Integrating OSLO semantics in word processors

DWIGHT VAN LANCKER, IDLab, Dep. of Electronics and Information Systems, Ghent University - imec, Belgium

NIELS VAN DURME and EVELINE VLASSENROOT, MICT, Dep. of Communication Sciences, Ghent University - imec, Belgium

RAF BUYLE, IDLab, Dep. of Electronics and Information Systems, Ghent University - imec, Belgium

PETER MECHANT, MICT, Dep. of Communication Sciences, Ghent University - imec, Belgium

PIETER COLPAERT, IDLab, Dep. of Electronics and Information Systems, Ghent University - imec, Belgium

ERIK MANNENS, IDLab, Dep. of Electronics and Information Systems, Ghent University - imec, Belgium

Documents issued by the government such as public tenders or policy documents often lack consistent semantics, which leads to ambiguities and misinterpretations. Take for example granting subsidies to companies. The conditions for entitlement to a subsidy are checked against the government's authentic data sources. However, the various governments and administrations have different definitions of, for instance, a small and medium-sized enterprise (SME), which can be derived from a European legal framework or a financial perspective. The absence of uniform definitions for these terms, results in a lot of duplicate efforts for both the government and the entrepreneur. To tackle the problem of semantics, Flanders has an interoperability program, Open Standards for Linked Organizations (OSLO) whose primary goal is to ensure that systems exchanging data can use a common vocabulary. However, despite the results made by OSLO, they do not reach policy-makers working mainly on the legal and organizational levels. To close this gap and make semantic agreements available at these levels, we developed two tools. With OSLO Lookup, we provide a simple user interface which lets users query the semantics assets, while the OSLO365 plugin allows embedding the semantic assets in a Microsoft Word document. To assess the relevance and usability of these tools, servants of a local administration were interviewed. This paper outlines that semantic agreements that are mainly used on data level, can provide added value at an organizational and legal level as well.

CCS Concepts: • **General and reference** → **Computing standards**; • **Applied computing** → *E-government*; • **Information systems** → *Resource Description Framework (RDF)*.

Additional Key Words and Phrases: Interoperability, Linked Data, e-Government, Semantic Web, Word processors

---

Authors' addresses: Dwight Van Lancker, dwight.vanlancker@ugent.be, IDLab, Dep. of Electronics and Information Systems, Ghent University - imec, Belgium; Niels Van Durme, niels.vandurme@ugent.be; Eveline Vlassenroot, eveline.vlassenroot@ugent.be, MICT, Dep. of Communication Sciences, Ghent University - imec, Belgium; Raf Buyle, raf.buyle@ugent.be, IDLab, Dep. of Electronics and Information Systems, Ghent University - imec, Belgium; Peter Mechant, peter.mechant@ugent.be, MICT, Dep. of Communication Sciences, Ghent University - imec, Belgium; Pieter Colpaert, pieter.colpaert@ugent.be, IDLab, Dep. of Electronics and Information Systems, Ghent University - imec, Belgium; Erik Mannens, erik.mannens@ugent.be, IDLab, Dep. of Electronics and Information Systems, Ghent University - imec, Belgium.

---

1 INTRODUCTION

In 2015, the founder of the World Economic Forum, Klaus Schwab, announced that the world was undergoing its fourth industrial revolution. He defines the Fourth Industrial Revolution as "a world in which virtual and physical systems of manufacturing cooperate with each other in a flexible way at the global level"[7]. One of the pillars of this revolution is *data*, which is used at various levels, e.g. systems use data to communicate to each other, but also at the political level as policy-makers use data-based tools to justify their decisions. In the past few years, governments have had to invest enormously to keep up with digitization. In Flanders, they did this by launching a five-year plan, called "Vlaanderen Radicaal Digitaal" (VRD), which runs until 2024. An already great result of the VRD program is called *Mijn Burgerprofiel*, a smart digital assistant, which aims to provide access to multiple public services through a single entry point [2]. The ultimate goal of the VRD initiative is for the Flemish government to shift towards a digital and information-driven government.

As data has become a key pillar in the European digital economy, Europe has come up with a strategy to unlock the full potential of data, which should lead to the rise of European data spaces. Europe defines a data space as a "decentralized infrastructure for trustworthy data sharing and exchange in data ecosystems based on commonly agreed principles". In their strategy, they released the first set of guidelines, collectively called the *Data Governance Act*, which aim to facilitate data sharing across member states and between different sectors [5, 6]. An essential aspect of this architecture is *interoperability* because this will be necessary for an efficient and sustainable exchange of data. Figure 1 shows the building blocks that are part of a European Data Space, with a *Vocabulary Provider* building block in the top-right corner, whose task it is to provide data standards, so that data can flow between the various components in the data space.

In Flanders, the question of Vocabulary Provider is already answered with the Flemish Interoperability program, called Open Standards for Linked Organizations (OSLO). OSLO was aligned with the European Interoperability Framework [2] (EIF), which acts as a set of guidelines focused on the interoperability of digital public services. This framework introduces four layers of interoperability: legal, organizational, semantic and technical. The legal layer refers to aligned legislation between different organizations and policy areas. The second layer, organizational, refers to aligned business processes between different public administrations. The semantic interoperability layers refers to the definition of data content. For example, when two information systems are semantically interoperable, the meaning of the exchanged data is preserved and understood between the systems. The last layer, technical interoperability, aims at the different protocols, formats and interfaces to exchange data [3]. The main focus of OSLO is on the latter two interoperability layers. Take the smart digital assistant example from above, by using OSLO, semantic and technical interoperability was ensured, making it possible to easily integrate the information systems from different administrations [2].

However, by focusing on the semantic and technical layers, the semantic agreements made by OSLO are locked on the data level, although they can provide added value at other levels as well. This resulted in the following research question: "**How can we support certain target groups, e.g policy-makers, at the organizational and legal level with agreements made at the semantic level?**".

In this paper, we aim to close this gap by making semantic assets generated by OSLO, available to target audiences at the organizational and legal layers by two tools.

---

[1]https://github.com/International-Data-Spaces-Association/idsa
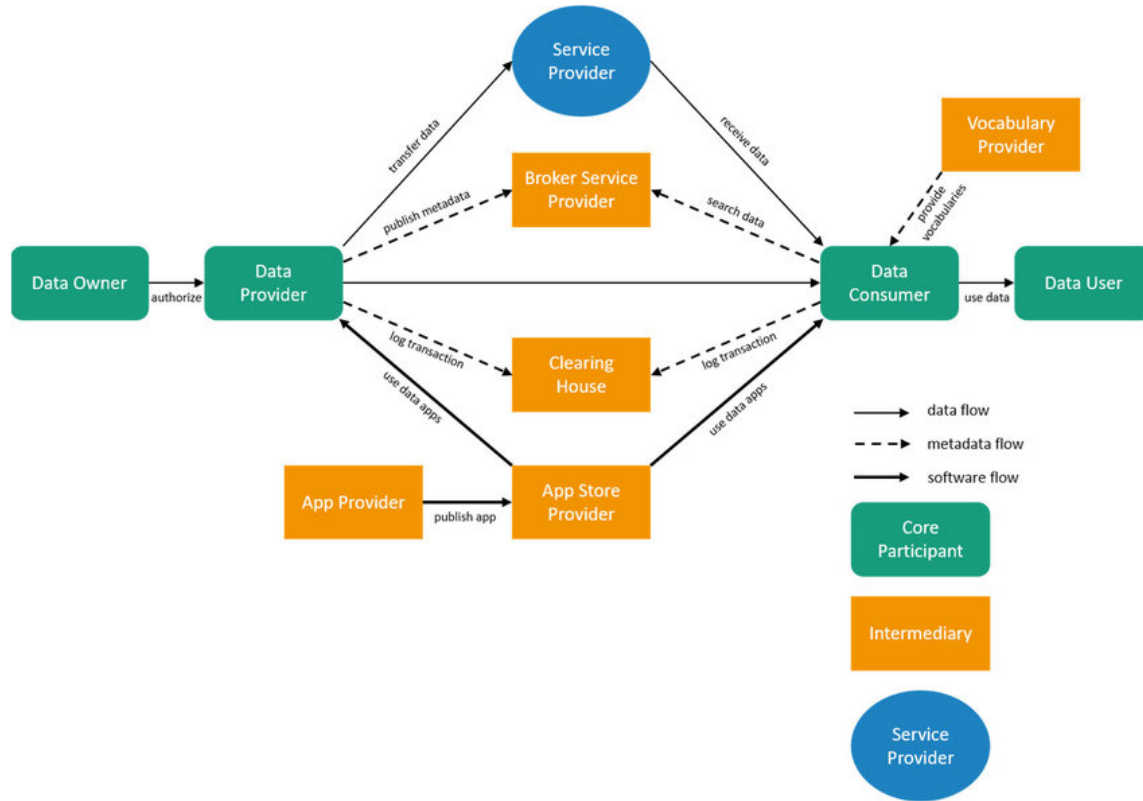[2]https://ec.europa.eu/isa2/sites/default/files/eif_brochure_final.pdf

Fig. 1. Overview of the architecture of an European Data Space[1]

2 BACKGROUND

At the time of writing, Flanders has 300 local governments which provide over 800 different products and services, often using applications from various software vendors. Data coming from those different systems often have a different data model, are in a different format, and do not necessarily have the same semantics. This results in the creation of vertical data silos and makes it challenging to exchange high-quality data. To resolve these interoperability problems, in 2012, an interoperability initiative was founded, called *Open Standards for Linked Organizations* (OSLO) [1]. Since then, OSLO has facilitated the co-creation of data standards in more than 100 domains, capturing over 4000 definitions. The realization of data standards in Flanders is done by following a well-defined process and methodology, called "OSLO Process & Method" [3]. The OSLO process is aligned with international best practices of ISA, W3C, and OpenStand, ensuring qualitative data standards and a governance structure. Through various workshops together with domain experts, a domain model is established as a UML (Unified Modelling Language) class diagram. Using UML to create the data model, a view is provided of which relations exist between various objects in the domain, and how data should be exchanged in general. Based on this UML data model, a vocabulary is generated in both a human and machine-readable format, by re-using the principles of Linked Data [4].

---

[3]https://data.vlaanderen.be/cms/Proces_en_methode_voor_de_erkenning_van_datastandaarden_v1.0.pdf
[4]https://www.w3.org/DesignIssues/LinkedData.html

Linked Data is defined as a set of design principles to share machine-readable information on the Web, as defined by the founder of the World Wide Web, Tim Berners-Lee in 2006. The first design principle is about identifying things and states that unique resource identifiers (URIs) should be used as names for things. The second principle expands the first by stating that HTTP URIs should be used, so that users can look up those names. The third principle is about information coupled to the HTTP URIs. When a user enters an HTTP URI in their browser, they should get useful information in return. That information should be described by using a standardized format, called the Resource Description Framework (RDF). RDF is a directed graph that contains statements. These RDF statements consist of three parts, a subject, a predicate, and an object, and are therefore also called a triple. For example: "John knows Jane", where John is the subject, "knows" the predicate, and Jane the object. Each part in the triple is represented by a URI, whereas the object can also be a literal value (e.g. string, integer, …). The last design principle is about linking to other data by including HTTP URIs to other entities in your data, allowing users to discover more information on the Web.

All properties and classes on an OSLO UML data model are assigned URIs, uniformly identifying each term within the domain model. These URIs are not assigned at random, but by aligning with already existing data standards. For example, the term representing a person was assigned the URI http://www.w3.org/ns/person#Person, re-using the URI and definition the term "person" was assigned by W3C. In some cases, no URI exists to assign to the term, as their definition differs too much, and in that case, a new URI is created. For example, in Flanders, to exchange data about a person, there is a mandatory property that contains the full name of a person. However, in international standards, this does not exist, and therefore a new URI was created: https://data.vlaanderen.be/ns/persoon#volledigeNaam.

3 ARCHITECTURE

OSLO targets the semantic and technical level, meaning the semantic agreements are mainly used to add context to data. Take for example the following code sample:

Listing 1. JSON data snippet

```
{
    "place": "Brussels"
}
```

Listing 2. JSON-LD data snippet

```
{
    "@context" : {
        "place": "https://schema.org/location"
    },
    "place": "Brussels"
}
```

Listing 1 shows a JSON object containing a property "place" with value "Brussels". However, one could argue about the definition of the property "place". This problem is solved by listing 2. This code snippet is a Linked Data serialization, called JSON-LD, and adds semantics to the data by assigning URIs, with "place" being assigned the URI

"https://schema.org/location" in this example. Users can enter that URI in their browser to discover which definition is attached to this URI and by extension the property "place".

As Linked Data is the solution to add semantics to data and resolve ambiguities and misinterpretations, it is only used at the data level. People at the legal and organizational levels almost never come in contact with these semantic agreements, although they can provide added value to legal documents for example.

3.1 OSLO Lookup

In order to make the semantic agreements available to target groups at the organizational and legal level, we created a simple user interface, called *OSLO Lookup*, which allows users to enter multiple keywords and check if they were defined within OSLO. The code of OSLO Lookup is publicly available on GitHub [5].

Although this is an excellent first step toward semantic agreements being used at the organizational and legal levels, constantly using the OSLO Lookup functionality to check if a word exists in OSLO can be pretty cumbersome. Therefore, we looked for a way in which we could make the semantic agreements available within the work environment. This was achieved by developing an add-in for Microsoft Word, called *OSLO365*.
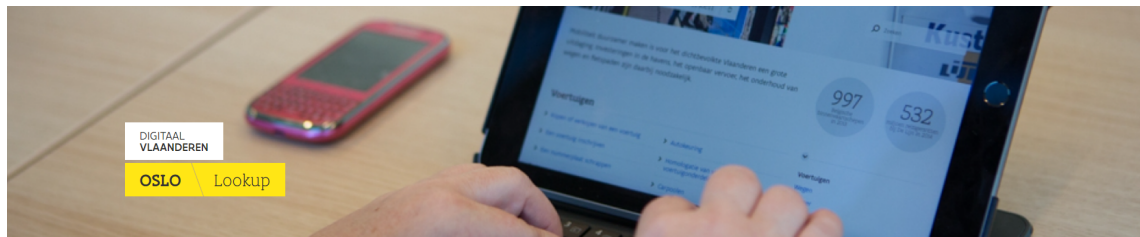
Fig. 2. Overview of the UI of OSLO Lookup

3.2 OSLO Lookup integrated in OSLO365

The Office Word add-in has multiple functions to assist the user in adding semantics to their document. Firstly, we integrated the OSLO lookup functionality in the add-in, allowing users to enter a keyword in a search field and relevant results will be shown to the user, as shown on figure 3. To make it even more user-friendly, users can also select a word in the document and the add-in will automatically look up the selected word. To actually add semantics to their documents, users can add the definition of a word along with their URI as a foot or endnote to the document, as shown in figure 4. By embedding the URI and its definition in the document, other readers or contributors do not have to deal with ambiguities anymore.

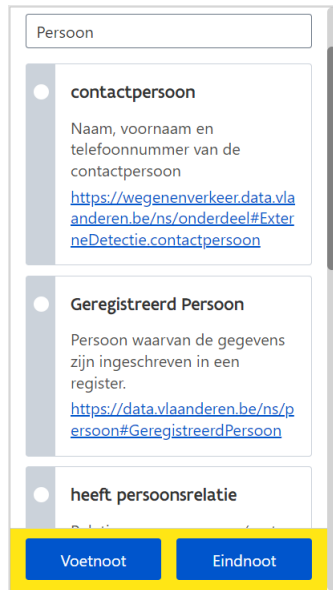[5]https://github.com/informatievlaanderen/OSLO-Lookup

Fig. 3. The OSLO add-in shows all the results for a keyword or a selected word in the document. In this example, results are shown for the keyword "Persoon" (Person in English)

Fig. 4. The OSLO add-in allows to embed an OSLO definition in the Word document as a foot or endnote. In this example, a definition for the word "persoon" is embedded as an endnote in the document, alongside with its URI.

This functionality above is a great first way to add semantics to Word documents but can be time-consuming for large documents. Therefore, the add-in also has a function to scan the whole document. When done scanning, the add-in will guide the user through his Word document and indicate for which words it has results. The user can individually decide for each word what definition to insert into the document. The last functionality of this add-in is a dictionary. As a user, you want to easily find a commonly used definition. Therefore, users can add definitions to their dictionary making it immediately available if they click on the dictionary. Figure 6 shows that users can access their dictionary to check which definitions are saved, insert them as a foot or endnote or remove them if necessary. On top of this, if a user searches a word or lets the add-in scan the document, definitions added to the users his dictionary will be marked in yellow and shown as the first result, as shown in figure 5. The code of the OSLO365 plugin is publicly available on GitHub [6].

4 RELEVANCE AND USABILITY ASSESSMENT

To assess the relevance and usability of the developed OSLO365 plug-in, we took an explorative ethnographic research approach. This approach involved interviewing multiple local servants at an administration of a local mid-sized municipality in Flanders. Using a semi-structured interview guide, we talked to six public servants from different departments with distinct roles and responsibilities, asking them to perform their daily tasks and to define specific terms or concepts using the plug-in. The "think aloud" protocol [4] enabled us to identify information that is concentrated on during problem-solving as well as information used to facilitate problem resolution. When asked to define a particular concept (e.g., 'Gemeente' [Municipality]), most respondents interacted with the plug-in thesaurus function - expecting a

[6]https://github.com/informatievlaanderen/OSLO365-plugin
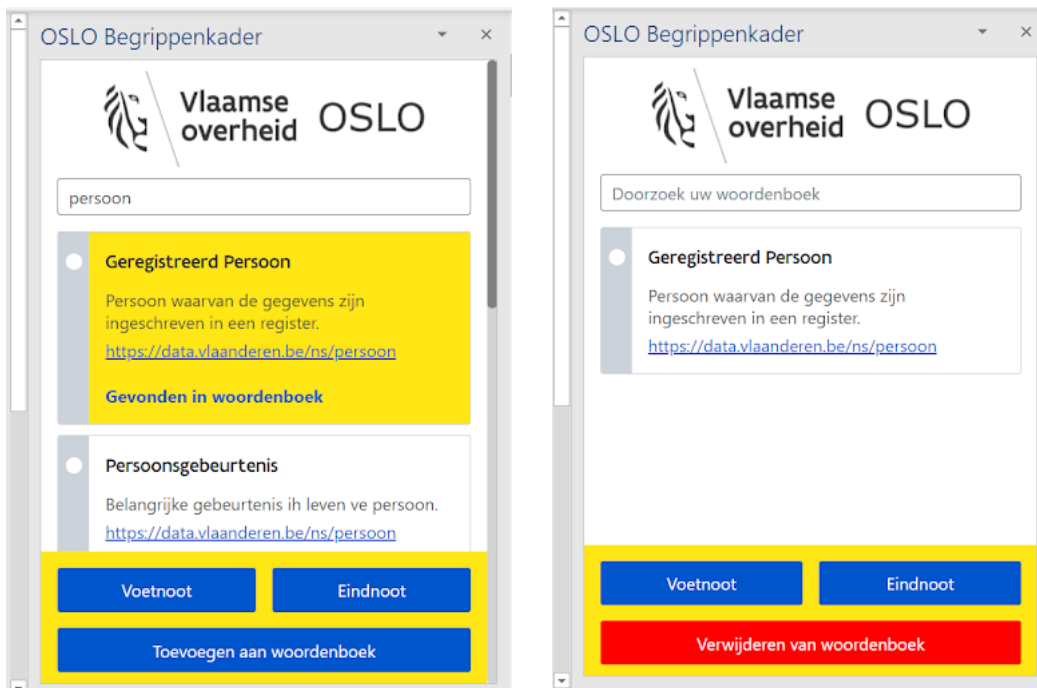
Manuscript submitted to ACM

Fig. 5. Definitions saved in the dicionary are marked in yellow when they appear in the search results

Fig. 6. Users can consult their dictionary and remove a definition from their dictionary

full dictionary containing all OSLO vocabulary terms - rather than the therefore intended search function. The thesaurus function was generally well-received, with most civil servants finding it helpful in defining frequently used terms or concepts. The search functionality was also met with positive feedback, with respondents stating that it felt natural and intuitive to use. The functionality 'document scan' was the most criticized affordance; after scanning a document, the plug-in would only allow public servants to navigate through a list of matched terms by clicking the buttons 'previous' or 'next', with one public servant stating: "I am flabbergasted that such a useful feature as the document scan is implemented in such a non-efficient way.". The functionality however also received praise as it was identified as a useful feature to make people aware of how ambiguous and unclear their day-to-day jargon could be for others outside of their domain. In general, respondents told us that usability is key and that the plug-in should be supported in the whole administration of the municipality for them to use it individually in their daily tasks. The plug-in proved to be relatively easy to use, as no respondent had any particular issues when working with it.

5 DISCUSSION AND FUTURE WORK

To increase interoperability, Flanders has invested intensely in the OSLO program, with OSLO targeting the semantic and technical interoperability layer, which provides a great added value at the data level. However, for target audiences at the organizational and legal level, there is a barrier to use these semantic agreements, as they rarely come into contact with the actual data level. This results in the fact that the efforts made by OSLO often stay at the semantic and technical interoperability layers, although they can provide an added value at the other layers as well.

In this paper, we introduced two tools in order to smooth the flow of semantic agreements to the legal and organizational level. With OSLO Lookup, we show that we can easily provide an application to unlock the semantic agreements at the data level and make them available at other levels as well. With the Microsoft add-in, OSLO365, we make the semantic agreements available for Microsoft Word users, who can look up semantic agreements and incorporate them in their Word documents. Doing this will decrease the number of ambiguities or misinterpretations by readers of these documents. The interviews with the public servants of a local administration demonstrate that these tools allow semantic assets to be used at the organizational and legal levels, and therefore answers our research question. The interviewees also mentioned that in order to move away from their known and trusted working patterns, it is essential for them to be made aware of the benefits that such a move could provide for them. This will be a key working point for OSLO in the future.

As OSLO is a Flemish initiative and all semantic assets have their definitions written in Dutch, which makes both OSLO Lookup and the Word add-in only usable in a Dutch context. Nevertheless, the code of these tools is publicly available on GitHub, so that it can be a source of inspiration for initiatives in other languages.

In addition to assigning URIs to terms in a UML data model, OSLO also uses URIs to uniquely identify data objects. For example, each address in Flanders has a unique HTTP URI. Using the Linked Data principles, entering this HTTP URI in the browser, useful information about the address will be returned. The useful information can be both human and machine-readable and is compliant with the OSLO data model for addresses. In the future, we will be doing research on how we can evolve to evidence-based documents by also making these data URIs available for the tools. For example, if someone enters an address in their document, the add-in should be able to retrieve the HTTP URI of that address and add it to the document. What remains to be developed then, is the organizational aspect of implementing such a tool within a public administration.

6   ACKNOWLEDGMENTS

REFERENCES

[1] Raf Buyle, Laurens De Vocht, Mathias Van Compernolle, Dieter De Paepe, Ruben Verborgh, Ziggy Vanlishout, Björn De Vidts, Peter Mechant, and Erik Mannens. 2016. OSLO: Open standards for linked organizations. In *Proceedings of the international conference on electronic governance and open society: Challenges in Eurasia.* 126–134.

[2] Raf Buyle, Mathias Van Compernolle, Dieter De Paepe, Jens Scheerlinck, Peter Mechant, Erik Mannens, and Ziggy Vanlishout. 2018. Semantics in the wild: a digital assistant for Flemish citizens. In *Proceedings of the 11th International Conference on Theory and Practice of Electronic Governance.* 1–6.

[3] European Commission and Directorate-General for Informatics. 2017. *New European interoperability framework : promoting seamless services and data flows for European public administrations.* Publications Office.   https://doi.org/doi/10.2799/360327

[4] Riitta Jääskeläinen. 2010. Think-aloud protocol. *Handbook of translation studies* 1 (2010), 371–374.

[5] Hendrik Mildebrath. 2021. Data Governance Act.   https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/690674/EPRS_BRI(2021)690674_EN.pdf

[6] Lars Nagel and Douwe Lycklama. 2021. Design Principles for Data Spaces - Position Paper.   https://doi.org/10.5281/zenodo.5105744

[7] Klaus Schwab. 2017. *The fourth industrial revolution.* Currency.