

Using Chatbots and Semantics to exploit Public Sector Information

Efthimios Tambouris

University of Macedonia, Egnatia 156, Thessaloniki, Greece, tambouris@uom.gr

Abstract: During the past 20 years, electronic Government (eGov) has become a political priority worldwide. A part of eGov concerns providing Public Sector Information (PSI), lately as open data, to reduce red tape, empower citizens and facilitate economic growth. Recently, a number of standards (namely Core Vocabularies) have been proposed for PSI provision but remain unexploited. Emerging technologies including Big, Open, Linked Data (BOLD) and chatbots provide a fertile ground for improving PSI provision. In this context, the aim of this paper is two-fold. First, to motivate the use of chatbots for providing PSI published according to the Core Vocabularies as BOLD. Second, to develop an architecture blueprint and usage scenarios thus suggesting the use of chatbots is feasible and useful. The ultimate aim of this research is to enable citizens to obtain personalised, easy-to-use PSI, resulting in red tape reduction for citizens and businesses.

Keywords: Chatbot, Core Vocabularies, CPSV, PSI

Acknowledgement: The ideas presented in the paper are rooted in the OpenGovIntelligence project (<http://OpenGovIntelligence.eu>), which is funded by the European Commission within the H2020 Programme under grant agreement no. 693849. The author would like to thank the anonymous reviewers for their important comments and suggestions. The author would also like to thank Anestis Stamatis and Dimitris Zeginis for their assistance in deriving the architecture blueprint.

1. Introduction

During the past 20 years, electronic Government (eGov) has become a political priority worldwide (EC, 2010; USA 2009). Large amounts of funding have been and are still allocated in Public Service and Public Sector Information (PSI) provision (Gartner, 2011). Lately, PSI is also provided widely as Open Data (Janssen et al., 2012).

At the same time, technology and policies evolve rapidly. An emerging area is the Semantic Web and particularly Big, Open, Linked Data (BOLD) PSI (Janssen et al., 2015). BOLD PSI enables machine-understandable data that can be linked with other data on the Web. In addition, recently, EU standards (namely Core Vocabularies) have been proposed for providing BOLD PSI but have not been yet fully exploited. Another rapidly evolving area is chatbots, which are increasingly

used to facilitate interaction and PSI provision (Shawar & Atwell, 2007). However, although research on integrating chatbots and linked data exists (Augello et al., 2009), in our knowledge there are still no chatbots that exploit BOLD PSI based on Core Vocabularies. This can be particularly useful for young people as they spent significant time in texting using their smartphones. Social media data value and impact (Panagiotopoulos et al., 2017; Isett et al., 2017) could be further increased.

In summary, despite technological advances, citizens and businesses are spending considerable time in obtaining PSI while public servants in many public authorities are still overwhelmed with providing information rather than exploiting their knowledge skills.

The aim of this paper is two-fold. First, to motivate the use of chatbots for providing PSI published according to the Core Vocabularies as BOLD. Second, to develop an architecture blueprint and usage scenarios thus suggesting the use of chatbots is feasible and useful. The ultimate aim of this research is to enable citizens to obtain personalised, easy-to-use PSI, resulting in red tape reduction for citizens and businesses.

The rest of this paper is organised as follows. In Section 2 we present background material on Public Service Provision to motivate the use of chatbots for providing PSI. In Section 3 we present an architecture blueprint along with usage scenarios thus suggesting the use of chatbots is feasible and useful. Finally, in Section 4 we present the conclusions and future work.

2. Public Service Provision Informative Phase

To a large extent, eGov refers to the provision of better public services using ICTs. Consequently, Public Service Provision (PSP) is at the heart of many eGov initiatives. In this section, we outline main developments in PSP in three areas, namely stages, information systems and supporting models to motivate the use of chatbots for providing PSI published according to the Core Vocabularies.

2.1. Stages

Public Service Provision (PSP) can be divided into two stages (Peristeras, 2006). The first is the *informative stage* and includes all interactions between a potential service consumer (usually citizen or business) and a service provider (public authority) aiming to inform the former about service details. During this phase, the citizen/business asks questions like *Why? What? Who? Where? When? How?* about a public service. This information can specify whether or not a citizen is eligible to execute a service, the input (e.g. administrative documents) that must be submitted for the service to be invoked, the costs related to the execution of the service, etc.

When the answers to these questions have been provided, this phase is completed and the citizen/business usually proceeds to the *performative phase* i.e. *service invocation*. This phase involves all actions relevant to the execution of a public service.

The informative phase is very important for public administration efficiency and citizen/business satisfaction (Tambouris et al., 2012). Despite its importance however it is often overlooked by governments. In addition, it is often considered as straightforward to provide information online. This is reflected in various stage models that suggest "information provision" is the first stage of sophistication and can be fully accomplished simply by "providing information online", e.g. as plain text in a website. On the other hand, targetization (often perceived as the ultimate stage of sophistication), is related to providing personalisation during the performative phase. Personalisation however is also highly relevant to the informative phase particularly since public services often have numerous versions depending on a number of factors. For example, issuing a driving license has many versions depending on age, whether the application is for a new license or renewing a license or replacing a lost license etc. In each version, the input documents that should be submitted may be different whereas the cost may differ as well. Therefore, we believe that obtaining personalised information is highly relevant to the informative phase.

In this paper, we concentrate on the informative stage that commonly includes providing online public service descriptions and other relevant information.

2.2. Information Systems

The majority of national eGov portals provide public service descriptions. These are often in the form of public service catalogues. In previous research, three different ways used for providing public service descriptions were identified, namely unstructured websites, structured websites and dialogue-based systems (Tambouris et al., 2012).

Unstructured websites contain information about public service(s) in the form of text. The text can be excessive however there is no structure in the content other than the menu of the website. Structured websites also contain information about public service(s) in the form of text. This time however information is provided according to a template that is common for all public services. Finally, dialogue-based systems guide the visitor through a number of online questions. The visitor is answering questions and, at the end of the dialogue, is provided with personalized information, e.g. on whether they are eligible for a public service (e.g. social benefit), the required input documents, costs, responsible authority, etc. These platforms can be similar regarding visitor experience but may differ on the underlying technologies used. An academic example is a prototype that has been developed using open source technologies that enables citizens to get informed on the required documents for declaring their properties in a cadastral (Tambouris et al., 2009).

An illustrative practical example of dialogue-based systems is [benefits.gov](https://www.benefits.gov/)¹. This website includes an online tool, called "benefit finder". Benefit finder enters an online dialogue with the citizen asking a number of questions, with some of them depending on the citizen's answer in a

¹ <https://www.benefits.gov/>

previous question. In this way, the citizen's circumstances are explicitly considered. The end-result is a list of benefits that the citizen is entitled to along with the list of organisations that provide these benefits. The use of a dialogue-based system has some profound advantages over the use of long texts: the former is more easy to use, faster and more accurate.

Chatbots is another promising technology (Jurafsky & Martin, 2017) for interacting with public authorities anytime, anywhere to obtain detailed, personalised and accurate information. Chatbots are not new but have recently attracted considerable attention. Examples include chatbots for: (a) providing information on password retrieval and supporting asylum seekers by helping immigrants to complete the appropriate form with the right data (USA); (b) leading citizens to the right office (Mexico); (c) helping parliament's offices to answer citizens' questions (Japan); (d) providing digital assistance (Signapore); (e) answering questions about the city by employing IBM Watson (New York) (Shawar & Atwell, 2007; Boten et al., 2006).

2.3. Models

Structured websites and dialogue-based systems are often based on a public service model. Many public service models have been proposed in the scientific literature as well as by national initiatives and eGov interoperability frameworks (Peristeras et al., 2009). Despite these efforts, until recently no model has been universally accepted.

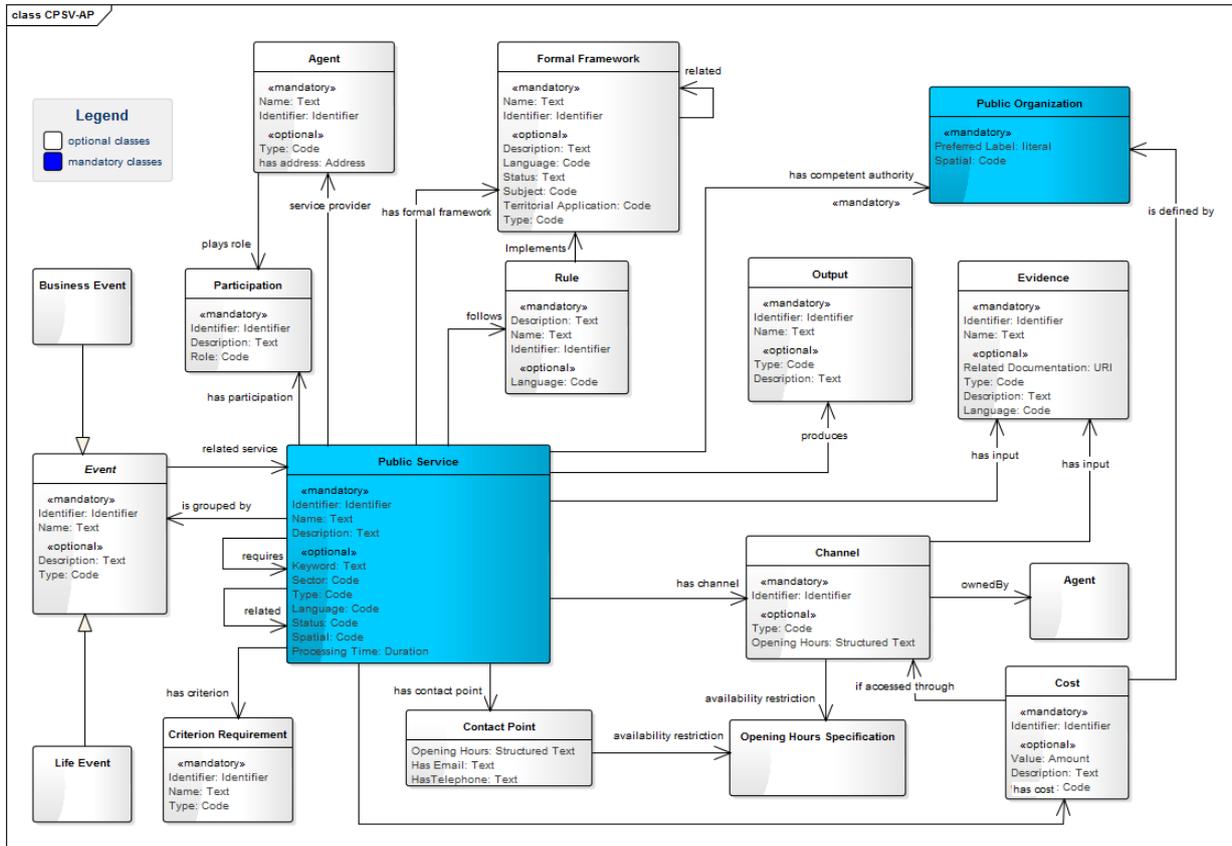
Recently, the European Union has launched the Core Vocabularies, an initiative aiming to overcome interoperability problems by introducing simple models for important entities, which are adopted in all Member States. Core Vocabularies developed so far include, amongst others, the Core Person (EC, 2012), Core Organisation (EC, 2016a), and Core Public Service (EC, 2016b).

The Core Public Service Vocabulary (CPSV) captures the fundamental characteristics of a service offered by a public administration. Lately, the CPSV Application Profile (CPSV-AP) ver. 2 has been launched that specifies a Linked Open Data (LOD) of CPSV. This is illustrated in Fig. 1 using UML notation. Some Member States (e.g. Italy²) have used CPSV as a basis to develop their own models for public service, which covers national needs but still remains compatible with CPSV.

The CPSV-AP can be used, amongst others, to enable public authorities to publish descriptions of their public services using Linked Open Data (LOD). The use of LOD suggests service descriptions can be linked to other data hence providing more rich information. Recent research reports challenges when publishing public service descriptions using CPSV-AP (Gerontas et al., 2018). The results reveal CPSV-AP is promising; however, there is a lack of a coherent method for using CPSV-AP including supporting ICT tools.

² <http://www.essepuntato.it/lode/owlapi/lang=en/http://www.dati.gov.it/onto/cpsvapid>

Figure 1: UML diagram of the CPSV-AP 2.0 (from [European Commission 2016b])



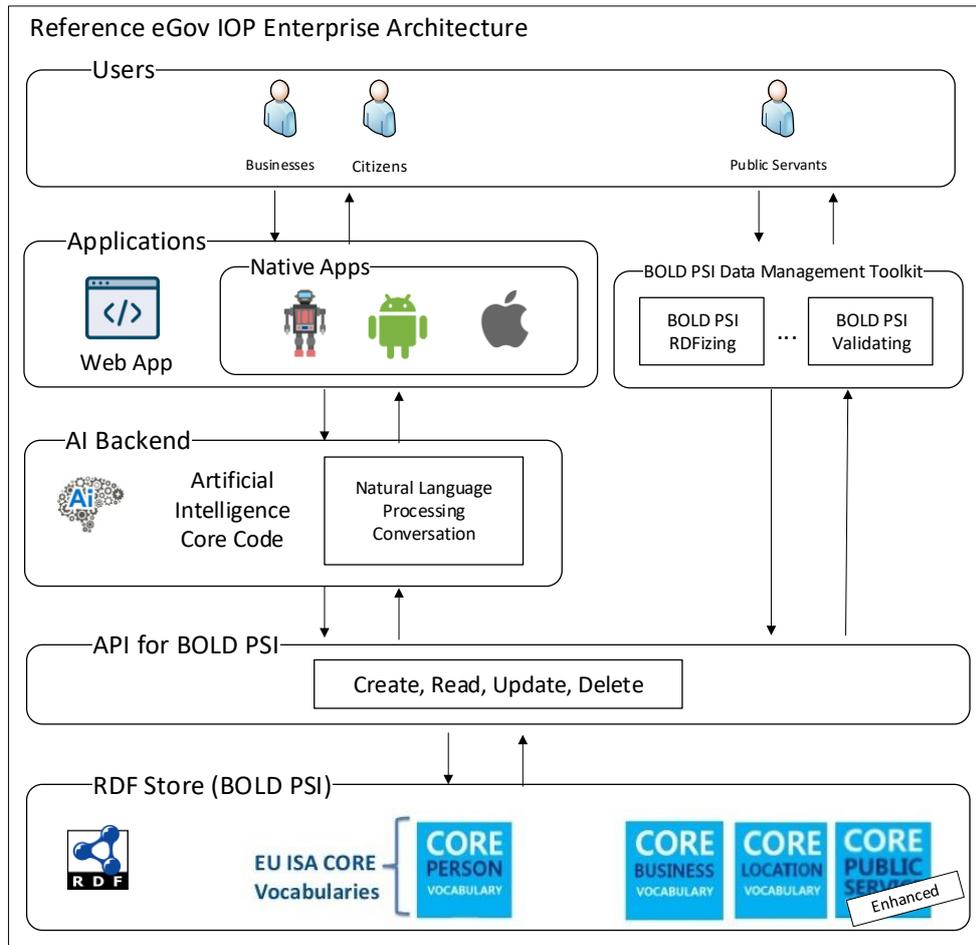
3. Using Chatbots on top of CPSV-AP

Up to now, there is no implementation of chatbots that consume linked data published according to the Core Vocabularies. In this section we present an architecture blueprint of a system enabling the use of chatbots to interact with public service descriptions published using CPSV-AP. In Fig. 2, users can be citizens and businesses in need for PSI (such as details about public authorities and public services, e.g. contact details, costs, relevant legislation etc.) and public servants maintaining the relevant data.

The users will be able to interact with the system using different Applications. We expect that the Web will be the main interface. However, we will also experiment with other environments, such as iOS, Android devices and actual robots. The users will employ in a dialogue with the chatbot through speaking or typing. The system will accept the input as Natural Language and respond accordingly. For this purpose, we will experiment with a number of existing backend platforms, libraries and open resources for chatbots supporting multiple languages, if needed.

The actual information to be provided to users will be stored as Linked Data (i.e. RDF) following the Core Vocabularies and similar standards. For managing the data we will improve existing tools and develop new ones. The resulting BOLD PSI Data Management Toolkit will support all stages of the publication process.

Figure 2: Architecture blueprint



Using this infrastructure, users (citizens and/or business) will be able to ask and obtain answers on any Public Sector Information, which is structured and stored according to the Core Vocabularies standards as RDF.

For example, *citizens and businesses* will be able to obtain information about the cost of services, administrative documents to be provided, contact points, opening hours etc. In addition, *policy makers* will be able to investigate different process reengineering and simplification scenarios, e.g. which public services will be affected if one administrative document is abolished.

In order to facilitate the communication between the backend components and the RDF store we will specify and develop an API. This API for BOLD PSI will enable programmers without RDF knowledge to process the RDF Store.

4. Conclusions and Future Work

Recently, the Core Vocabularies have been introduced as light, standard models that can be used by public authorities when providing information or developing systems. In addition, linked open data versions of Core Vocabularies have been also agreed upon by European Union Member

States. At the same time, chatbots is a technology with high potential in the public sector. Integrating chatbots with linked data has started to attract attention due to the combined benefits of the fields (Augello et al., 2009).

Integrating chatbots with BOLD PSI is a promising but also ambitious step. The results could potentially be significant for the public sector in many perspectives. Citizens and businesses could have in their fingertips easy to use application for obtaining PSI. This would make compliance to legislation easier for citizens and businesses hence reducing administrative burden, which is an EU strategic goal. The use of chatbots through dedicated web pages but also applications like Facebook and twitter could significantly change the way we communicate with the public sector to obtain information. Public servants who are currently working on providing information will be able to concentrate on jobs where their skills are better exploited.

In this paper, we presented the motivation and architecture blueprint of a project aiming to develop chatbots on top of Public Sector Information that is published according to the Core Public Service Vocabulary - Application Profile (CPSV-AP) standard. Next steps of the project involve designing and implementing the software following an agile development methodology, hence having many small development cycles, involving relevant stakeholders and regularly evaluating progress and results.

References

- Augello, A., Pilato, G., Vassallo, G., & Gaglio, S., (2009) *A semantic layer on semi-structured data sources for intuitive chatbots*, Proceeding of the Second International Workshop On Intelligent Interfaces For Human-Computer Interaction (IIHCI), pp. 760-765, 2009.
- Boden C., Fischer J., Herbig K., Spierling U. (2006) *CitizenTalk: Application of Chatbot Infotainment to E-Democracy*. In: Göbel S., Malkewitz R., Iurgel I. (eds) *Technologies for Interactive Digital Storytelling and Entertainment*. TIDSE 2006. Lecture Notes in Computer Science, vol 4326. Springer, Berlin, Heidelberg.
- European Commission (ISA Programme), (2012). Core Person Vocabulary Version 1.0.0. Available at: <<https://joinup.ec.europa.eu/release/core-person-vocabulary/100>> [Accessed Jan. 2018]
- European Commission (ISA Programme), (2016a). Core Public Organization Vocabulary, Version 1.0.0. Available at: <<https://joinup.ec.europa.eu/solution/core-public-organisation-vocabulary/releases>> [Accessed Jan. 2018]
- European Commission (ISA Programme), (2016b). D04.01 – Core Public Service Vocabulary Application Profile 2.0. Available at: <https://joinup.ec.europa.eu/rdf_entity/http_e_f_fdata_ceuropa_ceu_fw21_f7661a4a0_bbd9_b4e98_bb04a_b967c6a1053f9> [Accessed Jan 2018].
- European Commission, (2010). *Europe 2020: A strategy for smart, sustainable and inclusive growth*, available at: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:2020:FIN:EN:PDF>, last accessed on Jan. 2018.
- Gartner, (2011). *Forecast: IT services 2008–2015*, 2Q11 update.

- Gerontas A., Tambouris E., & Tarabanis K., (2018) On Using the Core Public Sector Vocabulary (CPSV) to Publish a “Citizens Guide” as Linked Data, *Proceedings of the 19th Annual International Conference on Digital Government Research (dg.o 2018)*, ACM, [to appear].
- Isett, K.R., Head, B.W. & Vanlandingham, G., (2017). *Building Holistic Evidence for Social Media Impact*. Public Administration Review, 77 (4), pp. 489–495.
- Janssen, M. & van den Hoven, J. (2015). *Big and Open Linked Data (BOLD) in government: A challenge to transparency and privacy?* Government Information Quarterly, 32(4), pp.363–368.
- Janssen, M., Yannis, C. & Zuiderwijk, A. (2012). *Benefits, Adoption Barriers and Myths of Open Data and Open Government*. Information Systems Management, 29(4), pp.258–268
- Jurafsky, D. & Martin, J. H., (2017). *Dialog Systems and Chatbots*, in Speech and Language Processing, available at: <https://web.stanford.edu/~jurafsky/slp3/29.pdf> [Accessed June 2018].
- Panagiotopoulos, P., Bowen, F. & Brooker, P., (2017). *The value of social media data: Integrating crowd capabilities in evidence-based policy*. Government Information Quarterly, 34(4), pp. 601–612. Available at: <https://doi.org/10.1016/j.giq.2017.10.009>.
- Peristeras, V., (2006) *The Governance Enterprise Architecture-GEA-for Reengineering Public Administration*. PhD Thesis, University of Macedonia, Greece.
- Peristeras, V., Tarabanis, K. & Goudos, S.K. (2009). *Model-driven eGovernment interoperability: A review of the state of the art*. Computer Standards and Interfaces, 31(4), pp.613–628.
- Shawar, B. A., & Atwell, E. (2007). *Chatbots: Are they Really Useful?* LDV Forum 22, pp. 29-49.
- Tambouris, E., Kliafas, A., Kalampokis, E. & Tarabanis, K. (2009) “Reducing administrative burden through online dialogues: The case of declaring a property to the Hellenic Cadastre”, *Proceedings of the 8th International Conference on Electronic Government*, Trauner Druck, pp. 3-10.
- Tambouris, E., Outsetari, V. & Tarabanis, K. (2012) “On Public Service Provision Informative Phase: A dialogue-based Model and Platform”, *Proceedings of the International Conference on Theory and Practice of Electronic Governance (ICEGOV 2012)*, ACM, pp.132-141.
- USA Government, (2009). *American Recovery and Reinvestment Act*, available at: <http://www.gpo.gov/fdsys/pkg/BILLS-111hr1enr/pdf/BILLS-111hr1enr.pdf> [Accessed Jan 2018].

About the Author

Efthimios Tambouris

Efthimios Tambouris is an Associate Professor of Information Systems and eGovernment at the Department of Applied Informatics at the University of Macedonia, Thessaloniki, Greece. Before that, he served at various Research Centers and the ICT Industry. He holds a Diploma in Electrical Engineering from the National Technical University of Athens (NTUA), Greece, and an MSc and PhD from Brunel University, UK. During the last two decades, he has initiated, coordinated and participated in numerous EU-funded research projects. He is an expert for the European Commission and CEN. He has more than 150 scientific publications.