Towards interoperability in the public sector

Raf Buyle

Ghent University - imec, IDLab, Department of Electronics and Information Systems
Technologiepark 15, 9052 Ghent, Belgium
raf.buijle@ugent.be

Abstract. Due to a lack of interoperability public administrations are unable to share and reuse structured information across different applications. The hurdles are the lack of adequate semantic standards, scarcity of web-oriented architecture, government austerity and reluctance caused by budget constraints. In this paper, I outline a process and method to tackle these hurdles and apply them on Base Registries as part of my doctoral research. The outcome, based on the principles of the semantic web, is valuable for researchers and public administrations that aim to raise interoperability in complex data ecosystems.

Keywords: Interoperability; e-Government; Base Registries, Semantic Web

1 Problem statement

Interoperability is the ability of organisations to share information and knowledge, through the business processes they support, by means of the exchange of data between their ICT systems [9]. We notice two primary drivers for interoperability in existing literature [17, 26]. Firstly, citizens and entrepreneurs expect a coherent customer experience from their government as they became accustomed to by services in the private sector [17]. Secondly, our digital economy embraces new ecosystems including 'the government as a platform' where business expect public-private partnerships to arise. Governments struggle to deliver integrated, interconnected and cross-sectoral services due to sectoral specialisation or "departmentalisation" [16]. Governments provide several hundreds[1] of products; their service delivery is supported by specialised applications from different software vendors. The information in the software solutions are often modelled from a single perspective and therefore cannot be shared and reused across different applications and processes, causing data silos [5]. To integrate these applications, data needs to be transformed, which causes high costs. Given government budget cuts, applications often remain data islands and citizens and businesses have to provide the same information over and over. To overcome existing data islands caused by information systems that have no or limited external connectivity, we need to address multiple interoperability levels; namely on the legal, organisational, semantic and technical level [8]. Interoperability Frameworks assume

[1] http://doc.esd-toolkit.eu/ServiceList/

some kind hierarchy in terms of maturity with regard to layers of interoperability [19]. In other words, organisational interoperability can only be achieved when standards for semantic and technical interoperability have successfully been implemented. Therefore, my research mainly targets semantic and technical interoperability.

Semantic interoperability focusses on the meaning of data elements, such as a resource accessible by a Universal Resource Identifier[2] (URI), and the relationship between the things they identify. It includes developing vocabularies to describe data exchanges and ensures that data elements are understood in the same way by communicating parties [8]. Semantic interoperability covers also the syntactic aspect which refers to the grammar and format [9], such as HTML or XML. Unfortunately "several technical limitations and practical challenges preventing easy adoption remain unsolved", including unjustified benefits and considerable effort expectations [25]. However, a government administration that has the *political power* and *will* when introducing e-government systems is vital to success [20], this includes internal organisational politics involving organisational members as well as external politics concerning how the government organisation relates to its council [23]. Technical Interoperability is often centred on (communication) protocols and the infrastructure needed for those protocols to operate [27]. Due to the size of the public sector and fragmentation of ICT solutions, this resulted in various interface specifications and communication protocols. This legacy is a major obstacle for interoperability [9].

The problem statement of my doctoral research is: what *processes (events to produce a result)* and *methods (how to complete these events)* are suited for raising semantic and technical interoperability within an operational public sector context. I study this problem both from the technical and political point of view in the context of Base Registries. A base registry is a trusted authentic source of information under the control of an appointed public administration or organisation appointed by the government [9].

2 The relevance of targeting interoperability

Because of budget cuts, public administrations have to do more with considerably less. Interoperability can lead to lower costs [8] and produce savings, but at the same time it requires an initial investment [10]. To secure these investments and interoperability there is a demand [19, 3] for a stable, governed standard, which is "a technical document designed to be used as a rule, guideline or definition. It is a consensus-built, repeatable way of doing something"[3]. Interoperability addresses the need for corporation between administrations, the exchange of information to accomplish with legal conditions or political engagements and to share and reuse information which leads to an improved public service delivery and lower cost [3].

[2] https://www.w3.org/TR/uri-clarification/
[3] https://www.cen.eu/work/ENdev/whatisEN/Pages/default.aspx

My research is valuable for (semantic web) researchers in the public sector but also in other domains including transport [6], finances [21] and life sciences [2].

3 Related work

Projects on interoperability such as StUF, INSPIRE, ISA² as well as CSMICS are struggling with semantic interoperability. These struggles play out in various domains. We distinguish:

1. context-neutral, re-usable and extensible data models [11] which are embedded in
2. a stable, governed standard and accompanied by
3. technical guidelines that specify how these could be implemented in an operational context.
4. on an organisational level political support is essential, for collecting sponsoring and gaining authority and engagement [5].

In this section we discuss these four projects on interoperability using numbers to refer to the aforementioned domains of 'struggle'. *Standaard Uitwisseling Formaat (StUF)*[4] is a canonical data exchange model for information exchange within the Dutch government, introduced in 1996. According to a study[5] of the City of Den Haag, StUF is overspecified and not extendable, which makes it harder to reuse (1), and also has a lack of technical guidance (3). *Infrastructure for SPatial InfoRmation in the European Community (INSPIRE)* is a programme that focuses on the interoperability of geographical information for environmental policy making within Europe. Since 2004 INSPIRE is a directive which sets the legal framework in Europe [15]. The INSPIRE data specifications[6] are legally binding and accompanied by Technical Guidelines that specify how legal obligations could be implemented[7]. The data specifications tend to be overspecified because it was designed for a specific domain, which makes them harder to reuse (1). The INSPIRE programme is investigating how the Linked Data and RDF[8] can facilitate cross-sector interoperability. *The ISA² programme*, which is running since 2016, focusses on the interoperability of public services across Europe, in specific on Core vocabularies which cover the semantics of a set of generic concepts. ISA defines a "Core Vocabulary" as a simplified, reusable, and extensible data model that captures the fundamental characteristics of an entity in a context-neutral fashion [12]. The Core vocabularies provide both an RDF and XML schema. The CORE vocabularies are not legally binding (4). Detailed Technical Guidelines (3) could speed-up the adoption. *Collaborative development of a common semantic model for interlinking Cancer Chemoprevention linked data sources (CSMICS)* defines a (1) re-usable data model for cancer chemoprevention, using RDF

[4] http://www.gemmaonline.nl/index.php/StUF_Berichtenstandaard
[5] https://www.sig.eu/files/nl/11_Eindrapport_DenHaag_StUF_standaard.pdf
[6] http://inspire.ec.europa.eu/data-specifications/2892
[7] http://inspire.ec.europa.eu/Technical-Guidelines2/Metadata/6541
[8] http://inspire.ec.europa.eu/news/linking-inspire-data-draft-guidelines-and-pilots

as the data model. The bottom-up (4) "meet-in-the-middle" approach involves the stakeholders at the different phases of the development [28]. This approach facilitates interoperability and contributes to the re-use of biomedical ontologies.

4 Research Question(s)

My research questions consider the problem statement from a technical and political point of view. The main question in my research is:

(1) *how can governments develop a scalable technique for raising and implementing semantic and technical interoperability, applied to Base Registries within an operational public sector context?*

This question has two perspectives. On the one hand, we have a technical viewpoint:

(2) how to define technical guidance to business analysts and developers to maintain semantic agreements, provide persistent unambiguous identifiers and design an interface which can be easily interpreted by clients?

On the other hand, we have the political context:

(3) how to build consensus among different public administrations and rewire public sector programs which often are under the authority of a different governmental level?

5 Hypotheses

My research is based on the following presumptions:

(1) The design principles of the Semantic Web[9] can facilitate interoperability within the public sector by adding context and useful links, using the Resource Description Framework[10] (RDF) as a data model for Base Registries.

(2) Due to government austerity, decisions in relation to semantic agreements must be traceable, transparent and consistent at all levels. Therefore the form of the specifications and guidelines must be aligned to the different types of stakeholders (e.g. technical, business, policy) to facilitate a levelled discussion.

6 Preliminary results

My research took into account the process of reaching and implementing of semantic agreements in the Open Standards for Linked Organizations program (OSLO). OSLO is an interoperability program in the Region of Flanders, which brings together expertise from different business domains and governmental levels, independent of a specific thematic project. The Flemish Government developed a domain model in line

[9] https://www.w3.org/DesignIssues/LinkedData.html
[10] https://www.w3.org/2001/sw/wiki/RDF

with international standards including ISA and INSPIRE[11] enriched by data extensions to comply with the local context [5]. The formal specification is published at data.vlaanderen.be[12]. The thematic working groups, with 88 participants from the public and private sector, demonstrated that it is possible to reach semantic agreements and overcome the political hurdles. These agreements are documented using the Unified Modeling Language™ (UML[13]). In a next step, the UML model is enriched with tags, which allows mapping the properties to RDF vocabulary terms. The UML model along with the mappings are then automatically transformed into an RDF model[14] [7]. The formal specification is then published, including a JSON-LD[15] context[16] which allows embedding the semantic agreements in JSON services. These JSON documents can now be interpreted[17] as Linked Data. This method indicated that the semantic agreements, reached at the business level, can be preserved. In addition, the Flemish Government has developed a URI [18] standard for persistent identifiers based on principles of W3C, ISA and the Netherlands and applied it to addresses[18] which can be dereferenced using the HTTP protocol. Moreover, several pilots [4,5] and a Base Register where over 4 million addresses and their geographical coordinates were published showed that the design principles of the Semantic Web could facilitate technical and semantic interoperability using RDF as a data model.

7 Approach

I will evaluate and improve my approach via action-research, which aims to contribute both to the practical concerns of people and to the goals of social science by joint collaboration within a mutually acceptable ethical framework [22]. I will conduct my research in the Flemish public sector in Belgium. Belgium is a federal state with three communities, three regions, and four language areas. My approach to addressing the research questions is to focus on two deliverables: the *processes* and *methods* suited for raising interoperability by researching and improving the OSLO programme within the context of Base Registries.

The data specification *process* will be aligned with the principles[19] of international standardisation bodies; due process, broad consensus, transparency, balance and openness. The current development activities of OSLO already follow a transparent process: all records of decisions[20] and discussions[21] are publicly accessible These

[11] http://inspire.ec.europa.eu/
[12] http://data.vlaanderen.be/ns/
[13] http://www.omg.org/spec/UML/
[14] https://github.com/Informatievlaanderen/OSLO-EA-to-RDF
[15] https://www.w3.org/TR/json-ld-syntax/#the-context
[16] http://data.vlaanderen.be/context/adres.jsonld
[17] https://www.w3.org/TR/json-ld-syntax/#interpreting-json-as-json-ld
[18] http://data.vlaanderen.be/id/adres/2179183
[19] https://open-stand.org/about-us/principles/
[20] https://informatievlaanderen.github.io/OSLO/

activities will be formalised and the different process steps adapted to fit the different stakeholders in the specification process, including domain experts, business- and technical analysts. The *method* pursues an implementation of the design principles of linked data[22] as asserted by Tim Berners-Lee in 2006. Existing public sector information systems store data in relational databases and often use Extensible Markup Language[23] schemas to exchange data. These schemas, intended to exchange data, can not be easily adapted or extended [16]. In my research, I will focus on how RDF, which is an extendable data model, can be adopted in the public sector and how the semantic agreements reached between domain experts, automatically[24] can be transformed into an RDF model preserving the semantic agreements. To allow structured and semi-structured data to be mixed, exposed, and shared across different applications[25], it is crucial that the specifications are resolvable on the Web. Therefore, I will research how existing software architectures can be rewired to a Representational State Transfer (REST) style, which outlines how to construct network-based software applications having the same characteristics as the Web: simplicity, evolvability, and performance [24]. The key innovation lies in combining a bottom-up consensus-based approach with a formal top-down approach which anchors the decisions within a formal government body, using linked data as a blueprint.

8 Evaluation plan

I will evaluate the success of my research by applying the process and method in public sector initiatives within the context of the re-use of address information. Address information has an important value for the public and private sector. Public data often has a location-based component. "… It is estimated that 80% of the informational needs of government policymakers are related to geographic location" [13]. I will benchmark the output variables that affect the Successful Implementation of ICT Projects in Government [14], using the following criteria:

- *Cost reduction*: I will evaluate the reduced number of technical and semantical conversions of addresses between applications and estimate the financial benefits in relation to the total integration cost,
- *The quality of service delivery*: I will measure the increase in re-use of address-information by comparing the decrease of requested information citizens provide, in relation to the service complexity and the customer satisfaction.
- *Technological benefits:* I will research potential semantic conflicts using the ISA²-method [11] for classification of schema-level conflicts,
- *Improved efficiency*: I will conduct qualitative research by interviewing the stakeholders in the public and private sector, including perceived benefits.

[21] https://github.com/Informatievlaanderen/OSLO/issues
[22] https://www.w3.org/DesignIssues/LinkedData.html
[23] https://www.w3.org/TR/REC-xml/
[24] https://github.com/Informatievlaanderen/OSLO-EA-to-RDF
[25] https://www.w3.org/2001/sw/wiki/RDF

9 Reflections

This paper describes my ongoing research on semantic and technical interoperability in the public sector, in the context of base registries. In contrast to StUF, and INSPIRE, my approach combines the *process* to reach semantic agreements by broad consensus and an end-to-end *method* based on the principles of linked data to maintain the semantic agreements on within a public sector context. My method allows datasets to be linked into a public sector knowledge graph governed by a public body. Preliminary results indicate that it is possible to reach semantic agreements[26] and overcome the political hurdles within an operational public sector context by using a meet-in-the-middle approach. Moreover, the pilot project on the address base registry demonstrated the feasibility of rewiring existing information systems to REST-style architectures. These first findings indicate that my method could raise semantic and technical interoperability within an operational public sector context.

Acknowledgements. I would like to thank my supervisors Erik Mannens and Peter Mechant for their support, Laurens De Vocht, Mathias Van Compernolle, Pieter Colpaert and Dieter De Paepe for their writing advice. My research is supported by Informatie Vlaanderen and Ghent University.

References

1. Akatkin, Y. et al. Towards the digital government in Russia: integrative approach. In Proceedings of the International Conference on Electronic Governance and Open Society: Challenges in Eurasia. pp. 41-48 (2016).
2. Aoki-Kinoshita, K. et al. Implementation of linked data in the life sciences at BioHackathon 2011. Journal of biomedical semantics, 6(1), 3, p. 11 (2015).
3. Audit Vlaanderen, Thema-audit gemeentelijke belastingen en retributies. p. 81 (2016).
4. Buyle, R. et al. Local Council Decisions as Linked Data: a proof of concept. In Proceedings of the 15th International Semantic Web Conference. Kobe, Japan. (October 17-21, 2016).
5. Buyle, R. et al. OSLO: Open Standards for Linked Organizations. Proceedings of the 2016 3rd International Conference on Electronic Governance and Open Society: Challenges in Eurasia'. (November 21-22 2016).
6. Colpaert, P. et al. Open Transport Data for maximising reuse in multimodal route planners: a study in Flanders. IET Intelligent Transport Systems (2017).
7. De Paepe, D., Thijs, G., Verborgh, R., Mannens, E., Buyle, R.: Automated uml-based Ontology Generation in oslo². In: Proceedings of the 14th eswc: Posters and Demos (2017).
8. European Commission, ISA: European Interoperability Framework (EIF) for European public services, Annex 2. pp. 2-21 (2010).

[26] http://data.vlaanderen.be

9. European Commission, ISA: European Interoperability Framework (EIF). p. 5-37 (2017).
10. European Commission, ISA: The ISA programme: Overcoming eBarriers to European public services. p2 (2011).
11. European Commission, ISA: Handbook for using the Core Vocabularies. pp. 8-10 (2014).
12. European Commission, Process and methodology for developing semantic agreements. p. 7 (2013).
13. Garson, D., Biggs R. Analytic Mapping and Geographic Databases, SAGE university paper, 87 (1992).
14. Gichoya, D. Factors affecting the successful implementation of ICT projects in government. the Electronic Journal of e-government, 3(4), pp. 175-184 (2005)
15. Infrastructure for Spatial in formation in Europe, Drafting Team "Data Specifications" – deliverable D2.5: Generic Conceptual Model, p. 5 (2007).
16. ISA programme by PwC EU Services EESV, Case study: How Linked Data is transforming eGovernment. pp 1-5 (2013).
17. Janssen, M., & Van Veenstra, A. F. Stages of growth in e-government: An architectural approach. The electronic journal of e-government, 3(4), pp. 193-200 (2005).
18. Juha H. Persistent identifiers – an overview. Technology Watch Report - Standards in Metadata and Interoperability (2010).
19. Kubicek, H., Cimander, R. Three dimensions of organizational interoperability. European Journal of ePractice, 6, pp. 1-12 (2009).
20. Novakouski. et al. Interoperability in the e-Government Context (2012).
21. O'Riain, S. et al. Linked data driven information systems as an enabler for integrating financial data. Information Systems for Global Financial Markets: Emerging Developments and Effects. pp. 239-270 (2011).
22. Rapoport, R. Three dilemmas in action research: with special reference to the Tavistock experience. Human relations, 23(6), pp. 499-513 (1970).
23. Rocheleau, B. Politics, accountability, and governmental information systems. Public information technology: Policy and management issues. pp. 20-52 (2003).
24. Fielding, R., Taylor, R. Architectural styles and the design of network-based software architectures. Doctoral dissertation: University of California, Irvine. pp. 151 (2000)
25. Tatarintseva, O. et al. Is Your Ontology a Burden or a Gem?–Towards Xtreme Ontology Engineering. Proc. ICTERI, p. 65-81 (2011).
26. United Nations Department of Economic and Social Affairs. United Nations E-Government Survey 2012. United Nations, p. 69 (2012).
27. Van der Veer, H., Wiles, A. European Telecommunications Standards Institute: Achieving Technical Interoperability – the ETSI Approach (2006).
28. Zeginis. D. et al. Collaborative development of a common semantic model for interlinking Cancer Chemoprevention linked data sources. The Semantic Web Journal. p. 14 (2012).