of municipalities and associated stakeholders. We estimate market saturation through domain-related feature extraction from publically available data sources [4]. We rank these features based on their effectiveness and determine the value of these features through recorded customer actions. We do this by using a Lasso regression to predict the amount of business within the sectors of all municipalities in Switzerland. We train our model with historic business listings which are publically available. Comparing our test data set with the training data set, we are able to predict the amount of business within a domain with high reliability.

We imagine this application to be a component to integrate in a PaaS system allowing municipalities, business associations, political parties, financial service providers and businesses that primarily serve a local demand to integrate and build upon this service in respect to their own system.

### A. Application

The ability to model local business demand based on structurally similar municipalities serves a wide range of consumers. The aspiring local business owners having access to this knowledge would have a foundation to better understand the potential market they can service as an entrepreneur. Financial service providers, acting as facilitators can use the service to guide their clients to discover the most

suitable market without creating over capacity in a sector. Further, due to the PaaS system's structure, they can integrate the service into their own systems and use it as additional feature to model credit worthiness of the local business proposals. Municipalities striving to maintain and strengthen the local business offerings can use the service to quantitatively estimate the impact obsolescence based demographic changes in Switzerland's general public which is a challenging problem especially for the country's rural municipalities. The municipalities themselves, political parties and local business associations can use the service to attract the most suitable businesses and promote the municipality in terms of ease of doing business and attractiveness of a local market.

With this service offering, stakeholders can better understand when supply and demand changes, thus local markets become increasingly saturated within certain business sectors. With this information available, they can display this information within their own IT systems, to create content on municipality website and use the insights as a base for a political agenda to influence local politics.

### B. Model configuration

Our goal is to use a set of features from publically available data sources [4], extract, load and transform (ETL) them into a machine readable format, and then estimate how large the capacity (market saturation in number of businesses) of a municipality is. These businesses will be classified in groups such as accountants, carpenters, plumbers, general practitioners (doctors), pharmacies, bakeries or butchers. The current state of our less granular model however, can only classify between raw material production, manufacturing, and services. These sectors municipality distribution follows a power law distribution with many observations having low count values, i.e. strongly positively skewed (see Fig. 1). Thus, applying an ordinary linear regression model is problematic. To overcome this problem, the independent variable $y$ is log-transformed before applying a linear regression. Our model assumes that the expected value $E(y) = \mu$ can be modeled as a linear combination of the dependent variables

$$\boldsymbol{x} \in \mathbb{R}^p : \log(\mu) = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p \qquad (1)$$

Furthermore in order to build a model with good accuracy and easy interpretation, the power of Least Absolute Shrinkage and Selection Operator (LASSO) is exploited, a shrinkage and variable selection method for linear regression. Thus, LASSO linear regression performs automatically variable selection and yields sparse models, i.e. models involving only a subset of the variables and are easier to interpret. Additionally, cross-validation is used to select the model with the best generalization to a given test data set.

## V. Analytical Capacity

### A. Dataset

To predict the number of businesses in municipalities, we used a dataset extracted directly from the Swiss Federal Statistical Office (FSO), the largest public open-data provider in Switzerland [8]. The data set is divided into two parts, one for training and the other for testing. The data file which were available to us after the ETL process consisted of 37 socio-demographic and economic factors describing all Swiss municipalities for the years 2011 and 2012. An excerpt of the features (all decimals) is shown in Table 1.

| Characteristic | Description |
|---|---|
| Population | Residents |
| | Population per km² |
| | Population change in % |
| | Population change through migration/birth |
| | Foreign nationals in % |
| | Age of residents: 0 – 19 yrs / 20 – 64 yrs / > 64 yrs |
| Area | Total area |
| | Settlement / Agricultural area in % |
| | Forest stock area in % |
| | Unproductive area in % |
| Mobility | Public transportation |
| | Private vehicle number |
| Work | Enterprise sectoral distribution |
| | Unemployment rate in % |
| Housing | Residential density |
| | Ownership ratio |
| | New housing per 1000 residents |
| | Vacant housing |
| Education | No post-compulsory education |
| | Secondary education |
| | Tertiary education |
| Political votes | FDP / CVP / SP / SVP / right wring / others |

Table 1. Socio-demographic and economic factors.

### B. Performance

In our preliminary model configuration, we were able to predict the number of businesses in the three business sectors in Switzerland's municipalities. Table 2 summaries the results of our preliminary model performance.

| | Goodness of the Fit | |
|---|---|---|
| **Performance** | **Training** | **Testing** |
| Sector 1 | 68.0% | 62.0% |
| Sector 2 | 71.8% | 71.5% |
| Sector 3 | 78.8% | 78.3% |

Table 2. Model performance results

We were able to reach up to 78.3% precision in our projections for Sector 3. The models difference in training and testing ranges from 0.3% to 6%.

## VI. Preliminary Results

Our predictions for the three sectors in Switzerland allows to reliably estimate the capacity of businesses in Switzerland from a macroeconomic perspective. Fig. 1 shows the training and test results of Sector 2

(manufacturing) enterprises ranked by number of enterprises in the top 1'000 municipalities.

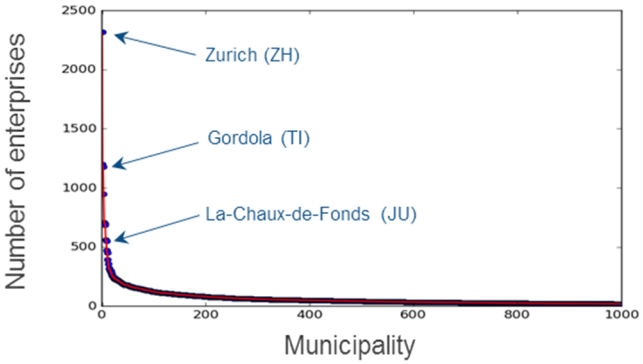


Figure 1. Manufacturing business demand estimation

The socio-demographic and economic features (see Table 1) we used to model enterprise numbers were evaluated by our model. They correlated positively, negatively and some of them had no-effect on the quantity of businesses in the sectors. Table 3 shows their coefficients, ranked according the most positive and most negative effect on the number of businesses in the above mentioned model for Sector 2 (manufacturing).

| **Sector 2 regression coefficients** | |
|---|---|
| **Most positive (decreasing)** | **Most negative (increasing)** |
| • Settlement area in %: 0,051 | • No post-compulsory education: -0,470 |
| • Housing ownership ratio: 0,035 | • Tertiary education: -0,063 |
| • Public transportation: 0,033 | • Population change through birth: -0,034 |
| • Age of residents: 20 – 64: 0,032 | • Population change through migration: -0,028 |
| • Population change in %: 0,029 | • Secondary education: -0,022 |

Table 3. Coefficient and effect ranking for Sector 2

## VII. Discussion and Conclusions

Modeling local market saturation is a computationally intensive task which requires the implementation of advanced machine learning concepts, especially as the number of input factors increases and the model complexity rises. To this point, we were not able to access the number of businesses in a more granular framework such as *International Standard Industrial Classification of all Economic Activities (ISIC)* nor were we able to limit our estimation to SMEs or quantify full time equivalents (FTE), instead of the number of enterprises. Since 99% of all businesses are considered SMEs, our results remain valid. However, increasing the granularity of business classification is required to allow stakeholders to actively use the service. The practical proof of work in a PaaS infrastructure remains to be established. The potential for this service however has been demonstrated and the regression results were robust. As such, the framework can effectively allow stakeholders with interest in local market saturation of business sectors to internalize the computational framework and build their own services on top of them.

## VIII. Summary and Future Work

In this paper, we proposed a PaaS framework which allows local stakeholders to estimate the number of businesses which should serve a local market. In the future, we plan to extend the space of input features and extend and compare machine learning techniques in order to improve the model's scope and performance. Further, we plan to implement our calculation in a PaaS system (e.g. IBM Blue Matrix) and test the user's acceptance with a local financial service provider and a business association in Switzerland.

## References


[1] D. J. Storey, "Understanding the Small Business Sector," University of Illinois at Urbana-Champaign's Academy for Entrepreneurial , Urbana, 1994.

[2] L. B. Barnes and S. A. Hershon, "Transferring power in family business," *Harvard Business Review,* vol. 54, no. 4, pp. 105-114, 1976.

[3] D. Hilley, "Cloud Computing: A Taxonomy of Platform and Infrastructure-level Offerings," Georgia Institute of Technology, 2009.

[4] S. S. Djoko, T. A. Pardo and M. Cook, "A Framework for Benchmarking Open Government Data Efforts," in *2014 47th Hawaii International Conference on System Science*.

[5] A. S. Khairul, A. O. Mohd and S. Shahida, "Interoperable SOA-Based Architecture for E-Government Portal".

[6] S. M. Lin, J. Ansell and G. Andreeva, "Predicting default of a small business using different definitions of financial distress," *Journal of the Operational Research Society,* vol. 63, pp. 539-548, 2012.

[7] C. Verbano and K. Venturini, "Managing Risks in SMEs: A Literatrue Review and Research Agenda," *Journal of Technology Management and Innovation,* pp. 186-197, 2013.

[8] S. Bundesamt für Statistik, "Statistik der Unternehmensstruktur," Bundesamt für Statistik, Neuchâtel, 2016.

[9] N. Taewoo, "New Ends, New Mean, but Old Attitudes: Citizens' Views on Open Government and Government 2.0," in *Proceedings of the 44th Hawaii International Conference on System Sciences*, 2011.

[10] P. M. Kreiser, D. L. Marino, D. F. Kuratko and M. K. Weaver, "Disaggregating entrepreneurial orientation: the non-linear impact of innovativeness, proactiveness and risk-taking on SME performance," *Frontiers of Entrepreneurship Research,* vol. 40, p. 273–291, 2012.

[11] I. E. Altman, G. Sabato and N. Wilson, "The value of qualitative Information in SME Risk management," *Journal of Technology Management & Innovation,* vol. 3, p. 8, 2009.

[12] J. Wang and D. Hang, "Customer-Centered e-Government Service Quality Evaluation: Framewrok and Case Study," in *International Colloquium on Computing, Communication, Control, and Management*, 2009.