

biblio.ugent.be

The UGent Institutional Repository is the electronic archiving and dissemination platform for all UGent research publications. Ghent University has implemented a mandate stipulating that all academic publications of UGent researchers should be deposited and archived in this repository. Except for items where current copyright restrictions apply, these papers are available in Open Access.

This item is the archived peer-reviewed author-version of:

The Public Sector DNA on the web: semantically marking up government portals

Raf Buyle, Laurens De Vocht, Dieter De Paepe, Mathias Van Compernelle, Geraldine Nolf, Ziggy Vanlishout, Björn De Vidts, Erik Mannens, and Peter Mechant

In: Smart Descriptions & Smarter Vocabularies, 2016.

https://www.w3.org/2016/11/sdsvoc/SDSVoc16_paper_1

To refer to or to cite this work, please use the citation to the published version:

Buyle, R., De Vocht, L., De Paepe, D., Van Compernelle, M., Nolf, G., Vanlishout, Z., De Vidts, B., Mannens, E., and Mechant, P. (2016). The Public Sector DNA on the web: semantically marking up government portals. *Smart Descriptions & Smarter Vocabularies*

The Public Sector DNA on the web: semantically marking up government portals.

Raf Buyle¹, Laurens De Vocht¹, Dieter De Paepe¹, Mathias Van Compernelle²,
Geraldine Nolf³, Ziggy, Vanlishout³, Björn De Vidts³, Erik Mannens¹, Peter Mechant²

¹ Data Science Lab, Ghent University - iMinds, Ghent, Belgium.

{firstname.lastname}@ugent.be

² MICT, Ghent University - iMinds, Ghent, Belgium.

{firstname.lastname}@ugent.be

³ Informatie Vlaanderen, Brussels, Belgium.

{firstname.lastname}@kb.vlaanderen.be

Abstract. Base registries contain core public sector data. They are fundamental building blocks in supporting interaction between government and private sector. To enable the private sector to discover, adopt and use information from base registries (e.g. addresses of organizations and public services), the government needs a distribution model. Therefore, the Flemish government is working on a technical strategy to add markup to government portals to embed their ‘DNA’, semantic annotations, on third-party private sector platforms, to dissolve the existing governmental silos and to provide better public services. In this context, this paper reviews a potential strategy to ‘open up’ base registries that combines the best of both worlds: bridging between the schema.org and the European ISA Core vocabularies.

1 Introduction

Today, citizens and entrepreneurs are consuming government information and services on private platforms. For example, when an internet user searches for a specific museum in the City of Ghent, Google will show the address, opening hours and related information in a pane on the right hand side of the search results, without the need for navigating to the government portal itself. Siri by Apple and Cortana by Microsoft are good examples of third-party platforms that provide information about government services. Asking Cortana “the opening hours of the city hall in Paris”, she answers “City Hall, Saint-Jean Hall is open until 7 pm” and shows a map and the contact information. In the future similar services should be able to handle more complex or detailed questions about public services e.g. “funding for roof insulation”. Search engines such as Google and Bing, use machine-readable data embedded in human-readable HTML documents. This data is published in line with the schema.org vocabularies, used by over 10 million sites¹. These provide a single

¹ Welcome to Schema.org: <http://www.schema.org> (last accessed July, 04, 2016)

schema across a wide range of topics (including people, places, events, products, opening hours) and are the only broad vocabulary used by more than a quarter of the pages indexed by the major search engines [3]. The majority of ‘internet bots’ visiting the government portals are ‘crawling software’ used by search engine bots or marketing research tools [3]. We expect that in the future, the share of crawlers used for data mining will increase quickly.

The administrations have no control on (the quality of) the information about public services on these third-party platforms. However, we believe that governments can create the right circumstances to safeguard the quality of information by allowing and enabling third parties to build services upon authoritative public datasets such as Base Registries. Base registries contain core public sector data and are generic building blocks supporting interaction between the government and the private sector. For the public sector, they are a fundamental footprint like DNA for living organisms.

2 Motivation

The vision of the European Commission is that governments will be connected, networked and fully joined-up, interacting with each other and with private actors [1]. This requires dissolving governmental data silos and moving towards a "whole-of-government" approach [2]. The European Commission defines a base registry (BR) as a trusted authentic source of information under the control of an appointed public administration or organization appointed by the government².

To afford the private sector to use information from base registries, we need to have a semantic distribution model. The success of schema.org proves that adding machine-readable in human-readable HTML documents³ is a feasible strategy. Therefore the Flemish Government is working on a strategy to add markup to government portals to promote information about their public services on third-party platforms and, as a side effect, to boost visibility of government information in the search engine result pages.

Recommendations that specify how administrations, businesses, and citizens should communicate with each other within the EU and across borders are defined in the European Interoperability Framework (EIF⁴). Because the Flemish Government finds it crucial to support interoperability across multiple interoperability levels we consider combining schema.org and the European ISA Core vocabularies. Flanders is aligning the semantics of the base registries on the Core Vocabularies of the Interoperability Solutions for Public Administrations Programme, which promotes semantic in-

² Access to base registries: base registry definition (2012), EUROPEAN COMMISSION Interoperability Solutions for Public Administrations Programme, page 4. http://ec.europa.eu/isa/actions/documents/isa_1.2_d1.2_base_registry_definition.pdf

³ HTML Microdata: <https://www.w3.org/TR/microdata/> (last accessed July, 04, 2016)

⁴ Towards interoperability for European public services: http://ec.europa.eu/isa/documents/isa_annex_ii_eif_en.pdf

teroperability among the European Union Member States⁵. A set of commonly agreed Core Vocabularies supported by the European Member States provides a starting point⁶. The interoperability programme of the Flemish Government, “Open Standards for Linked Organizations” also referred to as “OSLO²” focuses on the semantic level and extends the ISA CORE Vocabularies in order to facilitate the integration of the Flemish base registries with one another and to safeguard their implementation in business processes of both the public and private sector [5].

3 Preliminary Results

Using the network of public portals in Flanders, semantically marking up web content in line with the EU ISA CORE vocabularies should promote government building blocks and create leverage to adopt these data and identifiers in private sector initiatives. Our solution uses both schema.org and the ISA CORE vocabularies⁷ (Core Public Service Vocabulary, Core Public Organization Vocabulary, Core Location Vocabulary, ...) to publish information via (government) portals.

In addition, we will relate both resource descriptions so they each might provide additional information about the subject resource." This implies that 'ISA CORE Location' may provide additional properties or more accurate properties about for example a 'schema:addressLocality' that are not available in schema.org.

The scenario described below will be evaluated (see also Figure 1).

1. Publishing information on public organizations (e.g addresses and opening hours) in line with the schema.org vocabulary on the website of the Flemish Government and the website of a Local Government and will evaluate if the information is indexed search engine results page. Furthermore, we will evaluate the impact on the structured information⁸ and how it affects the page ranking as a side effect.
2. Test how search engines handle more volatile information, specifically if updates of government information propagate at the level of the structured information in popular search engines.
3. Add the ISA CORE vocabularies and the `rdfs:seeAlso` which relates both resources and evaluate if crawlers can find connections between both representations.
4. Prove that this comes with no extra work, and will prototype a plugin for Drupal which will add the machine readable information.
5. Demonstrate that has is aimed at creating a political basis to implement the best practices of this study in Flanders.

⁵ Improving semantic interoperability in European eGovernment systems: http://ec.europa.eu/isa/actions/01-trusted-information-exchange/1-1action_en.htm

⁶ e-Government Core Vocabularies: https://joinup.ec.europa.eu/asset/core_vocabularies/description

⁷ Core Vocabularies :http://ec.europa.eu/isa/ready-to-use-solutions/core-vocabularies_en.htm

⁸ Introducing the Knowledge Graph: things, not strings: <https://googleblog.blogspot.be/2012/05/introducing-knowledge-graph-things-not.html>

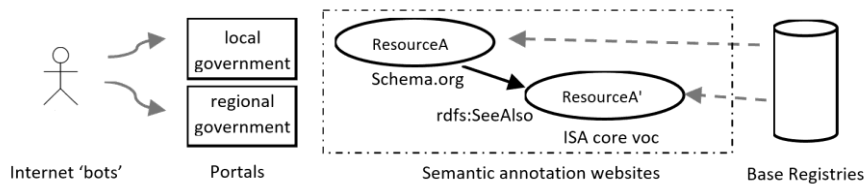


Fig. 1. Evaluating the adoption of data in Base Registries on private sector platforms

4 Conclusions and Next Steps

We are gathering feedback on (1) the implementation of the presented resulting method, (2) the expected leverage regarding adoption and quality of the base registry data on private sector platforms and (3) how to measure the effects. Governments in Flanders will only implement the proposed publication strategy if they perceive its usefulness and ease-of-use. In the case of schema.org, the main breakthrough of the adoption of schema.org was the extensive support in content management systems (ease-of-use) including Drupal and Wordpress. The main success of Schema.org is found in the support of the most important search engines (usefulness).

Therefore, the Agency for Information in Flanders will ask the European Commission (ISA²) to evaluate and promote this strategy for government portals in Europe, aiming at a vast scale from the beginning. To support the adoption in the Region of Flanders, we will build generic software components, supporting (local) governments to integrate this strategy in their portals.

References

1. Botterman Maarten, Millard Jeremy et al. (2009). Value for citizens: A vision of public governance in 2020. Report for the European Commission.
2. The European eGovernment Action Plan 2011-2015: Harnessing ICT to promote smart, sustainable & innovative.
3. Guha, R.V., Brickley, D., MacBeth, S.: (2015) Schema.org: Evolution of Structured Data on the Web. Queue - Structured Data Queue Homepage archive, Volume 13 Issue 9, page 10
4. Rama Raol, A. et al.: (2016) Advanced Traffic Management with Bot Detection and Geo Location Monitoring. Imperial Journal of Interdisciplinary Research (IJIR), Vol-2, Issue-5, 2016, page 1824
5. Van Compernelle, M., De Vocht, L., Goedertier, S., Loutas, N., Verborgh, R., Mannens, E., ... & Van de Walle, R. (2016). State-of-the-Art Assessment on the Implementations of International Core Data Models for Public Administrations. In Proceedings of the 9th International Conference on Theory and Practice of Electronic Governance (pp. 327-337).