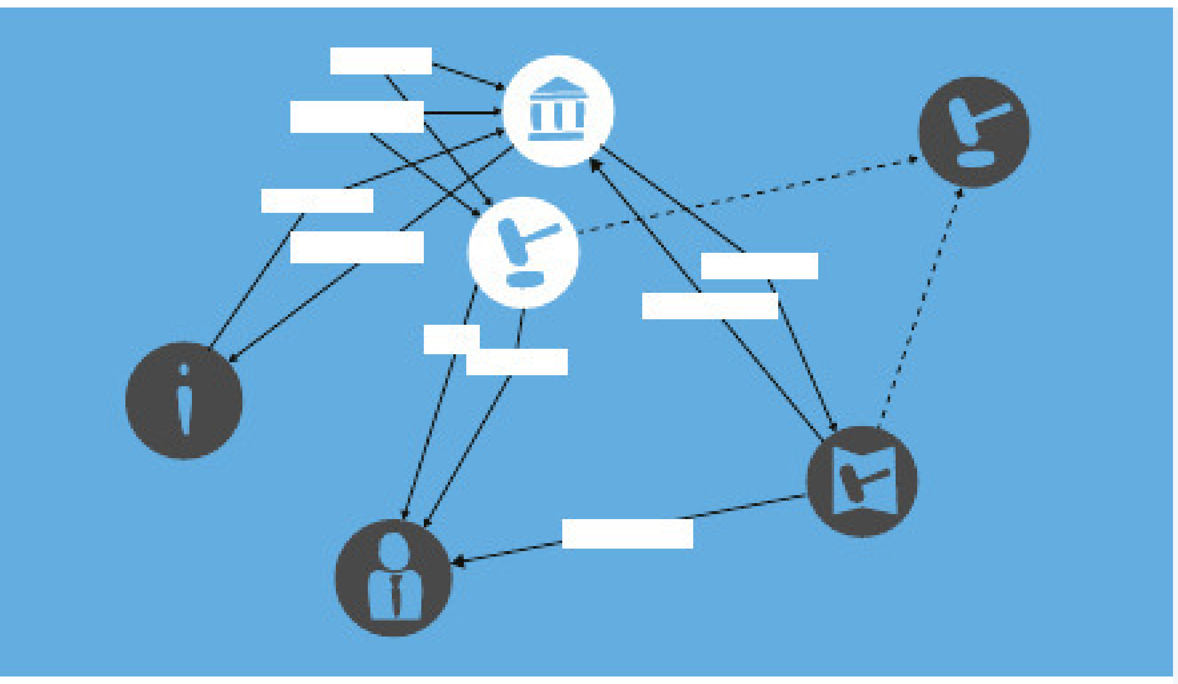
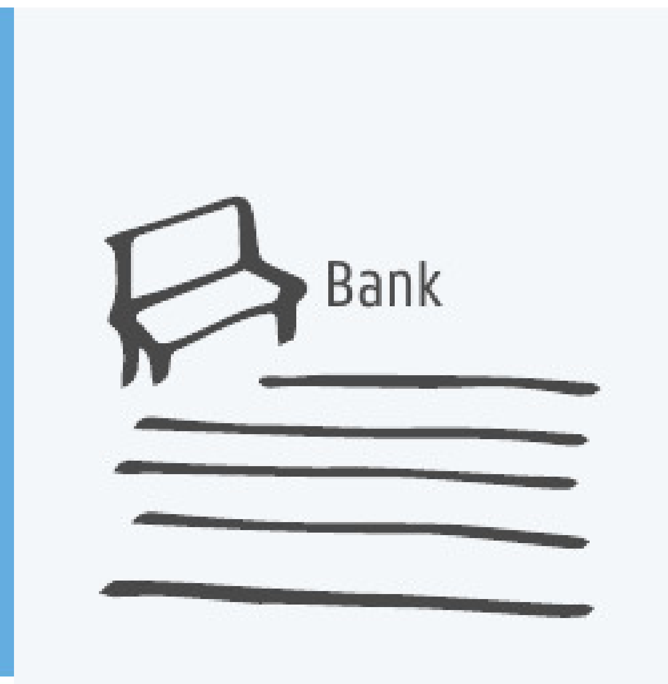
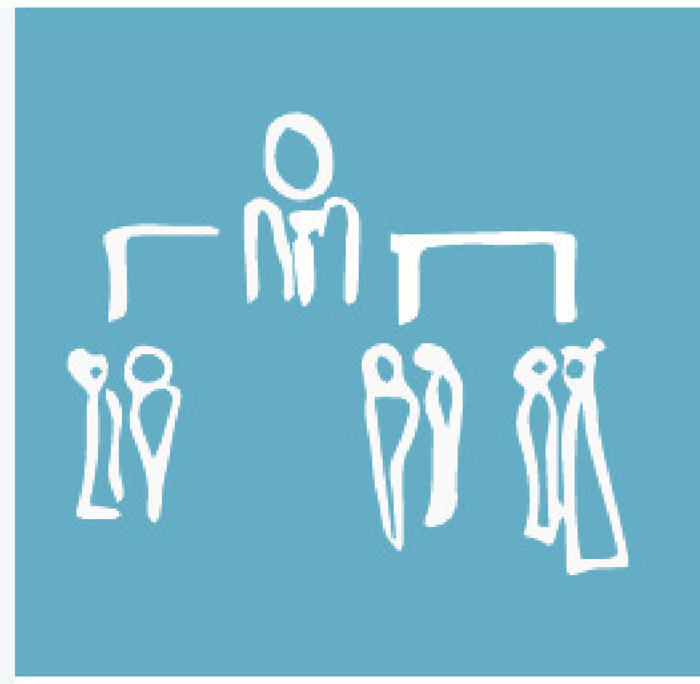
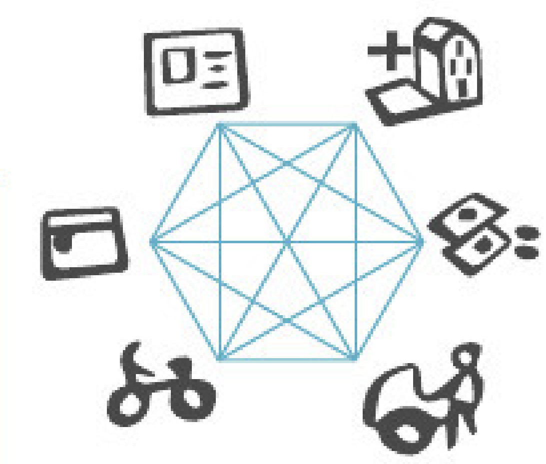
FACULTY OF ENGINEERING
AND ARCHITECTURE







https://data.vlaanderen.be/ns/openbaar-domein/infrastructuur#Zitbank

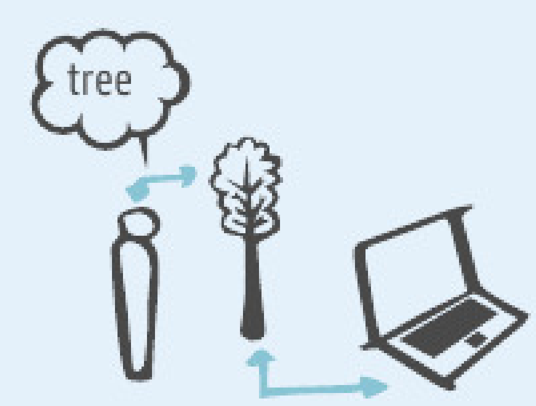


Raising Semantic and Technical Interoperability in the Public Sector

Raf Buyle

https://data.vlaanderen.be/ns/openbaar-domein/infrastructuur#fietsdelen

Doctoral dissertation submitted to obtain the academic degree of
Doctor of Information Engineering Technology



https://data.vlaanderen.be/ns/openbaar-domein/vegetatie#Boomvorm

Supervisors
Prof. Erik Mannens, PhD* - Peter Mechant, PhD**

* Department of Electronics and Information Systems
Faculty of Engineering and Architecture, Ghent University
** Department of Communication Sciences
Faculty of Political and Social Sciences, Ghent University

July 2021

GHENT
UNIVERSITY

FACULTY OF ENGINEERING
AND ARCHITECTURE

Raising Semantic and Technical Interoperability in the Public Sector

Raf Buyle

Doctoral dissertation submitted to obtain the academic degree of
Doctor of Information Engineering Technology

Supervisors
Prof. Erik Mannens, PhD* - Peter Mechant, PhD**

* Department of Electronics and Information Systems
Faculty of Engineering and Architecture, Ghent University
** Department of Communication Sciences
Faculty of Political and Social Sciences, Ghent University
July 2021

GHENT
UNIVERSITY

ISBN 978-94-6355-504-3
NUR 982, 986
Wettelijk depot: D/2021/10.500/52

Members of the Examination Board

Chair

Prof. Patrick De Baets, PhD, Ghent University

Other members entitled to vote

Pieter Colpaert, PhD, Ghent University
Prof. Lieselot Danneels, PhD, Ghent University
Prof. Lieven De Marez, PhD, Ghent University
Francesca Gleria, PhD, Provincia autonoma di Trento, Italy
Prof. Seth van Hooland, PhD, Université libre de Bruxelles & Europese Commissie

Supervisors

Prof. Erik Mannens, PhD, Ghent University
Peter Mechant, PhD, Ghent University

PREFACE

Find a group of people who challenge and inspire you,
spend a lot of time with them, and it will change your life...

— Amy Poehler

If someone asked me to create my ideal business card, it would mention '*Chief Glue Officer*', as my core competence is to build bridges between people, ideas and open knowledge. My dissertation is the result of collaborating with stunning academic researchers and major practitioners in the public sector. It builds on previous discoveries in the domain of e-government. Therefore I would like to quote Sir Isaac Newton; "*If I have seen further it is by standing on the shoulders of Giants".*

The previous thesis I wrote dates back to 2000 and addressed the development of state-of-the-art test-equipment for digital video broadcasting over satellite. My master thesis was at the crossroads of electronics, hardware description languages and embedded software. It revealed my passion for technology which I developed since I was seven years old and dismantled various old television sets. I kicked off my professional career as an R&D engineer in the domain of biometrics, designing a retina scanner.

In 2002 I took my first steps in the exciting world of e-government, designing geographical information systems. I worked in various roles in an ICT consulting company, including as software engineer as well as general director. The role that I was really fond of was e-government advisor, as I had the feeling I was able to contribute to a better and more transparent society. In this adventure, I initiated a project in 2009 together with ir. Filip Meuris and ir. Luk Van Beneden to standardise and exchange Master Data of citizens and organisations, which we applied in several cities in Flanders.

Although I had already developed a passion for data standardisation in the public sector, it was Filip Meuris from the Intercommunale Leiedal who really fired-up my e-government passion during a Smart City congress in Cambridge in 2012. Later that year, I joined the Flemish ICT organisation, an umbrella organisation for ICT practitioners in the public sector. Together with Eddy Van der Stock, we bootstrapped a bottom-up crowd-funded initiative to standardise public-sector information; 'Open Standards for Linked Organisations' (OSLO). The sponsors where local governments, the Flemish Government, and ICT service suppliers. The Internet technology and Data Science Lab (IDLab) was a beacon fire from day one in the OSLO programme.

In 2015 I joined ‘Digitaal Vlaanderen’, where I continued my work on standardisation and supported the incubation of state-of-the-art projects that accelerate the digital transformation. Since 2015, I'm the chair of the OSLO interoperability programme in the region of Flanders, coördinating a vibrant community of over 400 authors collaborating on Open Data Standards. Also, I am a co-founder of the Belgian interfederal interoperability initiative in 2019.

While working on these exciting projects and being actively involved in Open Knowledge Belgium, I met prof. dr. Erik Mannens and dr. Peter Mechant. Erik asked me to start a joint PhD at Internet

technology and Data science Lab (IDLab) and the Research Group for Media, Innovation and Communication (MICT), which was the start of a stunning adventure. I have combined the role of researcher with my standardisation activities in the public sector in Flanders. As a privileged observer, analyst and critic, I gathered my data via action research. This refers to the fact that I was involved as co-practitioner in the setting under study and thus combined theory and practice. I had the opportunity to broaden my local and international network, and even had the opportunity to write a paper with the inventor of the Web; prof. dr. Sir Tim Berners-Lee. Many researchers and major e-government practitioners at the local, regional, interfederal and European governments have — knowingly and unknowingly — contributed to my dissertation, for which I am incredibly thankful.

Now comes the most complex part, being grateful without becoming too sentimental and most importantly, without forgetting someone. First and foremost, my research would not have been possible without the inspiration, support and backup of my supervisors professor. dr. Erik Mannens and dr. Peter Mechant. Erik's presence is motivating, and his enthusiasm for the work our team is doing is exceptional. Peter's academic writing skills are extraordinary, his eye for detail and substantiated feedback was crucial throughout my entire PhD journey. Thanks to this setting of a joint PhD and my promotors, I had the opportunity to combine IDLab's state-of-the-art technicality with the core competence of MICT; namely gathering scientific insights on how people interact with technology.

To all members of the jury: prof. dr. Patrick De Baets, prof. dr. Lieven De Marez, prof. dr. Lieselot Danneels, prof. dr. Seth van Hooland dr. Francesca Gleria and dr. Pieter Colpaert, I wish to express my sincerest gratitude for thoroughly reviewing the work and results presented in this dissertation.

Further thanks to all my colleagues at IDLab-KNoWS, MICT and Digitaal Vlaanderen who were the emulsifier to make the e-government mayonnaise.

Also thank you to the many people I had the pleasure to research and publish with and who have become too numerous to name, but here’s a tip to the hat to dr. Bert Van Nuffelen, Björn De Vidts, Brecht Van de Vyvere, Dieter De Paepe, Dwight Van Lancker, Geert Thijs, Geraldine Nolf, Geroen Joris, Jens Scheerlinck, Katrien Mostaert, dr. Philippe Michiels, dr. Ruben Taelman, prof. dr. Ruben Verborgh, prof. dr. Serena Coetzee, Stefan Lefever, prof. dr. Sir Tim Berners-Lee, Veerle Beyaert, Veronique Volders and Ziggy Vanlishout.

I am very grateful to dr. Laurens De Vocht, who guided me in the first steps of my PhD adventure. Also, I would like to thank dr. Esther De Loof for proofreading significant parts of my dissertation.

My profound gratitude also goes out to my fellow researchers; dr. Pieter Colpaert, Mathias Van Compernolle and Eveline Vlassenroot for their support, inspiration, company, and the occasional beer. Also, I would like to thank Juanita Devis for designing a stunning personalised cover for this dissertation.

I want to thank my family for their support and endless patience while living with a midlife PhD student, more specific during the times I was around, but my thoughts were not. Thanks to my mom Vera and my dad Roger for their lasting support during my entire career. Finally, the last and most important words of thanks for the moral support and love all these years goes to the most fantastic and beautiful woman in my life; Nadja and our children Charlotte and Joana who supported me along the entire way of this extraordinary journey.

Raf Buyle

Ghent, February 1, 2021

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

API	Application Programming Interface
BAG	Basic registry of Addresses and Buildings
CEN	European Committee for Standardisation
CIM	Context Information Management
CPU	Central Processing Unit
CRAB	Central Reference Address Registry
EC	European Commission
EIF	European Interoperability Framework
EPC	Energy Performance Certificate
ETSI	European Telecommunications Standards Institute
FIWARE	Future Internet Ware
FOAF	Friend of a Friend
GIPOD	Generic Information Platform for the Public Domain
GIS	Geographic Information System
GDPR	General Data Protection Regulation
GPS	Global Positioning System
GUID	Globally Unique IDentifier
HATEOAS	Hypermedia As The Engine Of Application State
HTML	HyperText Markup Language
HTTP	HyperText Transfer Protocol
ICT	Information and Communication Technology
IDLab	Internet technology and Data science Lab
IS	Information Systems
HIS	Healthcare Information Systems
ISG	Industry Specification Group
IDT	Innovation Diffusion Theory
IMS	Information Model Structure
INSPIRE	INfrastructure for SPatial InfoRmation in the EC
IoT	Internet of Things
ISA	Interoperability Solutions for European public Administrations

JSON	JavaScript Object Notation
JSON-LD	JavaScript Object Notation for Linked Data
LBLOD	Local Council Decisions as Linked Data
LEZ	Low-Emission Zone
LDF	Linked Data Fragments
LOV	Linked Open Vocabularies
LRD	Large-scale Reference Database
LTS	Linked Time Series
MaaS	Mobility as a Service
MiB	MebiByte
NBI	NorthBound Interfaces
NGSI	Next Generation Service Interfaces
NGSI-LD	Next Generation Service Interfaces as Linked Data
NPO	Nonprofit organization
NSS	Node Solid Server
OGC	Open Geospatial Consortium
OSLO	Open Standards for Linked Organisations
OWA	Open World Assumption
PSI	Public Sector Information
REST	Representational State Transfer
RDF	Resource Description Framework
RSD	Road Sign Database
SBI	SouthBound Interfaces
SDW	Spatial Data on the Web
SHACL	Shapes Constraint Language
SKOS	Simple Knowledge Organization System vocabulary
SOAP	Simple Object Access Protocol
SOSA	Sensor, Observation, Sample, and Actuator
SPARQL	SPARQL Protocol and RDF Query Language
SQL	Structured Query Language
SSN	Semantic Sensor Network ontology
SSO	Single Sign-On
StUF	Standaard Uitwisseling Formaat
TAM	Technology Acceptance Model
TRI	Technology Readiness Index
TPB	Theory of Planned Behaviour
TR	Technology Readiness
TRAM	Technology Readiness and Acceptance Model
UML	Unified Modeling Language
URI	Universal Resource Identifier

URL	Uniform Resource Locators
VMM	Flanders Environment Agency
WHO	World Health Organization
WFS	Web Feature Service
WKT	Well-Known Text
WWW	World Wide Web
W3C	World Wide Web Consortium
XML	Extensible Markup Language

SUMMARY

Governments in Flanders provide over eight hundred public services in different domains such as building permits, subsidies, public welfare, and daycare. The back-office processes and service delivery of these services are supported by specialised Information Systems (IS) from different software vendors. Because the data in these IS is modelled from a single thematic perspective, it is difficult or impossible to share and reuse them among all services. This causes unnecessary frustrations in everyday life of both citizens and businesses as they are required to provide the same information more than once to their government. The smart use of available information about citizens by public administrations is referred to as the once-only principle. Also, the transformation of society towards a digital economy is leading to changing roles where barriers between public and private actors are blurring. This happens in a context where information and IS are combined with new technologies such as live data from (mostly mobile) physical devices.

The holy grail in e-government — that all IS can exchange information without bespoke integration efforts — is the topic of the majority of the debates in the public sector. 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. We notice two primary drivers for

interoperability. Firstly, citizens and entrepreneurs expect a coherent user experience from their government as they became accustomed to by services in the private sector. Secondly, our digital economy embraces new ecosystems, including 'the government as a platform' where business expects public-private partnerships to arise. But governments keep struggling to deliver integrated, interconnected, and cross-sectoral services due to sectoral specialisation or 'departmentalisation'. Data and information in the region of Flanders in Belgium are fragmented across 300 municipalities, the regional administration, the federal administration and the private sector. To raise interoperability and facilitate innovation among these actors, robust, coherent, and universally applicable data standards are essential.

For centuries, standards have been fueling innovation, catalysing the growth of markets and protecting the health and safety of citizens. In the sixteenth century, nuts and bolts were hand-crafted in matching pairs. In 1800, Henry Maudslay invented the screw-cutting lathe, which allowed to produce screws with standardised thread. As they became interchangeable, it was possible to create interchangeable machine parts which enabled the Industrial Revolution. This turning point can be compared to the invention of the World Wide Web (WWW) by Tim Berners-Lee. Alike any nut and bolt which adheres to the standards can be put together, electronic documents that are formatted in HyperText Markup Language (HTML) and transferred using the HTTP (HyperText Transfer Protocol) can be exchanged via the Web. This created a digital revolution with new forms for social and economic enterprises and a new scope and efficiency for markets. Throughout this dissertation, we research how the building blocks of the WWW contribute to raise interoperability and provide better public services. The WWW interconnects documents and services and runs on top of the Internet.

In 2013, the Open Standards for Linked Organisations (OSLO) initiative raised awareness for interoperability, but it faced barriers at the technical and semantic level as there was no end-2-end approach for implementing data standards for both public and private partners. Therefore, our research mainly targets semantic and technical interoperability. Technical interoperability refers to the technical issues to link IS, while semantic interoperability refers to the meaning and structure of information.

The goal of this PhD is to study 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. We research this challenge both from the technical and political point of view in the context of e-government in the region of Flanders in Belgium. Our 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 within a public sector context. Our method allows datasets to be linked into a public sector knowledge graph governed by a public body.

We research the implementation of a Linked Base Registry for Addresses by unfolding the process followed towards raising semantic interoperability based on the Linked Data principles. While implementing the address vocabulary, we stumble on competing international semantic standards and difficult choices on how to extend them to fit the local context. To *enable business analysts and developers to maintain semantic agreements* and to cope with these challenges, it is crucial to have a transparent process and methodology for developing semantic agreements and a governance structure for making and institutionalising pivotal decisions in place. This can be realised through a policy framework for technical as well as domain-specific topics, alike to the OSLO programme in Flanders. This policy also includes a URI standard for *persistent unambiguous identifiers* that supports government administrations by providing guidance that ensures that HTTP URIs are future proof. The results show that the method of Linked Data can indeed increase semantic and technical interoperability and can lead to better adoption of data, such as addresses, in the public and private sector. Also, we delve into technical interoperability and research what methods are suited for publishing Open Data time series in a sustainable, predictable, and cost-effective way in the context of Smart Cities. We analyse the REST architectural style that resembles the human Web, which builds upon hyperlinks, and a set of architectural constraints that facilitate architectural elasticity. The *uniform interface* simplifies the architecture and empowers software clients to evolve separately. Also, as all messages are self-descriptive and responses include links to possible actions, they can be *easily interpreted by (non-human) clients, making out-of-band documentation needless*.

We outline the process and method for developing Open Data Standards . Also, we address the interoperability hurdles at the different governmental levels and examine how the OSLO programme tackled these obstacles. We have determined that both the *bottom-up and top-down approach* were vital to create the *necessary political support at the different governmental levels*. The bottom-up approach, where different government levels and private partners are collaborating on interoperability, is crucial for *building consensus*. This bottom-up process is combined with formal top-down governance. Our results show that it is possible to reach semantic agreements and overcome the political hurdles within an operational public sector context. Throughout this thesis we demonstrate that the design principles of the Semantic Web can facilitate interoperability within the public sector by adding context and useful links, using the Resource Description Framework (RDF) as a data model.

Also, we demonstrated that 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.

Our researched process and method for raising semantic and technical interoperability within an operational public sector context were embraced by the 'Steering Committee of Flemish Information and ICT-policy' in 2018. This government-embraced governance raised the development and adoption of data standards in the region of Flanders. At the time of writing this dissertation, the number of ratified data standards by the Flemish Government has grown to more than ninety. The community has grown to over four hundred authors from the public sector, private sector, and academia, all collaborating on the development of data standards. As of 2019 our process and method was also adopted on the Belgian interfederal level.

SAMENVATTING

Overheden in Vlaanderen bieden gemiddeld meer dan achthonderd publieke dienstverleningen aan in verschillende domeinen zoals bouwvergunningen, subsidies, welzijnszorg en kinderopvang. De backoffice processen en de dienstverlening van deze diensten worden ondersteund door gespecialiseerde Informatiesystemen (IS) van verschillende softwareleveranciers. Omdat de gegevens in deze IS vanuit één enkel thematisch perspectief zijn gemodelleerd, is het moeilijk of onmogelijk om ze te delen en te hergebruiken tussen de verschillende diensten. Dit leidt tot onnodige frustraties in het dagelijkse leven van zowel burgers als bedrijven omdat zij dezelfde informatie meer dan eens aan hun overheid moeten verstrekken. Het slimme gebruik van beschikbare informatie over burgers door overheidsdiensten wordt het ‘once-only principe’ genoemd (vraag niet wat je al weet). De transformatie van de samenleving zet de overgang in naar een digitale economie met veranderende rollen waarbij de grenzen tussen publieke en private actoren vervagen. Dit gebeurt in een context waarin informatie en informatiesystemen worden gecombineerd met nieuwe technologieën zoals live data van (overwegend mobiele) fysieke apparaten.

De heilige graal in e-government, dat alle IS informatie kunnen uitwisselen zonder integratie-inspanningen op maat, is het onderwerp van het merendeel van de debatten in de publieke sector.

Interoperabiliteit is het vermogen van organisaties om informatie en kennis te delen, over informatiesystemen en bedrijfsprocessen heen. Er zijn twee primaire drijfveren voor interoperabiliteit. Ten eerste verwachten burgers en ondernemers een coherente gebruikerservaring van hun overheid, zoals zij dat gewend zijn van diensten in de private sector. Ten tweede omarmt onze digitale economie nieuwe ecosystemen, waaronder ’de overheid als een platform’, waar bedrijven verwachten dat publiek-private partnerschappen zullen ontstaan. Maar overheden blijven worstelen om geïntegreerde, onderling verbonden en sectoroverschrijdende diensten te leveren als gevolg van sectorale specialisatie of ‘departementalisering’ (verkokering). Gegevens en informatie in het Vlaamse Gewest in België zijn versnipperd over 300 gemeenten, de gewestelijke administratie, de federale administratie en de private sector. Om de interoperabiliteit te verhogen en innovatie tussen deze actoren te vergemakkelijken, zijn robuuste, coherente en universeel toepasbare datastandaarden van essentieel belang.

Eeuwenlang hebben normen en standaarden innovatie aangewakkerd, de groei van de economie gekatalyseerd en de gezondheid en veiligheid van burgers beschermd. In de zestiende eeuw werden moeren en bouten met de hand gemaakt in passende paren. In 1800 vond Henry Maudslay de schroefdraaddraaibank uit, waardoor het mogelijk werd schroeven met gestandaardiseerde schroefdraad te produceren. Aangezien ze onderling uitwisselbaar werden, werd het mogelijk om verwisselbare machineonderdelen te maken, wat de industriële revolutie mogelijk maakte. Dit keerpunt kan worden vergeleken met de uitvinding van het World Wide Web (WWW) door Tim Berners-Lee. Net zoals elke moer en bout die aan de normen voldoet, in elkaar kan worden gezet, kunnen elektronische documenten die zijn geformatteerd in HyperText Markup Language (HTML) en worden verzonden met behulp van HTTP (HyperText Transfer Protocol), worden uitgewisseld via het Web. Hierdoor ontstond een digitale revolutie met nieuwe sociale en economische initiatieven. In deze dissertatie onderzoeken we hoe de bouwstenen van het WWW bijdragen aan het verhogen van de interoperabiliteit en het leveren van betere openbare diensten. Het WWW verbindt documenten en diensten met elkaar en is geïmplementeerd bovenop het Internet.

In 2013 heeft het OSLO initiatief (Open Standaarden voor Linkende Organisaties) het bewustzijn voor interoperabiliteit vergroot, maar het stuitte op belemmeringen op technisch en semantisch niveau, aangezien er geen end-to-end benadering was voor het implementeren van gegevensstandaarden voor zowel publieke als private partners. Daarom is ons onderzoek vooral gericht op technische en semantische interoperabiliteit. Technische interoperabiliteit heeft betrekking op de technische aspecten van het koppelen van IS, terwijl semantische interoperabiliteit betrekking heeft op de betekenis en structuur van informatie.

Het doel van dit doctoraat is te bestuderen welke processen (*acties om tot een resultaat te komen*) en methoden (*hoe deze acties te voltooien*) geschikt zijn om semantische en technische interoperabiliteit binnen een operationele overheidscontext te verhogen. We onderzoeken deze uitdaging zowel vanuit technisch als politiek oogpunt in de context van e-government in het Vlaamse Gewest in België. Onze aanpak combineert het *proces* om semantische overeenkomsten te bereiken door brede consensus en een end-to-end *methode* gebaseerd op de principes van gelinkte data om de semantische overeenkomsten binnen een publieke sector context consistent te houden. Onze methode maakt het mogelijk datasets te koppelen in een 'Vlaamse Knowledge Graph' voor de publieke sector die wordt beheerd door een overheidsinstantie.

We onderzoeken de implementatie van een gelinkt Basisregister voor Adressen door het proces om semantische interoperabiliteit te verhogen op basis van de Linked Data principes te ontleden. Tijdens de implementatie van het vocabularium voor adressen stuiten we op concurrerende internationale semantische standaarden en uitdagende keuzes over hoe deze uit te breiden en aan te passen aan de lokale context. Om *analisten en ontwikkelaars in staat* *te stellen semantische afspraken te handhaven* en deze uitdagingen het hoofd te bieden, is het van cruciaal belang te beschikken over een transparant proces en een transparante methodologie voor de ontwikkeling van semantische afspraken en over een governancestructuur voor het nemen en institutionaliseren van cruciale beslissingen. Dit kan gerealiseerd worden door een beleidskader voor zowel technische als domeinspecifieke onderwerpen, gelijkaardig aan het OSLO programma in Vlaanderen. Dit beleid omvat ook een URI-standaard voor *persistente eenduidige identificatoren* die publieke administraties ondersteunt door richtlijnen te bieden die ervoor zorgen dat HTTP URI's toekomstbestendig zijn. De resultaten tonen aan dat de methode van Linked Data inderdaad de semantische en technische interoperabiliteit kan verhogen en kan leiden tot een betere adoptie van gegevens, zoals adressen, in de publieke en private sector. Ook verdiepen we ons in technische interoperabiliteit en onderzoeken we welke methoden geschikt zijn om tijdreeksen van Open Data op een duurzame, voorspelbare en kosteneffectieve manier te publiceren in de context van Smart Cities. We analyseren de REST-architectuurstijl die lijkt op het menselijke Web, dat voortbouwt op hyperlinks, en een reeks architecturale beperkingen die architecturale elasticiteit vergemakkelijken. De *uniforme interface* vereenvoudigt de architectuur en stelt in staat zich afzonderlijk te ontwikkelen. Aangezien alle berichten zelfbeschrijvend zijn en de antwoorden links bevatten naar mogelijke acties, kunnen ze *gemakkelijk worden geïnterpreteerd door* (*niet-menselijke*) *clients, waardoor documentatie buiten de technische koppelingen overbodig wordt*.

Wij schetsen het proces en de methode voor de ontwikkeling van Open Data Standaarden. Ook gaan we in op de interoperabiliteitsbelemmeringen op de verschillende overheidsniveaus en onderzoeken we hoe het OSLO programma deze drempels heeft aangepakt. We hebben vastgesteld dat zowel *de bottom-up als de top-down benadering* van vitaal belang waren om *de nodige politieke steun op de verschillende overheidsniveaus te creëren*.

De bottom-up benadering, waarbij verschillende overheidsniveaus en particuliere partners samenwerken aan interoperabiliteit, is van cruciaal belang voor het *bereiken van consensus*. Dit bottom-up proces wordt gecombineerd met formele top-down governance. Onze resultaten tonen aan dat het mogelijk is om semantische overeenkomsten te bereiken en de politieke hindernissen te overwinnen binnen een operationele publieke sector context. In dit proefschrift tonen we aan dat de ontwerpprincipes van het Semantisch Web interoperabiliteit binnen de publieke sector kunnen vergemakkelijken door context en nuttige links toe te voegen, gebruikmakend van het Resource Description Framework (RDF) als datamodel.

Ook hebben we aangetoond dat beslissingen met betrekking tot semantische afspraken op alle niveaus traceerbaar, transparant en consistent moeten zijn. Daarom moet de vorm van de specificaties en richtlijnen worden afgestemd op de verschillende soorten stakeholders (bijv. technisch, zakelijk, beleid) om een genivelleerde discussie mogelijk te maken.

Ons onderzochte proces en methode om semantische en technische interoperabiliteit binnen een operationele overheidscontext aan de orde te stellen, werden in 2018 omarmd door de 'Stuurgroep Vlaams Informatie- en ICT-beleid'. Deze door de overheid onderschreven governance verhoogde de ontwikkeling en adoptie van datastandaarden in Vlaanderen. Op het moment van schrijven van deze dissertatie is het aantal geratificeerde datastandaarden door de Vlaamse overheid gegroeid tot meer dan negentig. De community is gegroeid tot meer dan vierhonderd auteurs uit de publieke sector, de private sector en de academische wereld.

CHAPTER 1. INTRODUCTION: RESEARCH AIM, APPROACH AND CONCEPTS

Research is to see what everybody else has seen,
and to think what nobody else has thought.

— Albert Szent-Gyorgyi.

This first chapter sets the tone of this dissertation by positioning the research within the broader field of e-government. First, it introduces the starting point of the dissertation: due to a lack of interoperability, public administrations are unable to share and reuse structured information across different IS and policy domains. The hurdles are the lack of adequate semantic standards, scarcity of Web-oriented architecture, government austerity, and reluctance caused by budget constraints. Second, this

introductory chapter presents the central aim of the dissertation and its research question: how can governments develop a scalable technique for raising and implementing semantic and technical interoperability within an operational public sector context? In addition, the relevance of this dissertation is addressed. Next, the chapter provides an overview of the research approach and primary concepts of the dissertation: interoperability, integrated public-sector governance, data standards, Linked Data, and the Semantic Web. Finally, an outline of the structure of the dissertation is presented. This chapter is based on the paper 'Towards interoperability in the public sector' [9].

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 [22]. We notice two primary drivers for interoperability in existing literature [40, 60]. Firstly, citizens and entrepreneurs expect a coherent user experience from their government as they became accustomed to by services in the private sector [40]. Secondly, our digital economy embraces new ecosystems, including 'the government as a platform' where business expects public-private partnerships to arise. Governments struggle to deliver integrated, interconnected, and cross-sectoral services due to sectoral specialisation or 'departmentalisation' [22]. 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 is often modelled from a single perspective and therefore, cannot be shared and reused across different applications and processes, causing data silos [11]. To integrate these applications, data needs to be transformed, which causes high costs. During the installation of the new local councils in 2019, 130 Flemish municipalities saved 67 labour

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

days when publishing 8055 new public mandates as interoperable data as no more manual copying and pasting of local mandates to regional databases needed to happen [63]. As more municipalities join in and new types of local decisions — such as mobility — become available, we furthermore expect this profit to increase in the future. 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 IS that have no or limited external connectivity, we need to address multiple interoperability levels; namely on the technical, semantic, organisational, and legal level [20, 21]. First, technical interoperability refers to the technical issues to link IS. Second, semantical interoperability refers to the meaning and structure of information. Third, organisational Interoperability refers to processes optimised for data exchange. And lastly, legal Interoperability — the top-level — focusses on an aligned and digital-friendly legislation. According to the European Interoperability Framework (EIF), Interoperability Frameworks assume some kind of hierarchy in terms of maturity with regard to layers of interoperability [41]. In other words, organisational interoperability can only be achieved when standards for semantic and technical interoperability have successfully been implemented. In 2013, the Open Standards for Linked Organisations (OSLO) initiative raised awareness for interoperability, but it faced barriers at the technical and semantic level[2] as there was no end-to-end approach for implementing data standards for both public and private partners. Therefore, our research initially mainly targeted semantic and technical interoperability.

[2] https://www.v-ict-or.be/assets/5384d510ce3fb57c500006ad/OSLO1.1-specificaties.pdf

1.1 The aim and relevance of the dissertation

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 studied this problem both from the technical and political point of view in the context of e-government in the region of Flanders in Belgium. Since the focus of this dissertation is mainly on semantic and technical interoperability, the 'users' who are the subject of this thesis are mainly intermediaries such as information managers, analysts and software developers in public administrations and public sector software service providers.

Because of budget cuts, public administrations have to do more with considerably less. Interoperability can lead to lower costs [21] and produce savings, but at the same time, it requires an initial investment [23]. To secure these investments and interoperability, there is a demand [2, 41] 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].

This dissertation builds on previous discoveries in the domain of e-government. The most recent and leading paradigms on how ICT impacts government include New Public Management (NPM) and Digital Era Governance (DEG). The NPM adheres to the assumption that the public sector should become more efficient and can be improved by adopting business concepts from the private sector [36, 40]. These concepts include performance, internal competition and incentivisation [36, 18]. In the NPM paradigm, ICT is used to create merely an incremental and intra-organizational efficiency improvement [7]. According to the DEG paradigm, this intra-organizational focus has led to siloes [17]. DEG claims there is a need for reintegration based on

---

[3] https://www.cen.eu/work/ENdev/whatisEN/Pages/default.aspx

the needs of citizens; therefore new technologies and new business models are necessary [19].

DEG puts digital technologies and platforms at the centre of the government [19]. Also, different stakeholders are involved that rely on digital platforms and open standards [16]. The role of the government will be to focus on strategy, processes and governance, which enable the private sector to organize itself [30]. Hence, this dissertation fits the DEG paradigm rather than limiting the challenges to savings and efficiency in the context of NPM.

Interoperability addresses the need for cooperation between administrations, the exchange of information to accomplish with legal conditions or political engagements, and to share and reuse information which leads to improved public service delivery and lower cost. [2] This dissertation has been written for (Semantic Web) researchers in the public sector but also in other domains, including transport [19], finances [51] and life sciences [1].

The aim of this section is not to be exhaustive on the different standardization initiatives, but to provide telling examples. Projects on interoperability such as StUF, INSPIRE, ISA² as well as CSMICS are struggling with semantic and technical interoperability. These struggles play out in various domains. We distinguish:

1. context-neutral, re-usable, and extensible data models [24] which are embedded in
2. a stable, governed standard and are accompanied by
3. technical guidelines that specify how these could be implemented in an operational public sector context
4. on an organisational level where political support is essential, for collecting sponsoring and gaining authority and lasting engagement [11].

In this section, we discuss these aforementioned four projects on interoperability using numbers to refer to the 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 extensible, 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 policymaking within Europe. Since 2004 INSPIRE is a directive which sets the legal framework in Europe [37]. The INSPIRE Data Specifications[6] are legally binding and accompanied by technical guidelines that specify how these legal obligations could be implemented[7]. The data specifications tend to be overspecified because it was designed for a specific domain, which again makes it harder to reuse (1). The INSPIRE programme is investigating how 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 [25]. The Core Vocabularies provide both an RDF and XML schema. The Core Vocabularies are not legally binding though (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 as the data model. The bottom-up (4) 'meet-in-the-middle'

---

[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

approach involves the stakeholders at the different phases of the development [65]. This approach facilitates interoperability and contributes to the re-use of biomedical ontologies.

As government administrations struggle to raise interoperability and face hurdles on the semantical and technical level, the main research question of this dissertation is:

How can governments develop a scalable technique for raising and implementing semantic and technical interoperability, within an operational public sector context?

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

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 (non-human) clients?

On the other hand, we have the political context:

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

As RDF is — unlike XML which is often used as a non-semantic exchange format in the public sector — an extensible knowledge representation data model, and the Open Standards for Linked Organisations (OSLO) initiative faced barriers[9] as there was no end-to-

[9] https://www.v-ict-or.be/assets/5384d510ce3fb57c500006ad/OSLO1.1-specificaties.pdf

end approach for implementing data standards, my research is based on the following presumptions [7]:

> **The design principles of the Semantic Web[10] can facilitate interoperability within the public sector by adding context and useful links, using the RDF[11] as a data model.**
>
> **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.**

My research took into account the process of raising and implementing semantic and technical agreements in the OSLO program. OSLO is an interoperability program in the region of Flanders,

RESEARCH QUESTION

How can governments develop a scalable technique for raising and implementing semantic and technical interoperability, within an operational public sector context?

which brings together expertise from different business domains and governmental levels, independent of a specific thematic project. The Flemish Government developed several domain models in line with international standards, including ISA and INSPIRE[12] enriched by data extensions to comply with the local context [11]. The formal

[10] https://www.w3.org/DesignIssues/LinkedData.html
[11] https://www.w3.org/2001/sw/wiki/RDF
[12] http://inspire.ec.europa.eu/

specification is published at data.vlaanderen.be[13]. The thematic working groups, with over 400 authors from the public and private sector, demonstrated that it is possible to raise interoperability and overcome the political hurdles.

1.2 Approach and main concepts

This dissertation used an inductive approach, data was gathered via action-research, which refers to the involvement of researchers as co-practitioners in the setting under study and the attention paid to the context where the events took place [11, 42, 55]. I aim 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 [52]. Additional data was gathered via desk research. I've conducted my research in the Flemish public sector in Belgium. Belgium is a federal state with three communities, three regions, and four language areas. Flanders is the northern federated state of Belgium, covers an area of 13,522 km² and has over 6 million inhabitants [12].

Our approach to addressing the research questions is to focus on two outcomes: the *processes* and *methods* suited for raising interoperability by researching and improving the OSLO programme within the context of e-government. OSLO started as an initiative of a mediating non-profit organization 'Vlaamse ICT Organisatie' (V-ICT-OR), an interest group of public servants active as IT practitioner at local government level in Flanders. Among the initial sponsors were Flemish ICT service providers, major cities, and the Flemish regional Government administration. In 2015 the ownership and governance of the initiative was transferred to the Flemish Government and embedded in 'Steering Committee of Flemish Information and ICT-

---

[13] http://data.vlaanderen.be/ns/

policy', which is empowered by a decree[14]. In 2019 the OSLO process and method was adopted on the Belgian interfederal level[15] too (the federal level, the communities, regions and language areas in Belgium).

The OSLO data specification *process* is aligned with the principles[16] of international standardisation bodies, i.e., due process, broad consensus, transparency, balance, and openness. The current development activities of OSLO already follow a transparent process: all records of decisions[17] and discussions[18] are publicly accessible. These 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 the implementation of the design principles of Linked Data[19], as asserted by Tim Berners-Lee in 2006. Existing public sector IS store data in relational databases and often use Extensible Markup Language[20] (XML) schemas to exchange data. These schemas, intended to exchange data, cannot be easily adapted or extended [38]. In my research, I focus on how RDF, which is an extensible data model, can be adopted in the public sector and how the semantic agreements reached between domain experts, automatically[21] 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[22], 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

---

[14] http://docs.vlaamsparlement.be/pfile?id=1213278
[15] https://github.com/belgif/review/tree/master/Process
[16] https://open-stand.org/about-us/principles/
[17] https://informatievlaanderen.github.io/OSLO/
[18] https://github.com/Informatievlaanderen/OSLO/issues
[19] https://www.w3.org/DesignIssues/LinkedData.html
[20] https://www.w3.org/TR/REC-xml/
[21] https://github.com/Informatievlaanderen/OSLO-EA-to-RDF
[22] https://www.w3.org/2001/sw/wiki/RDF

applications having the same characteristics as the Web, a.o., simplicity, evolvability, and performance [30]. 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.

I evaluate the success of my research by applying the process and method in public sector initiatives within the context of base registries, legislative data, e-government portals, public sector data on the Web, sensor data, and personal data. I benchmark the output variables that affect the Successful Implementation of ICT Projects in Government [32], using the following criteria:

> ***Cost reduction*: I 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 measure the increase in re-use of public sector information (PSI) by comparing the decrease of requested information citizens provide, in relation to the service complexity and customer satisfaction.**
>
> ***Technological benefits:* I research potential benefits using the Linked Data approach, to lower the cost of data publishing, and raise the availability of public endpoints.**
>
> ***Improved efficiency*: I conduct qualitative research by interviewing the stakeholders in the public and private sector, including perceived benefits.**

1.2.1 Interoperability levels and integrated governance

This section addresses the different interoperability levels as defined by the EIF, namely the legal, organizational, semantic, and technical level (Fig. 1.1) [20].

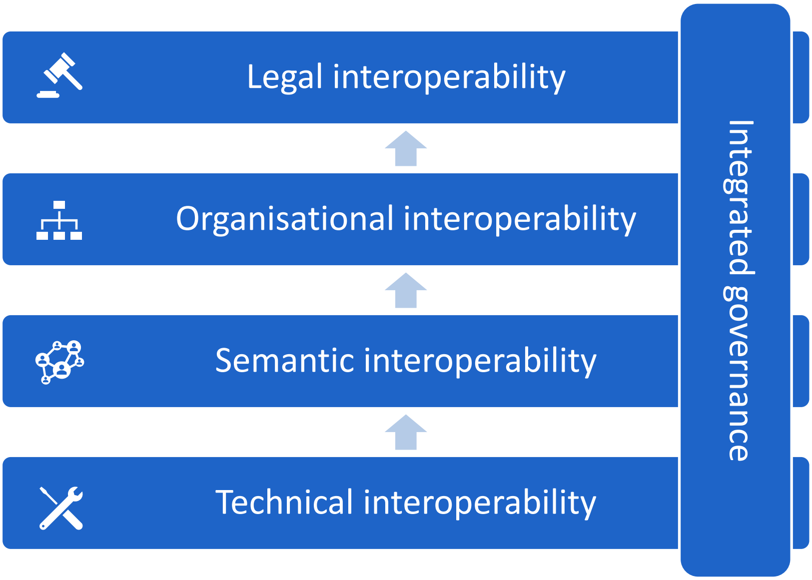


Fig. 1.1. Overview of the various interoperability levels [20].

Semantic interoperability focusses on the meaning of data elements, such as a resource accessible by a Universal Resource Identifier[23] (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, which we discuss in **Chapter 4** [21]. Semantic interoperability also covers the syntactic aspect, which refers to the grammar and format [22], such as HTML or XML.

---

[23] https://www.w3.org/TR/uri-clarification/

Technical Interoperability is often centred on (communication) protocols and the infrastructure needed for those protocols to operate [61]. 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 [22]. We observe this variety of interface specifications — that interconnect systems and services — not only in legacy applications but also in new initiatives such as the Internet of Things (IoT) paradigm. IoT infrastructures — that harvest information both from the physical world (sensors) and interact with their environment (actuators) — also face major challenges on interoperability, which we discuss in Chapter 5 [33].

Organisational Interoperability refers to aligned business processes across the different public sector administrations or organisations that act on behalf of the public sector [20]. This is also crucial in new initiatives such as Smart Cities where the aim is to combine vast amounts of (sensor) data for better decision making, by creating a sustainable network of sensors and actuators. The different actors in the ecosystem must agree on a Service Level Agreement to coordinate responsibilities and expectations [21, 44]. According to the European Commission (EC) Directory General for Communications Networks, Content and Technology (DG CONNECT)[24] boundaries between public and private services will fade. Therefore, I explore organisational interoperability in a public-private context, more specific how we can reshape the relation between citizens, their personal data, and the applications they use in the public and private sector. In Chapter 6, I focus on how to give citizens true control over their data and raise organisational interoperability by rewiring existing IS to a 'me-central' architecture, which can reshape the relationship between

---

[24] https://ec.europa.eu/digital-single-market/en/dg-connect

citizens, their personal data, and the applications they use in the public and private sector.

Legal Interoperability focusses on the various barriers in legislation when exchanging data across policy domains, different governmental levels or in a public-private context [20]. In Chapter 7, we discuss how interoperability can be raised by embedding it in the legislative phase of local council decisions. Also, clauses in agreements between government administrations and software vendors create hurdles to reuse the data outside the information system, which we address in Chapter 5.

Public administrations face both technical limitations and practical challenges to raise interoperability, including unjustified benefits and considerable effort expectations [59]. A government administration that has the *political power* and *will* when introducing e-government systems is vital to success [46], which includes internal organisational politics involving organisational members as well as external politics concerning how the government organisation relates to its council [53]. As multiple organisations are involved, integrated public service governance is crucial, which focusses on organization structures, roles, responsibilities, and a transparent decision-making process, which we address in Chapter 2 [20]. Interoperability frameworks — such as the EIF — assume a hierarchy in the interoperability levels. The two top levels (Fig. 1.1) can only be implemented successfully when both semantic and technological interoperability are in place [41].

As such, my research mainly targets semantic and technical interoperability and a generic process and method to raise interoperability, which we discuss in Chapter 3. We demonstrate this end-to-end approach — using the design principles of Linked Data — in a high-impact governmental integration project in Chapter 8.

1.2.2 Data standards

Standards have been vital for innovation, market growth, and health & safety for centuries [12, 45]. By the end of the 18th century, the introduction of standards for machine parts was the key turning point of the industrial revolution. Until then, nuts and bolts were hand-crafted in matching pairs. It was the invention of the screw-cutting lathe that allowed to produce screws with standardised thread [12, 45]. As nuts and bolts became interchangeable, it was possible to combine machine parts and facilitate new innovations. This can be paralleled with the invention of the WWW by Tim Berners-Lee, which fuelled digital innovation, by making documents and knowledge interchangeable on a global level [56]. The platform-independent documents are formatted using the HTML and transferred using the HTTP.

The literature distinguishes 'de jure' and 'de facto' standards [27, 31]. De jure standards emerge through consensus via an industry standards body or by a standards organisation such as the International Organization for Standardization, while de facto standards refer to processes where all or nearly all potential adopters use the same interoperability agreements and turn it into a setting that is hard to deviate from [8, 12, 58].

The European Union defines a 'standard' as "a technical specification, adopted by a recognised standardisation body, for repeated or continuous application…" [26]. Although most standardisation organisations promote Open Standards, there seems to be no general definition in the literature of the precise meaning of 'open' in the context of data standards [14, 57]. According to Open Stand — the principles where impactful standardisation initiatives build upon — 'open' refers both to the availability of specifications as well as the openness of the development process itself [50].

Since 2009 there is a strong demand for Open Data Standards and transparent governance in Flanders within the context of e-government, that resulted in OSLO standardisation initiative [10, 36].

1.2.3 Linked Data and the Semantic Web

The holy grail in e-government — that all IS can exchange information without bespoke integration efforts — is the topic of the majority of the debates in the public sector [36]. In this section, I briefly sketch the origin of the key concepts that I build upon in my dissertation to raise interoperability.

Throughout this dissertation, I research how the building blocks of the WWW contribute to raise interoperability and provide better public services. The WWW interconnects documents and services and runs on top of the Internet [4]. In my opinion, the Web — invented in 1989 by Sir Tim Berners-Lee — is definitely the most important innovation of the modern era. The world has never been so interconnected, information is at our fingertips, from applying for a scholarship, getting directions and even start a revolution using social media [34].

The Web — which is a network of information resources — builds upon three mechanisms: Uniform Resource Locators (URLs), the HTTP, and the HTML. First, URLs such as https://data.vlaanderen.be/standaarden/ identify resources on the Web and also locate these resources [47]. Second, HTTP provides a standardised protocol that allows Web servers and clients to communicate using a simple request/response message paradigm [49]. Third, HTML allows to mark-up documents and link to other resources via their URI. When following a link, the browser will dereference the URL and download the resource [48]. In practice, https://data.vlaanderen.be/standaarden/ is a unique identifier for a resource, but also allows to dereference it, e.g., get a human-readable HTML representation of the resource. The web page can contain links which refer to other resources. The Web has a software architecture which is designed for internet-scale and builds upon the fundaments of a distributed hypermedia application. To raise scalability, web applications follow the REST architectural style. The REST architectural style builds upon hyperlinks, and a collection of architectural constraints that facilitate architectural elasticity and allow to navigate an application without preexisting knowledge [28, 62]. These architectural constraints are a blueprint of the behaviour of a well-designed web application and are further discussed in **Chapter 5** [28].

The Web evolved from connecting documents and people to connecting data by enriching web resources with semantics. In 2001, Tim Berners-Lee et al. stated that *"The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation"* [5]. The Semantic Web is a Web of data that allows people and machines to discover more data just by following the links [3]. According to Wood et al., the term Linked Data refers to a *"set of best practices for publishing and connecting structured data on the Web using international standards of the W3C"* [64]. Throughout this dissertation, we will refer to the Linked Data design principles as asserted by Tim Berners-Lee in 2006 (Fig. 1.2). In this section, we apply them to the Linked Base Registry for Addresses (see also **Chapter 4**) [13].

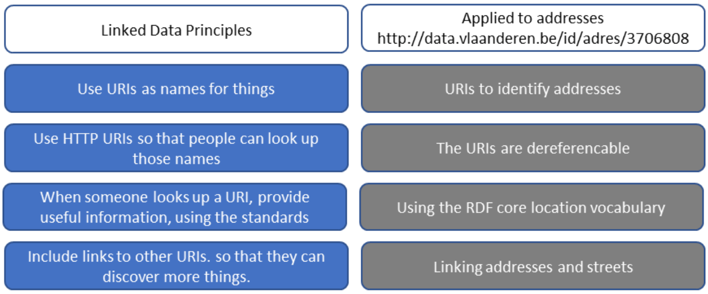


Fig. 1.2. Design principles of Linked Data as asserted by Tim Berners-Lee and the specific implementation in the Linked Base Registry for Addresses in Flanders, see also Chapter 4 [3, 10].

The first principle states to "Use Uniform Resource Identifiers (URIs) as names for things". All addresses were given a universally unique identifier which can be looked up via the Web: http://data.vlaanderen.be/id/adres/3706808.

The second principle mentions to "use HTTP URIs so that people can look up those names". Dereferenceable URIs allow machines to browse the Web of data, as humans browse the Web of documents.

The third principle poses that "When someone looks up a URI, one should provide useful information, using the standards RDF and SPARQL". RDF is the data model for Linked Data. According to W3C, "*RDF extends the linking structure of the Web to use URIs to name the relationship between things as well as the two ends of the link (this is usually referred to as a 'triple')"*[25]. This model is all about interoperability and allows data to be shared across different applications. The Linked Base Registry for Addresses is using RDF as a data model and applies the ISA Core Location Vocabulary[26] which provides an extensible, context-neutral vocabulary. The endpoint of the Base Address Registry is implemented as a SPARQL endpoint, which allows querying the RDF data sources via the Web. SPARQL is a recursive acronym for SPARQL Protocol and RDF Query Language. In Chapter 5, we will discuss alternative endpoints as well.

The final principle states to "Include links to other URIs so that they can discover more things". Links to other resources make data on the Web more meaningful and useful. This allows humans and machines to discover more information, just by following the links. Also, other organizations can link their datasets to the addresses, making their data more useful. In Flanders, the domicile of a citizen and the

[25] https://www.w3.org/2001/sw/wiki/RDF
[26] https://joinup.ec.europa.eu/asset/core_location/description

Energy Performance Certificate (EPC) are both linked to a physical home address.

1.3 Publications

The work presented in this dissertation is based on peer-reviewed publications in international scientific journals and conference proceedings. They reflect most of the work I conducted and reveal the evolution of my point of view on interoperability in the public sector. The following lists provide an overview of all publications I (co-)authored during my PhD. This list contains three journals (being first author in two and first Ghent University author in the other one), 8 conference publications (first author in 6) and one journal article under review. Also, my paper 'OSLO: Open standards for linked organizations' has won the best paper award on the international conference on electronic governance and open society.

As stated in Section 1.2 (Approach and main concepts), my research data was gathered via action-research [11, 42, 55]. The articles originated from my work on the OSLO initiative, which I've co-initiated in 2011 at the level of the local governments in Flanders. In 2015, I took the lead on the OSLO interoperability programme at the Flemish Government. Furthermore, I was involved in several standardisation initiatives of the EC since 2011, including the ISA and INSPIRE programme. Also, I am a co-founder of the Belgian interfederal interoperability initiative which started in 2019, which adopted the OSLO process and method too.

1.3.1 Publications in international journals

- Buyle, R., Vanlishout, Z., Coetzee, S., De Paepe, D., Van Compernolle, M., Thijs, G., Van Nuffelen, B., De Vocht, L., Mechant, P., De Vidts, B., & Mannens, E. (2019). Raising interoperability among base registries: The evolution of the Linked Base Registry for Addresses in Flanders. Journal of Web

Semantics, 55, 86-101. DOI: https://doi.org/10.1016/j.websem.2018.10.003

- Coetzee, S., Vanlishout, S., Buyle, R., Beyaert, V., Siebritz, L., & Crompvoets, J. (2019). Changing stakeholder influences in managing authoritative information–the case of the Centraal ReferentieAdressenBestand (CRAB) in Flanders. Journal of Spatial Science, 1-23. DOI: https://doi.org/10.1080/14498596.2019.1650301

- Buyle, R., Van Compernolle, M., Vlassenroot, E., Vanlishout, Z., Mechant, P. and Mannens E. (2018). “Technology readiness and acceptance model“ as a predictor for the use intention of data standards in smart cities. *Media and Communication*, *6*(4), 127-139. DOI: http://dx.doi.org/10.17645/mac.v6i4.1679.

- Buyle, R., Van de Vyvere, B., Julian Andres Rojas Melendeza, Van Lancker, D., Vlassenroot, E., Van Compernolle, M., Lefever, S., Colpaert, P., Mechant, P., & Mannens, E. (2021). A sustainable method for publishing interoperable open data on a Web Scale. *Data* (under review).

1.3.2 Publications in conference proceedings

- Buyle, R., Taelman, R., Mostaert, K., Joris, G., Mannens, E., Verborgh, R., & Berners-Lee, T. (2019, November). Streamlining governmental processes by putting citizens in control of their personal data. In International Conference on Electronic Governance and Open Society: Challenges in Eurasia (pp. 346-359). Springer, Cham. DOI: https://doi.org/10.1007/978-3-030-39296-3

- Coetzee, S., Du Preez, J., Behr, F. J., Cooper, A. K., Odijk, M., Vanlishout, S., Buyle, R., Jobst, M., Chauke, M., Fourie, N., & Schmitz, P. (2019). Collaborative custodianship through collaborative cloud mapping: Challenges and opportunities. In *ICC2019, the 29th International Cartographic Conference* (pp. 1-10). DOI: https://doi.org/10.5194/ica-proc-2-19-2019

- Buyle, R., Van Compernolle, M., De Paepe, D., Scheerlinck, J., Mechant, P., Mannens, E., & Vanlishout, Z. (2018, April). Semantics in the wild: a digital assistant for Flemish citizens. In *Proceedings of the 11th International Conference on Theory and Practice of Electronic Governance* (pp. 1-6). DOI: 10.1145/3209415.3209421

- Buyle, R. (2017). Towards interoperability in the public sector. In *ISWC2017, the 16e International Semantic Web Conference* (Vol. 1931, pp. 1-8).

- De Paepe, D., Thijs, G., Buyle, R., Verborgh, R., & Mannens, E. (2017, May). Automated uml-based ontology generation in oslo 2. In European Semantic Web Conference (pp. 93-97). Springer, Cham. DOI: https://doi.org/10.1007/978-3-319-70407-4_18

- Buyle, R., De Vocht, L., Van Compernolle, M., De Paepe, D., Verborgh, R., Vanlishout, Z., ... & Mannens, E. (2016, November). OSLO: Open standards for linked organizations. In *Proceedings of the international conference on electronic governance and open society: Challenges in Eurasia* (pp. 126-134). DOI: http://dx.doi.org/10.1145/3014087.3014096.

- Buyle, R., Colpaert, P., Van Compernolle, M., Mechant, P., Volders, V., Verborgh, R., & Mannens, E. (2016). Local Council Decisions as Linked Data: a proof of concept. In *15th International Semantic Web conference* (Vol. 1690). RWTH. http://ceur-ws.org/Vol-1690/paper71.pdf

1.3.3 W3C standardisation

- Buyle, R., De Vocht, L., De Paepe, D., Van Compernolle, M., Nolf, G., Vanlishout, Z., ... & Mechant, P. (2016). The Public Sector DNA on the web semantically marking up government portals. In *Smart Descriptions & Smarter Vocabularies (SDSVoc)* (pp. 1-4). The World Wide Web Consortium (W3C). https://www.w3.org/2016/11/sdsvoc/SDSVoc16_paper_1

1.4 References

[1] 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).

[2] Audit Vlaanderen, Thema-audit gemeentelijke belastingen en retributies. p. 81 (2016).

[3] Berners-Lee, T. (2011). Design issues: Linked data (2006). http://www.w3.org/DesignIssues/LinkedData.html

[4] Berners-Lee, T., Cailliau, R., & Groff, J. The worldwide web. Computer Networks andisdn Systems, 25(4–5):454–459, 1992.

[5] Berners-Lee, T., Hendler, J. A., & Lassila, O. (2001). The Semantic Web Scientific American, 284 (5): 34—43

[6] *Bizer, C., Heath, T., & Berners-Lee, T. (2011). Linked data: The story so far. In Semantic services, interoperability and web applications: emerging concepts (pp. 205-227). IGI Global.*

[7] Boston, J. (1991). The theoretical underpinnings of public sector restructuring in New Zealand. Reshaping the state, 1-26.

[8] Brunsson, N., Rasche, A., & Seidl, D. (2012). The dynamics of standardization: Three perspectives on standards in organization studies. Organization Studies, 33(5-6), 613-632.

[9] Buyle, R. (2017). Towards interoperability in the public sector. In *ISWC2017, the 16e International Semantic Web Conference* (Vol. 1931, pp. 1-8).

[10] Buyle, R., Colpaert, P., Van Compernolle, M., Mechant, P., Volders, V., Verborgh, R., & Mannens, E. (2016). Local Council Decisions as Linked Data: a proof of concept. In The International Semantic Web Conference (ISWC). CEUR Workshop Proceedings (CEUR-WS. org).

[11] Buyle, R., De Vocht, L., Van Compernolle, M., De Paepe, D., Verborgh, R., Vanlishout, Z., De Vidts, B., Mechant, P., Mannens, E. (2016). OSLO: Open Standards for Linked Organizations. In *Proceedings of the International Conference on Electronic Governance and Open Society: Challenges in Eurasia* (pp. 126-134). ACM.

[12] Buyle, R., Van Compernolle, M., Vlassenroot, E., Vanlishout, Z., Mechant, P., & Mannens, E. (2018). "Technology readiness and acceptance model" as a predictor for the use intention of data standards in smart cities. *Media and Communication*, 6(4), 127-139.

[13] Buyle, R., Vanlishout, Z., Coetzee, S., De Paepe, D., Van Compernolle, M., Thijs, G., Van Nuffelen, B., De Vocht, L., Mechant, P., De Vidts, B., &

Mannens, E. (2019). Raising interoperability among base registries: The evolution of the Linked Base Registry for Addresses in Flanders

[14] Chesbrough, H., Vanhaverbeke, W., & West, J. (2006). Open Innovation: Researching a New Paradigm. Oxford: Oxford University Press.

[15] Colpaert, P. et al. Open Transport Data for maximising reuse in multimodal route planners: a study in Flanders. IET Intelligent Transport Systems (2017).

[16] Danneels, L. (2017). Transforming government: the way towards digital era governance.

[17] Drechsler, W. (2005). The rise and demise of the new public management. Post-autistic economics review, 33(14), 17-28.

[18] Dunleavy, P. (1994). The globalization of public services production: can government be 'best in world'?. Public Policy and Administration, 9(2), 36-64.

[19] Dunleavy, P., & Margetts, H. (2015). Design principles for essentially digital governance.

[20] European Commission, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. European Interoperability Framework Implementation strategy. http://eur-lex.europa.eu/resource.html?uri=cellar:2c2f2554-0faf-11e7-8a35-01aa75ed71a1.0017.02/DOC_1&format=PDF (accessed 26.08.17).

[21] European Commission, ISA: European Interoperability Framework (EIF) for European public services, Annex 2. pp. 2-21 (2010). https://ec.europa.eu/isa2/sites/isa/files/isa_annex_ii_eif_en.pdf.

[22] European Commission, ISA: European Interoperability Framework (EIF). p. 5-37 (2017).

[23] European Commission, ISA: The ISA programme: Overcoming eBarriers to European public services. p2 (2011).

[24] European Commission, ISA: Handbook for using the Core Vocabularies. pp. 8-10 (2014).

[25] European Commission, Process and methodology for developing semantic agreements. p. 7 (2013).

[26] European Commission. Regulation (EU) 1025/2012 of the European Parliament and of the Council of 25 October 2012 on European standardisation, amending Council Directives 89/686/EEC and 93/15/EEC and Directives 94/9/EC, 94/25/EC, 95/16/EC, 97/23/EC, 98/34/EC, 2004/22/EC, 2007/23/EC, 2009/23/EC and 2009/105/EC of the European Parliament and of the Council and repealing Council Decision 87/95/EEC and Decision No 1673/2006/EC of the European Parliament and of the

Council. Official Journal of the European Union (OJ), 19(12-33), 316 (2012).

[27] Farrell, J., & Simcoe, T. (1996). Choosing the rules for formal standardization. University of California, Berkeley, working paper.

[28] Fielding, R. Architectural styles and the design of network-based software architectures, University of California, Irvine Irvine, 2000. https://www.ics.uci.edu/~fielding/pubs/dissertation/fielding_dissertation.pdf.

[29] Fielding, R., Taylor, R. Architectural styles and the design of network-based software architectures. Doctoral dissertation: University of California, Irvine. pp. 151 (2000)

[30] Fishenden, J., & Thompson, M. (2013). Digital government, open architecture, and innovation: why public sector IT will never be the same again. Journal of public administration research and theory, 23(4), 977-1004.

[31] Funk, J. L. (2001). Global competition between and within standards: the case of mobile phones. London: Palgrave Macmillan.

[32] Gichoya, D. Factors affecting the successful implementation of ICT projects in government. the Electronic Journal of e-government, 3(4), pp. 175-184 (2005).

[33] Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. Internet of Things (IoT): A vision, architectural elements, and future directions, Futur. Gener. Comput. Syst. 29 (2013) 1645–1660. doi: 10.1016/j.future.2013.01.010.

[34] Harb, Z. (2011). Arab revolutions and the social media effect. M/C Journal, 14(2).

[35] Hautekiet, P. (2009). Interbestuurlijk e-government: aandachtspunten en aanbevelingen. In Opsomer, B. (Eds.), Memorandum 2009 - 2014 - aanbevelingen voor een krachtig bestuurlijk beleid (edition, pp. 171 - 196). Brussels: VLABEST

[36] Hood, C. (1991). A public management for all seasons?. Public administration, 69(1), 3-19.

[37] Infrastructure for Spatial in formation in Europe, Drafting Team ”Data Specifications“ – deliverable D2.5: Generic Conceptual Model, p. 5 (2007).

[38] ISA programme by PwC EU Services EESV, Case study: How Linked Data is transforming eGovernment. pp 1-5 (2013).

[39] 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).

[40] Kernaghan, K. (2000). The post-bureaucratic organization and public service values1. International Review of administrative sciences, 66(1), 91-104.

[41] Kubicek, H., Cimander, R. Three dimensions of organizational interoperability. European Journal of ePractice, 6, pp. 1-12 (2009).

[42] Lewin, K. (1946). Action research and minority problems. Journal of social issues. 2,4, 34-46.

[43] Li, F., Vögler, M., Claeßens, M., & Dustdar, S. Efficient and scalable IoT service delivery on cloud, in: 2013 IEEE Sixth Int. Conf. Cloud Comput., IEEE, 2013: pp. 740–747. doi: 10.1109/CLOUD.2013.64.

[44] Margetts, H., & Dunleavy, P. (2013). The second wave of digital-era governance: a quasi-paradigm for government on the Web. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 371(1987), 20120382.

[45] Mills, S. (2013). International standards and the IEEE standards association in the emerging global economy. IEEE Aerospace and Electronic Systems Magazine, 28(7), 6-11.

[46] Novakouski. et al. Interoperability in the e-Government Context (2012).

[47] Network Working Group. RFC 1738: Uniform Resource Locators (URL), 1994. https://tools.ietf.org/html/rfc1738.

[48] Network Working Group. RFC 1866: Hypertext Markup Language - 2.0, 1995. https://tools.ietf.org/html/rfc1866.

[49] Network Working Group. RFC 2616: HyperText Transfer Protocol – HTTP/1.1, 1999. http://www.ietf.org/rfc/rfc2616.txt.

[50] Open Stand (2012) OpenStand: Principles for The Modern Standard Paradigm. Available at: http://openstand.org/ (accessed: 7.5.2018).

[51] 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).

[52] Rapport, R. Three dilemmas in action research: with special reference to the Tavistock experience. Human relations, 23(6), pp. 499-513 (1970).

[53] Rocheleau, B. Politics, accountability, and governmental information systems. Public information technology: Policy and management issues. pp. 20-52 (2003).

[54] Roe, J. W. (1916). English and American tool builders. New York, London, McGraw-Hill.

[55] Saunders, M., & Lewis, P. 2000 Research Methods For Business Students. Harlow Financial Times. Prentice Hall, 147-148.

[56] Sheridan, J. (2010). The nuts and bolts of opening government data. Digital Content Quarterly, 4(2), 6-7. Retrieved from

https://sca.jiscinvolve.org/wp/files/2010/04/sca_dcquarterly_02_apr10-final.pdf

[57] Simcoe, T. (2006). Open standards and intellectual property rights. Open innovation: Researching a new paradigm. In Chesbrough, H., Vanhaverbeke, W., & West, J. (Eds.), Open Innovation: Researching a New Paradigm (pp. 161-183). Oxford: Oxford University Press.

[58] Stango, V. (2004). The economics of standards wars. Review of network economics, 3(1).

[59] Tatarintseva, O. et al. Is Your Ontology a Burden or a Gem?–Towards Xtreme Ontology Engineering. Proc. ICTERI, p. 65-81 (2011).

[60] United Nations Department of Economic and Social Affairs. United Nations E-Government Survey 2012. United Nations, p. 69 (2012).

[61] Van der Veer, H., Wiles, A. European Telecommunications Standards Institute: Achieving Technical Interoperability – the ETSI Approach (2006).

[62] Verborgh, R. (2014). *Serendipitous web applications through semantic hypermedia* (Doctoral dissertation, Ghent University).

[63] Volders, V. (2020). *Vlaanderen Radicaal Digitaal II: Lokale Besluiten als Gelinkte Open Data (LBLOD)*. Unpublished manuscript, Flemish Agency for Local and Provincial Government and the civic integration policy, Brussels, Belfgium.

[64] Wood, D., Zaidman, M., Ruth, L., Hausenblas, M. (2014). Linked Data - Structured Data On The Web, Manning, p.4. https://livebook.manning.com/#!/book/linked-data/chapter-1/99

[65] 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).

1.5 Structure of the dissertation

This dissertation consists of ten chapters. Chapter 1 serves as an introduction to the thesis by presenting its aim, relevance, central research question and main concepts.

Chapter 2 discusses the adoption criteria for Open Data Standards. We research criteria for implementing data standards in the public sector by analysing the factors that affect the adoption of data governance. Additionally, this chapter gives an overview of the history of data standards and sketches the concept of openness.

Chapter 3 outlines the process and method for developing Open Data Standards. We address the interoperability hurdles at the different governmental levels and examine how the bottom-up OSLO programme tackled these obstacles. Finally, we detail a generic process and method and provide practical insights on how to raise interoperability.

Chapter 4 researches how to raise semantic interoperability. We examine the evolution of the Linked Base Registry for Addresses in Flanders. Also, this chapter introduces the benefits of Linked Data and argues the increase interoperability leads to a better adoption of government information in the public and private sector.

Chapter 5 delves into technical interoperability. We research what methods are suited for publishing Open Data time series in a sustainable, predictable, and cost-effective way in the context of Smart Cities. This chapter argues that the method of Linked Data raises technical interoperability, lowers the publishing cost, and raises the availability of the endpoints.

Chapter 6 scrutinises organisational interoperability by exploring how governmental processes can be streamlined by putting citizens truly in control of their data. We apply the decentralised Solid ecosystem to two high-impact public sector use cases. This chapter argues that Solid allows reshaping the relationship between citizens, their personal data, and the applications they use in the public and private sector. We detail how these processes can be streamlined by building upon existing web standards and methods such as Linked Data and decentralisation.

Chapter 7 discusses how to raise legal interoperability. This chapter examines a method to manage Local Council Decisions as Linked Data (LBLOD). We argue that the method makes the legislation process more efficient, raises the quality of the decisions and lowers the barriers for reuse.

Chapter 8 researches interoperability in high-impact governmental integration projects. This chapter outlines a method for raising semantic interoperability between different IS and actors. We examine how semantic agreements are maintained and implemented end-to-end using the design principles of Linked Data.

Finally, in Chapter 9, after a review of the research questions, the key findings of the dissertation are summarised and the main contributions are listed. Also, limitations and directions for future research are addressed.

CHAPTER 2. ADOPTION CRITERIA FOR OPEN DATA STANDARDS

Innovation is not disruptive, consumer adoption is.

— Jeff Bezos.

This chapter discusses the adoption criteria for Open Data Standards. We research criteria for implementing data standards in the public sector by analysing the factors that affect the adoption of data governance. Additionally, this chapter gives an insight in the history of data standards and sketches the concept of openness. This chapter is based on the paper, ’An assessment of the ”Technology Readiness and Acceptance Model” (TRAM) as a predictor for the use intention of data standards in Smart Cities’ [10].

2.1 Introduction

2.1.1 Data standards

Flemish municipalities provide over 800 public services in domains such as building permits, subsidies, public welfare, and day-care. The back-office processes and service delivery of these services are supported by specialised IS from different software vendors [7]. Because the data in these IS is modelled from a single thematic perspective, it is difficult or impossible to share and reuse them among all services [18]. This causes unnecessary frustrations in everyday life of citizens and businesses as they are required to provide the same information more than once to their government [25]. The smart use of available information about citizens by public administrations is referred to as the once-only principle [25]. Also, the transformation of society towards a digital economy is leading to changing roles where barriers between public and private actors are blurring [24]. This happens in a context where information and IS are combined with new technologies such as live data from physical devices [77]. Smart Cities have a comprehensive commitment to innovation in technology, management, and policy according to Nam and Pardo (2011). In 2012 Flemish cities started a grassroots initiative to overcome this fragmented data landscape and implement 'once-only' via the Open Standards for Linked Organisations programme (OSLO). The initiative was launched as a private-public partnership in the region of Flanders in Belgium, co-funded by the cities, the regional government of Flanders, and Information & Communication Technology (ICT) service suppliers [8]. The goal of OSLO is to raise interoperability in the region of Flanders. 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 [26].

For centuries, standards have been fuelling innovation, catalysing the growth of markets and protecting the health and safety of citizens [48]. Alike any nut and bolt which adheres to the standards can be put together, electronic documents that are formatted in HTML and transferred using the HTTP can be exchanged via the Web [61, 66]. This created a digital revolution with new forms for social and economic enterprise and a new scope and efficiency for markets [6]. The safety of citizens is often a driver for standardisation. On July 30th, 2004, an immense explosion took place in the city of Ghislenghien in Belgium. The blast, with a perimeter of 6km, instantly killed 24 people and more than 132 were injured. The disaster was caused by a leakage of a high-pressure gas pipe, damaged by a drilling machine [23]. Following this incident, the Flemish Government agreed on a common standard for exchanging information on cables and pipes and a single-point-of-access was established to automate the process to provide utility data in support of groundworks [78]. This standardisation process resulted in a reduction of claims and incidents, and in significant time and financial savings [41].

The literature differentiates between de jure and de facto standards [27, 28]. De facto standards refer to processes that aim at uniformity, where all or nearly all potential adopters use the same interoperability agreements and turn it into a mode that is hard to deviate from [5], such as the native Microsoft Word 'doc' and 'docx' file format for storing and exchanging text documents. By contrast, de jure standards are those that emerge through consensus. Consensus may be reached informally or formally expressed through an industry standards body or by a standards organisation such as the International Organization for Standardization [70]. EU-Regulation No 1025/2012 defines a 'standard' as "a technical specification, adopted by a recognised standardisation body, for repeated or continuous application…".

Most standard setting organisations promote the adoption of Open Standards [68].As mentioned in Section 1.2.2, the term 'open' refers both to the availability of specifications as well as the openness of the development process itself [51].

2.1.2 A historical perspective of standardisation initiatives in Flanders

According to Steen and Wayenberg the complex state structure of Belgium is reflected in the organisation of local governments as well [71]. Merely 12 cities have more than 50.000 citizens and 30% of the cities have less than 10.000 citizens. As the public servants in the cities vary between 1 in the municipality of Herstappe up to 6.900 in the City of Antwerp, the local governments administrations organisation is diverse [71]. Data and information in the region of Flanders are fragmented across 300 municipalities, the regional administration, the federal administration, and the private sector. To achieve interoperability among these actors robust, coherent, and universally applicable data standards are essential [26]. Since 2009 there is a demand for Open Data Standards and transparent governance [32]. The region of Flanders has a vast track record on information governance since 2009 which originated in the governance of geospatial data [11]. In 2012 the Flemish municipalities initiated an interoperability initiative ‘Open Standards for Local Governments’ (Open Standaarden voor Lokale Overheden - OSLO) to facilitate the re-use of information across all IS [9]. They initiated thematic working groups with participants from governments, industry, and academia and they agreed on reusable data specifications, which facilitate sharing and re-using information across IS. In 2015 the steering committee for Flemish Information and ICT Policy was installed. This committee is empowered by decree and engages the regional government, the cities, academia, and industry via a so-called Triple Helix approach [37].

2.1.3 Data standards in Smart Cities

In Smart Cities an amalgamate set of devices is deployed, that generate different types of (real-time) data. These peripheral devices are connected to IS via existing communication networks. The mapping of traffic flows is an epitome case for interoperability in Smart Cities. For example, the quality of service, air, and noise can be derived from traffic models. Without proper agreements, different sound meters and cameras are connected to the same post because the sensor data is only suitable for a specific application. An example of a widespread data standard is the data exchange standard for exchanging traffic information (DATEX2) that is managed by the European Committee for Standardisation (CEN) [2]. The standard was extended to also publish the availability of parking lots in Smart Cities. DATEX2 was rewired to a Linked Data format, that allows different sensor datasets to be interlinked on a semantic level and become machine-readable [14]. In the region of Flanders, best practices related to publishing data in an interoperable and sustainable way were ratified by the thirteen biggest cities and the regional government in an Open Data Charter [69].

2.2 Theoretical background and hypothesis development

2.2.1 Acceptance models for data standards

The goal of this chapter is to explore the user's attitude towards data standards in the public sector and the factors that affect the adoption. Since the focus of this dissertation is mainly on semantic and technical interoperability, the 'users' who are the subject of this chapter are mainly intermediaries such as information managers, data analysts and software developers. When addressing the higher interoperability levels, it becomes relevant to involve end-users such as citizens as well. This connects to the main research question “How can governments develop a scalable technique for raising and implementing semantic and technical interoperability, within an operational public sector context?", as the adoption of data standards in the public sector is crucial to raise interoperability.

The identification of the factors that cause people to accept new technologies has been researched heavily over the past decades [35, 74, 36]. Acceptance models in relation to the adoption of data standards emerged in the health and e-commerce sector [12, 52, 41]. As e-commerce websites need to be seamlessly integrated with back-office applications of the suppliers, which provide information on the price and availability of their products, data standards that lower the integration cost and avoid vendor lock-in are crucial. Chen (2003) researched the adoption and diffusion of standards in the context of e-business [12]. The adoption framework builds upon Rogers' (1983) Innovation Diffusion Theory (IDT) [62]. Rogers [62] defines innovation as "*an idea, practice, or object that is perceived as new by an individual or other unit of adoption*" (p. 12). Chen [12] identified the challenge of "*separating individual and organizational decisions*" (p. 277). Also, this research shows that standards are often embedded in software components, which makes it hard to distinguish the adoption of the standards from the de-facto adoption of the tools. In the health sector, information standards are crucial to create patient-centric records and exchange them between healthcare providers. Lin proposed a framework for evaluating the adoption of data standards in hospitals [42]. The adoption framework also builds upon the IDT. Lins' framework identified industrial competitions and government involvement, system integrity, top management attitudes, technological capability of the staff, and organisation scale as influencers of the adoption of health data standards in hospitals. Pai analysed the introduction of Healthcare Information Systems (HIS) [53]. A HIS is "*a set of standards based on healthcare diagnosis, symptoms, cause, healthcare target, and measurements*" (p. 651) [53]. These HIS provide the hospital staff with integrated healthcare plans. This research builds upon one of the most widely accepted frameworks to predict and explain the adoption of IS: the Technology Acceptance Model (TAM) [20]. TAM asserts that Perceived Usefulness (PU) and Perceived Ease of Use (PEU) have a determining impact on the intended and actual use of technology [22]. PU is defined as the probability to which a user believes that a HIS will improve his or her job performance. PEU refers to the degree to which the eventual user foresees the target system will be free of effort [19]. Pai [53] integrates TAM with the HIS Success Model [21] and analyses three interrelated dimensions that have an impact on PU and PEU: Information Quality, Service Quality, and System Quality. This study concludes that the proposed dimensions have a positive influence on the use intention via the mediating constructs PU and PEU. Alike, Chen's [12] research combines the impact of standards and technology. Mueller [49] researches the elements that influence the intention to accept and use IT standards and focuses on the individual. The study researches the acceptance using the TAM and the Theory of Planned Behaviour (TPB) [1]. TPB [1] states that "Attitudes toward the behaviour, subjective norms with respect to the behaviour, and perceived control over the behaviour are usually found to predict behavioural intentions with a high degree of accuracy" (p. 206). Mueller [49] discusses the moderating role of the personality of the individual. People with a high score on 'openness' are likely to adopt innovations.

2.2.2 Technology Readiness and Acceptance Model

The problem statement of this chapter is: **Cities are struggling due to the fragmentation of data and services across federal, regional, and local administrations**. Our research question considers the relationship between individual characteristics of decision makers and their intention to use data standards. The insights of this paper are valuable for organisations and government administrations that aim to speed up the adoption of Open Standards to raise interoperability in complex ecosystems. Also, it provides valuable observations for researchers that aim to study and predict the use intention of Data standards.

To find a predictor for the use of data standards in Smart Cities, we chose a deductive approach based on existing research. First, we derived the concept of a 'data standard'. Second, we gained a deeper understanding of the acceptance research stream. Muellers' [49] research indicated that the acceptance of standards can be embedded in TAM. This shaped the idea of building upon TAM, alike Pai [53] and Mueller [49]. The innovative aspect of this study is that we research the moderating role of people's individual characteristics [12, 49] in the context of the adoption of data standards.

As TAM was initially developed to predict technology adoption in settings where organizational objective mandate adoption, the model has limitations for users which are more free to choose between several alternatives [43]. Lin et al. [43] (p. 642) argue that "a model incorporating some individual difference variables is a necessary first step toward identifying and qualifying the psychological processes of the perceptions of a technology's value".

A model that considers individual differences is the Technology Readiness construct (TR). Parasuraman [54] defines the TR-construct as "people's propensity to embrace new technologies for accomplishing goals in life and work" (p. 308). The construct addresses four sub-dimensions which predict people's technology-related behaviour: optimism and innovativeness, which can boost TR, and discomfort and insecurity, which may reduce it [54].

The limitation of TAM is that it was initially designed to predict technology adoption in work environments, which makes it less applicable in contexts where the consumer has a higher autonomy [43]. The user's perception towards the usefulness and ease of use is determined by prior experience [58]. Therefore Lin et al. [43] broaden the applicability of TAM by augmenting it with the TR individual-specific construct into the TRAM. The findings of TRAM emphasise the impact of the user/individual characteristics and former experiences on the use intention. Also, the impact of usefulness and ease of use dominates the decision-making process of adoption behaviour, which can explain why a high TR score does automatically result in a high adoption.

2.2.2.1 Development of hypothesis

2.2.2.1.1 Hypothesis

In this chapter, we investigate the potential to use the TRAM-model (see Figure 2.1) to predict the use intention of data standards in Flanders. This will be done by making use of an adapted version of the TRAM-model as it was developed by Lin, Shih & Sher [43]. This model is based on TAM [19] and Technology Readiness Index (TRI) [55], see Figure 2.1.

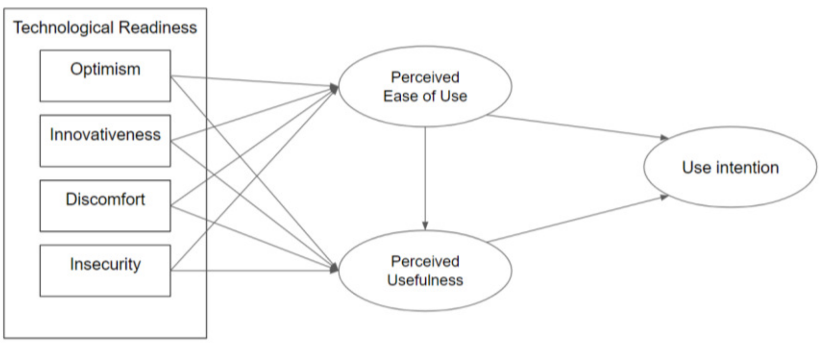


Fig. 2.1. Theoretical model based on TRAM [43].

Optimistic people generally expect that “good rather than bad things will happen to them” (p. 219) [64]. How they approach the world will have an impact on their attitude towards risk perception and acceptance in relation to technology [15]. Parasuraman [15] argued that optimism relates to “a positive view towards technology and trust that it will offer people more efficiency, flexibility, and control” (p. 311). Also, he concludes that this has a positive impact on technological readiness (TR). According to Lin et al. [43] PU and PEU have reconciling effects between TR and the use intentions. Based on these insights, Hallikainen [31] argued that optimism has a positive influence on both the PEU and the PU of digital services in the business-to-business healthcare sector. Building upon this prior research, we propose the following hypothesis:

- H1a: Optimism has a positive influence on perceived ease of use of data standards.
- H1b: Optimism has a positive influence on the perceived usefulness of data standards.

Garcia & Calantone [29] state that 'innovativeness' is generally used to assess the 'newness' of an innovation, where innovative products are labelled with a high degree of newness. Users that are characterised as 'innovative' adopt new ideas earlier than others (p. 22) [62]. Parasuraman [54] introduces the technological dimension and refers to "a propensity of being a technology pioneer and influencer" (p. 311). Venkatesh & Bala [75] identify a direct positive link between TR and the adoption of business process standards. Building upon these insights, we propose the following hypothesis:

- H2a: Innovativeness has a positive influence on perceived ease of data standards.
- H2b: Innovativeness has a positive influence on the perceived usefulness of data standards.

Discomfort attributes are defined as "a perceived lack of control regarding technology and the sense of being overwhelmed by it" (p. 311) [54]. Mukherjee and Hoyer [50] argue that the high-complexity features of technology products have a negative impact on product evaluation because of the user's learning cost. Albeit both studies hint towards a negative impact on the PEU and PU, some recent studies could not find a correlation [79, 30]. Building upon TRAM, we propose the following hypothesis:

- H3a: Discomfort has a negative influence on perceived ease of use of data standards.
- H3b: Discomfort has a negative influence on the perceived usefulness of data standards.

Insecurity "implicates a distrust of technology and the disbelief about its ability to work properly" (p. 311) [54]. Albeit TRAM hints towards a negative impact on the PEU and PU, some recent studies could not find a correlation [79, 30]. Building upon the insights of TRAM, we propose the following hypothesis:

- H4a: Insecurity has a negative influence on perceived ease of use of data standards.
- H4b: Insecurity has a negative influence on the perceived usefulness of data standards.

Scholars have been researching the relation between PU and PEU on UI, according to the initial TAM model, and reported that PU and PEU positively influence use intention [19]. However, studies on the use intention of data standards are very limited. Nevertheless Pai [53] refers to a HIS as a set of standards and his study demonstrates that PEU positively affects users' intention to use the HIS. Therefore, we propose the following hypothesis:

- H5a: The perceived usefulness has a positive influence on the intention to use data standards.
- H5b: The perceived ease of use has a positive influence on the intention to use data standards.

There are researchers [57] who have studied the relationship between perceived ease of use and perceived usefulness. In the context of data standards, both are surmised to be closely linked as the argument is such that a user who perceives data standards as 'easy to use' should in turn develop a tendency to perceive it as useful. Therefore, we hypothesize that:

- H6: The perceived ease of use has a positive influence on perceived usefulness of data standards.

2.2.2.1.2 Control variable

We have added the 'decision maker' as a control variable, by asking the respondents whether they qualify themselves as someone who takes decisions when it comes to purchasing or implementing new ICT principles. According to Mazis [46] decision makers are more receptive to novel information than non-decision makers.

2.3 Method: data collection and measurement scales

Data for this research was collected in June 2018 among people working in the public and private sector or as academics. An online questionnaire was developed in English and translated into Dutch. A qualitative pre-test was done by 20 respondents. The pre-test indicated that some of the questions about data standards were too conceptual. Also, the terms ‘technology’ and ‘standards’ proved to be too broad. Therefore, the questions were adjusted and definitions were added. Survey respondents were recruited using the snowball method [3]. This resulted in 338 responses, which after excluding unfinished answers and unanswered questions left us with 205 usable respondents. As respondents were recruited using the snowball method, we don’t have an indication of the response rate.

The study adopts measure items of TR from Parasuraman and Colby [55] consisting of a 16-item measurement instrument evaluating an individual’s propensity to adopt and use new technologies at work. The four dimensions of TRI, i.e., optimism, innovativeness, insecurity, and discomfort, consist of four measure items each. Moreover, four measure items of PEU and PU were adopted from Venkatesh and Bala [75] (see Table 1). Use intention for data standards and decision maker (or not) are measured using manifest variables. A seven-point Likert scale ranging from ‘1= Strongly agree’ to ‘7= Strongly disagree’ was used for TR, PEU, PU, and use intention.

Table 2.1. The questionnaire.

Construct	Questions	Cronbach’s Alpha
Optimism	1. New technologies contribute to a better quality of life	0.800
	2. Technology gives me more freedom of mobility	
	3. Technology gives people more control over their daily lives	
	4. Technology makes me more productive in my personal life	

Innovativeness	1. Other people come to me for advice on new technologies	0.807
	2. In general, I am among the first in my circle of friends to acquire new technology when it appears	
	3. I can usually figure out new high-tech products and services without help from others	
	4. I keep up with the latest technological developments in my areas of interest	
Discomfort	1. When I get technical support from a provider of a high-tech product or service, I sometimes feel as if I am being taken advantage of by someone who knows more than I do	
Insecurity	1. People are too dependent on technology to do things for them	0.678
	2. Too much technology distracts people to a point that is harmful	
	3. Technology lowers the quality of relationships by reducing personal interaction	
Perceived ease of use	1.Learning to work with data standards would be easy for me	0.931
	2. I find it easy to work with data standards to do what I want it to do.	
	3. It is easy for me to become skillful at using data standards.	
	4. I find it easy to use data standards.	
Perceived usefulness	1. The use of data standards in my job, enables me to accomplish tasks more quickly.	0.886
	2. The use of data standards in my job, increases my productivity.	
	3. The use of data standards in my job, makes it easier to do my job.	
	4. The use of data standards in my job, is very useful.	
Use intention	In the future months, I will make use of data standards in my job.	
Decision maker	I see myself as someone who takes decisions when it comes to purchasing or implementing new ICT principles.	

2.4 Data analysis and results

2.4.1 Descriptive statistics

In all, 205 respondents completed the questionnaire (21% female and 79% male). 1% has an age of less than 24, 22% an age between 25 and 34, 40% between 35 and 44, 27% between 45 and 54, 10% above 55 years. Regarding the respondent's educational level, 27% have a bachelor's degree, 66% have a master's degree, 4% have a PhD degree. Only 3% only have a degree of secondary education (see Figure 2.2).

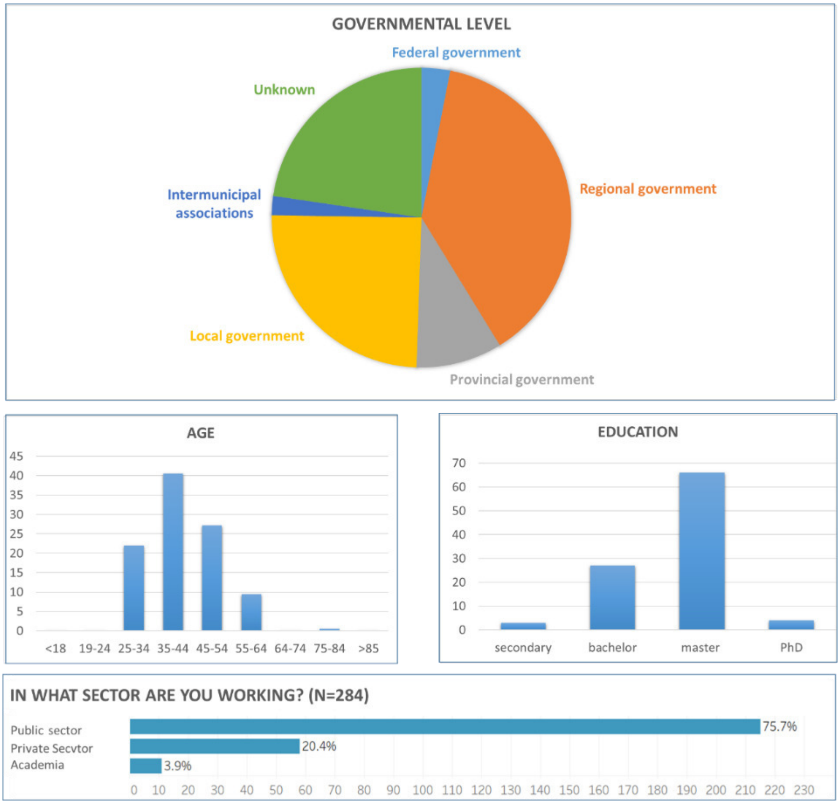


Fig. 2.2. A sample profile of the respondents (sector, age, education and governmental level).

For the professional experience, 78% of the respondents are active in the public sector, 19% in the private sector and 3% in academia. From all the respondents active in the public sector; 3% is active in the federal government (including the federal Digital Transformation Office), 37% in the regional government, 9% in the provincial government, 24% in the local government, and 2% in intermunicipal associations. Finally, 24% of the respondents did not provide this information (question was not a required one) (see Figure 2.2).

Results show that 56% of the respondents in this sample saw themselves as a decision maker in their organisation on purchasing or implementing new ICT principles or technology. 88% of our respondents working in the public sector reported that their organisation makes use of data standards.

T-tests showed no difference between gender and organisation (public, sector, academia) when it comes to making decisions. Also, we detected that people that qualify themselves as a decision maker, are significantly (on the 0.05 level) more innovative, than the respondents who indicated that they are a non-decision maker.

2.4.2 **Validity and reliability**

The validity of the TRAM approach was tested using convergent validity and discriminant validity. A measurement model with seven latent constructs and 26 observed variables was fit using lavaan version 0.6-2.1268 [63] in R version 3.4.3 [59]. For the model fit assessment, we evaluated the Robust Root Mean Square Error of Approximation (RMSEA) and the Robust Comparative Fit Index (CFI). CFI should be larger than .95, RMSEA values should be .05 or lower to indicate a good fit. Small deviations from these standards are, however, acceptable [45].

Reliability was measured based on the Cronbach's Alpha score of the constructs. As a rule of thumb, a Cronbach's Alpha rho greater than 0.7 is considered acceptable. We can conclude that the values show acceptable reliability (see Table 2.1).

2.4.3 Results

The overall measurement model provides an adequate fit with chi^2= 419.110 (df = 259); p<0.000, CFI = 0.925 and RMSEA = 0.055. Standardised regression loadings for all measures exceeds 0.60 except for seven items. Based on these low factor loadings (below 0.6), which indicate that the items are not valid and would, therefore, falsify results, we decided to eliminate four items of which the loadings were extremely low. Low factor loadings can be problematic because questions with low loadings do not measure the intended element. After these modifications, the final model demonstrated an acceptable fit with chi^2= 278.790 (df = 174); p<0.000., CFI = 0.948 and RMSEA = 0.054. In Figure 2.3 shows the structural model.

Table 2.2 provides an overview of the hypothesis results. Hypothesis H1a and H1b are rejected because the correlation is not statistically significant. Optimism concerns the positive attitude towards technology such as one's perceived level of control, flexibility, convenience, and efficiency [54], while it is essential that people can assure that the technologies are under their control [16]. The results show that whether someone is a technological optimist is not related to the PEU and the PU of data standards. Other factors might be more relevant.

As expected, we obtained a positive relationship between innovativeness and both PEU and PU (H2a and H2b). This entails that innovativeness has a positive influence on perceived ease and usefulness of data standards. This can be explained by the fact that innovative people are more open to new ideas in general [39]. An individual's level of innovative attitude has been shown to be a key element in his/her acceptance of new technologies [4]. Innovative individuals are eager to learn new technologies and to understand and to use them while doing their tasks. This increases their technology acceptance rate automatically [72]. We assume that innovative people are more familiar with new technological concepts, such as data standards.

Table 2.2. The hypothesis test results.

	Hypotheses	**Estimate**	**Std. Error**	**Z-values**	**P-values**	**Std. All**	**Decision**
H1a	Optimism → Perceived ease of use	0.114	0.156	0.734	0.463	0.065	Not supported
H1b	Optimism → Perceived usefulness	0.109	0.146	0.748	0.454	0.062	Not supported
H2a	Innovativeness → Perceived ease of use	0.397	0.126	3.147	0.002	0.287	Supported
H2b	Innovativeness → Perceived usefulness	0.297	0.122	2.440	0.015	0.216	Supported
H3a	Discomfort → Perceived ease of use	-0.002	0.067	-0.035	0.972	-0.003	Not supported
H3b	Discomfort → Perceived usefulness	0.203	0.063	3.199	0.001	0.236	Supported
H4a	Insecurity → Perceived ease of use	-0.154	0.169	-0.908	0.364	-0.081	Not supported
H4b	Insecurity → Perceived usefulness	-0.317	0.165	-1.924	0.054	-0.169	Not supported
H5a	Perceived usefulness -> use intention	0.095	0.089	1.064	0.287	0.081	Not supported
H5b	Perceived ease of use -> use intention	0.317	0.089	3.560	0.000	0.271	Supported
H6	Perceived ease of use → perceived usefulness	0.311	0.074	4.205	0.000	0.313	Supported

Hypothesis H3a is not supported because the correlation is not statistically significant. Hypothesis 3b is supported and implies that discomfort is positively correlated with PU of data standards. It implies that if people are uncomfortable with technology, they will be more likely to perceive data standards as being useful. These results are not consistent with previous literature where discomfort negatively influenced PU [34]. This seems counterintuitive. However this discomfort could lead to new solutions that mitigate the discomfort. Also, people feeling more uncomfortable with technology may have been accustomed to using existing technologies which do not meet their needs and therefore perceive data standards as useful [38].

Hypotheses H4a and H4b are both rejected because the correlation is not statistically significant. This means that there are other predictors that influence this PEU and PU of data standards [38].

Hypothesis H5a is rejected because the correlation is not statistically significant. In line with the findings of Lin et al. [43] we see that hypothesis H5b is supported, demonstrating the positive influence of PEU on the intention to use data standards. This proves that the 'user-friendliness' of data standards is associated with the use intention. Factors that contribute to higher perceived unfriendliness of data standards may be for example the conceptual or intangible characteristic of data standards or the implementation cost. Because of this high cost, (potential) users of data standards could lose focus on the advantages and the ease of use of the data standards. In other words, barriers such as cost reduce the perception on the ease of use of data standards, allowing the users to develop a negative attitude. In turn, this leads to an unwillingness to engage in the use of data standards [57].

Lastly, hypothesis H6 is supported. It is widely acknowledged that PEU contributes to PU [36, 44, 65, 74]. This is based on the theoretical argument that some user-friendly technologiess could be perceived as useful, but not all useful technologies are user-friendly [30]. PU is influenced by the PEU, which means that if data standards are perceived as easy to use, they are also perceived as more useful [38].

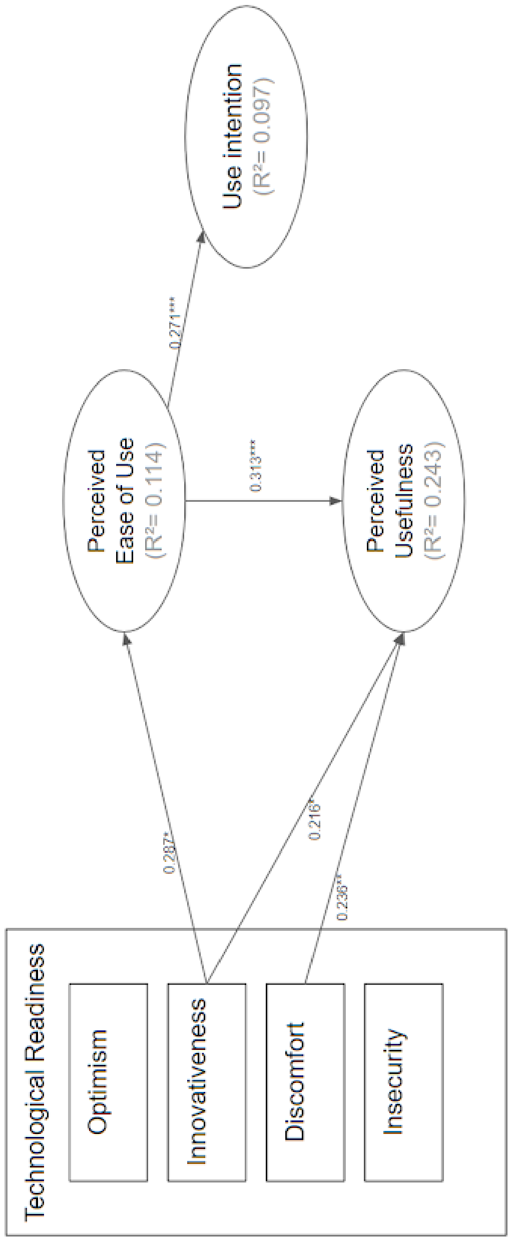


Fig. 2.3. Structural model (standardised paths) of the total sample.

We can conclude that a low effect size is measured for PEU (R^2= 0.114). Figure 2.3 shows that PEU is driven by one determinant (innovativeness) derived from the TRAM model. Another low effect size is measured for use intention (R^2= 0.097) as we see that there is only one determinant (PEU) that contributes to this construct. Finally, we see a moderate effect for PU (R^2= 0.243) predicted by innovativeness and discomfort.

2.5 Discussion and conclusion

In this chapter, we investigated the potential to use the TRAM-model to predict the use intention of data standards, in specific we looked at the relationship between individual characteristics of the decision and non-decision makers in Flanders and their intention to use data standards. This study was the first to apply the TRAM model on the use intention of data standards. Also, we applied the TRI 2.0 scale, a recently developed scale by [55]. This is a more recent version of the TRI 1.0, characterised by the fact that the new questionnaire is shorter and more adapted to current technology developments such as the Internet, the use of smartphones and, e.g., apps that are used on these platforms.

We detected a positive correlation between the respondent’s perception on the ease of use of data standards and the perceived usefulness. Also, our analysis indicates that the respondent’s perceptions on data standards are positively correlated with their intentions to use it. The study also indicated the positive correlation between perceived ease of use and the use intention of data standards.

This research is subject to several limitations that need to be considered. First and foremost, we saw that one characteristic of the TRAM model (innovativeness), predicts perceived ease of use and perceived usefulness of data standards within our sample of respondents. The low effect sizes show that the TRAM model is not a good fit for this context. The characteristics of our respondents indicate that they are a homogeneous group of people active in information management and have a high level of education. These different kinds of homogeneity may have biased our results similar to the study of Godoe & Johansen [30]. Moreover, creating the scores for PU and PEU was often based on purely subjective judgement of respondents as some did not have prior experience with (implementing) data standards, as such this subjective appraisal of performance and effort do not necessarily reflect objective reality [19]. Second, Although the pre-test indicated that some of the questions about data standards were too conceptual, and although the questions were adjusted and working definitions for concepts such as data standards were added, respondents showed a wide diversity in interpreting the concepts that were polled, thus lessening the reliability of the results. Third, given the low availability of literature on the relation between personality traits and adoption of data standards, a more qualitative approach might have been more helpful as it allows a more exploratory and broader research approach.

Our results indicate that respondents that score high on innovativeness have a higher intention to use data standards. According to Melas et al. [47], it is essential to target these early adopters first, as they can influence their peers and the diffusion process. The diffusion process is the crucial stage where “more members of the social system also adopt the same innovation” (p. 87) [33]. To speed up the adoption of Open Standards and raise interoperability in complex ecosystems, we should focus on these early adopters. Our research results show that personality traits are only influential to a lesser degree in terms of adopting data standards. Albeit the TRAM-model reveals that innovativeness is an important influencer for the use intention of data standards, we expect that other parameters which are not included in the model might have an impact on the use intention, such as organisational factors and potential network effects, because data standardisation is a multi-stakeholder activity as well (e.g., coordination between agencies, the context of policy framework, …). The governance model in Flanders, that finds its roots in geospatial e-services and standards, can be characterised as a mix of hierarchical and network governance [11]. Network coordination has an important impact on addressing complex problems [56]. Therefore, we suggest to also explore the effects of network governance in order to speed up the adoption of Open Standards to raise interoperability in complex ecosystems. We suggest researching the impact of organisational impediments (e.g., lack of support from top management) and economic impediments [52]. As Lee & Yu [40] suggest, raising the organisational competencies (e.g., providing user-friendly tools, training and success stories) heightens the perceived ease of use and use intention. Also, our research suggests that the characteristics of the data standard (complexity, cost, relative advantage, and impact) might influence the adoption [17].

2.6 References

[1] Ajzen, I. (1991). The theory of planned behavior. *Organizational behavior and human decision processes*, 50(2), 179-211.

[2] Badii, C., Bellini, P., Cenni, D., Difino, A., Nesi, P., & Paolucci, M. (2017). Analysis and assessment of a knowledge based smart city architecture providing service APIs. *Future Generation Computer Systems*, 75, 14-29.

[3] Biernacki, P., & Waldorf, D. (1981). Snowball sampling: Problems and techniques of chain referral sampling. *Sociological methods & research*, 10(2), 141-163.

[4] Brancheau, J. C., & Wetherbe, J. C. (1990). The adoption of spreadsheet software: testing innovation diffusion theory in the context of end-user computing. *Information systems research*, 1(2), 115-143.

[5] Brunsson, N., Rasche, A., & Seidl, D. (2012). The dynamics of standardization: Three perspectives on standards in organization studies. *Organization Studies*, 33(5-6), 613-632.

[6] Brynjolfsson, E., & Kahin, B. (Eds.). (2002). *Understanding the digital economy: data, tools, and research*. Cambridge: MIT press.

[7] Buyle, R. (2017). Towards interoperability in the public sector. In *Proceedings of the 16e International Semantic Web Conference* (pp. 1-8). Vienna, Austria: Springer.

[8] Buyle, R., Colpaert, P., Van Compernolle, M., Mechant, P., Volders, V., Verborgh, R., & Mannens, E. (2016). Local Council Decisions as Linked Data: a proof of concept. In *The International Semantic Web Conference (ISWC). CEUR Workshop Proceedings* (CEUR-WS. org).

[9] Buyle, R., De Vocht, L., Van Compernolle, M., De Paepe, D., Verborgh, R., Vanlishout, Z., ... & Mannens, E. (2016). Oslo: open standards for linked organizations. In *Proceedings of the International Conference on Electronic Governance and Open Society: Challenges in Eurasia* (pp. 126-134). ACM.

[10] Buyle, R., Van Compernolle, M., Vlassenroot, E., Vanlishout, Z., Mechant, P., & Mannens, E. (2018). "Technology readiness and acceptance model" as a predictor for the use intention of data standards in smart cities. *MEDIA AND COMMUNICATION*, *6*(4), 127–139.

[11] Chantillon, M., Crompvoets, J., & Peristeras, V. (2017). The governance landscape of geospatial e-services—The Belgian case. ISPRS International *Journal of Geo-Information*, 6(9), 282-307.

[12] Chen, M. (2003). Factors affecting the adoption and diffusion of XML and Web services standards for E-business systems. *International Journal of Human-Computer Studies*, *58*(3), 259-279.

[13] Chesbrough, H., Vanhaverbeke, W., & West, J. (2006). *Open Innovation: Researching a New Paradigm*. Oxford: Oxford University Press.

[14] Colpaert, P. (2017). *Publishing transport data for maximum reuse* (Doctoral dissertation, Ghent University).

[15] Costa-Font, J., Mossialos, E., & Rudisill, C. (2009). Optimism and the perceptions of new risks. *Journal of risk research*, 12(1), 27-41.

[16] Dabholkar, P. A. (1996). Consumer evaluations of new technology-based self-service options: an investigation of alternative models of service quality. *International Journal of research in Marketing*, 13(1), 29-51.

[17] Damanpour, F. & Schneider, M. (2009). Characteristics of Innovation and Innovation Adoption in Public Organizations: assessing the Role of Managers. *Journal of Public Administration Research and Theory*, 19(3), 495-522.

[18] Davies, J., Harris, S., Crichton, C., Shukla, A., & Gibbons, J. (2008). Metadata standards for semantic interoperability in electronic government. In *Proceedings of the 2nd international conference on Theory and practice of electronic governance* (pp. 67-75). ACM.

[19] Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: a comparison of two theoretical models. *Management science*, 35(8), 982-1003.

[20] Davis, F. D. (1985). *A technology acceptance model for empirically testing new end-user information systems: Theory and results* (Doctoral dissertation, Massachusetts Institute of Technology).

[21] Delone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information systems success: a ten-year update. *Journal of management information systems*, 19(4), 9-30.

[22] Del-Aguila-Obra, A. R., Padilla-Meléndez, A., & Abouseada, A. (2014). *Use and Acceptance of Social Technologies by Internet Banking Services Users in the Middle East*. Paper presented at XXIV Congreso Nacional de ACEDE, Castelló.

[23] De Soir, E., Zech, E., Versporten, A., Van Oyen, H., Kleber, R., Mylle, J., & van der Hart, O. (2015). Degree of exposure and peritraumatic dissociation as determinants of PTSD symptoms in the aftermath of the Ghislenghien gas explosion. *Archives of public health*, 73(1), 21.

[24] European Commission. (2013). *A vision for public services.* Luxembourg: Publications Office.

[25] European Commission. (2014). *Study on eGovernment and the Reduction of Administrative Burden*. Luxembourg: Publications Office.

[26] European Commission (2017). *ISA: European Interoperability Framework (EIF)*. Luxembourg: Publications Office.

[27] Farrell, J., & Simcoe, T. (1996). Choosing the rules for formal standardization. University of California, Berkeley, working paper.

[28] Funk, J. L. (2001). *Global competition between and within standards: the case of mobile phones*. London: Palgrave Macmillan.

[29] Garcia, R., & Calantone, R. (2002). A critical look at technological innovation typology and innovativeness terminology: a literature review. *Journal of product innovation management*, 19(2), 110-132.

[30] Godoe, P. and Johansen, T.S., 2012. Understanding adoption of new technologies: Technology readiness and technology acceptance as an integrated concept. *Journal of European Psychology Students*, 3(1), pp.38–52.

[31] Hallikainen, H., & Laukkanen, T. (2016, June). How technology readiness explains acceptance and satisfaction of digital services in B2B healthcare sector?. In *PACIS 2016 proceedings.* (pp. 294-306). New York: Elsevier.

[32] Hautekiet, P. (2009). Interbestuurlijk e-government: aandachtspunten en aanbevelingen. In Opsomer, B. (Eds.), *Memorandum 2009 - 2014 - aanbevelingen voor een krachtig bestuurlijk beleid* (edition, pp. 171 - 196). Brussels: VLABEST

[33] Hoffmann, V. (2007). The diffusion of innovations – the Hohenheim concept. In Hoffmann, V (Eds.), *Knowledge and Innovation Management Module Reader* (pp. 87-96). Hohenheim: Hohenheim University.

[34] Igbaria, M., Schiffman, S. J., & Wieckowski, T. J. (1994). The respective roles of perceived usefulness and perceived fun in the acceptance of microcomputer technology. *Behaviour & Information Technology*, 13(6), 349-361.

[35] Keen, P. G. (1981). Information systems and organizational change. *Communications of the ACM*, 24(1), 24-33.

[36] King, W. R., & He, J. (2006). A meta-analysis of the technology acceptance model. *Information & management*, 43(6), 740-755.

[37] Kolehmainen, J., Irvine, J., Stewart, L., Karacsonyi, Z., Szabó, T., Alarinta, J., & Norberg, A. (2016). Quadruple helix, innovation and the knowledge-based development: lessons from remote, rural and less-favoured regions. *Journal of the Knowledge Economy*, 7(1), 23-42.

[38] Kuo, K. M., Liu, C. F., & Ma, C. C. (2013). An investigation of the effect of nurses' technology readiness on the acceptance of mobile electronic medical record systems. *BMC medical informatics and decision making*, 13(1), 88-101.

[39] Kwang, N. A., & Rodrigues, D. (2002). A Big-Five Personality profile of the adaptor and innovator. *The Journal of Creative Behavior*, 36(4), 254-268.

[40] Lee, S., & Yu, J. (2015). Comparative study of BIM acceptance between Korea and the United States. *Journal of Construction Engineering and Management*, 142(3), 1-9.

[41] Lieberman, J., Ryan, A. (2017). OGC Underground Infrastructure Concept Study Engineering Report http://www.opengis.net/doc/PER/uicds

[42] Lin, C. H., Lin, I. C., Roan, J. S., & Yeh, J. S. (2012). Critical factors influencing hospitals' adoption of HL7 version 2 standards: An empirical investigation. *Journal of medical systems*, 36(3), 1183-1192.

[43] Lin, C. H., Shih, H. Y., & Sher, P. J. (2007). Integrating technology readiness into technology acceptance: The TRAM model. *Psychology & Marketing*, 24(7), 641-657.

[44] Lin, C. H., Shih, H. Y., Sher, P. J., & Wang, Y. L. (2005, July). Consumer adoption of e-service: Integrating technology readiness with the technology acceptance model. In *Technology management: A unifying discipline for melting the boundaries* (pp. 483-488). IEEE.

[45] Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural equation modeling*, 11(3), 320-341.

[46] Mazis, M. B. (1972). Decision-making role and information processing. *Journal of Marketing Research*, 9(4), 447-450.

[47] Melas, C. D., Zampetakis, L. A., Dimopoulou, A., & Moustakis, V. (2011). Modeling the acceptance of clinical information systems among hospital medical staff: an extended TAM model. *Journal of biomedical informatics*, 44(4), 553-564.

[48] Mills, S. (2013). International standards and the IEEE standards association in the emerging global economy. *IEEE Aerospace and Electronic Systems Magazine*, 28(7), 6-11.

[49] Mueller, T., Dittes, S., Ahlemann, F., Urbach, N., & Smolnik, S. (2015, January). Because everybody is different: towards understanding the acceptance of organizational IT standards. In *System Sciences (HICSS), 2015 48th Hawaii International Conference* (pp. 4050-4058). IEEE.

[50] Mukherjee, A., & Hoyer, W. D. (2001). The effect of novel attributes on product evaluation. *Journal of Consumer Research*, 28(3), 462-472.

[51] Open Stand (2012) OpenStand: Principles for The Modern Standard Paradigm. Available at: http://openstand.org/ (accessed: 7.5.2018).

[52] Ouma, F. K. (2014, May). Impediments to interagency statistical information sharing amongst government agencies in Uganda: A G2G adoption. In *IST-Africa Conference Proceedings,* 2014 (pp. 1-11). IEEE.

[53] Pai, F. Y., & Huang, K. I. (2011). Applying the technology acceptance model to the introduction of healthcare information systems. *Technological Forecasting and Social Change*, 78(4), 650-660.

[54] Parasuraman, A. (2000). Technology Readiness Index (TRI) a multiple-item scale to measure readiness to embrace new technologies. *Journal of service research*, 2(4), 307-320.

[55] Parasuraman, A., & Colby, C. L. (2015). An updated and streamlined technology readiness index: TRI 2.0. *Journal of service research*, 18(1), 59-74.

[56] Provan, K. G., & Kenis, P. (2008). Modes of network governance: Structure, management, and effectiveness. *Journal of public administration research and theory*, 18(2), 229-252.

[57] Ramayah, T., & Ignatius, J. (2005). Impact of perceived usefulness, perceived ease of use and perceived enjoyment on intention to shop online. ICFAI *Journal of Systems Management* (IJSM), 3(3), 36-51.

[58] Rao, A. R., & Monroe, K. B. (1988). The moderating effect of prior knowledge on cue utilization in product evaluations. *Journal of consumer research*, 15(2), 253-264.

[59] R Core Team (2017). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*. Retrieved from https://www.R-project.org/.

[60] Reed, C., Buehler, K., & McKee, L. (2015). OGC consensus: how successful standards are made. ISPRS International *Journal of Geo-Information*, 4(3), 1693-1706.

[61] Roe, J. W. (1916). *English and American tool builders*. New York, London, McGraw-Hill.

[62] Rogers, E.M. (2003). *Diffusion of innovations* (5th ed.). New York: Free Press.

[63] Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1-36.

[64] Scheier, M. F., & Carver, C. S. (1985). Optimism, coping, and health: assessment and implications of generalized outcome expectancies. *Health psychology*, 4(3), 219-247.

[65] Schepers, J., & Wetzels, M. (2007). A meta-analysis of the technology acceptance model: Investigating subjective norm and moderation effects. *Information & management*, 44(1), 90-103.

[66] Sheridan, J. (2010). The nuts and bolts of opening government data. *Digital Content Quarterly*, 4(2), 6-7. Retrieved from https://sca.jiscinvolve.org/wp/files/2010/04/sca_dcquarterly_02_apr10-final.pdf

[67] Simcoe, T. (2003). Committees and the creation of technical standards. Unpublished working paper, University of California at Berkeley.

[68] Simcoe, T. (2006). Open standards and intellectual property rights. Open innovation: Researching a new paradigm. In Chesbrough, H., Vanhaverbeke, W., & West, J. (Eds.), Open Innovation: Researching a New Paradigm (pp. 161-183). Oxford: Oxford University Press.

[69] Smart Flanders (2018). Open Data Charter: 20 principes. *Stedenbeleid - het Agentschap Binnenlands Bestuur.* Retrieved from https://smart.flanders.be/open-data-charter/

[70] Stango, V. (2004). The economics of standards wars. Review of network economics, 3(1).

[71] Steen , T., & Wayenberg, E. (2003). Local governance in Flanders. In B. Denters , O. van Heffen , J. Huisman, & P.-J. Klok (Eds.), *The rise of interactive governance and quasi-markets* (pp. 261–276). Dordrecht: Kluwer Academic Publishers.

[72] Turan, A., Tunç, A. Ö., & Zehir, C. (2015). A theoretical model proposal: Personal innovativeness and user involvement as antecedents of unified theory of acceptance and use of technology. In *Proceedings of the 4th International Conference on Leadership, Technology, Innovation and Business Management* (pp. 43-51). Yıldız: Elsevier.

[73] Van Dijk, J. A. G. M., Peters, O., Ebbers, W., (2008). Explaining the acceptance and use of government internet services: A multivariate analysis of 2006 survey data in the Netherlands. *Government Information Quarterly*, *25*(3), 379-399.

[74] Venkatesh, V. (2000). Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Information systems research*, 11(4), 342-365.

[75] Venkatesh, V., & Bala, H. (2012). Adoption and impacts of interorganizational business process standards: Role of partnering synergy. *Information Systems Research*, 23(4), 1131-1157.

[76] Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, 27(3), 425-478.

[77] Viale Pereira, G., Cunha, M. A., Lampoltshammer, T. J., Parycek, P., & Testa, M. G. (2017). Increasing collaboration and participation in smart city governance: a cross-case analysis of smart city initiatives. *Information Technology for Development*, 23(3), 526-553.

[78] Vlaams Parlement (2008). Decreet houdende de ontsluiting en de uitwisseling van informatie over ondergrondse kabels en leidingen. *Belgisch Staatsblad,* 178(135), 23945.

[79] Walczuch, R., Lemmink, J., & Streukens, S. (2007). The effect of service employees' technology readiness on technology acceptance. *Information & management*, 44(2), 206-215.

CHAPTER 3. PROCESS AND METHOD FOR DEVELOPING OPEN DATA STANDARDS

If you want to go fast, go alone. If you want to go far, go together.

— African Proverb.

.

This chapter outlines a process and method for developing Open Data Standards. We address the interoperability hurdles at the different governmental levels and examine how the bottom-up OSLO programme tackled these obstacles. Finally, we detail a generic process and method and provide practical insights on how to raise interoperability. This chapter is based on the paper: ’OSLO - Open Standards for Linked Organizations’ [3].

3.1 Introduction

One of the most widespread e-government best practices is the 'once-only principle[27]' which states that citizens and businesses have to provide administrative information only once to a public administration, avoiding administrative burden. To achieve this, administrations must be able to share and reuse this information across different applications and processes.

A good example are local governments in Flanders, which provide over 800[28] different products and services. To support their processes and service delivery, they use back-office applications from different software vendors. These domain specific applications are organized as vertical processes, requesting administrative data from citizens and business which often cannot be reused by other applications, causing data silos.

The OSLO program transformed IT-service delivery efforts in the region of Flanders (Belgium) in fundamental ways. Its strategy focuses on semantic agreements and machine-readable data which softens the existing data silos on various governmental levels and facilitates the once-only principle. This paper reports on the development, methodology, and the outcome of the OSLO program.

OSLO started in February 2012 and the first phase has ended in 2015. The project was the result of a public-private partnership initiated bottom-up by the Flemish Organization for ICT in Local Government (V-ICT-OR), and co-funded by Flemish ICT service providers and Flemish government administrations. The project was also supported by a wider community, including Local, Regional and Federal administrations, non-profit organizations, academic partners, and the EC program Interoperability Solutions for European public administrations (ISA).

---

[27] https://joinup.ec.europa.eu/community/once-only-principle/home
[28] http://productencatalogus.vlaanderen.be/fiches

In Flanders there are various governmental levels with their own jurisdiction, presented simplified in Figure 3.1. Each level has various data sources and applications:

- The *authoritative* (official, established) sources about people and enterprises are the federal (national) sources in Belgium (I).
- The *authentic* sources (embedded in policies) obliged and supported by the regional government (II) with address information and geographical locations are a regional responsibility.
- At the local level, 300 municipalities provide a variety of services to enable public service delivery for citizens and business (III). Local governments consume the authentic data from the federal and regional government and are often responsible for the creation and maintenance of authoritative data at the other administrative levels (I, II). A lack of interoperable information products at local level has led to redundant and repeated data.

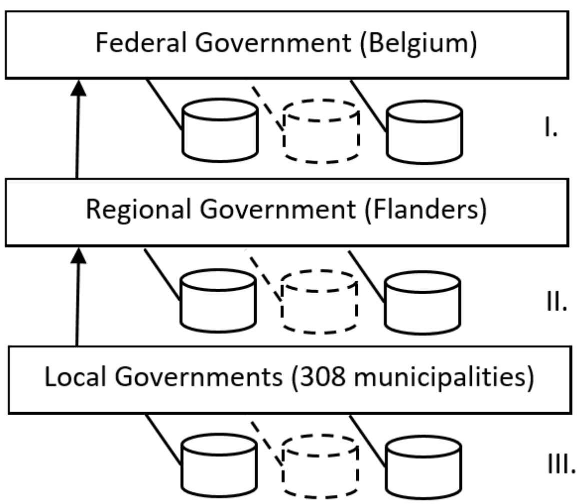


Fig. 3.1. Simplified view of the various governmental levels in Flanders, Belgium.

OSLO focusses on a broader framework, in line with the EIF. At the start of the program the governance was at the level of membership organizations (ICT, secretary, financial managers) of local governments, regional, and federal administrations, referred to as the steering committee.

The program focused on semantic interoperability. Semantic interoperability “enables organizations to process information from external sources in a meaningful manner. It ensures that the precise meaning of exchanged information is understood and preserved throughout exchanges between parties”[29]. The project had two main tracks: (i) gain ‘political support and adoption’ and (ii) develop the ‘semantic agreement’. Political support is essential, for collecting sponsoring and gaining authority and engagement. Semantic agreement is expressed in a domain model.

Semantic interoperability facilitates information exchange without a specific translation step. Two organizations are semantically interoperable if they know how to interpret data from each other and can reuse each other’s data directly. Public administrations should support and monitor these information management processes which should lead to better interoperability. For example, providing direct feedback such as an interoperability score when a new dataset has been published can help the adoption of the available vocabularies [4].

OSLO provides three variants of its domain model:

- a human readable specification, covering the domain model itself, examples and a conformance statement.
- a technical implementation serialized to XML.
- an implementation following the RDF.

[29] http://ec.europa.eu/isa/documents/isa_annex_ii_eif_en.pdf, p28

3.1.1 Overview

This chapter provides an in-depth insight into the process and methodology of developing better public services in Flanders. We expect that this process and methodology can be applied by other administrations in order to facilitate their Open Government Data policy and to support the development of citizen centred e-government services.

We explain the importance of interoperability in Section 3.1 and motivate why it is crucial to focus on different interoperability levels. More specific, we will explain how both ‘political support and adoption’ and ‘semantic agreements’ are cornerstones for ensuring interoperability. Section 3.2 describes the research goal and the applied method, followed by a description of how OSLO created the conditions to reach ‘political support and adoption’ in Section 3.3. Here we also discuss the process to reach ‘semantic agreement’ and the domain model, which is the starting point for the implementation. Section 3.4 illustrates the different characteristics of OSLO. After a short discussion in Section 3.5, this chapter ends with Section 3.6 where we present the conclusions.

3.1.2 Background

Public data often has a location-related component: “It is estimated that 80% of the informational needs of local government policymakers are related to geographic location.” [8].

In many cases the location is the anchor to which other information or data is linked, for example: construction permits for residential houses or environmental permits for industrial areas. Despite the existence of these sources (Figure 3.1), the non-availability of interoperable information products related to public services led to local governments and their software suppliers being unable to connect to these data sources. The processes that drive these products are often digitalized in separate systems. Due to this, shadow databases arise which lead to lots of redundant and repeated data. The quality of this information, ultimately delivered through e-government or other public services to citizens, is not as good as the authoritative data sources and certainly not as guaranteed or supported. This leads to a fragmented view of the public service concept which impacts the quality and the efficiency of public services. This fragmented view is a major obstacle for the development of citizen and businesses centred services because data sources were developed as independent products, each modelling information differently. This causes unnecessary translation steps which triggers multiple investments for interlinking data. Citizens benefit from once-only information delivery approaches, public administrations should not request information from citizens and businesses that already has been provided in another context, thus increasing government effectiveness and efficiency, and decreasing administrative burden [7, 15].

In Europe, various frameworks have emerged to safeguard interoperability in the deployment of e-government services, both at national and at European level [9]. Methodologies for linking government data as such are not new: many guidelines considering applications, methodology, coverage, and quality exist [20]. In particular the Interoperability Solutions for public Administrations (ISA) programme, now in its second chapter, promotes semantic interoperability among the European Union member states[30].

3.1.3 Goals of the programme

OSLO is an interoperability facilitator. Data cannot pass by default through different applications, because each application models the 'real world' from a (slightly) different perspective, therefore OSLO:

- transforms the delivery of public and government services so that citizens and businesses have to provide

[30] http://ec.europa.eu/isa/actions/01-trusted-information-exchange/1-1action_en.htm

their information only once and it is shared across the different applications and informations systems.

- aggregates information from different national, regional, and local e-government IS or combines existing services to create new ones.
- creates machine-readable public service descriptions that are reusable (following the Linked Open Government Data paradigm) and enable functionalities such as automated service discovery and composition[31].

With an inventory[32] of problems related to the exchange of information for local authorities in Flanders, the local government 'promoters' of OSLO created the necessary support at the local level and crowd funded the initiative. Among the initial sponsors were Flemish ICT service providers, major cities and a regional government Administration. In a parallel process, the promoters created a coalition of willing administrations at various government levels, by explaining the impact of those interoperability problems on citizens, businesses, and administrations. Next, collaboration with the ISA programme was realized[33] in order to create more stable standards (because the governance is at the EU level) and to create a more authoritative setting.

According to the EC Directory General for Communications Networks, Content and Technology (DG CONNECT)[34] boundaries between public and private services will fade. The increased connectivity of citizens and businesses, the possibility for people to work together, perform tasks, and distribute workload regardless of distance and boundaries as well as the availability of previously closed

---

[31] https://joinup.ec.europa.eu/asset/core_public_service/description

[32] http://contactinformatie.v-ict-or.be/documentation/OSLO_discussienota_inventarisatiefase%201_0.pdf

[33] https://joinup.ec.europa.eu/community/semic/news/flemish-oslo-standard-become-local-extension-isa-core-vocabularies

[34] https://ec.europa.eu/digital-single-market/en/dg-connect

information and data, implies that government tasks can also be performed — completely or in part — by citizens or companies. Potential near-future applications could involve the reuse of own data to have contact data delivered to the energy supplier, behind the scenes, without having to fill in yet another form, or to validate if one has the required vaccines before traveling.

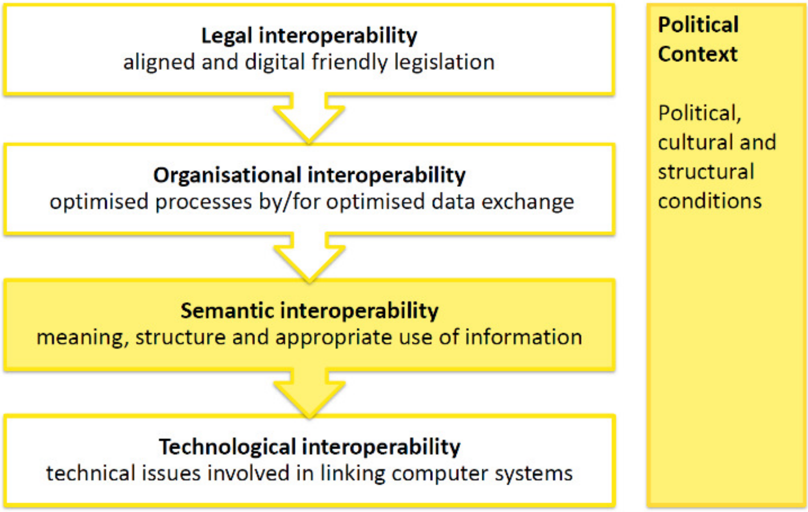


Fig. 3.2. European Interoperability Framework[35]

The ISA programme promotes interoperability across multiple interoperability levels between member state's borders and public service sectors, see Figure 3.2. One of its key components is the EIF[36]. EIF is a set of recommendations which specify how administrations, businesses, and citizens communicate with each other within the EU and across borders. These interoperability levels are defined as legal, organizational, semantic, and technical within a political context.

[35] http://ec.europa.eu/isa/documents/isa_annex_ii_eif_en.pdf, p. 21.
[36] http://ec.europa.eu/isa/documents/isa_annex_ii_eif_en.pdf, p. 3

In this chapter we focus on the semantic interoperability and the political context in Flanders as these levels are in line with the OSLO programme as described in the introduction.

3.2 Research goal and method

This study aims to represent and validate this first OSLO programme which delivers a canonical data model and a process to arrive at an agreement for facilitating better semantic interoperability. In this study, we describe the involved stakeholders, the process, our methodology, the implementation of OSLO as three PoCs (Proof of Concept), and the final output & outcome of the OSLO programme. We also elaborate on the conditions and contextual factors that influenced and shaped the implementation of OSLO.

This study used an inductive approach as data was gathered via action-research [12]. This approach refers to the involvement of researchers as co-practitioners in the setting under study and the attention paid to the context where the events took place [13]. Additional data was gathered via desk research.

3.3 Unpacking the semantic process

In this section we will compare the process and methodology of developing the OSLO semantic agreement with the approach defined by the ISA programme. We will discuss the 'stakeholders' and the steps to reach the semantic agreements among these stakeholders (the 'specifications process'). The 'methodology' describes how the model will be developed. The model is a starting point for the 'implementation'.

Our methodology was based on ISA best practices: 'Process and methodology for developing Core Vocabularies'. 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 [17]. These Core Vocabularies cover the semantics of a small set of generic concepts which support semantic interoperability among public administrations in Europe and were a starting point for new semantic agreements in the context of administrations in Flanders. Meanwhile these vocabularies are alse being adopted outside Europe, including the IMI Core Vocabulary in Japan [37].

The ISA programme also provided guidelines for consensus building on semantic agreements among stakeholders and a methodology for developing semantic agreements [16], which is based on the "Process and methodology for developing Core Vocabularies".

3.3.1 Stakeholders

The ISA guidelines first identify the various stakeholders which are involved in developing and maintaining semantic agreements. They identify the 'authority'. In case of OSLO our approach was bottom-up rather than top-down. The authority[38] consisted of representatives of the membership organizations (ICT, secretary, financial managers) of local governments, regional, and federal administrations.

According to ISA, "the activity is undertaken by a group of organizations that have decided to build shared services that require their IS to operate", referred to as 'members'. In case of OSLO, the members[39] were the consortium partners who funded the project, referred to as the 'steering committee', i.e. among them V-ICT-OR (NPO), local governments, the regional governments, application developers, and a start-up.

---

[37] https://joinup.ec.europa.eu/sites/default/files/ckeditor_files/files/3-IMI%20project%20in%20Japan%20L.pdf

[38] https://www.v-ict-or.be/kenniscentrum/overleg/e-gov-competentiecentrum

[39] https://www.v-ict-or.be/assets/5384d510ce3fb57c500006ad/OSLO1.1-specificaties.pdf, p 113.

The third group identified in the ISA guidelines is 'the wider community' that have an interest in the agreements, in case of OSLO this group consists of academic partners and ISA[40].

3.3.2 Specification process

The 'specifications process' describes the roles of the stakeholders and the steps to reach the semantic agreements among these stakeholders.

In the OSLO specification process, the steering committee which represents the 'authority' agreed among the stakeholders on the working groups, which were grouped per thematic agreed domain. OSLO focuses on three thematic domains: contact information, localization, and public services. Each domain has a dedicated working group. In each domain the relevant entities, relations, and attributes were discussed and iteratively refined and formalized. The steering committee validated each iteration of the domain model, specifications, and the vocabulary.

The ISA specification process describes two variants. The first is for complex activities and includes a domain working group and two or more data entity subgroups. The second variant is targeted at simple projects that build upon an existing domain model. In this case the working groups are merged into one working group. We aligned the process of the development of each domain in the model to the second ISA specification process [16].

In case of new entities that had no vocabularies (such as describing the relationship between a natural person and a registered organization) or when multiple entities with complex relations were involved (such as public services) break-out sessions were organized to

---

[40] https://www.v-ict-or.be/assets/5384d510ce3fb57c500006ad/OSLO1.1-specificaties.pdf, pp 96-98.

zoom in on specific objects with other participants. This process aligns with the first version of the ISA specification process [17].

To ensure that anyone can freely use and distribute the results of the project, the OSLO specification and all related documentation[41] were published under an Open License. Due to historical reasons in Flanders — related to Flemish Legislation — the OSLO specification and all related documentation were published under the Flemish Open Data License; "Modellicentie Gratis Hergebruik - v1.0". This License is compatible with the "Creative Commons Attribution licentie" which would also have been an excellent choice.

3.3.3 Development methodology

In this section we describe how the model was developed. The model served as the starting point for the 'implementation' phase. The ISA methodology consists of five phases:

1. The domain working group focuses on use cases that enable them to derive the requirements.
2. The domain working group develops a rough-cut Domain Model, based on the requirements of step one.
3. The data entity subgroups refine the domain model by adding attributes and linking to existing vocabularies.
4. The domain working group integrates the results of the data entity subgroups into the global domain model.
5. A conformance statement is created. The ISA methodology used a 'meet-in-the-middle' approach [21], focusing on stakeholders' commonalities rather than on their differences.

Immediately after the kick-off of the project, OSLO working groups created an inventory of the challenges and use cases related to the exchange of information for local public administrations in Flanders. This resulted in three main modelling domains of interest: (i)

---

[41] https://purl.org/oslo

persons and organizations, (ii) locations, and (iii) public services, see Figure 3.3. The specification for OSLO was developed by a multidisciplinary working group, with a total of 58 people from 28 organizations (all of them are listed in the specification). The working groups followed the same approach, with one working group per topic, integrating the domain working group and the data entity subgroups.

For the conformance criteria OSLO defined a conformance statement with different levels of engagement, represented by stars, aiming to lower the threshold:

- One star requires a human-readable mapping to OSLO.
- Two stars requires the mapping needs to be machine readable.
- Three stars requires the data to be in line with the OSLO vocabulary.
- Four stars adds requirements on the provenance of the data.
- The fifth and final star requires HTTP content negotiation[42], in which the client can specify the response format (e.g. HTML, RDF/XML, Turtle).

3.3.4 Implementation

The starting point for the implementation is the domain model (as mentioned in Section 3.3), delivered by the process described in Section 3.2.

The OSLO semantic agreement focuses on three domains*:* *contact information*, *localization*, and *public services*. Each of the models are local extensions of the ISA Core Person, Business, Location, and Public Service Vocabularies created at European level in the context of ISA. These four Core Vocabularies are simplified, reusable, context-neutral, and extensible specifications for information exchange.

---

[42] https://www.w3.org/TR/ld-glossary/#content-negotiation

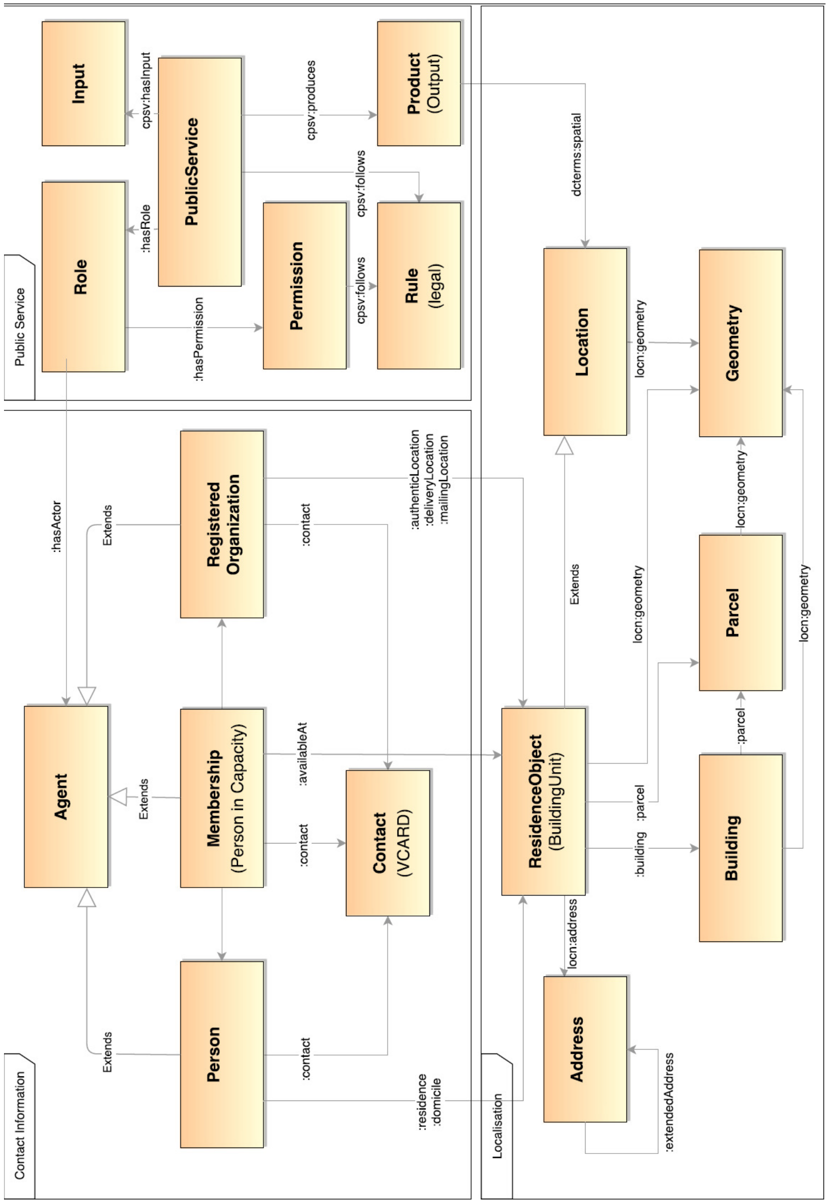


Fig. 3.3. OSLO focuses on three domains: contact information, localization and public services.

Terms appointing the structure of data and representing real world or abstract concepts might have an ambiguous meaning or multiple interpretations. ‘What do we consider as an address? Is it a residence or a domicile, or the place where someone works?’. The context determines the meaning of each term. Contact information of a person might contain other data, depending on his/her capacity, e.g., responsible in an enterprise, representative of an organization, or as natural person. Figure 3.3 shows the conceptual domain model capturing all the important entities.

Contact information. OSLO introduced a new concept to describe the relationship between a natural person and a (registered) organization: ‘in de hoedanigheid van ...’, of which the closest English translation is ‘in the capacity of ...’. It is related to and best captured by the concept of Membership in the ISA Organization Vocabulary, though it is not exactly the same: a citizen is not simply a citizen as in a different organization each citizen might take up different roles or functions which are obviously linked in a very particular way to the fact that a person is ‘member’ of an organization. The concept of ‘hoedanigheid’ (capacity) intends to capture this subtle nuance. ISA will start incorporating this concept this year in the Core Powers and Mandates Vocabulary [11]. The concept that expresses each ‘capacity’ a person has, functions as a unique identifiable object, and is a specialization of the broad concept of ‘Agent’ (as it occurs in the FOAF vocabulary). It is enriched with contact information. Each capacity a person takes up might come with different ways of how to reach this person. This contextual information is captured by the OSLO model too.

One of the most interesting aspect of the OSLO domain model, is the modelling of persons, organizations, and roles. At first, it may seem that a person ‘in acertain capacity’ is strongly connected to a person having a certain role in a Public Service. However, there is a strict distinction between those two concepts. For example, when someone (Person) picks up a certain mandate in a local government, this mandate will be a specification, an instance, of one of the capacities a person is in (Membership). This describes the relationship between the person and the organization (local government). Along with this mandate there will be certain roles to be able to carry out the public service of this particular local government. A role has one or more permissions which are embedded in rules (a legal framework). Opposite to the Membership/Capacity, the Role is bound to a specific Public Service rather than an organization (local government). An example of such a service might be the delivery of passports. This distinction enables describing a Public Service and all the necessary roles involved without the need to immediately couple it to an instance of a person or an organization. Secondly, when the person in capacity, the mandate holder, takes up a certain role, the instance of the role will be linked to the Membership/Capacity through the *has Actor* property. A Person, Membership, and registered Organization are specializations of the generic concept Agent. All three of them can be linked to a Contact, which is a VCARD description on how one can be contacted, and their physical location, the ResidenceObject.

Localization. The localization models the Physical Location of a person. This involves the Physical Location which can be described by its Geometry and in case the Location is a ResidenceObject (a BuildingUnit) also an address (with possible extensions).

Public Service. A Public Service is modeled as a black box system requiring a certain input document Input and delivering a certain Product as an output, e.g., requesting an ID results in the delivery of a new ID. The Input and Product entities capture the metadata of these products and can wrap electronic documents, e.g. as XML. The Product and Input are also bound to a specific location (administrative region). Certain instances of Agent (thus of Person, Membership, or Organization) have a role in the Public Service via the Role. Each Role has been granted some Permissions. Both the Public Service and Permission are following certain legal Rules.

Formalization. OSLO offers documentation for various target audiences and has a knowledge base with details on the specification (both human and machine readable):

- Specification targeted towards developers and policy makers: OSLO 1.1[43].
- Knowledge Base[44].
- Mapping guidelines[45].
- RDF and XML serialization of the vocabulary, managed on the GitHub and published in its own namespace with a fixed prefix[46].
- RDF namespace: purl.org/oslo/ns/localgov# resolves RDF version of OSLO.

3.3.5 Outcome

The OSLO programme brought expertise together from different business domains and governmental levels, independent of a specific thematic project. This group set up an interoperable model in line with EU standards ISA and INSPIRE[47] with specific local enrichments that support the processes of the different governmental levels. The model paved the way for a policy framework with the Flemish Government[48].

- OSLO is listed on several platforms for optimal dissemination.
- as an asset on the EU platform for semantic assets, i.e., Joinup[49].
- the OSLO namespace prefix is registered at prefix.cc[50].
- in Linked Open Vocabularies (LOV[51]).

---

[43] http://purl.org/oslo
[44] https://www.v-ict-or.be/kenniscentrum/projectfiches/OSLO/OSLO-2
[45] https://github.com/v-ict-or/oslo-mapping-guidelines
[46] https://github.com/v-ict-or/oslo_xml_schemas
[47] http://inspire.ec.europa.eu/
[48] https://docs.vlaamsparlement.be/docs/stukken/2015-2016/g522-1.pdf
[49] http://purl.org/oslo
[50] http://prefix.cc/oslo
[51] http://lov.okfn.org/dataset/lov/vocabs/oslo

Three initiatives in administration and research projects served as pilots for a preliminary implementation of the OSLO standard: the ‘shared catalog for local public administrations’, the crossroad database for Enterprises (VKBO), and the LBLOD proof of concept.

- The ‘shared catalog for local public administrations’ pilot [5] wanted to disclose contact information related to products and services effectively between governments and towards citizens through a common interface. An extension on OSLO was developed as a convergence between various stakeholders in local government data. The extension enriched OSLO vocabulary with three new entities: ***Channel, Activity*** and ***Product***.
- The crossroads database for Enterprises (VKBO), interlinked with a snapshot of the Linked Base Registry for Addresses as a Linked Open Data proof of concept[52], had its data modeled according to the OSLO vocabulary. The datasets were used for the evaluation and validation of data quality of OSLO among other vocabularies [6]. The applied methodology focusses on an approach for assessing the mappings instead of the RDF dataset itself, as mappings reflect how the dataset will be formed when generated. This methodology executes semi-automatic mapping refinements, which are based on the results of the quality assessment. In the dataset, we found four violations, afer manually refining the mapping definitions (according to the first mapping assessment’s results), only 7% of the range violations remained. OSLO reuses ontologies but does not cause violations as the combination is harmonized following the Web Ontology Language (OWL)[53] restrictions. It only caused few errors and those are among the most frequently encountered errors with vocabularies in general, for example ‘mapping a URI as a literal and vice versa’. OSLO is also used to annotate data of businesses and their locations

52 http://lddemo.datasciencelab.be/query/#startFragment=http://ewi.mmlab.be/ba/all

53 https://www.w3.org/TR/owl-ref/

in the COMBUST project, a platform for reliable business data[54].

- The LBLOD proof of concept, demonstrates a method to manage LBLOD, and aims to create a new base registry for mandates. This project from the Flemish Agency for Domestic Governance, has used the OSLO methodology. The project extends OSLO with two new concepts: one for metadata of decisions made at a local governmental level and one for describing public mandates. By publishing decisions that are automatically in a machine-readable format, in line with international vocabularies, they are suitable for reuse by third parties (Linked Open Data) without additional efforts [2].

3.4 Characteristics of OSLO

Semantic interoperability is the key to create appealing citizen-centered e-government services and better reusable Open Government Data. As explained in Section 3.1.1. the lack of interoperable information products leads to a fragmented view of the public service concept. This is a major obstacle for creating citizen-centric services. By publishing information in line with international vocabularies, information becomes more suitable for reuse by third parties.

OSLO, and in particular its vocabulary specification, empowers a technology independent, generic representation of contact information, localization and services provided by public administrations. By its nature and design, OSLO is generic enough to be applied in a wide range of scenarios, not just for its original purpose. However, as illustrated by the EIF (see Figure 3.2), it requires the necessary political support at the different governmental levels in Flanders to rollout OSLO. Many public services delivered at the local

[54] https://www.iminds.be/en/projects/combust

level are dependent on data and processes from the regional government, therefore consensus at the different levels is needed.

Below we discuss the different elements (see Table 3.1) we have identified the influences that characterize the context of the creation of the Open Standard.

Ownership: the ownership of OSLO changed during the process. It started as an initiative of a mediating non-profit organisation, i.e., an interest group of public servants active as IT practitioner at local government level in Flanders. At the end of the first phase (2015) the ownership was transferred to the Flemish Government.

Vocabulary alignment: the vocabulary is aligned with European initiatives: ISA and INSPIRE. Because the OSLO semantic agreement is built upon international vocabularies and the methodology guides the stakeholders towards a reusable machine-readable format, we ensure that Open Government Data can be reused by third parties without the need for expensive mappings and transformations.

Adoption: although the municipalities' awareness is rising for a common agreed data standard, those who were not familiar with and aware of the potential prior to the OSLO programme, started to integrate some elements at their local information system management. OSLO encountered commonly known challenges regarding its adoption [14]. There is a transition phase involved for public administrations and organizations when deciding to implement OSLO. At the local government level, OSLO is being adopted in public tenders, facilitated by a whitepaper with a conformance statement [1]. The Flemish vice-minister president supported OSLO and embedded it in the strategy for the Digital Flanders Agency to stepout of thinking in data silos, as mentioned in the Policy letter of Administrative Affairs Department [10].

Adaptation: in spite of the 'once-only' and 'whole-of-government'[55] principles, there remain problems to overcome the lack of integration in e-government projects. According to the United Nations Government Survey, information 'silos' are created by departmentalism and lack of coordination; "The problem lies not with the technology but in the political challenge of rewiring a range of public sector programs delivered by different levels of government" [18]. To soften these silos in the OSLO programme, agreements on various governmental levels were also essential. Both OSLO and ISA methodologies focus on commonalities rather than on differences. The process allows participants to focus on use cases in an early stage, instead of defending their definitions based on their (domain specific) implementations. To ensure that published data can remain accurate, consistent across data sources and up-to-date, OSLO facilitates modelling public and governmental data (belonging to citizens). The uptake in the long term relies on easier access to authoritative and other data sources following the OSLO semantic agreement. This enables aligning data with authoritative sources and exchanging data among the variety of data sources. To realize this aspect of the interoperability vision we need machine-readable data, with standards that are supported beyond the (single) government, i.e., semantic interoperability.

Governance: as described in the specification process (3.2), a permanent steering committee was installed, which represents the 'authority'. The steering committee validates the installation of a new thematic working group, each new vocabulary, each review of a vocabulary and the conformance statements. During the development period a broad coalition, mainly based on goodwill, participated in explicit use cases, e.g., the pilots, authentic base registries. This convinced other public administrators to further support such efforts

---

[55] http://glossary.usip.org/resource/whole-government-approach

and recommend it at the regional and federal level. Although there were some early adopters within local administrations, this self-steering committee based on voluntarism lacks power to embed the standard in the legislation. All these experiences lead to a follow-up trajectory. How OSLO could approach certain of the encountered roadblocks and the shift in critical success factors is explained in Section 3.5.

Table 3.1. Characteristics of OSLO (described in this chapter)

	OSLO
Ownership	An interest group of public servants active as IT practitioner at local government level in Flanders: V-ICT-OR (non-governmental body, non-profit organization).
Vocabulary alignment	Alignment with EU initiatives: ISA, INSPIRE
Adoption	Adoption in public tenders at the local government level and embedded it in the policy of the Digital Flanders Agency
Adaptation	Focus on the commonalities rather than on differences, agreements on various governmental levels.
Governance	Self-steering approach with one chair / facilitator and business owners invited as experts.

3.5 Discussion and conclusion

This research introduced Open Standards for e-government vocabularies and guidelines for governments in Flanders. Bottom-up organized working groups delivered a reusable formal specification and serialization of domain specific models. As a result, public administrations and private partners can model people, organizations, public services, and locations (including addresses and buildings) for data exchange.

Information is often not reusable in multiple contexts because information (intensive) processes are implemented in a legally binding context or a specific organizational context within a public administration. OSLO enforces the principle: ‘first clarify and then digitize’. This principle is put forward in most digitalization projects but often there is a lack of political basis and support to adapt the necessary rules to cope with this principle.

In this chapter, we’ve focussed on **the second viewpoint of our research question** “*how to build consensus among different public administrations and rewire public sector programs which often are under the authority of a different governmental level?*”.

Both the bottom-up and top-down approach were important to create the necessary political support. OSLO was built on consensus, rather than on a legal framework. This unique situation where different government levels are working towards consensus, can stimulate future uptakes of particular core data models by other administrations [19].

The OSLO Programme increased awareness and the ISA-based methodology led to semantic convergence creating a foundation to develop interoperable e-government services in Flanders. As tested in the “shared catalogue for local public administrations” pilot, this affords providing information from different government levels through a common interface.

Because OSLO is now embedded in the strategy of the regional government, we expect this could change the characteristics, as discussed in Table 3.1. As the ownership is transferred to the Flemish Government, a governmental organization will be the ‘authority’ instead of a non-profit organization. This implies a transfer of the governance and life-cycle management of the ontology to the regional government.

During our research in 2016 we expected, in terms of adaptation, the development of OSLO compliant products at the higher regional government, could overcome the lack of OSLO-compliant authoritative sources and applications. We expected a vigorous commitment and accountability from the regional government could speedup the further adaptation at local governments, regional administrations, and the intergovernmental data sharing as well. The latter often implies an adaptation of the organizational processes. Product owners who manage (authoritative) information sources at regional level could be in charge of the working groups, which can enable a governance shift from a grassroots local approach towards a central governance system co-funded by the different policy domains at the Flemish level. This could secure a more sustainable funding to support local and horizontal regional governments in their transition towards Open Standards and generic building blocks, and could help to speed-up the adoption of OSLO. The role of the regional government could then be facilitating a harmonized information exchange policy where standardization in terms of infrastructure, semantics, and data formats will play a crucial role.

At the time of writing this dissertation in 2021, the number of ratified data standards by the ‘Steering Committee of Flemish Information and ICT-policy’ has grown to more than ninety[56]. There is a vast adoption of OSLO standards at local and regional government including high-impact projects such as 'the Flemish Citizen Profile', 'Local Council Decisions as Linked Open Data' and 'management of the public domain'. Also, we see a broadening to initiatives in the private sector such as Mobility as a Service (MaaS). In addition, this transition resulted in a more structural funding that is a combination of a grant from the Flemish government and funding from various government administrations. The estimated budget of the OSLO interoperability programme in 2021 will exceed 1500 man-days, not including the work of the vast community who cocreate the standards in the various business domains. This series of events exceeded our expectations and confirms our hypothesis that a formal government-embraced governance can raise the adoption of data standards.

---

[56] https://data.vlaanderen.be/standaarden/

3.6 References

[1] Audenaert, M., D'haenens, T. 2014. *Overheidsdata, standaardisatie en goed ondernemerschap*. V-ICT-OR. https://www.v-ict-or.be/assets/5450fdffce3fb537ef00014f/Overheidsdata%20en%20goed%20ondernemerschap.pdf

[2] Buyle, R., Colpaert, P., Van Compernolle, M., Mechant, P., Volders, V., Verborgh, R., Mannens, E. 2016. 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.*

[3] Buyle, R., De Vocht, L., Van Compernolle, M., De Paepe, D., Verborgh, R., Vanlishout, Z., … Mannens, E. (2016). OSLO - open standards for linked organizations (pp. 126–134). Presented at the Electronic Governance and Open Society: Challenges in Eurasia (EGOSE, St.Petersburg, Russia: Association for Computing Machinery (ACM).

[4] Colpaert, P., Van Compernolle, M., De Vocht, L., Dimou, A., Vander Sande, M., Verborgh, R., Mechant, P., Mannens, E.2014. Quantifying the interoperability of open government datasets. *Computer*. 10, 50–56.

[5] De Vocht, L., Van Compernolle, M., Dimou, A., Colpaert, P., Verborgh, R., Mannens, E., Mechant, P., Van de Walle, R. 2014. Converging on semantics to ensure local government data reuse. In *Proceedings of the Fifth International Conference on Semantics for Smarter Cities.* 1280. CEUR-WS. Org, 47–52.

[6] Dimou, A., Kontokostas, D., Freudenberg, M., Verborgh, R., Lehmann, J., Mannens, E., Hellmann, S., Van de Walle, R. 2015. Assessing and refining mappingsto rdf to improve dataset quality. *The Semantic Web-ISWC 2015.* Springer, 133–149.

[7] Gallo, C., Giove, M., Millard, J., Kare, R., Thaarup, V. 2014. *Study on eGovernment and the Reduction of Administrative Burden*. European Commission, p.149.

[8] Garson, G.D., Biggs, R.S. 1992. Analytic Mapping and Geographic Databases, *SAGE university paper*, volume 87, p2.

[9] Guijarro, L. 2009. Semantic interoperability in eGovernment initiatives. *Computer Standards & Interfaces.* 31,1, 174–180.

[10] Homans, L., *Beleidsbrief bestuurszaken 2015-2016*. 2015. In *Flemish Parliament* (522), p. 25. https://docs.vlaamsparlement.be/docs/stukken/2015-2016/g522-1.pdf

[11] ISA. 2014. *ISA2 Work Programme: Annex I, Detailed Action Descriptions,* p. 78.

[12] Lewin, K. (1946). Action research and minority problems. *Journal of social issues*. 2,4, 34-46.

[13] M.N.K. Saunders, P. Lewis. 2000 Research Methods For Business Students. *Harlow Financial Times*. Prentice Hall, 147-148.

[14] Nekvasil, M., Sv´atek, V. 2013. Towards savvy adoption of semantic technology: From published use cases to category-specific adopter readiness models. *Web Semantics: Science, Services and Agents on the World Wide Web.* 21, 61–74.

[15] Pardo, T.A., Nam, T., Burke, G.B. 2011. E-government interoperability: Interaction of policy, management, and technology dimensions. *Social Science Computer Review,* p. 10.

[16] Interoperability Solutions for European Public Administrations (ISA). 2013. *Process and methodology for developing semantic agreements*, p.7.

[17] PwC EU Services EESV. 2012. *Process and methodology for core vocabularies*. Tech. rep., Technical report, Interoperability solutions for European Public Administrations, p. 5.

[18] United Nations Department of Economic and Social Affairs. 2012. *United Nations E-Government Survey 2012.* United Nations, p. 69.

[19] Van Compernolle, M., De Vocht, L., Goedertier, S., Loutas, N., Verborgh, R., Mannens, E., Mechant, P., Peristeras, V., 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. ICEGOV 2015-2016,* 327-337.

[20] Wood, D. 2011. *Linking government data*. Springer, p.112.

[21] Zeginis, D., Hasnain, A., Loutas, N., Deus, H.S. Fox, R., Tarabanis, K. 2012. Collaborative development of a common semantic model for interlinking Cancer Chemoprevention linked data sources. *The Semantic Web Journal*, p.14.

CHAPTER 4. SEMANTIC INTEROPERABILITY

Data is a precious thing and will last longer than the systems themselves.

—Sir Tim Berners-Lee.

This chapter researches how to raise semantic interoperability. We examine the evolution of the Linked Base Registry for Addresses in Flanders. Also, this chapter introduces the benefits of Linked Data and argues the increase of interoperability which leads to a better adoption of government information in the public and private sector. This chapter is based on the paper, 'Raising interoperability among base registries: The evolution of the Linked Base Registry for Addresses in Flanders.' [16].

4.1 Introduction

The transformation of society towards a digital economy created a new context leading to changing roles where boundaries between public and private services are blurring [25]. This enables the government and the private sector to collaborate and share responsibilities [22]. In Belgium, the Flemish Government is undertaking ambitious reforms to further transform public services. The policy letter of the Vice Minister President states that Open Data is 'the norm' and that the government should focus on economic added value and close collaboration with the private sector [35]. The Flemish Government is already working with the private sector to co-create and co-finance the development and maintenance of Open Data sources. The Large-scale Reference Database (LRD) is an official Flemish data source with precise and detailed location information on buildings, parcels, roads, watercourses, and railroads, identifying millions of objects in Flanders and providing a source for address positions. The LRD is the result of a public-private partnership between the Flemish Government and the utility sector, with a substantial setup cost of EUR 93 Mio and an annual maintenance budget of EUR 7 Mio [21].

Since the LRD became Open Data in November 2015, more than 2500 bulk downloads (Figure 4.1Fig. 4.1) and over 2.9 Mio requests were successfully processed each month. Re-use of government data is considered to be an enabler of Open Government [40]. The **problem statement** of this chapter is: *even though the Government of Flanders embedded the re-use of PSI in legislation [7, 11] and published a data portal[57] containing over 4000[58] Open Datasets, interconnecting and interpreting these sources of information remains challenging for businesses, citizens and public administrations alike*.

[57] http://opendata.vlaanderen.be/
[58] http://opendata.vlaanderen.be/dataset

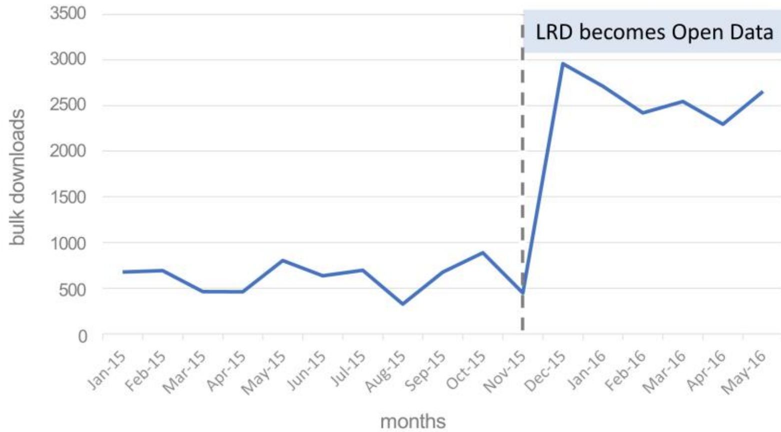


Fig. 4.1. The impact of opening LRD data (in terms of downloads), showing the increase of downloads due to opening up the dataset (L. De Wolf, Digital Flanders Agency).

Public administrations struggle to deliver interconnected and cross-sectoral services due to sectoral specialisation or ‘departmentalisation’ [40]. Due to interoperability problems, including adequate semantic standards and scarcity of web-oriented architecture, private partners struggle to reuse PSI. 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 [29]. PSI is often modelled from a single perspective and therefore cannot be integrated with other information sources, applications, and business processes [15]. The lack of standards causes high costs due to data transformations and mapping [40]. To overcome these hurdles, we need to address multiple interoperability levels; namely the legal, organisational, semantic, and technical level [23].

This has led to a demand for stable, governed data standards [42, 4], which are “technical documents designed to be used as a rule, guideline, or definition. They are consensus-built, repeatable ways of doing something”[59]. The Flemish Government launched an interoperability programme: Open Standards for Linked Organisations (OSLO) [23] which builds upon the principles of the EIF [28]. Interoperability Frameworks assume a hierarchy in terms of maturity with regard to layers of interoperability [42]. This means organisational and legal interoperability can only be achieved when standards for semantic and technical interoperability have successfully been implemented. Therefore, OSLO addresses both semantic and technical interoperability.

Incompatibilities between legislation in different policy domains and legal frameworks make working together more complex. Legal Interoperability refers to aligned legislation between different organisations. Organisational interoperability refers to aligned business processes between public administrations. This implies integrating business processes and related data exchange. [23, 28].

Semantic interoperability focusses on the meaning of data elements. As inter-organisational IS only work when they communicate with other systems and interact with people, it includes developing vocabularies to describe data exchanges and ensures that data elements are understood in the same way by different parties when communicating [23, 34]. According to EIF, semantic interoperability also covers the syntactic aspect which refers to the grammar and format, such as HTML or XML.

Technical Interoperability is often centred on (communication) protocols and the infrastructure needed for those protocols to operate [28]. OSLO builds upon the design principles of Linked Data. The term ‘Linked Data’ refers to data, which is published on the Web and, apart

[59] https://www.cen.eu/work/ENdev/whatisEN/Pages/default.aspx

from being machine-readable, it is also linked to other external datasets [1], using the RDF[60] as a flexible and extensible data model.

The main goal of this chapter is to unfold the process of reaching semantic and technical interoperability among base registries based on the principles of Linked Data. The address registry is presented as a case study of a base registry. The EC defines a base registry (BR) as a trusted and authoritative source of information which can and should be digitally re-used by others, where a single organisation is responsible and accountable for the collection, use, updating, and preservation of information. 'Authoritative' here means that a base registry is considered to be *the* 'source' of information, i.e., it shows the correct status, is up-to-date and is of the highest possible quality and integrity [28].

The Flemish Government administration aligns its base registries with this definition but introduces three additional requirements:

- The base registries are part of a semantic coherent system of uniform identified base-objects and relations, which are in line with the OSLO standards.
- A base registry reuses the identifiers of the base-objects in other base registries.
- The base registries are obliged to maintain the Lifecycle and History of the base objects.

Public data often has a location-based component. According to Garson and Biggs [33] "It is estimated that 80% of the informational needs of local government policymakers are related to geographic location" (p. 87). The central address registry is one of the Flemish base registries and of significant value to the public and private sector. The address registry includes geographical coordinates [55] and is released

[60] https://www.w3.org/2001/sw/wiki/RDF

under the Flemish Open Data licence[61] which enables re-use, including commercial re-use, for free. Examples of re-use of this authoritative source in the private sector are a more accurate address-position in navigation applications than commercially available data and a better address quality in administrative processes than locally managed datasets resulting in a lower mail bounce and thus lower costs.

In 2013, a pilot project 'Interconnecting Belgian National and Regional Address Data' was carried out in the context of the ISA programme of the EC. The pilot published data from the Belgian federal level and the three Belgian regions as Linked Data. Results of this pilot indicated that the public sector had not yet tapped into the full potential of its address registries. The obstacles: (i) address data fragmentation, (ii) heterogeneous address data formats, and (iii) a lack of common identifiers. The research reported in this chapter reveals how these obstacles can be overcome so that the full potential of address registries and other base registries can be realised [17]. Interoperability in the public sector is influenced by internal as well as external politics. Internal, organisational politics includes dealing with issues involving organisational members. External organisational politics is about how public administrations relate to their council, board, or other organisations [49]. With the majority of the government data having a location-based component, it is important for the Government that all government administrations and partners link to the 'authoritative' addresses avoiding redundant and incomplete shadow-databases.

Information management specialists often lack knowledge about how to deal with political aspects of information management, and consequently are ineffective. The outcome of this chapter is valuable for researchers, public administrations, and public sector ICT

[61] https://overheid.vlaanderen.be/open-data-bij-de-vlaamse-overheid#modellicenties

service suppliers that aim to raise interoperability in complex data ecosystems.

Our **research questions** consider the problem statement from a semantical, technical, and organisational point of view:

- How can public administrations raise semantic interoperability to ensure data from Base Registries can be interpreted by all communicating parties?
- How to maintain the agreed semantics on a technical level and design a programmable interface that can be well interpreted by both humans and machines?
- How to rewire existing processes and data products to become interoperable within an operational public sector context?

The **contributions** of this chapter, which build upon the principles of the Semantic Web, are valuable for semantic Web researchers and organisations that aim to publish interoperable authoritative data sources. The remainder of this chapter is organised as follows: Section 4.2 discusses relevant related work: we discuss the nature of the Flemish, Dutch, and Danish address registries. We illustrate the Linked Data strategy in Flanders and outline a comparative study between the current URI strategies in Europe. Section 4.3 gives an overview of the address vocabulary and discusses the 'fitness for use' of the prevailing European address vocabularies for modelling the address registry in Flanders. Section 4.4 outlines the key success factors of a real-time Linked Data architecture. We will point out how the addresses, mined from the municipalities, are published at the SPARQL-endpoint in 'nearly real-time'. On top of that, we will elaborate on the deployment strategy which should allow other agencies to reuse some of the technical components when refactoring their own base registries. Section 4.5 discusses how the Linked Base Registry for Addresses facilitates the adoption of Addresses as Linked Data in the public and private sector. Section 4.6 provides an evaluation of the Linked Data approach in the address registry. Finally, Section 4.7

presents conclusions on how to raise interoperability in the public sector and an outline for future work.

4.2 Background and related work

This section discusses the nature of the Flemish, Dutch, and Danish address registries. By providing an insight into the events that influenced the evolution of these address registries, we can untangle the different interoperability levels. Next, we outline the Linked Data strategy in Flanders and describe the process and methodology we have used to create a reusable vocabulary for addresses. Finally, we provide an evaluation of the current URI strategies in Europe which lead to the Flemish URI-strategy.

4.2.1 The Central Reference Address Database in Flanders (CRAB)

In 2011, the Flemish government and the Flemish Geographical Information Agency (now Digital Flanders Agency), developed an authentic source for addresses, referred to as 'Central Reference Address DataBase' (CRAB), containing well over 4.5 million addresses as well as their address positions (xy-coordinates). According to the European INSPIRE directive, the overall concept of an address data specification is that it has a 'locator', e.g., a house number that enables a user to distinguish it from neighbouring addresses, and a geographic position, which enables an application to locate the address spatially. To identify an address in Flanders, it must be associated with a number of 'address components' represented by a spatial identifier. These components are defined in the CRAB decree as *streetname, house number and box number, postal code, and municipality* [8, 10].

Key influences are:

- The municipalities are responsible for the creation of addresses and management of the address components.
- Governments in Flanders are obliged to use the address registry and to provide feedback in case they detect an error.

- At the European level, the INSPIRE Directive aims to create a data infrastructure for the purposes of EU environmental policies and policies or activities which may have an impact on the environment [38]. By the end of 2017[62] member states are obliged to provide harmonised address information by means of compatible services.

In the past two decades, many events influenced the development of the Central Reference Address Registry (CRAB) in Flanders. Figure 4.2 depicts the milestones evaluated using the EIF [12, 15]. EIF is a set of recommendations which specify how administrations, businesses, and citizens can communicate with each other across borders and within the EU. These interoperability levels are defined as legal, organisational, semantic, and technical within a political context.

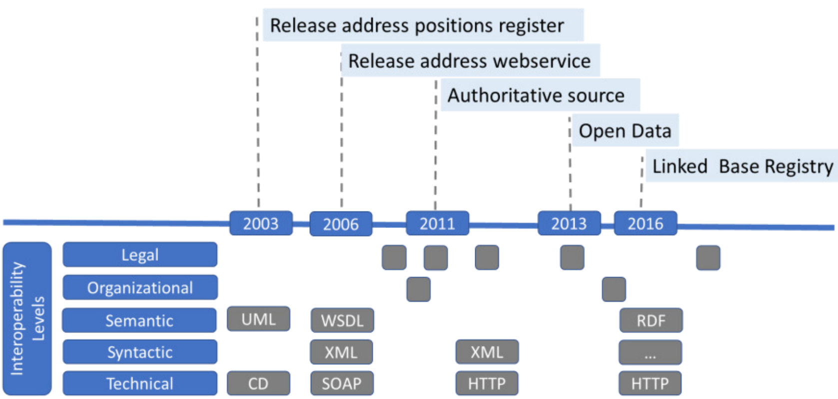


Fig. 4.2. Key influences on the address registry showing the impact on all interoperability levels, evaluated using the EIF [23] and inspired by an ISA evaluation framework in "How Linked Data is transforming eGovernment" [40].

When the Flemish GIS (Geographic Information System) administration (Ondersteunend Centrum GIS Vlaanderen, now Digital

62 http://inspire.ec.europa.eu/inspire-roadmap/61

Flanders Agency) was founded in 1995, there was no authoritative source for addresses. Although local governments have been responsible for the creation and maintenance of street names since 1977 [6], the lack of formal rules for addressing resulted in duplicate street names within the same municipality, misspelling of street names, and many flavours of house numbers and box numbers to identify individual apartments in a building. The first initiatives towards a harmonised data model at regional and Federal level were initiated in 1999 (Geocodi). An important milestone was the release of the 'address positions database' in 2003, distributed to administrations in Flanders using a CD-rom as carrier, accompanied by a formal data specification in UML [48] ratified by a decision of the Flemish Government. In 2006 the address database became available via the Internet, by means of Simple Object Access Protocol -webservices (SOAP). A shared strategy for managing the lifecycle of addresses among the regional and local administrations (VLAR-address, 2008) paved the way towards an authoritative data source. The following year the INSPIRE directive from the EC [38] was converted into Flemish legislation [9]. The regulations are embedded into the CRAB [8, 10] and SDI [53] Decree, which are laws of the Flanders Region and the Flemish Community and set the scene for a legal interoperability level. The same year, an important hurdle on semantic interoperability was taken: a formal agreement on shared semantics was reached which eventually led to shared semantics at Federal and regional levels in 2015. These events that intervened on all interoperability levels eventually led to the formal approval of the address registry as the first authentic data source in the region of Flanders.

Driven by the *once-only* principle embedded in Flemish registration [54] an authoritative source (also referred to as an 'authentic' source) builds upon following principles: (i) administrations are obliged to use the authentic source to avoid requesting information from citizens more than once, raise data consistency and reduce administrative burden; (ii) the source is recognized by the Flemish government; and (iii) administrations are obliged to report errors to the administrator of the source. In 2013, under an impulse of the European Directive on the re-use of PSI [37] the CRAB address registry became available as free Open Data. In 2014 The Flemish Government decided to focus on interoperable base registries [35], as defined in Section 4.1.

The Flemish government is managing the address registry conformant to best practices in all EIF-domains [26], more specific at the level of technical, semantic, and organisational interoperability also conformant to legal requirements. At the organisational level, an important step was taken in 2015 by merging the GIS and e-Gov administrations into a new Information Agency. At the legal level, Digital Flanders Agency has the ambition to embed the concept of Base Registries in a decree. The new agency launched the programme "OSLO" that focuses on the semantic interoperability level and extends the ISA Core and INSPIRE Vocabularies in order to facilitate the integration of base registries with one another and their implementation in business processes of both the public and private sector. Finally, in 2016 the Digital Flanders Agency rewired the Authentic Source for addresses and published it as Linked Open Data. The next step was taken in 2018, when the Federal and Regional administrations were obliged to use regional address registries (BeSt-Add[63]) in all their processes. The EC wanted to overcome the fragmentation of the address data. The aim was to tackle following hurdles: (i) address data fragmentation caused by the different isolated databases at the various government levels and the lack of a single access point; (ii) heterogeneous address specifications and standards; and (iii) a lack of common well-formed identifiers which are the cornerstone to link the different addresses (Colas et al., 2013). In 2011, the EC initiated a pilot project in Belgium, a federal state with three communities, three regions, and four language areas. The goal of the

[63] https://overheid.vlaanderen.be/CRAB-Belgie-BeSt-Add

project was to interconnect the Belgian and Regional Address Data. The pilot was built upon the principles of Linked Data (Fig. 4.3) to make addresses more interoperable and lovable in Belgium. The term Linked Data refers to a set of best practices for publishing and connecting structured data on the Web using international standards of the W3C [56]. The design principles[64] as asserted by Tim Berners-Lee in 2006 (Fig. 4.3) were adopted by the address pilot project [17].

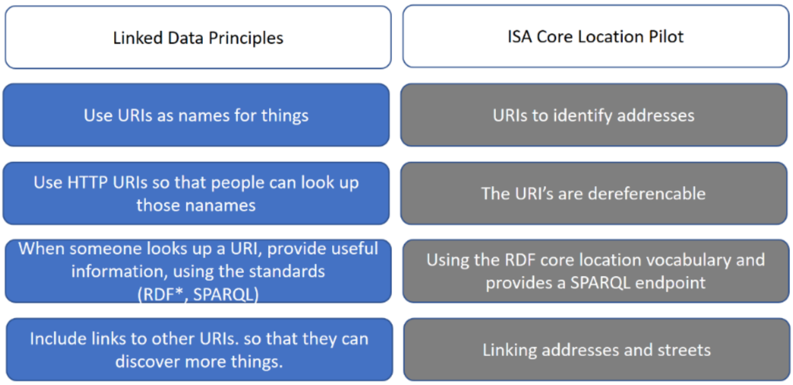


Fig. 4.3. Design principles of Linked Data and the specific implementation in the Core Locations Pilot, using the design principles as asserted by Tim Berners-Lee as a framework and mentioned in by ISA in ”How Linked Data is transforming eGovernment“ [40].

The first principle states “Use URIs as names for things”. All addresses and streets were given a universally unique identifier which can be looked up via the Web.

The second principle points to the “use HTTP URIs so that people can look up those names”. To create these stable identifiers best practices from the ISA programme for creating persistent URIs [3] were applied: including avoiding stating ownership in the URI, avoiding version numbers, and implementing HTTP response code 303 to redirect from the real object to a document which describes the address or street.

---

64 https://www.w3.org/DesignIssues/LinkedData.html

The third principle: “When someone looks up a URI, provide useful information, using the standards RDF and SPARQL” is all about interoperability. RDF is the data model for Linked Data. According to W3C, “RDF extends the linking structure of the Web to use URIs to name the relationship between things as well as the two ends of the link (this is usually referred to as a ‘triple’). Using this simple model allows structured and semi-structured data to be mixed, exposed, and shared across different applications.”[65] This pilot is using RDF as data model and the ISA Core Location Vocabulary[66] provides the schema. The Core Location Vocabulary is a context-neutral, extensible data model derived from the INSPIRE address representation. We will discuss the differences between both vocabularies in Section 4.3. This vocabulary was set up as a canonical data format to bridge between the different non-interoperable address data models at the federal level and the three Belgian regions. The pilot implemented a SPARQL endpoint, which allowed querying the RDF data sources via the Web.

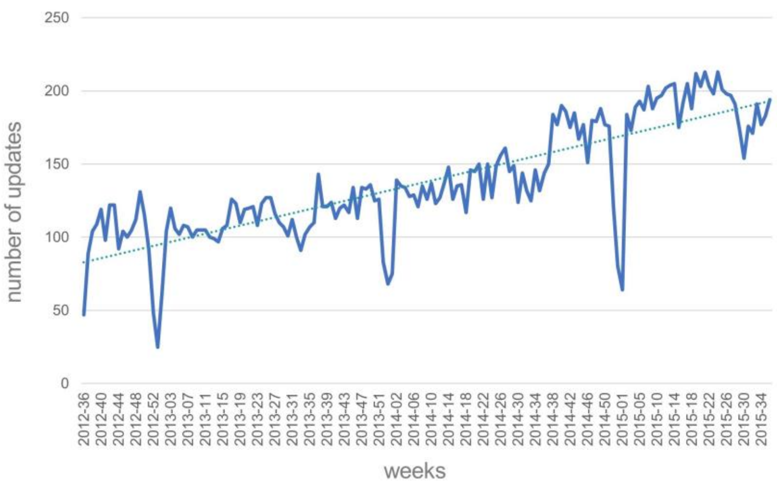
The last principle states “Include links to other URIs so that they can discover more things”. A dataset becomes more useful when it links to other resources, which allows humans and machines to discover more information following the links. Other parties on the Web can link to the addresses, making their data more useful. The addresses core dataset itself does not link to other resources but enables other parties on the Web to link to the addresses, making their data more useful. In Flanders, the domicile of a citizen and the EPC are for example both linked to an address.

Because both resources are linked to the same URI, the government administration is able to give citizens insight in the energy efficiency of their domicile by just following the links.

---

[65] https://www.w3.org/2001/sw/wiki/RDF
[66] https://joinup.ec.europa.eu/asset/core_location/description

This pilot demonstrated the viability of a Linked Data infrastructure for addresses in Flanders, which can provide interoperability across the regions in Belgium. Despite this proof-of-concept paved the way towards a sustainable implementation for



addresses in Flanders, some important complexities towards addresses still needed to be tackled, including the different numbering conventions.

Fig. 4.4. Overview of the number of updates in the CRAB address registry each week (J.Laporte, Digital Flanders Agency).

As the CRAB address registry is embedded into the core processes of government administrations and administrations are obliged to report errors, the number of updates is growing over 30% per year (see Fig. 4.4). Figure 4.5 depicts the increasing use of the address registry, based on the transactions on the address Web Feature

Service[67] (WFS). The dips in the graphs are holidays which naturally cause a lower activity on the address registry.

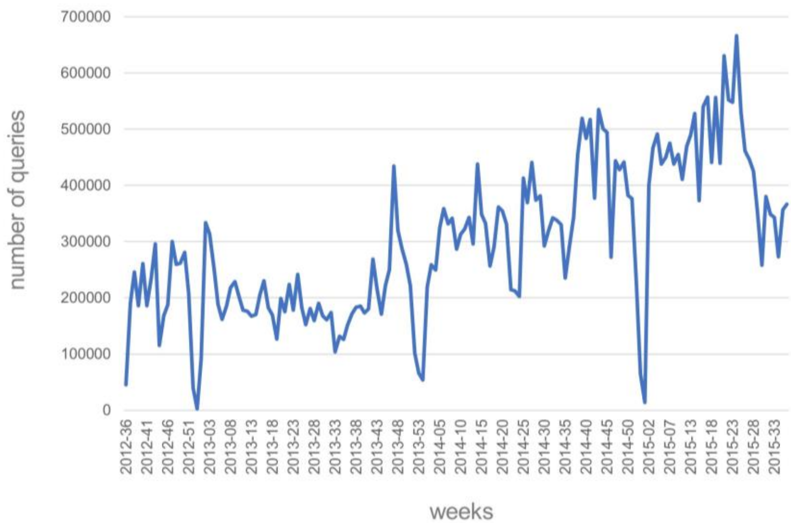


Fig. 4.5. the increasing use of the product based on registered users (J.Laporte, Digital Flanders Agency).

4.2.2 The Address Registry in The Netherlands

The Netherlands are going through a similar process. The 'Cadastre, the Dutch Land Registry and Mapping Agency' (Kadaster) is at the helm of administrative and spatial data on addresses and buildings. The municipalities are responsible for the creation of addresses and management of the addresses components and have to register updates within four days [14] in the 'Basic registry of Addresses and Buildings' (BAG). The Cadastre has shifted the focus from the address object to the physical objects that are addressable. These addressable objects can be 'residence objects' (dwellings or offices),

---

67 http://www.opengeospatial.org/standards/wfs

mooring places, or places for the permanent placement of mobile homes. The residence objects refer to buildings, individual apartments, or offices within a large complex. These objects are registered in the base registry for addresses and buildings. The address is a property of a physical object and consists of following components as defined in the BAG object-catalogue [13]: address locator name, the thoroughfare name (street name, 'naam openbare ruimte'), and the municipality. Since 2012, the address registry is published as Open Data. The Dutch Cadastre published the base registry with cadastral information as Linked Open Data in 2016. The address registry was published as Linked Open Data in 2017[68].

In collaboration with Geonovum, responsible for developing and maintaining standards for geographical information, administrations in the Netherlands joined forces to develop an Open Data strategy [32]. They have brought together experts from the public and private sector in the steering committee 'Platform Linked Data Netherlands'[69] (PILOD). The PILOD-platform has published an experimental version of the Dutch Building and Address registry as Linked Open Data[70]. A crucial step in the development was that all addresses and streets were given a persistent universally unique identifier which can be looked up via the Web. They developed a similar URI strategy both building upon best practices of ISA and INSPIRE [47]. Whereas the pilot in Flanders is extending the RDF using the ISA Core Location Vocabulary as a basis, the Dutch pilot on addresses has developed an extension[71] which is more in line with the existing BAG data model.

---

[68] https://data.pdok.nl/datasets
[69] http://www.pilod.nl/wiki/Wie_we_zijn
[70] http://www.pilod.nl/wiki/Bag_dataset
[71] http://lod.geodan.nl/vocab/bag/index.html

4.2.3 The Danish Address Registry

The Danish Address Programme is a sub-programme of the Basic Data Programme[72] which is a collaboration between a number of National governmental bodies, the associations of Local and regional governments. The Danish Agency for Data Supply and Efficiency[73] is the responsible authority for addresses in Denmark as well as for the Danish address registry, which was a part of Building and Dwelling Registry [18] but is now an independent registry[74]. The Danish administrations have focused on eliminating shadow databases by identifying base registries' which they refer to as 'GRUNDDATA'.

The Danish government has a long history in the standardisation of addresses which goes back to 1978 when a standard address structure was introduced [43] followed by the first address registry in 1980. The first address structure consisted of a municipal code, street code, address number, floor, and door identifiers. In 1990 the addresses were rewired from 'Address as an Attribute' in the different registers (such as the central person register and central business register) to an 'Address as a common asset' where the address becomes a fully-fledged object [43] which allows other registries to link to the Address Registry. These addresses were geo-referenced and harmonised voluntarily by the local municipalities. This included a harmonisation of the 'property data registers' and the 'municipality technical base maps' towards a building and dwelling registry.

The 98 municipalities have the practical authority for the assignment of addresses (Danish Enterprise and Construction Authority, 2010). In 2000 the Danish Government appointed the building and dwelling register as a base register which was published in

[72] http://grunddata.dk/english/
[73] http://sdfe.dk
[74] http://danmarksadresser.dk/

2002 as Open Data. Although registries, data models, and ministerial initiatives existed, there were no official guidelines for the infrastructure model [36]. In 2004, the report "Basisdata, forståelsesramme og analysemodel til kategorisering af basisdata" [41] was published, which focused on data quality and the unambiguous connections between the registries. [36, 41]. Additional requirements were introduced [36, 43]:

- The base registries are part of a semantic coherent system of uniformly identified base objects and relations. The data model has similarities with the conceptual data models behind INSPIRE.
- Core Objects contain provenance information, which defines when the information was registered and for which period of time it is valid, often referred to as bi-temporal data.
- A base registry should reuse the identifiers of the base objects of other base registries. The addresses have persistent identifiers, which are a globally unique identifier (GUID).
- The base registries are obliged to maintain the Lifecycle and History of the base objects. A telling example is the data about a building that is under construction or has been demolished.

4.2.4 The Linked Data strategy in Flanders

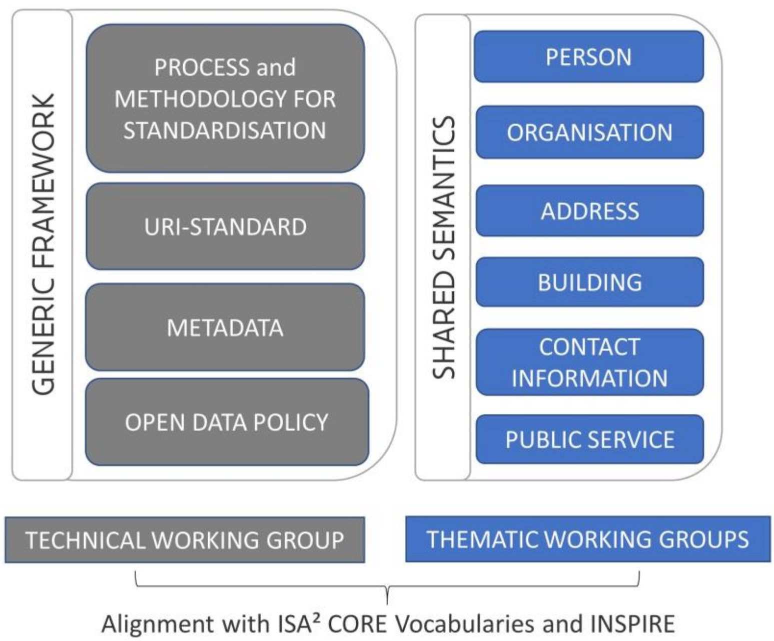
The Flemish government is focusing on a sustainable strategy for linked base registries. In the previous section, we have learned that it is essential to focus on all interoperability levels. In this section, we will focus on the semantic interoperability strategy and how it influenced the development of the CRAB in Flanders.

The OSLO programme, which started as a grassroots initiative at the level of the local governments[75], increased awareness on the

[75] https://joinup.ec.europa.eu/community/semic/news/flemish-oslo-standard-become-local-extension-isa-core-vocabularies

need for semantic interoperability. OSLO created an ontology in three main domains of interest: (i) persons and organisations [15]; (ii) locations; and (iii) public services[76], in a setting where stakeholders are focusing on their similarities rather than on their differences [57]. This was achieved by implementing a process and methodology for developing semantic agreements, based on the ISA methodology [30]. In 2015 the ownership of OSLO and the responsibility for the governance and life-cycle management of the ontology was transferred to the regional government[77], which started the follow-up project OSLO.

OSLO provides a policy framework for technical topics, including the URI-strategy (Fig. 4.6), and domain-specific topics including a context-neutral model for addresses.



[76] https://purl.org/oslo
[77] https://overheid.vlaanderen.be/OSLO-Wat-is-OSLO2

Fig. 4.6. Overview the OSLO working groups [51].

The steering committee (representing the 'authority') agreed upon the specification process and the selection of the different working groups. Each thematic working group develops a context-neutral vocabulary, by extending EU standards (ISA and INSPIRE) with specific local concepts that support the processes of the different governmental levels in Flanders. Working groups discuss relevant entities, relations, and attributes, which are iteratively refined and formalized. At the end of the process the steering committee validates the specifications and the vocabulary. The working group consists of over 70 experts representing local governments, Flemish administrations, telecom providers, utility companies (water, energy), the real estate sector, and non-profit organisations. To derive the requirement for the high-level domain model, the participants start by specifying use cases. In a next step the working group or a subgroup defines the attributes. The umbrella working group integrates the results into the global domain model. Finally, a conformance statement is created.

4.2.5 Overview of the Danish, Dutch, and Flemish Address Registries

Comparing the strategy and development in Denmark with the Netherlands and Flanders, Table 4.1. Overview of the Danish, Dutch and Flemish Address Registries shows that all initiatives focus on a semantic coherent system of uniform identified base objects and relations and that all the registries are available as Open Data. The Danish Government is adopting the principles of Linked Data[78] to raise semantic interoperability. At the time of writing, the Danish Address Registry is not published as Linked Data yet.

[78] https://arkitektur.digst.dk/model-rules-english

Table 4.1. Overview of the Danish, Dutch and Flemish Address Registries

	Danish	Dutch	Flemish
Open Data	YES	YES	YES
Reuse of vocabularies	INSPIRE	patchy	ISA and INSPIRE
Unique Identifier	GUID	URI	URI
Linked Data	NO	YES	YES

4.2.6 URI strategy in Flanders in relation to W3C, ISA, and other EU member states

In this section, we will compare the different prevailing URI standards which influenced the Flemish strategy (as seen in Fig. 4.7). The OSLO URI working group has developed a URI standard for persistent identifiers that supports government administrations in Flanders by providing guidance that ensures that HTTP URIs are future proof. By providing a standard, an accompanied guidelines document and individual guidance on request, Digital Flanders Agency ensures a consistent URI structure for all publishers. The URI strategy is based on principles from W3C documents[79], ISA guidelines[80], and a comparison of the URI strategy in the Netherlands [46], the latter being inspired by the UK recommendation ‘Designing URI Sets for the UK Public Sector’ [19,20].

[79] http://www.w3.org/TR/cooluris/

[80] https://joinup.ec.europa.eu/community/semic/document/10-rules-persistent-uris

Interoperability Solutions for European Public Administrations (I)	United Kingdom (II)	The Netherlands (III)	Flanders, OSLO² (IV)
http://{domain}/{type}/{concept}/{reference}	http://{domain}/{type}/{concept}/{reference}	http://{domain}/{type}/{concept}/{reference}	http://{domain}/{type}/{concept}(/{reference})*
Domain: the combination of the host and the relevant sector.	**Domain**: based around the data.gov.uk domain, split by sectors as sub-domains.	**Domain**: a combination of an internet domain and a path.	**Domain**: a thematic subdomain is optional
Type: id or item for real world; doc for documents; def for concept definitions; set for datasets; a string specific to the context	**Type**: URI type (id for Identifier URI, doc for Document URI), Representation URI (def for Ontology URI, set for Set URI)	**Type**: id for identifier of a real world; doc for documents; def for definition of a term in an ontology	**Type**: describes the nature of the referenced resource; id: non-information resource; doc: document describing a non-information resource; nir: used instead of id and doc when using hash-based ids; list; raw; ns; meta
Concept: collectionID, the type of real world object or the name of the concept scheme.	**Concept**: A word or string to capture the essence of the real-world 'Thing' that the set names.	**Concept**: indicates the kind of thing the URI identifies. Usually it is a class name.	**Concept**: specifies the type of the resource. This can be interpreted using a simple trick: {resource} is a {category} {class}
Reference: refers to a specific item, term or concept.	**Reference**: A string that is used by the set publisher to identify an individual instance of concept. The reference should match the way that it is used in normal use.	**Reference:** usually a string which is used in the existing register already.	**Reference**: a repeating pattern that allows for hierarchical structure. Each element follows one of the following structures: {base-reference}; {base-reference}/{version}

Fig. 4.7. Comparing the URI strategy of ISA, United Kingdom, the Netherlands, and Flanders. The recommended pattern is (a variation on) 'http://{domain}/{type}/{concept}/{reference}'. The domain is the combination of the host and the relevant sector. It is a matter of choice whether the sector is defined as a sub-domain of the host or as the first component of the path.

The W3C group 'Cool URIs for the Semantic Web' discusses the 303-redirect (a way to redirect web applications to a new URI) and hash-URI approach to identify real-world objects and the use of content negotiation but does not suggest or impose any kind of URI structure. Some examples do include 'id' and 'doc' that are also used in other URI strategies that are evaluated in this chapter. The document does not suggest a specific approach regarding redirect versus hash-URIs, as the best approach is case dependent.

Three approaches are described: (i) when using redirects: http://www.example.com/id/alice 303-redirects to http://www.example.com/doc/alice, (ii) in case using fragment identifiers: http://www.example.com/about#alice is automatically truncated to http://www.example.com/about, and the last approach (iii) is combining both approaches, so that there is a 1-to-1 mapping between a real-life object and its describing page: http://www.example.com/bob#this.

As Fig. 4.7 shows, the compared strategies mostly share the same approach. The most important differences are related to the interpretation of 'type' and 'concept'. The ISA 'type' can be one of the following values: *id* or *item* for real-world objects; *doc* for documents that describe those objects; *def* for concept definitions; *set* for datasets; or *a string* specific to the context, such as 'authority'[81] or 'dcterms'.[82] The Dutch 'type' indicates the kind of the URI: *id:* identifier of a real-life object in a registry; *doc:* documentation about the real-life object by this registry, and *def:* definition of a term in an ontology. The Dutch 'type' is in line with the UK strategy, also using *id, doc* and *def* [19]. The OSLO 'type' describes the nature of the referenced resource. It has to be chosen from the following list: *id:* when referencing a non-information resource; *doc:* document (that describes a non-

[81] http://publications.europa.eu/code/en/en-130100.htm
[82] http://dublincore.org/documents/dcmi-terms

information resource and, the target of a 303-redirect after resolving an id resource); or *ns:* taxonomies, ontologies or vocabularies. Additionally, the *id* class does not demand the described resource to originate from a registry. Finally, the Flemish OSLO standard recommends that the 303-redirect approach is used (as is recommended by ISA). Resources in the *ns* class are allowed to use fragment identifiers. The UK URI strategy refers to the W3C Cool URI rules when it comes to resolving URIs and considers both options: fragment identifiers and 303-redirects.

The ISA concept is a collectionID, the type of real-world object identified (e.g., a road), the name of the concept scheme (e.g., 'language'). Finally, the reference refers to a specific item, term, or concept. The Dutch concept indicates the kind of thing the URI identifies, usually being a class name. Lastly, the {reference} is usually a string which is used in the existing registry already. When using dates to refer to versions, they advise using the W3C DateTime format. The OSLO concept specifies the category of the resource. This becomes part of the identifier as follows: {resource} is a {concept} {type}. For example, ‘https://data.vlaanderen.be/doc/adres/3706808’[83]can be read as: ‘…/id/adres/3706808’ is an identifier for an address’.

ISA allows a collectionID to be used as well, while this is not allowed in the Flemish rules. We have detected no conflict with the standard in the Netherlands. The new part is referred to as the ‘reference’, a repeating pattern that allows for a hierarchical structure. Each single reference element follows either one of the following structures: (i) {base-reference} or (ii) {base-reference}/{version}. The base reference allows for specification of a resource. It should be interpreted in the context of all preceding references. Finally, the version can represent a specific absolute date (in W3C format), a relative date (e.g., ‘latest’) or a version (e.g., 1.2). Versions are only

---

[83] http://data.vlaanderen.be/id/adres/3706808

allowed where different versions of resources are allowed to co-exist. ISA provides little details about the specific format of the reference. There are no conflicts between ISA and the Flemish rules.

4.3 Core Vocabularies for Addresses

This section discusses the prevailing European ISA and INSPIRE Address Vocabularies and evaluate their 'fitness for use' for modelling the CRAB address registry. This 'rewiring' of the CRAB happened as part of the development of a building registry which required an important revision of the address registry. Finally, we discuss the rewiring of the address model in depth. The resulting model bridges to other administrative domains like Persons, Businesses, and Public Services.

4.3.1 INSPIRE Data Specification for the Spatial Data theme Addresses

Driven by 'the Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007' to establish an INSPIRE, the 'INSPIRE Data Specification for the spatial data theme Addresses' was released in 2009. The data specification was developed by the Thematic Working Group on Addresses, in which the Flemish Agency for geographical information (now Digital Flanders Agency) participated.

INSPIRE [39] defines an address as the "Location of properties based on address identifiers, usually by road name, house number, and postal code" and serves different purposes including location, identification, jurisdiction, sorting & ordering, and emergency response (e.g., CRAB is used by the emergency services). The INSPIRE address defines following address components [39]: administrative unit name (the name of a country, municipality, or district), address area name (a non-administrative area or the name of a natural feature like a lake), thoroughfare name (a street name, the name of waterway) and postal descriptor. The vocabulary is documented using the Unified Modeling

Language[84] (UML)and the structure & constrains of the XML representation are described using the XML Schema Definition Language[85].

4.3.2 ISA Core Location Vocabulary

The ISA Core Location Vocabulary was developed following 'the Directive Decision 'COM/2008/0583 final - COD 2008/0185' of the European Parliament and the Council on ISA.

According the EC in 2011, there were two main principles which defined the Core Vocabularies [26]: (i) highly reusable: the specification is simple and captures basic and generic characteristics of an information entity, regardless of the context this entity is used, and (ii) extensible: domain specific specializations can be drafted on top of the core representation.

The Core Vocabularies working group (Location Task Force) defined the Core Location Vocabulary in 2012 as a minimum set of classes and properties for describing a location represented as an address, a geographic name, or a geometry [24]. A set of commonly agreed Core Vocabularies supported by the EU member states provides a concrete starting point for promoting semantic interoperability among European public administrations. The Core Location Vocabulary refers to the INSPIRE Data Specification for Addresses: the granular address properties (PO Box through to Post Code) are taken from the INSPIRE address guidelines. The vocabulary is documented both in an HTML[86] and a Turtle[87] specification that describes an RDF graph in a compact and natural text form[88]. An address that is provided using

---

[84] http://inspire.ec.europa.eu/data-model/approved/r4618-ir/html/index.htm?goto=2:1:1:1:7062

[85] http://inspire.ec.europa.eu/schemas/ad/3.0/Addresses.xsd

[86] https://www.w3.org/ns/locn

[87] https://www.w3.org/TR/turtle/

[88] https://www.w3.org/ns/locn.ttl

these properties will be INSPIRE conformant. The additional property of a full address is not part of the INSPIRE Address guidelines" [24].

4.3.3 A comparative survey of the ISA[89] and INSPIRE[90] address models

Although the ISA Core Location Vocabulary is driven by work on the INSPIRE directive [24], a mapping exercise revealed that INSPIRE and ISA have some significant differences. The ISA Core Location Vocabulary (Figure 4.8.) models its definition of 'Address' on AddressRepresentation from INSPIRE. AddressRepresentation (Figure 4.8.) is a composite datatype and not an entity as the ISA model suggests.

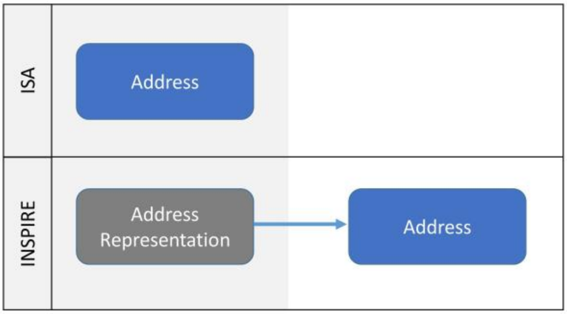


Fig. 4.8. Comparing the ISA and INSPIRE address model: Address and AddressRepresentation.

The fully-fledged INSPIRE address is intended to integrate the address life-cycle information with master-data management systems such as geographic information in support of Environmental Policy. The INSPIRE address representation and the ISA address are light-weight representations of an address. They are intended to be used in situations where the address is no more than an attribute of another

89 https://joinup.ec.europa.eu/asset/core_location/description
90 http://inspire-twg.jrc.ec.europa.eu/data-model/approved/r937/fc/

entity, e.g., a *Person*. In the ISA Core Location Vocabulary, an *Address* is intended to be used as an attribute of *Person*: http://data.vlaanderen.be/ns/persoon#gezinsadres. Moreover, it has no identifier of its own as there is only a pointer to the original structured address from which it was derived. Another problem is that ISA calls this entity/datatype *Address* and not *AddressRepresentation*. In doing so it created a schema-conflict with INSPIRE of the type 'homonym' [27]. The impact of using the same name for different concepts is that a user could expect the ISA gaddress to be a fully-fledged address instead of merely a lean representation without a unique identifier.

ISA added some attributes to INSPIRE's AddressRepresentation. *FullAddress* is one of them and appears to be an implementation of the 'ordered' constraint on INSPIRE: AddressRepresentation. Other attributes (*PO-Box*, *AdminUnitL1-2*) are specialisations of corresponding INSPIRE properties. As ISA limits itself to an address-as-an-attribute, certain properties of structured addresses are missing in the Core Location Vocabulary. Most noteworthy is *address position*. It can be argued that this property is also lacking in the INSPIRE:AddressRepresentation and that it is, therefore, permissible for ISA to leave it out of their definition. But as ISA renamed *AddressRepresentation* to *Address* one could ask if such an important aspect of structured addresses can be ignored. On the other hand ISA supports a *Location* entity/datatype which can have a geometry and address at the same time (if needed) and in that way an ISA:Address can be loosely coupled to a position. A relation with the parcel/building/building unit to which the address was assigned by the registering authority also lacks but is understandable as these addressable objects are currently not part of the Core Vocabularies.

On a more generic level, there are some additional differences to be mentioned, most important one being the definition of *identifier*. In the ISA Core an *identifier* is a composite datatype comprised of four attributes: the actual identifier or key and some metadata about the key, more specifically *source, date* and *type*. INSPIRE however leaves out the metadata and splits the key into its parts: a *namespace*, a *localid* and an optional *versionid*. Although *namespace* and *source* are somewhat overlapping concepts it seems that we have two different approaches to represent the *identifier*: one with metadata and an undivided key and one without metadata and a subdivided key. The INSPIRE programme is working on an RDF specification of the address data model. At the time of writing this article, the specification was still experimental[91].

4.3.4 OSLO address model

Based on the insights of Section 4.3.3, we conclude that the INSPIRE Address Specification is better adapted to the needs in the fields of all EIF levels, while the ISA Core Vocabularies provide a crucial integration with other administrative domains like Persons, Businesses and Public Services.

In order to achieve a feasible model, and to benefit from the best of both worlds, the Flemish Government has adapted the ISA address towards INSPIRE. The vocabulary is documented both in a HTML https://data.vlaanderen.be/ns/adres and a Turtle specification https://data.vlaanderen.be/ns/adres.ttl Most distinguishing feature of this rewiring was the instantiating of addresses as objects like BAG and OSLO, as discussed in Section 4.2.2. In the past only the address components were registered as separate objects. Part of this effort was a tighter alignment of the conceptual model to the INSPIRE:Address Specification. The INSPIRE model is very generic, designed to be applicable to all address definitions that can possibly exist in Europe. So, part of the alignment actually boiled down to mapping the more generic INSPIRE terms to existing terms in the Belgian context: http://data.vlaanderen.be/ns/adres#Adres (e.g., a Belgian

[91] http://inspire-eu-rdf.github.io/inspire-rdf-guidelines/

housenumber is actually a LocatorDesignator). Eventually, the resulting model was not only adopted on a regional level but also on a national level where it was published as Best-Add[92] as a result of a joint effort between the three regions and important federal institutions like the National Registry (of Persons), the postal services, the cadastral agency, and the National Geographical Institute amongst others.

Apart from making the generic aspects from the INSPIRE specification more concrete, the most difficult part of deriving a Core Vocabulary for addresses was modelling the clear distinction between the official or registered elements of an address and other more optional properties like sitenames or flooridentifiers. These were moved to a subclass AddressExtension: http://data.vlaanderen.be/ns/adres#Adresuitbreiding

Since the OSLO Address Vocabulary is to be considered as a local implementation of the ISA and INSPIRE specification, we aligned on the Belgian definition of Address: BeSt-Add.

However, to accommodate for foreign addresses we reused and extended the *Address* definition of http://www.w3.org/ns/locn#Address to denote an AddressRepresentation. The *locn* vocabulary was designed for capturing all kinds of addresses within the world. To keep the link with the Belgian addresses in a structured form, a relationship from the AdressRepresentation to the BelgianAddress is included.

---

[92] https://overheid.vlaanderen.be/CRAB-Belgie-BeSt-Add

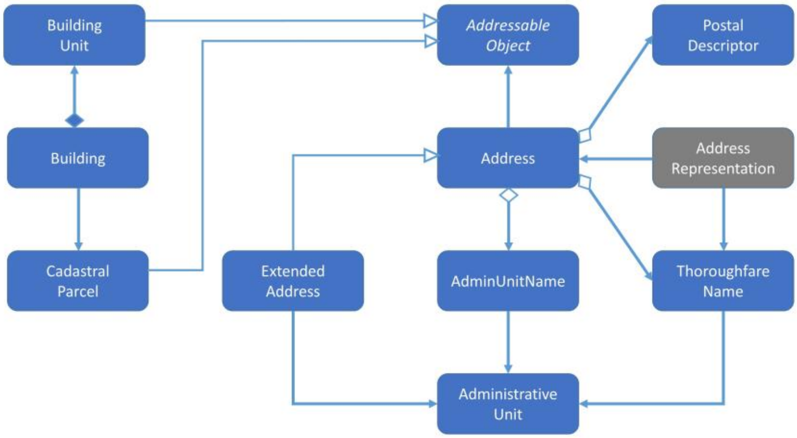


Fig. 4.9. OSLO Conceptual Address Model.

As we mentioned before we discovered that ISA and INSPIRE support different definitions of *identifier*. To accommodate both approaches we defined two subclasses for the key: simple key and composite key. We then added the metadata attributes from ISA. The concept of alternative identifier was incorporated since more than one identifier can be exchanged. For other generic elements like *geometry* http://data.vlaanderen.be/ns/adres#positie we aligned the OSLO Address Vocabulary to INSPIRE, although the requirement to use GML to serialise the geometry was replaced with a more generic construct also allowing well-known text representations of coordinate reference systems[93] (WKT) and such. The ISA Core Vocabularies nor the INSPIRE specifications mention much about metadata on an object-level. ISA explicitly leaves out metadata properties and INSPIRE mentions it only on a dataset-level although the available *lifecycleInfo* properties are in fact metadata.

[93] http://docs.opengeospatial.org/is/12-063r5/12-063r5.html

The OSLO Address Base Registry Vocabulary[94] is intended to exchange addresses. The application profile is a more elaborated vocabulary with additional restrictions (such as cardinality and code lists to be used) intended for a specific application. There is one application profile for addresses defined: namely the one for address registries. Examples of additional restrictions to the vocabulary are the municipality name and street name are obliged, and all code lists must be aligned on INSPIRE.

The address representation, which is an *Address* defined in https://www.w3.org/ns/locn, is extended with properties that are specific for Belgian Addresses, such as a *busnummer*, which is Dutch for *letterbox*, that identifies a dwelling and is identified by http://data.vlaanderen.be/ns/adres#Adresvoorstelling.busnummer. The *busnummer* is a specialisation of the W3C l*ocatorDesignator* http://www.w3.org/ns/locn#locatorDesignator that consists of a number or a sequence of characters that uniquely identifies the locator within the relevant scope. We define the *busnummer* property to have a domain of http://www.w3.org/ns/locn#Address and a range of http://www.w3.org/2001/XMLSchema#string. This can be compared to object-oriented programming which commonly defines a class address with an attribute called *busnummer* of the type String. The Range defines that the value of a *busnummer* must be a String as defined by http://www.w3.org/2001/XMLSchema#string.

The Model of the Flemish Address Register (https://data.vlaanderen.be/ns/generiek) applies the W3C PROV (https://www.w3.org/TR/prov-o/) ontology for modeling provenance. This structured metadata records the origin of addresses including when, why, and who created or modified the information. This allows users of the address register to evaluate if the information can be

[94] https://data.vlaanderen.be/id/applicatieprofiel/adresregister

trusted and to integrate the provenance information with other information sources.

4.4 Design principles of Linked Data applied to the Address Base Registry

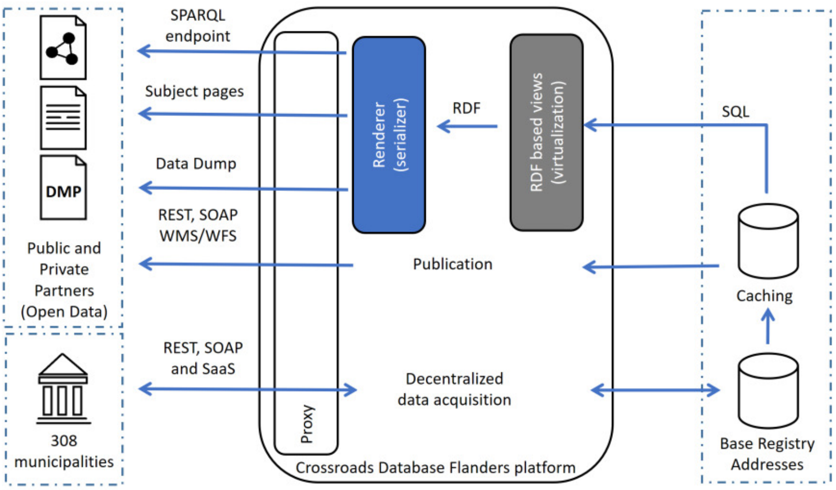


Fig. 4.10. Crossroads Database Flanders Architecture.

In 2008, the regional government of Flanders initiated the: ‘MAGDA’ data platform, which supports organisations to publish and access PSI. Due to the need of providing more integrated, interconnected and cross-sectoral services [40], Digital Flanders Agency started up a pilot project and applied the principles of Linked Data because they expected them to increase interoperability. In 2018, the MAGDA-platform was rebranded to Crossroads Database Flanders (CDF).

This section explains how the CDF-platform (Figure 4.10.) was enabled to publish Linked Data on the Web. We will discuss the reference components: the proxy server serving as the entry point of the Linked Data Infrastructure and the renderer which serialises the RDF or creates a human-readable HTML subject page and the RDF store. We will outline how the addresses, mined from the local communities, are published at the SPARQL-endpoint and HTTP services in 'nearly real-time'. Finally we will elaborate on the deployment strategy which allows other agencies to either reuse the complete setup as is or recombine the components to another setup with minimal effort.

4.4.1 Crossroads Database Flanders-platform

In Flanders, the administration responsible for the 'base registries' hosts a single point of service delivery via the data portal CDF which allows citizens and businesses to access all government supplied services, regardless of which authority or channel provides them. At this moment 300 local governments, 978 public partners, over 2000 private partners, and 1 out of 3 citizens are connected to the secure platform performing millions of requests on information objects about citizens, businesses, addresses, buildings, and their locations [45]. The first product released on the CDF-platform according to the Linked Data principles is the Central Reference Address Database, containing over 3 million addresses and their geographical coordinates. These addresses are synchronised in real time between 300 local governments and the Linked Base Registry for Addresses.

The realised Linked Data Infrastructure for CRAB applies traditional, known, Linked Data publishing techniques. The spill-over effect is the adoption of the URIs and vocabularies in traditional services including geographical WFS and SOAP Services, which facilitate interoperability between the different endpoints.

The solution is not dedicated towards solely publishing CRAB, but it is conceived with the ambition to support further exploitation beyond the owners of the CRAB registry. Other governmental bodies within the Flanders region are welcomed to reuse the complete solution or components from the solution. By embracing this ambition in the design phase future buy-in can be achieved. This ambition reflects in the chosen technologies that instantiate the solution. For each component commercial as well as open-source options are possible. An overview of the implementation details is published on the GitHub of the Flemish Government Administration: https://github.com/Informatievlaanderen/Data.Vlaanderen.be/tree/master/documentation.

4.4.2 Proxy-server

The proxy server is the entry point of the Linked Data Infrastructure. It implements the desired content negotiation. Information Flanders created a reference implementation that satisfies the URI strategy and the implementation guidelines of Information Flanders as a reusable component.

In addition to the effort gain, more coherency on the supported content negotiation is achieved. It is clear which headers are to be supported with which values. For instance, the reference implementation supports the content types ‘application/rdf+xml’, ‘text/turtle’, and ’text/html’. The implementation of the HTML subject page renderer is based on the mu.semte.ch[95] ecosystem. The reference implementation will provide answers to a special situation such as when the accept header is empty. Using content negotiation, the proxy will forward each valid URI to a subject page in human readable format (HTML) or to a machine-readable document (RDF in different serialisations). Human readable subject pages exist to provide useful information about the resource and to build a trust relationship with the user of the URI. The existence of a page that contains human interpretable data is key for the adoption of the URI as a trustworthy identifier, in particular for users who are less familiar with Linked Data. It is intended that the content of the human readable and the machine-readable formats are the same. But for reasons of easing the online exploration of the data, there might be slight differences to support

[95] https://mu.semte.ch/

better exploration of the data. For instance, data sources usually physically store relationships only in one direction.

For instance, a streetname belongs to a municipality. When exploring, the data users should not be restricted to follow only the directionality stored in the data source, but any direction should be offered. The provided reusable component for creating human-readable subject pages generates HTML (Figure 4.11.) according to the Flemish government style guide[96] and is easy configurable to create alternative views based on the provided RDF data. For the machine-readable subject pages two alternative solutions are being provided: users can perform direct queries on the RDF SPARQL endpoint or they can retrieve stored content on disk often referred to as a caching.

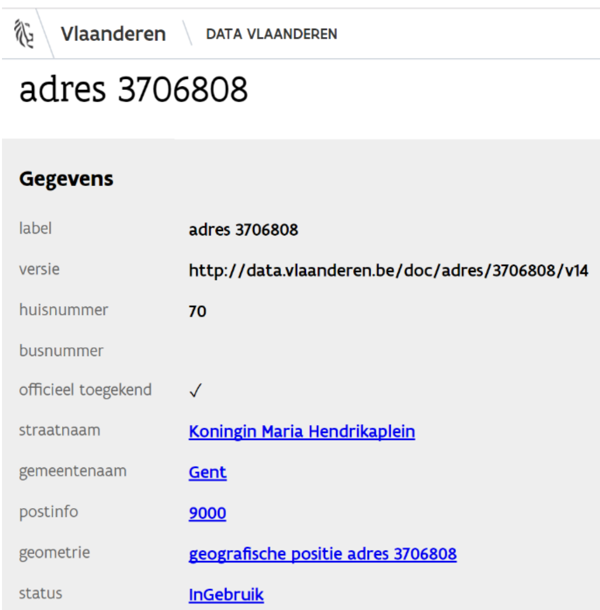


Fig. 4.11. Human readable representation of an address subject page: http://data.vlaanderen.be/id/adres/3706808/

96 http://www.vlaanderen.be/nl/vlaamse-overheid/webuniversum

4.4.3 The SPARQL-endpoint and API

Public and private partners can query the base registry using a service which accepts SPARQL queries. SPARQL[97] is a query language for the RDF, similar to the Structured Query Language (SQL) for relational data. We use the HTTP message header (e.g., 'application/rdf+xml' or 'text/html') which indicates the request format by the client application to provide the matching representation, often referred to as content negotiation. The renderer serialises the RDF or creates a human-readable HTML subject page. The RDF store is implemented as a virtual triple store as the Virtuoso[98] transformation engine transforms the data from the 'base registry cache' (SQL) on the fly to RDF triples. The advantage is that the address updates from the municipalities are published at the SPARQL-endpoint in 'nearly real-time' and are accessible via http://data.vlaanderen.be/sparql[99].

The API[100] of the Address Registry[101], which provides programmatic read and write access to Base Registries for Flanders is a classical XML-implementation for reasons of backwards compatibility, the semantics are in line with the RDF data model, and the service is using dereferenceable URIs as primary identifiers.

4.4.4 Deployment strategy

The address registry is deployed on a cloud infrastructure: Microsoft Azure[102]. All the components are available as runnable services using Docker. Docker is an open-source engine that automates the deployment of applications into containers [50]. This deployment strategy fits exactly the ambitions to provide a setup that is reusable by other governmental agencies. The infrastructure is setup using

---

97 https://www.w3.org/TR/rdf-sparql-query/
98 https://virtuoso.openlinksw.com/
99 http://data.vlaanderen.be/sparql
100 https://beta.basisregisters.vlaanderen.be/api/v1/adressen/200039
101 http://beta.basisregisters.vlaanderen.be/Help
102 https://azure.microsoft.com

Terraform[103]. The Terraform configuration describes the Docker Swarm[104] setup, that is responsible for executing and monitoring the services. One can opt to run the complete setup as is, but one can also recombine the components to another setup with minimal effort. For instance, replacing the ONTOP component with a commercial edition simply means replacing the docker reference of the ONTOP solution with the docker providing the commercial edition. The Docker layer creates a transparent system architecture with well-documented flexibility points. In addition the Docker layer makes the provided solution independent of the hardware used as it can be run on a Container As A Service solution offered by a cloud provider such as Azure, but it can also be run on a single machine or a developer machine.

4.5 Applications of the Linked Base Registry for Addresses

4.5.1 Adoption of Addresses as Linked Data in the private sector

An example of linked addresses as an interoperability facilitator is the case of Postbuzz[105] and Tenforce. This private initiative lets citizens and businesses discover what is buzzing in their neighbourhood including hyperlocal news. The location of the citizen, the businesses, the news, and events play a crucial role in the location-based services. On the other hand, Tenforce has developed a new service which builds upon the Generic Information Platform for the Public Domain[106] (GIPOD) and gathers all information concerning works or

---

[103] https://www.terraform.io/
[104] https://github.com/docker/swarm
[105] https://www.postbuzz.com
[106] https://overheid.vlaanderen.be/producten-diensten/generiek-informatieplatform-openbaar-domein-gipod

manifestations in the public domain. The GIPOD dataset contains URIs which identify addresses impacted by roadworks or manifestations (ongoing). Postbuzz provides personalized services by notifying businesses if their location cannot be reached, including detailed status information such as the start and end date of the manifestation. The personalized notifications are visualized as tiles (see Fig. 4.12).

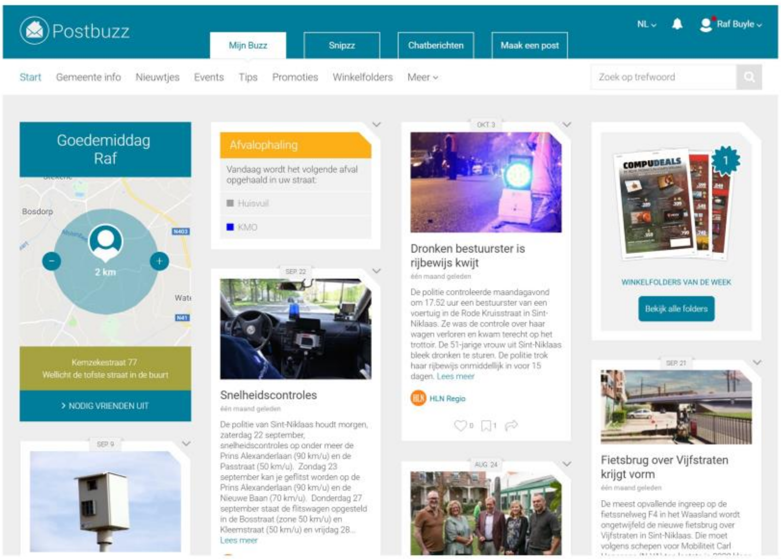


Fig. 4.12. Postbuzz lets citizens and businesses discover what is buzzing in their neighbourhood, including hyperlocal news.

Linked Address Data has the ability to combine data from the businesses and their addresses in the Postbuzz knowledge graph using the SPARQL endpoint with the impacted locations in GIPOD. When re-using PSI, until now Postbuzz had to look-up the coordinates of the provided addresses. In the past, the city of Ghent in Belgium annotated their news with the local address identifier of the Flemish government. Because of a lack of context, the developers of Postbuzz missed out this valuable clue. Linked address data provides context information via the dereferenceable address identifiers (URIs) including a reference to the vocabulary and links to other useful information such as other addresses in the same street.

4.5.2 Better adoption of Addresses within the public sector

The Flemish government administration applied the Linked Base Registry for Addressed in a high impact project 'citizens portal' that will allow 6.4 Million citizens to access public services. The citizens profile enables local and regional government administrations to integrate different public services, regardless of the channel they use. The starting point is the citizen that has a secure login on the portal, identified by its electronic identity. The address identifier is key for several administrative processes: birth, domicile, place of residence, and the death of the citizen & its relatives. Location-based information has rarely been used in transactional public services in Flanders because a lack of vocabularies that bridge between alphanumeric and geographical information. The RDF data model for addresses may facilitate the integration of addresses in several business domain services on a semantic level. While developing the citizens-portal, we noticed that developers are familiar with JSON, which has a very simple syntax but no inherent tied semantics. Developers want a simple, extensible way to create an API that gets the job done and does not design them into a corner[107]. As such, JSON-LD makes the bridge between JSON and the formal OSLO data specification. The semantics of the address properties in this business-oriented services are a subset of the address in the base registry. By adding a dereferenceable URI, the analyst or developer can discover and use additional data which is not provided in the API. The data can be retrieved by dereferencing the URI and selecting a representation through content negotiation. The semantic agreements reached at the business level of the 'citizens portal' project, are modelled in UML. The UML is automatically transformed into an RDF model and linked to other vocabularies including addresses. The formal specification is then published at

[107] https://json-ld.org/spec/latest/json-ld-api-best-practices/

data.vlaanderen.be, including a JSON-LD context which allows embedding the semantic agreements in business-oriented services.

To preserve the 'context' within these services, the shortcut terms in the service are mapped to the terms in the data.vlaanderen.be[108] RDF vocabularies. This principle is known as 'expanded term definition' and is accomplished by using a JSON-LD context [109]. The term 'verblijfsadres' (place of residence) is mapped to the vocabulary 'http://data.vlaanderen.be/ns/persoon#Verblijfplaats' (see Figure 4.13). These JSON documents can then automatically be interpreted[110] as Linked Data.

```
"@id": "https://data.vlaanderen.be/id/persoon/00000014182",
"@type": ["GeregistreerdPersoon", "Inwoner"],
"heeftVerblijfpaats": [
  {
    "@type": "Domicilie",
    "verblijfsadres": {
      "straatnaam": {
         "@id": "http://data.vlaanderen.be/id/straatnaam/7003",
         "naam": "Lubbeekstraat"
      },
      "huisnummer": "20",
      "gemeentenaam": {
         "@id": "http://data.vlaanderen.be/id/gemeentenaam/3210",
         "naam": "Lubbeek"
      },
      "land": {
         "@id": "http://publications.europa.eu/resource/authority/country/BEL",
         "naam": "België"
      }
    },
  }
]
```

Fig. 4.13. JSON-LD example used in the 'citizens portal' which links the citizens domicile to an address in the central reference address database.

[108] http://data.vlaanderen.be/ns/
[109] https://www.w3.org/TR/json-ld-syntax/#the-context
[110] https://www.w3.org/TR/json-ld-syntax/#interpreting-json-as-json-ld

4.6 Discussion

By focusing on Linked Data, the Flemish Information Agency aims to increase adoption of base registries by private and public partners via the presence on the Web. As a spill over, this architectural approach has important benefits for the internal organisation. In this section, we will reflect on the added value of Linked Data on addresses and on how OSLO fosters semantic interoperability. We use the design principles as asserted by Sir Tim Berners-Lee (Fig. 4.3) and an evaluation framework from the EC [44] to reflect on the implementation of Linked Data on addresses in Flanders.

4.6.1 Understandability

We have learned that the formal structure of addresses is difficult to understand. A telling case is Postbuzz, that missed out the context of addresses due to a lack of context. Linked address data provides context information via the dereferenceable address identifiers (URIs) including a reference to the vocabulary and links to other useful information such as other addresses in the same street. In addition, it is crucial that the vocabulary is supported by the community of stakeholders. OSLO created a setting where the stakeholders from the thematic working group ‘addresses and buildings’ focused on their commonalities rather than on their differences. In an early stage, consensus building and a meet-in-the-middle approach are essential for a broad support of a semantic standard [15]. A strong focus on Linked Data and on reusable ontologies has important benefits for the internal information household.

4.6.2 Scattered information

As 80% of the informational needs are related to geographic locations [33] and spatial data often uses domain specific standards, address information is scattered on different IS. By invoking the URI-standard, we have created universally unique and stable identifiers for addresses (first design principle). This allows linking addresses more easily with other datasets. Until now, the CRAB identifier was only unique on a system level, and implicit at the Flemish level. According to W3C's Library Linked Data Incubator Group, working towards a Linked Data architecture, "can help organisations improve their internal data curation processes and maintain better links between, for instance, digitised objects and their descriptions... even where data is not entirely open" [5]. Where an address in relation to a person to date used a different definition than the authentic source for addresses, extending and adopting the ISA Core Vocabularies enabled the Flemish Government to create a link between 'persons' and 'addresses' in an unambiguous way [24]. On top of the reuse of the vocabulary, persons can be linked to the authoritative address via a URI instead of duplicating the data.

4.6.3 Usability and interoperability

Because it is difficult to link or integrate location based information, the Flemish government has created an architecture (second principle) which allows retrieving the address information via the Web. This will allow users to link directly to an address using a lightweight HTTP service instead of duplicating the data in their information system. We expect this will avoid shadow databases, containing redundant and /or outdated information. OSLO is aligning the vocabularies of the different base registries, which are part of a semantic coherent system. This allows integrating address information more easily in the various business processes.

4.6.4 Different user needs

By using content negotiation mechanisms it becomes possible to serve different representations of a resource from the same URI. This allows clients to specify which version best fits their needs . The renderer serialises the RDF and creates both a human-readable HTML subject page and various machine-readable formats, including 'application/rdf+xml' and 'text/turtle'.

4.6.5 Machine-readability and reasoning

When looking up an address via the Web, we provide useful machine-readable information (third principle), using RDF as a data model. We have extended the RDF using the European ISA and Core Vocabularies. This enables addresses to become 'self-describing', which allows applications that are not familiar with the Flemish context to dereference the URI and find the definition [12]. Reuse of vocabularies can also lower the integration cost, particularly in case of reference data [2]. In addition, we have included links (fourth principle) to other URIs, so users can follow the links from the address to the street. Other parties on the Web can link to the addresses, making their data more useful. By adding a dereferenceable URI, we provide context to analysts or developers who can discover and use additional data which is not provided in the API.

4.6.6 Future work

The Flemish government has invested in different Linked Data distribution strategies that are aligned with the needs of their clients. A single-file data dump has a low server-side complexity but does not allow live querying on the Web. This introduces high costs for the clients. The SPARQL endpoint allows flexible live querying, but its availability is problematic [52]. Therefore, the Flemish Government administration will explore the possibilities of Linked Data Fragments[111] (LDF), a REST(ful) publishing strategy that allows efficient offloading of query execution from servers to clients through a lightweight partitioning strategy [52]. REST outlines how to construct network-based software applications having the same characteristics as the Web: simplicity, evolvability, and performance [31]. Figure 14 offers a uniform view[112] on the different HTTP interfaces for Linked Data in relation to the server and client effort based on the LDF [31] vision.

---

[111] http://linkeddatafragments.org/
[112] http://linkeddatafragments.org/concept/

Another obvious extension to this research is the publication of Linked Open Data which is linked to information under privacy regulations. Examples include questions such as: "is the Uniform Resource Locator referring to an object under privacy regulations also privacy sensitive information and how to cope with these challenges within an operational context?".

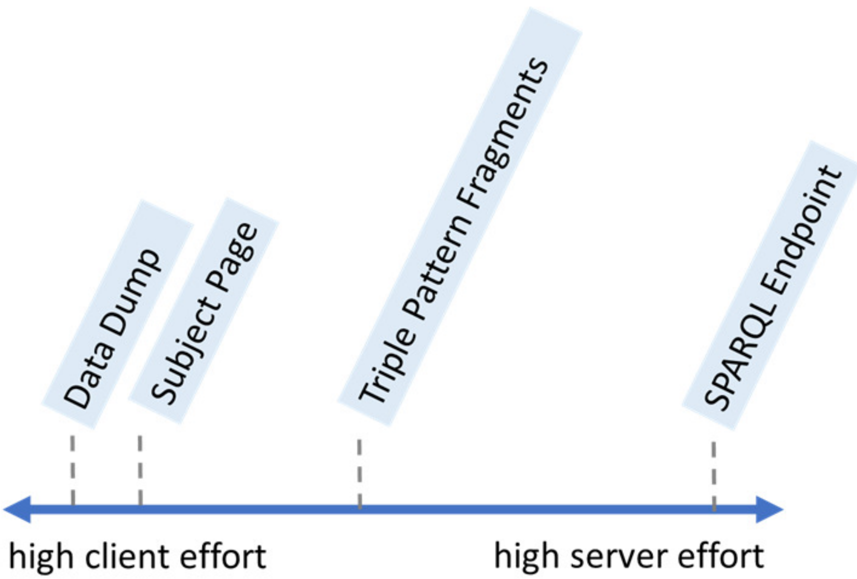


Fig. 4.14. An overview of the different HTTP interfaces for Linked Data in relation to the server and client effort [31].

4.7 Conclusions

In this chapter, we focus on the first viewpoint of our research question "*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?*". Although this chapter addresses multiple interoperability levels the main contribution on semantic interoperability.

In this research, we presented insights on the implementation of a Linked Base Registry for Addresses by unfolding the process followed towards raising semantic interoperability based on Linked Data principles. The Flemish Government administration aligned the Base Registry for Addresses with the design principles of Linked Data, because they expected this to increase interoperability. The approach goes beyond Linked Data initiatives which often take abstraction of the interoperability levels: namely on the *legal, organisational, semantic, and technical level* [23]. The private sector 'Postbuzz' initiative and the high impact project 'citizens profile' were the first to adopt the Linked Base Registry for Addresses in Flanders.

They indicate that Linked Data can indeed increase semantic and technical interoperability and can lead to the adoption of addresses in the public and private sector. For reasons of backward compatibility, the classical XML-webservices will be maintained for a certain amount of years. A spill-over effect of the Linked Data distribution is that the semantics of the XML-services were brought in line with the RDF data model and now have dereferenceable URIs as primary identifiers. While implementing the address vocabulary, we stumbled on competing international semantic standards and difficult choices on how to extend them to fit the local context. Therefore, it is crucial to have a governance structure for making and institutionalising pivotal decisions. This can be realised through a policy framework for technical as well as domain-specific topics, comparable to the OSLO programme in Flanders. This chapter identifies significant benefits in adopting the principles of Linked Data regarding base registries, not only by providing interoperability towards external stakeholders but also by fostering a more open architecture within the administration. A good example is the ability to unequivocally link *addresses* to other business objects, averting the creation of 'shadow' databases. We expect the insights from the linked address registry reported in this chapter to speed-up the process in other administrations that face the same complexity of publishing linked base registries.

4.8 References

[1] Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A systematic review of open government data initiatives. *Government Information Quarterly*, 32(4), pp. 399-418. https://doi.org/10.1016/j.giq.2015.07.006

[2] Archer, P., Dekkers, M., Goedertier, S., Hazard, N., Loutas, N. (2013). Business models for Linked Open Government Data: what lies beneath?, https://www.w3.org/2013/share-psi/workshop/krems/papers/LinkedOpenGovernmentDataBusinessModel.

[3] Archer, P., Goedertier, S., Loutas, N. (2012). D7.1.3 - Study on persistent URIs, with identification of best practices and recommendations on the topic for the MSs and the EC, ISA Programme. https://joinup.ec.europa.eu/sites/default/files/document/2013-02/D7.1.3%20-%20Study%20on%20persistent%20URIs.pdf

[4] Audit Vlaanderen. (2016). Thema-audit gemeentelijke belastingen en retributies. p. 81. http://www.auditvlaanderen.be/sites/default/files/atoms/files/Globaal_rapport_thema_audit_gemeentelijke_belastingen_en_retributies.pdf

[5] Baker, T. et al. (2011). Library linked data incubator group final report. report, W3C Incubator Group. P.25. https://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/

[6] Belgisch Staatsblad. (1977). Decreet tot bescherming van de namen van de openbare wegen en pleinen, B.S. 07.04.77, http://www.ejustice.just.fgov.be/cgi_loi/change_lg.pl?language=nl&la=N&cn=1977012832&table_name=wet

[7] Belgisch Staatsblad. (2007). Decree on the re-use of public sector information of 27/04/2007, B.S. 19.04.07. pp.56250-56256 http://ec.europa.eu/information_society/newsroom/cf/dae/document.cfm?doc_id=1701

[8] Belgisch Staatsblad. (2009). Decreet betreffende het Centraal Referentieadressenbestand van 8 mei 2009, B.S.01.07.2009, https://www.agiv.be/~/media/agiv/producten/crab/documenten/crab-decreet%20in%20belgisch%20staatsblad_090701.pdf

[9] Belgisch Staatsblad. (2011). Wet houdende instemming met het samenwerkingsakkoord van 2 april 2010 tussen de Federale Staat, het Vlaamse Gewest, het Waalse Gewest en het Brussels Hoofdstedelijk Gewest voor de coördinatie van een infrastructuur voor ruimtelijke informatie. B.S.26.04.2011, http://www.geopunt.be/~/media/geopunt/voor-

experts/documenten/afederale%20instemmingswet_tcm307-240907.pdf

[10] Belgisch Staatsblad. (2012). Decreet houdende wijziging van het CRAB-decreet van 8 mei 2009. B.S.12.06.12. http://reflex.raadvst-consetat.be/reflex/pdf/Mbbs/2012/06/12/121565.pdf

[11] Belgisch Staatsblad. (2015). Decreet tot wijziging van het decreet van 27 april 2007 betreffende het hergebruik van overheidsinformatie en het decreet van 18 juli 2008 betreffende het elektronische bestuurlijke gegevensverkeer, B.S. 30.06.15, http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=12909

[12] Bizer, C, Heath, T., Berners-Lee, T. (2009). Linked data-the story so far. Semantic Services, Interoperability and Web Applications: Emerging Concepts. pp.205-227. https://eprints.soton.ac.uk/271285/1/bizer-heath-berners-lee-ijswis-linked-data.pdf

[13] Bronhouders Afnemers Overleg. (2016). Objectenhandboek BAG. pp.6-12.http://www.kadaster.nl/documents/32706/37743/bag+objectenhandboek+2009+inclusief+besluiten+zorg-+en+studentencomplexen/389eb2fe-1c0f-450e-a56e-cac1b88d4111

[14] Bronhouders Afnemers Overleg. (2013). Processenhandboek BAG, basisregistraties adressen en gebouwen. p.13. http://www.kadaster.nl/documents/32706/37743/bag+processenhandboek+2013/cefe5fce-9330-4e4e-acae-f838e759e783, 2013

[15] Buyle, R., De Vocht, L., Van Compernolle, M., De Paepe, D., Verborgh, R., Vanlishout, Z., De Vidts, B., Mechant, P., Mannens, E. (2016). OSLO: Open Standards for Linked Organizations. Proceedings of the 2016 3rd International Conference on Electronic Governance and Open Society: Challenges in Eur-asia'. (November 21-22 2016). http://dx.doi.org/10.1145/3014087.3014096

[16] Buyle, R., Vanlishout, Z., Coetzee, S., De Paepe, D., Van Compernolle, M., Thijs, G., ... & Mannens, E. (2019). Raising interoperability among base registries: The evolution of the Linked Base Registry for Addresses in Flanders. *Journal of Web Semantics*, *55*, 86-101.

[17] Colas, C., Goedertier, S., Kourtidis, S., Loutas, N., Rubino, F. (2013). Core Location Pilot - Interconnecting Belgian National and Regional Address. pp.9-46. https://joinup.ec.europa.eu/svn/core_location/pilot/D5.2.1_Core_Location_Pilot-Interconnecting_Belgian_National_and_Regional_Address_Data.pdf

[18] Danish Enterprise and Construction Authority. (2010). Country Report Denmark: Addresses. pp. 1-11. http://danmarksadresser.dk/file/389699/0_CountryReport_2010_DK_EN.pdf

[19] Davidson, P. (2011). Designing URI Sets for Location. Chief Technology Officer Council. https://data.gov.uk/sites/default/files/Designing_URI_Sets_for_Location-V1.0_10.pdf

[20] Davidson, P. (2009). Designing URI Sets for the UK public sector. UK Chief Technology Officer Council. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/60975/designing-URI-sets-uk-public-sector.pdf

[21] De Wolf L., Van Hemelryck, H (2016). The Large-scale Reference Database (LRD), INSPIRE Conference 2016, p.15. http://inspire.ec.europa.eu/events/conferences/inspire_2016/pdfs/2016_psessions/28%20WEDNESDAY_PSESSIONS_A_16.00-18.00______pldwhvh160928av2_final_INSPIRE2016_GRB_met_comments.pdf

[22] eGovernment Sub-group (2009). Visions and priorities for eGovernment in Europe, Orientations for a post 2010 eGovernment Action Plan. p.4. https://www.aoc.cat/wp-content/uploads/2014/09/visions-priorities-egovernment-europe.pdf?x75273

[23] European Commission. (2010). European Interoperability Framework (EIF) for European public services: Annex 2. 2-21, https://ec.europa.eu/isa2/sites/isa/files/isa_annex_ii_eif_en.pdf

[24] European Commission. (2011). Core vocabularies. pp. 11, 50. https://joinup.ec.europa.eu/catalogue/distribution/core_vocabularies-business_location_person_v100_specification_in_pdf

[25] European Commission. (2013). A vision for public services. pp.1-6. http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=3179

[26] European Commission. (2013). eGovernment Core Vocabularies: The SEMIC.EU approach. p13. https://joinup.ec.europa.eu/sites/default/files/c1/23/4a/egovernment-core-vocabularies.pdf

[27] European Commission (2014). Handbook for using the Core Vocabularies. pp 8-10. https://joinup.ec.europa.eu/site/core_vocabularies/Core_Vocabularies_user_handbook/Handbook-for-using-the-Core-Vocabularies_v0.50.pdf

[28] European Commission (2017). European Interoperability Framework – Implementation Strategy, Annex 2. pp. 31-32. http://eur-lex.europa.eu/resource.html?uri=cellar:2c2f2554-0faf-11e7-8a35-01aa75ed71a1.0017.02/DOC_1&format=PDF

[29] European Commission, ISA (2017). New European Interoperability Framework (EIF). p. 5. https://ec.europa.eu/isa2/sites/isa/files/eif_brochure_final.pdf

[30] European Commission. (2013). Process and methodology for developing semantic agreements. p7. https://joinup.ec.europa.eu/sites/default/files/Process%20and%20methodology%20for%20developing%20semantic%20agreements.pdf

[31] Fielding, R., Taylor, R. (2000). Architectural styles and the design of network-based software architectures. Doctoral dissertation: University of California, Irvine. pp. 151. http://jpkc.fudan.edu.cn/picture/article/216/35/4b/22598d594e3d93239700ce79bce1/7ed3ec2a-03c2-49cb-8bf8-5a90ea42f523.pdf

[32] Folmer, E., Verdonk. Y. (2014). Linked data in beeld: Platform Linked Data Nederland. https://www.geonovum.nl/sites/default/files/Linked%20Data%20in%20beeld-2014-Locked.pdf

[33] Garson D., Biggs, R. (1992). Analytic Mapping and Geographic Databases, SAGE university paper, p. 87.

[34] Gottschalk, P. (2009). Maturity levels for interoperability in digital government. Government Information Quarterly, 26(1), 75-81. doi:10.1016/j.giq.2008.03.003

[35] Homans, L. (2015). Beleidsbrief Bestuurszaken 2015-2016, pp. 25-27. https://docs.vlaamsparlement.be/docs/stukken/2015-2016/g522-1.pdf

[36] Horst, N. L., Bjerre, S., Lind, M., Hvingel, L. (2014). The Basic Data Programme–A Danish Infrastructure Model for Public Data. Geoforum Perspektiv, pp. 36-38. https://journals.aau.dk/index.php/gfp/article/viewFile/637/518

[37] INSPIRE. (2003). Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information, http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:02003L0098-20130717

[38] INSPIRE. (2007). Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community. Published in the official Journal on the 25th April, http://inspire.ec.europa.eu/documents/directive-20072ec-european-parliament-and-council-14-march-2007-establishing

[39] INSPIRE (2013). Data Specification on Addresses. pp. VII, 6-7.http://inspire.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_AD_v3.0.pdf

[40] ISA programme by PwC EU Services EESV. (2013). Case study: How Linked Data is transforming eGovernment. pp 1-5. http://ec.europa.eu/isa/documents/case-study-data-integration-v2.pdf

[41] Jenssen, B. Basisdata: forståelsesramme og analysemodel til kategorisering af basisdata: Afrapportering fra Udvalg til nytænkning

vedrørende basisdata. Servicefællesskabet for Geodata,(2004). http://vbn.aau.dk/files/1815466/BHJ_Basisdata.pdf

[42] Kubicek, H., Cimander, R. (2009). Three dimensions of organizational interoperability. European Journal of ePractice, 6, pp. 1-12. http://www.ifib-consult.de/publikationsdateien/Kubicek_Cimander_ePractice_Journal_vol_6.pdf

[43] Lind, M. (2013). The Danish Address Program – Socio-economic Benefits from improved Addressing. pp. 4-13. http://odimpact.org/case-denmarks-open-address-data-set.html

[44] Loutas, N., Brulé, D., Wyns, B., Zajac, A. EU FISCAL DATA AS LINKED OPEN DATA. Publications Office of the EU. pp. 7-8 (2017). https://joinup.ec.europa.eu/sites/default/files/document/2017-07/eu_buget_as_lod_v1.0_0.pdf

[45] Mostaert, K., Buyle, R., Vanderstraete, T. (2016). Interoperability as a key enabler for citizen-centric location based services, INSPIRE. http://inspire.ec.europa.eu/events/conferences/inspire_2016/schedule/submissions/394.html

[46] Overbeek, H., Brentjens, T. (2013). Draft URI Strategy for the NL Public Sector. Ministry of the Interior of the Netherlands, Geonovum, https://www.w3.org/2013/04/odw/odw13_submission_14.pdf

[47] Overbeek, H., van den Brink, L. (2013). Towards a national URI-Strategy for Linked Data of the Dutch public sector. http://www.pilod.nl/doc/D1-2013-09-19_Towards_a_NL_URI_Strategy.pdf

[48] Peeters, K. (2011). Ministerieel besluit tot vaststelling van de CRAB-specificaties. B.S. 12.04.11, https://www.agiv.be/~/media/agiv/producten/crab/documenten/ministerieel%20besluit%20tot%20vaststelling%20van%20de%20crab-specificaties.pdf

[49] Rocheleau, B. (2003). Politics, accountability, and governmental information systems. Public information technology: Policy and management issues. pp. 20-52.

[50] Turnbull, J. (2014). The Docker Book: Containerization is the new virtualization. James Turnbull. https://dockerbook.com/TheDockerBook_sample.pdf

[51] Vanlishout, Z., Buyle R. Open Standards for Linked Organisations. (2016). INSPIRE conference. p.10. 2016, http://inspire.ec.europa.eu/events/conferences/inspire_2016/pdfs/2016_psessions/inspire_Linked-final.pdf

[52] Verborgh, R., Vander Sande, M., Colpaert, P., Coppens, S., Mannens, E., Van de Walle, R. (2014). Web-Scale Querying through Linked Data

Fragments. LDOW https://www.researchgate.net/profile/Ruben_Verborgh/publication/264274086_Web-Scale_Querying_through_Linked_Data_Fragments/links/53f498b10cf2fceacc6e918d.pdf

[53] Vlaamse Overheid. (2009). Decreet betreffende de Geografische Data-Infrastructuur Vlaanderen, Belgisch Staatsblad B.S. 28.04.09. http://www.ejustice.just.fgov.be/mopdf/2009/04/28_2.pdf#Page25

[54] Vlaamse Overheid. (2008). Decreet betreffende het elektronische bestuurlijke gegevensverkeer, Belgisch Staatsblad B.S. 29.10.08, http://www.ejustice.just.fgov.be/cgi_loi/change_lg.pl?language=nl&la=N&table_name=wet&cn=2008071823

[55] Vlaams Parlement. (2009). Ontwerp van decreet betreffende het Centraal Referentieadressenbestand, Stuk 2067, p.5. https://www.agiv.be/~/media/agiv/producten/crab/documenten/crab-voorontwerp-decreetg2067-1.pdf

[56] Wood, D., Zaidman, M., Ruth, L., Hausenblas, M. (2014). Linked Data - Structured Data On The Web, Manning, p.4. https://livebook.manning.com/#!/book/linked-data/chapter-1/99

[57] Zeginis, D., Hasnain, A., Loutas, N., Deus, S., Fox, R., Tarabanis, K. (2012). Collaborative development of a common semantic model for interlinking Cancer Chemoprevention linked data sources. The Semantic Web Journal. p. 14 http://www.semantic-web-journal.net/sites/default/files/swj263.pdf

CHAPTER 5. TECHNICAL INTEROPERABILITY

Computer science is no more about computers than astronomy is about telescopes.

— Edsger Dijkstra.

This chapter delves into technical interoperability. We research what methods are suited for publishing Open Data time series in a sustainable, predictable, and cost-effective way in the context of Smart Cities. This chapter argues that the method of Linked Data raises technical interoperability, lowers the publishing cost, and enhances the availability of the endpoints. This chapter is based on the paper, 'A sustainable Open Data platform for air quality data' [11].

5.1 Introduction

Today, the majority of the global population lives in urban areas and it is expected that this will increase to nearly 70% by 2050 [68]. Taking into account that in 1950, 70% of the world population was living in rural areas, these fast-growing megacities cause new sorts of problems such as complications in waste management, air pollution, and traffic congestions [14,68]. To avoid that this accelerated urbanisation turns into a crisis, cities must become 'Smart'. 'Smart' refers to a continuous comprehensive commitment to innovation in technology, management, and policy [56].

'Smart' cities however are not a new phenomenon. For example, the ancient city of Rome al-ready accommodated between five hundred thousand and 1 million inhabitants as a result of an advanced bureaucratic information system and efficient waste management [5]. Also, air pollution is not a new phenomenon: it was already observed in ancient times and recorded in ancient poems (e.g., by Horace (65 BC - 8 AD)) [51]. We can find the first regulations on air quality in Roman Law: "Aerem corrumpere non licet" (Air pollution is not allowed) [55]. Today, according to the World Health Organization (WHO) "air pollution represents the biggest environmental risk to health" [74]. The WHO also estimates that outdoor air pollution caused 3 million deaths in 2012. Due to urbanisation, ambient air pollution is concentrated in cities by cause of traffic, transport, domestic heating, and industrial emissions [59]. Air quality in Europe is regulated by Directive 2008/50/EC defining the threshold of the concentration of several pollutants. In order to make smarter decisions and raise awareness about air pollution, cities need to combine disparate data sources, including data on urbanisation, weather, and traffic [53]. The resulting insights can support better policymaking, including better urban planning decisions on where to build new roads, schools, or hospitals. Also, providing real-time air quality data to citizens can support them deciding, e.g., which route to take in order to avoid air pollution [13,40]. However, governments are struggling to create and maintain accessible public endpoints where high-value datasets can be published because of technological barriers which include availability, scalability, and the publishing cost [2].The EC [28] is fostering the re-use of high-value datasets, such as sensor data time series on air quality, by creating legislation that enforces (real-time) datasets to be published in a machine-readable format and to be automatically transferable through an Application Programming Interface (API). An API allows re-users to query data and combine data from several endpoints, without maintaining copies of all the data. Machine-readable data enables information to become self-describing, which avoids manual analysis to interpret the meaning and also avoids transformations.

As public city administrations cannot predict the load caused by re-users of any given dataset on their APIs, services often lack elasticity. A telling example is the launch of the 'solar map' (see https://apps.energiesparen.be/zonnekaart) in Flanders (the northern part of Belgium). The 'solar map' is an online application provided by the Flemish Energy Agency that shows the suitability for solar panels for any given roof and that calculates the payback period for investing in solar panels. The application builds on remote sensing data from the Flemish Information agency. During the public launch, the application went down because the services were under-dimensioned [71]. Such problems will escalate as the Web of sensors becomes a distributed, high volume, high velocity and heterogeneous mix of sensor and storage platforms [4].

While the concept of Open Data is widely motivated in terms of 'social' and 'economic' value [19,76] research on the cost of Open Data is scarce. It is a widespread argument that Open Government Data are data paid by the taxpayer and should be reusable for the general public [15]. However, government administrations need to make significant investments in collecting data, making it interoperable and publishing it for maximum re-use [43]. Therefore, our main ***research question*** addresses how data providers such as city governments can develop a sustainable method for publishing and archiving Open Data, in specific time series sensor data on air quality. This also includes the question on how governments can balance the cost between the data publisher and consumer.

Our research starts from the assumption that applying the method of LDF on public endpoints for air quality sensor data will lower the cost for publishing, raise availability thanks to a better caching strategy, and presupposes that the LDF approach is suitable to handle the business needs for an air quality sensor data endpoint.

This chapter is organised as follows: Section 5.2 discusses related work: we discuss the nature of air quality sensor data and explain why we need a different layer where sensor data can be connect-ed and interpreted by machines on a web-scale. We apply principles of Linked Data to sensor data. Finally, we elaborate on caching strategies to publish and archive data and on how the efforts between the publisher and the consumers can be balanced. Section 5.3 delineates a use case scenario and how it can be realised via a LDF approach. Section 5.4 benchmarks the Fiware and LDF architectures and assesses the cost for both data publisher and consumer. Section 5.5 discusses the findings of the benchmark in detail, which should allow government agencies and organisations to reuse the architectural components when refactoring their endpoints. We point out methods that can lower the cost for publishing and raise the availability of endpoints. We show how governments can balance the cost between the data publisher and consumer, and point-out how these insights can speed-up the realisations of a sustainable sensor network in terms of availability, scalability, and predictability.

5.2 Background and related work

5.2.1 Air Quality Sensor data

5.2.1.1 (Sensor) Data

Air quality in Europe is regulated by Directive 2008/50/EC which defines the threshold of the concentration of several pollutants, including fine particles (PM2.5), Sulphur dioxide (SO_2), Nitrogen dioxide (NO_2), PM10, and Carbon monoxide (CO) [29]. These pollutants are measured by Air Quality sensors. In the past, good quality sensors were only available in the high-end segment.

These high-cost monitoring installations are based on chemical analysers and are typically less densely deployed [59]. Recently, low-cost sensors, including electrochemical sensors (NO_2, CO, and SO_2 gas detection) and optical particle sensors for PM10, entered the consumer market at affordable prices [59]. A disadvantage of these sensors is that they must be calibrated due to the inferior mechanical and electrical tolerances. Also, their signal slowly changes independently of the measurements due to sensor drift [65]. To achieve a denser sensor-grid, low-cost sensors can be deployed at high volumes. The data can be calibrated by using the data of the less-dense high-end air quality sensors and can be combined with other valuable datasets, including weather and traffic.

To realise this, we need a 'layer' where sensor data can be connected and interpreted by machines. This layer will enable filling-out the blanks by interpolating nearby and historical data. Results can then be 're-gridded' to fit a uniform 'time-space' dataset, which can then be easily analysed.

Nittel [57] defines a sensor data stream as *"a time series of sensor measurements* m_{sj} *=<* t_i *,* l_{sj}*,* v_1*;* v_2*, ...*v_n *> generated by a sensor node* s_j*, based on one or more of its attached sensors"*. Both the timestamp t_i and the location of the sensor l_{sj} are crucial to interpret the sensed value v_n. The location can be a fixed value, in case of the high-end stationary air-quality sensors, or a variable value, derived from, e.g., Global Positioning System (GPS) when the low-cost sensors are mounted on a vehicle. Also, information related to the type (e.g., PM10 or NO2), the calibration parameters (including relative humidity and temperature) and the quality of the measurement is important [75]. It is expected that sensor streams will evolve to become spatially densely distributed and having a high-frequency sampling rate [57]. This will generate high volumes of data. Taken into account a couple of thousands air quality sensors and a sample each second, this represents over 126 billion samples collected per year, compared to another high volume dataset in the financial sector, this exceeds the number of bank transactions in Europe in 2018 [21].

5.2.1.2 Interoperability

When considering air quality sensor data, an important challenge is to identify and process the heterogeneous and mixed quality datasets (Hendler, 2014). Therefore, interoperability is crucial both for combining air quality data from different sources as well as for linking this data to other datasets such as traffic or weather data [17]. The EC [26] defines interoperability as the ability of organisations to share information and knowledge, through the business processes they support, by exchanging data between their ICT systems. To ensure that sensor data can be reused, various interoperability levels should be addressed; namely the legal, organisational, technical, and semantic level [24], see Fig 3.2.

As interoperability frameworks —including the EIF— assume a hierarchy in the interoperability levels, legal and organisational interoperability can only be implemented successfully when semantic and technical interoperability are in place [46].

First, as Smart Cities are networked ecosystems, organisations broaden their activities outside their policy domain which causes legislative barriers, introduces cost, and slows down innovation [27,73]. These barriers, referred to as ***legal*** interoperability, originate because of (i) non-interoperable legislation between different governmental levels such as municipalities and regional government, (ii) non-interoperable laws across different policy domains such as environmental regulations and mobility and lastly, (iii) clauses in agreements between governments and software vendors prohibiting the reuse of data.

Second, to create a sustainable sensor network, business processes among actors in the ecosystem must be aligned and documented, including service providers to agree on a Service Level Agreement framework [24,49]. These efforts on coordinated business processes, responsibilities, and expectations are referred to as ***organisational*** interoperability [26].

Third, ***technical*** interoperability covers the interconnection of applications and infrastructures, including interface specifications that interconnect systems and services [26]. In the IoT paradigm, objects that both harvest information from the physical world (sensors) and interact with their environment (actuators) are interconnected [36]. In these networks we distinguish northbound (NBI) and southbound interfaces (SBI). An SBI provides connectivity to the low-level components in the physical infrastructure such as sensors and actuators. Contrarily, an NBI provides connectivity with the other network nodes, regularly exposed as APIs. These APIs can shield the disparateness of the physical infrastructure and create a heterogeneous NBI, that further reduces the complexity of application development [50].

Sensors will not only generate an excessive amount of data but more importantly, also a great variety of data [36,62]. Hendler [41] refers to the phenomenon of *"trying to make sense out of a world that depends increasingly on finding data that is outside the user's control, increasingly heterogeneous, and of mixed quality"* as *'broad data'*. To face the challenges of *broad IoT data*, the principles of Linked Data enable data to become self-describing and machine-readable [62,66]. Machine-readable data allows autonomous agents to reason on the sensor data [18]. Linked Data builds upon the Web and uses typed links between data entities from disparate sources as explained in Section 4.2 [7, 8, 32, 45].

Finally, a deficit of semantic agreements causes multiple transformations on the different data models and syntaxes, which implies rewiring APIs and induces exorbitant costs [17]. The EIF refers to **semantic** interoperability as the meaning of information that is preserved and understood during the exchange between all communicating parties [24]. It includes both semantic interoperability, which refers to the meaning of the sensor data, and syntactic interoperability which specifies the grammar of the information such as XML or JSON [26]. The competing vocabularies which model the domain of air quality from slightly different viewpoints, including INSPIRE, NGSI-LD, and SSN/SOSA are discussed (see Table 5.1).

Table 5.1. Overview of the characteristics of three ubiquitous vocabularies [20,22,38,42].

Vocabulary	Use for	Wide industry/ community support	Ratified vocabulary for air-quality	Linked Data Support
INSPIRE	INSPIRE	NO	YES	NO
NGSI-LD	ETSI	YES	NO	YES
SSN/SOSA	W3C/OGC	YES	YES	YES

5.2.1.3 Semantic interoperability applied to air quality data

Space, time, and theme, are key dimensions for registering and analysing sensor data, as they make it possible to link the sensor data to other datasets [4]. The spatial component provides information about the location, the temporal attributes observe the time and time-zone, while the thematic attributes provide information about the sensor type [6]. In the context of this chapter, we will focus on air quality data that monitors several pollutants, including fine particles and Nitrogen dioxide.

European member states are obliged by law to report ambient concentrations and thus when thresholds are exceeded they need to inform the public [23]. The EC installed a legal framework "Infrastructure for Spatial Information in the European Community" (INSPIRE) that focuses on accessible and interoperable data. INSPIRE defines data specifications and implementing guidelines for exchanging air quality data, including a standardised description for sensors, sensor location, orientation, as well as the sensor's geometric, dynamic, and radiometric characteristics [22]. The conceptual schemas, which are part of the normative part of the standard, are defined in UML and in XMLschema. XMLschema is a description of a type of XML document which defines a set of rules for encoding documents in a format that is both human-readable and machine-readable.

In 2016, the EC requested the European Telecommunications Standards Institute (ETSI) to create an Industry Specification Group (ISG) to define a standardised API for Context Information Management (CIM) with Future Internet Ware (FIWARE) Next Generation Service Interfaces (NGSI) as a nominee. FIWARE is an open-source platform, supported by the EC. NGSI is a protocol to manage Context Information. The ISG delivered the Next Generation Service Interfaces as Linked Data (NGSI-LD) standard [20], which enables nearly real-time access to information from different distributed data sources. The NGSI-LD Information Model Structure (IMS) consists of two layers: a core Meta-model and a Cross-Domain Ontology which can be extended with domain-specific logic. The core Meta-model defines a minimal set of constructs which are the basic building blocks of the Cross-Domain Ontology including Entity, Relationship, Property, and Value [1,20]. The Cross-Domain Ontology describes concepts and constraints which provide consistency between the different IoT domains and applications. These concepts include Geographical properties, Temporal properties, and Time values [1,20]. The domain-specific logic can be extended with ontologies for a specific domain, including air quality, noise level, and water quality[113]. NGSI-LD requires a reimplementation of existing Linked Data domain models to fit the semantics of NGSI. At the time of writing, no ontologies for air quality were ratified by a standardisation organisation.

In 2017 the W3C and the Open Geospatial Consortium (OGC) Spatial Data on the Web (SDW) working group joined forces and developed a set of ontologies that annotate sensors, actuators, samplers, and their time series [38,42]. The ontologies include a lightweight core Sensor, Observation, Sample, and Actuator[114] (SOSA) ontology and the more expressive Semantic Sensor Network Ontology[115] (SSN) [38]. As such, SOSA provides a minimal core for SSN and ensures minimal interoperability. According to Haller, SSN and SOSA support various use cases including "satellite imagery, large-scale scientific monitoring, industrial and household infrastructures, social sensing, citizen science, observation-driven ontology engineering, and the Web of Things" [37]. The SSN and SOSA ontologies are available in line with the principles of Linked Data, which allow autonomous agents to reason on the capabilities, measurements, and provenance of an individual sensor or a network of sensors.

5.2.2 Data caching strategy

As the Sensor Web is distributed, multimodal (e.g., air quality, relative humidity, temperature, reference data), read-intensive and subject to large-scale load-variations, it becomes very brittle [4]. To lower a server's central processing unit (CPU) load — and thus the actual publishing cost — optimisations can be implemented via caching, which reduces traffic [10,12]. Caching stores data, which allows

[113] https://github.com/FIWARE/data-models/blob/master/specs/ngsi-ld_howto.md
[114] http://www.w3.org/ns/sosa/
[115] http://www.w3.org/ns/ssn/

handling future requests at a lower cost. In this section, we explain how Web caching — or HTTP caching — can reduce the need for client-server interaction.

The WWW has a software architecture which is designed for internet-scale across organisational boundaries and builds upon the principles of a distributed hypermedia application. To raise scalability the Web applications follow the REST architectural style — which is demarcated by a set of architectural constraints that enable caching — and is a blueprint of the behaviour of a well-designed Web application. The REST architectural style resembles the human Web, which builds upon hyperlinks, and a set of architectural constraints that facilitate architectural elasticity [33]. The three most essential constraints are: (i) uniform interface, (ii) client-server, and (iii) stateless and cache constraints.

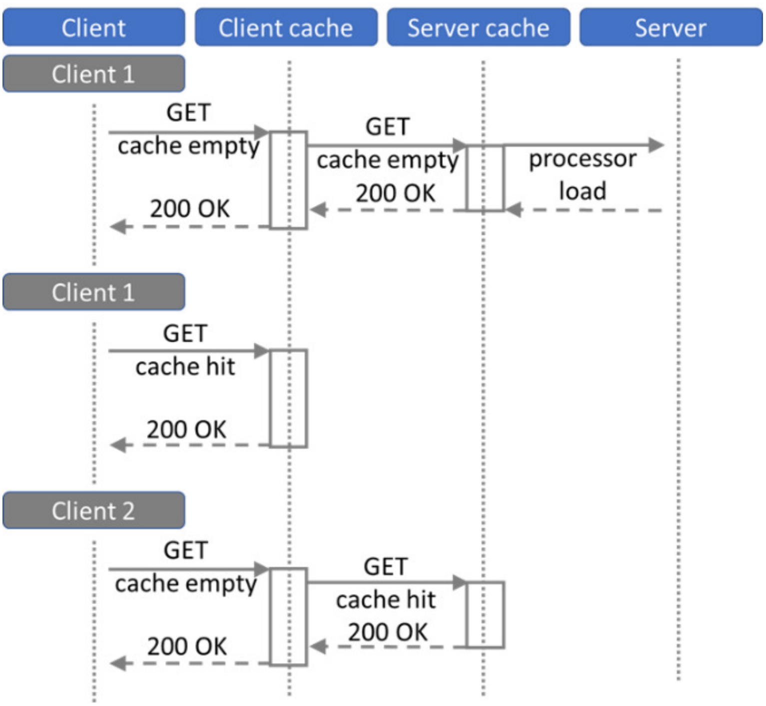


Fig. 5.1. Different stages of caching, no caching, client caching and server caching [33,61].

First, the uniform interface simplifies the architecture and empowers clients to evolve separately [34]. This key feature of the REST architecture unfolds in four sub constraints; namely (i) Uniform Resource Identifiers (URIs) — a generalisation of an HTTP URL to identify things on the Web — (ii) resource manipulation through representations — which implies that both the client and server can choose the representation of the resource such as JSON-LD, (iii) all messages are self-descriptive — they contain all the information that the client needs for interpreting the message such as indicating that the 'Content-type' is *'text/html'* —, and (iv) Hypermedia As The Engine Of Application State (HATEOAS) — which refers to the fact that a response should include links to possible actions. These hypermedia controls are comparable with links to forms, which makes out-of-band documentation needless [33,61].

Second, the client-server model implies several clients that communicate with a server. The client performs an action on a Web resource — using the HTTP protocol — by sending a request to the server [33].

Finally, Stateless and Cacheable prohibit the server to store the state of the client application [33]. This implies that every client request contains the context. This has two advantages; (i) as the server does not need to store the state of the client applications it can easily scale up, and (ii) the requests can be cached which lowers the load on the server (see Figure 5.1).

5.2.3 Balancing efforts between publisher and consumers

When servers encounter more complicated queries, they often respond with an error stating 'query too complex', or 'time out'. Examples of such complex queries with air quality data include route planning use cases where end-users want to be routed only through areas where a specific property of air pollution is lower than a certain threshold. Route planners should be able to do such querying, without the data publisher ever having thought about this specific use case. A similar case, illustrating a route planner which can evaluate any given query on the client-side, without having to rely on server-side functionality other than downloading the right LDF, was implemented by Colpaert [16]. The application builds upon the principles of the REST architectural style and is an implementation of LDF.

LDF is a conceptual framework that provides a uniform view on Linked Data interfaces [70], including SPARQL-endpoints, Linked Data documents, and data dumps. It makes the observation that all Linked Data interfaces have in common that they publish specific fragments of a dataset, whether very specific, such as with SPARQL endpoints, or very generic, such as in a single data dump. Also, in between solutions exist, where such a data dump is fragmented. A client can still answer individual queries by downloading the right subset of the knowledge graph. In order for a client to understand which fragments will be useful for answering a specific query, a server must document its fragmentation structure through hypermedia controls [70]. The LDF axis (data dumps to SPARQL endpoints) was first introduced for Triple Pattern Fragments, providing a low processing cost interface for answering Basic Graph Patterns.

To balance the effort between the data publisher and consumer for Air Quality Data, we limit the interface by applying a temporal and spatial fragmentation. When querying a dataset, an iterator allows traversing the data container, which is typically arranged as a tree or pipeline that divides the data stream into smaller parts that can be processed in parallel [35]. As we focus on self-describing and machine-readable data, we build upon the principles of Linked Data. Querying Linked Data is mostly associated with the SPARQL Protocol and the RDF Query Language, a semantic query language able to retrieve and manipulate the datasets. The approach of a dynamic iterator-based pipeline applied to process SPARQL Queries has been researched [39]. SPARQL endpoints implement a protocol on top of

HTTP — contrary to regular HTTP servers there are so many ways to express the same request that cache hits are likely to be very low — and therefore common HTTP caching cannot be used which has a negative impact on the scalability [30,69].

This approach leverages on HTTP caching and is therefore scalable. As time and space play a central role in air quality data, these are essential linking dimensions for the LDF [4]. Examples of iterators are *hydra:previous* and *hydra:next* which allow the client to iterate over the air-quality time series, retrieving the different LDF samples at a particular timestamp or the average during a specified time interval. These hypermedia controls are defined in the Hydra Core Vocabulary [47]. These iterators were applied to time and space dimensions by Colpaert, who extended[116] the Hydra ontology to describe a tile server that supports *osm:Way*, *osm:Relation* and *osm:Node* [16]. If an *osm:Way* has an overlap with a tile, links to bordering tiles will be added to the *hydra:Collection*[117].

5.3 Use case scenario

5.3.1 Delineating a clear use case

In order to delineate a clear use case, we conducted various semi-structured interviews, interviewing decision makers at (i) the Flanders Environment Agency (VMM), an agency of the Flemish government working towards a better environment in Flanders, (ii) the Agency for Facility Operations that is responsible for the Digital Archive Flanders, (iii) Digital Flanders Agency that is responsible for digitisation, and (iv) the international innovation hub imec City of Things that advances the state-of-the-art of smart city technology.

[116] https://openplanner.team/specs/2018-11-routable-tiles.html
[117] https://treecg.github.io/specification/

A use case was developed in which eighteen delivery vans of Belgium's leading postal operator were equipped with sensors to measure air quality on behalf of the University of Antwerp and imec (see Figure 5.2) [9].

Fig. 5.2. bpost van equipped with an air-quality sensor (by imec City of Things).

Based on the gathered data, it is possible to suggest citizens a healthier route with lower exposure to air pollutants. To realise this use case, multiple sensor data sources need to be queried. Therefore, we will also research a caching strategy that applies to Linked Data. The focus points are the northbound interfaces (NBI) for air-quality analysis, which do not require real-time data streams [44,52,64,67]. The purpose of this use case was to evaluate the caching strategy. Evaluating techniques to calibrate the data based on information such as the sensor noise level or location is not in the scope of this use case.

As SOSA and SSN are respectively a W3C recommendation and OGC implementation standards, and available as Linked Data, they are an excellent candidate to facilitate interoperability for air quality sensor data. Therefore, we have implemented SSN/SOSA for the use case scenario. However, we expect that with the support of the ECs and communities including the International Data Spaces Association[118] and TM Forum[119] that NGSI-LD could become a sustainable and interoperable standard for a wide variety of thematic domains.

5.3.2 Realising the use case via a LDF approach

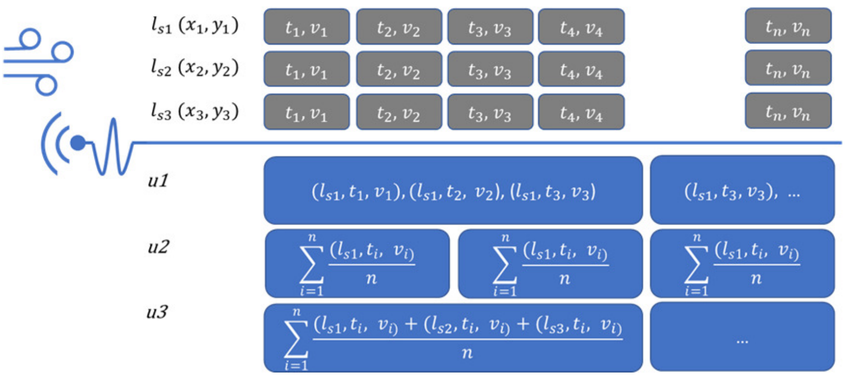


Fig. 5.3. The use cases (u1) 'the absolute sensor values in a time interval', (u2) 'the average sensor values per sensor' and (u3) 'the average sensor values within a bounding box', which consume measurements v_1 at a certain timestamp t_i from sensor nodes on a specific location l_{sj}.

As discussed in Section 5.2.3, the Linked Time Series (LTS) server decomposes the query into subqueries, which allows constructing a static iterator pipeline that has a predefined structure. In this section, we elaborate on how the use-case scenario can be realised using an LDF approach. We developed three tracks (see Figure 5.3) that implement this use case scenario titled 'the absolute sensor values in a time interval' (use case 1), 'the average sensor values per sensor' (use case 2), and 'the average sensor values within a bounding box' (use case 3), which consume measurements v1 at a certain timestamp ti from sensor nodes on a specific location lsj.

[118] https://www.internationaldataspaces.org/
[119] https://www.tmforum.org/press-and-news/fiware-foundation-tm-forum-launch-front-runner-smart-cities-program/

Use case 1: ’the absolute sensor values in a time interval’ gives the software client the option to request absolute sensor values in a time interval. The primary actor is the client of the LTS server. The main success scenario consists of four steps. First, the client determines the time interval in which it wants to receive sensor values. This time interval can, for example, be an hour, day, month or year. Second, the client sends its request. Third, the server responds with a fragment that contains absolute values within a time interval. Fourth, an optional step, as the fragment is provided with an attribute with a link to the previous fragment (a *hydra:previous* attribute), the client can request this by simply following the linkWe identified two extension scenarios: (i) the client requests the most recent update and will receive the most recent fragment and (ii) the server responds with an error message because no sensor values were found in the requested time interval.

– Use case 2: ’the average sensor values per sensor’ and allows to retrieve the average values of the air quality in a time interval. The primary actor is the client of the LTS server. The main success scenario consists of three steps. First, the client determines the time interval from which he wants to receive average sensor values. This time interval can be an hour, day, month or year. Second, the client sends its request. Third, the server responds with the desired fragment. We identified one extension scenario; when the server responds with an error message because no fragment was found for the requested time interval.

– Use case 3: ’the average sensor values within a bounding box’. The software client has the option to request a time series based on the average sensor values within a bounding box (an area defined by two longitudes and two latitudes). The primary actor is the client of the LTS server. The main success scenario consists of four steps. First, the client determines from which bounding box he wants to request a time series. Second, the client sends his request. Third, the server responds with the desired fragment. Fourth, an optional step, the fragment may contain an attribute with a link to the previous fragment (*hydra:previous*) and a link to the neighbouring fragments, which allows to retrieve these fragments by following the linkss.

5.4 Benchmark

5.4.1 Benchmark characteristics and approach

The goal of the benchmark is to test if the method of LDF on public endpoints for air quality sensor data will lower the cost for publishing and raise their availability due to a better caching strategy. We will compare the LTS client/server setup to the ubiquitous FIWARE QuantumLeap API [3]. Therefore, we will monitor the parameters that characterise a Web API. According to Vander Sande [61], Web APIs are characterised by (i) *the query response time, which refers to the rate at which tasks can be completed* such as the maximum number of requests a server can handle in a time interval, (ii) *the cost*, which indicates the amount of resources a single request consumes such as the load on the CPU and memory of both the client and server, (iii) *the cache reuse*, which is the ratio of items that are requested more than once from the cache, and (iv) *the bandwidth*, which is the required size of the HTTP communication.

In Section 5.3.2, we outlined three tracks to implement the use case that can be applied objectively to both architectural approaches. We will evaluate closed- and open-ended time intervals, as this has a significant impact on the caching strategy. To create an unbiased benchmark, we use the same backend for both architectures. We add an extra scenario where we request the most recent observation, which provides a baseline without caching. The third use case (use case 3, the average sensor values within a bounding box) can be reduced to use case 2 (the average sensor values per sensor) for this benchmark.

Hence, this leads to four benchmark scenarios:

- The most recent observation (b1).
- The absolute sensor values in a time interval that has not yet ended (b2).
- The absolute sensor values in a time interval that has ended (b3).
- The average sensor values in a time interval that has ended (b4).

To compare the LTS client/server setup with the FIWARE QuantumLeap API, we performed load testing on both Web APIs by using the emulab[120], which is a testbed that can be used for large networking and cloud experiments. The testbed consists of 160 pc3000[121] PC nodes with following characteristics: Dell PowerEdge 2850s with a single 3GHz processor and 2GB of RAM.

5.4.2 Testbed

As FIWARE is the preferred open-source platform by the EC, we will benchmark the LTS approach with the FIWARE QuantumLeap API. First, we discuss the end-to-end architecture including the FIWARE Quantum Leap API and the LTS, as illustrated in Figure 5.4 [48,54]. Next, we discuss the specific experimental setup, that uses the same backend for both architectures, to create an unbiased benchmark.

On the southbound, the IoT Agents (IoTa) facilitate the data stream from a sensor or a group of sensors to the context broker. These SBI interfaces are regularly using a native protocol. The Orion context broker is a building block of the FIWARE platform that decouples context producers and consumers. The broker facilitates updates, queries, or subscription to changes on context information. The clients that subscribe are notified when specific conditions arise, such as the change of the value of the air quality or the location [31]. The context elements — in this experiment Air-Quality data — are stored in a document-based MongoDB database.

First, we evaluate the FIWARE QuantumLeap API, which stores the data into a CrateDB time-series database.

The data can be queried via a REST API which serves the space-temporal FIWARE-NGSI v2[122] data [3]. The second component in our

---

[120] https://www.emulab.net/

[121] https://gitlab.flux.utah.edu/emulab/emulab-devel/-/wikis/Utah%20Cluster#pc3000s

[122] https://fiware.github.io/specifications/ngsiv2/stable/

experimental setup evaluates the LTS Server which is an implementation of LDF and is illustrated in Figure 5.4.

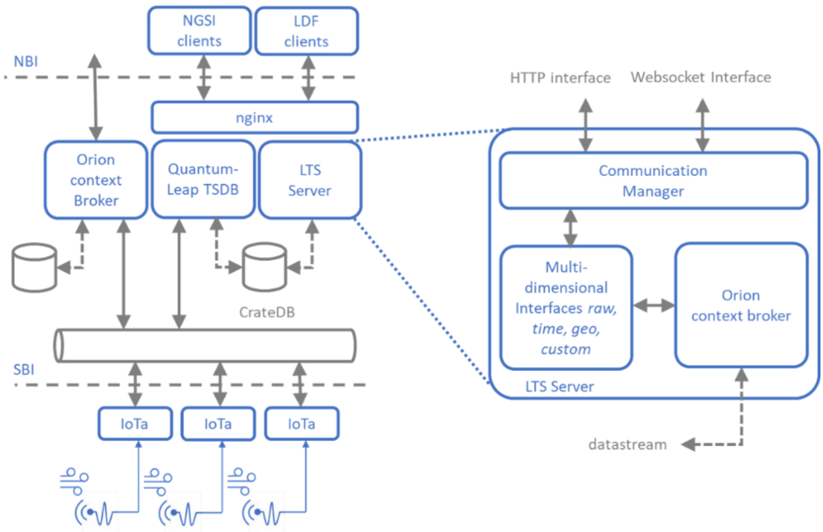


Fig. 5.4. Overview of the end-to-end testbed with the FIWARE Quantum Leap API (left) and the LTS Server (right) [48,54].

We distinguish three main building blocks; (i) the Data Event manager which facilitates the stream updates, (ii) the Multidimensional Interfaces which subscribe to specific events of the Data Event manager and calculate a predefined index, and (iii) the communications manager that facilitates the communication between the Multidimensional Interfaces and the clients [54]. In this experimental setup, both the Quantum Leap API and LTS server use the CrateDB database, to create an unbiased benchmark. The LTS server provides a LDF interface for publishing time series and uses the method of Multidimensional Interfaces (MI) to fragment and index the data [63,72]. MI ensures the discoverability of the fragments by annotating them with hypermedia controls, which are formalised in an RDF vocabulary. The vocabulary[123] introduces the concepts Range Fragments and Range Gates. Taelman [63] defines a Range fragment as 'an LDF that has an interval as a selector' — which is part of a predefined fragmentation strategy — and a Range Gate as 'a Linked Data interface through which Range Fragments can be selected by interval' and thus exposes a collection of Range Fragments [63]. These Ranges were applied to time and space dimensions of the Air-Quality data and both fragment and index the absolute and average sensor values in a time interval.

Also, we have added Nginx[124] to serve as a Web cache (or HTTP cache), which stores copies of requests, for both the FIWARE QuantumLeap API and LTS server.

We used Kubernetes — an open-source container-orchestration system — to package our testbed. One CPU, in Kubernetes[125], is equivalent to 1 AWS vCPU, 1 GCP Core, 1 Azure vCore, 1 IBM vCPU or 1 Hyperthread on a bare-metal Intel processor with Hyperthreading. The results are expressed in mebibyte (MiB) and millicpu. A mebibyte is equivalent to 1048 576 bytes. Kubernetes defines a metric called Millicores that is used to measure CPU usage. It is a CPU core split into 1000 units. To ensure that the benchmark is reproducible, we have published the repository[126] with source code and configuration scripts. The repository provides the necessary scripts and background information to deploy and benchmark the FIWARE QL API with the LTS Server API for timeseries on a Kubernetes cluster. We outline the main steps which are executed during the benchmark. First, we setup the Kubenetes cluster, a set of machines that run the containerized applications. Second, the scripts that deploy CrateDB, MongoDB, the Orion context Broker, Quantumleap TSDB, and the Nginx Web cache

[123] http://semweb.datasciencelab.be/ns/multidimensional-interface/
[124] https://www.nginx.com/
[125] https://kubernetes.io/
[126] https://github.com/brechtvdv/benchmark-quantumleap

are executed. Third, the metrics server that harvests the CPU and memory consumption of the server and clients is deployed. Fourth, we setup the data streams by creating a subscription between the Orion context Broker and the QuantumLeap TSDB, which ensures automatic updates. Fifth, a table is created in CrateDB, that stores the time series data. Sixth, data gets ingested every second to Orion. Seventh, an HTTP client is provided and configured to use the QL and LDF API. Finally, the 4 scenarios are executed on both the QL and LDF API and monitored using the metrics server.

5.4.3 Results

5.4.3.1 The most recent observation (b1)

This scenario benchmarks the request of the most recent observations (n = 100). Figure 5.6 shows that the CPU use of the FIWARE QL API and underlying database, with a load of ten clients, is a factor of four higher than the LTS Server API. The memory usage of the FIWARE QL API remains stable (Figure 5.5). At a load of four hundred clients, we notice that the query response time of the FIWARE Quantum Leap API increases to five seconds, at a higher load we get a timeout (Fig. 5.7). Figure 5.7 shows how an overview of the benchmark with the query response time (latency) of FIWARE Quantum Leap API. This is in contrast to the LTS Server API which — despite a rising query response time — still stands with a load of 1300 clients (Figure 5.8). The bandwidth per request of the FIWARE QL API is 4.5Kb, compared to 17.5Kb for the LTS Server API. The load on the LTS Server API clients is 10 millicpu, compared to 2 millicpu in the case of the clients of the FIWARE QL API.

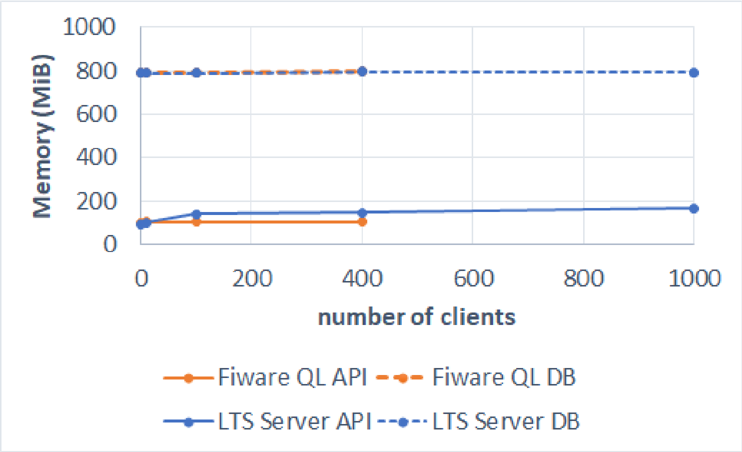


Fig. 5.5. Overview of the benchmark with memory cost needed to publish the most recent observations.

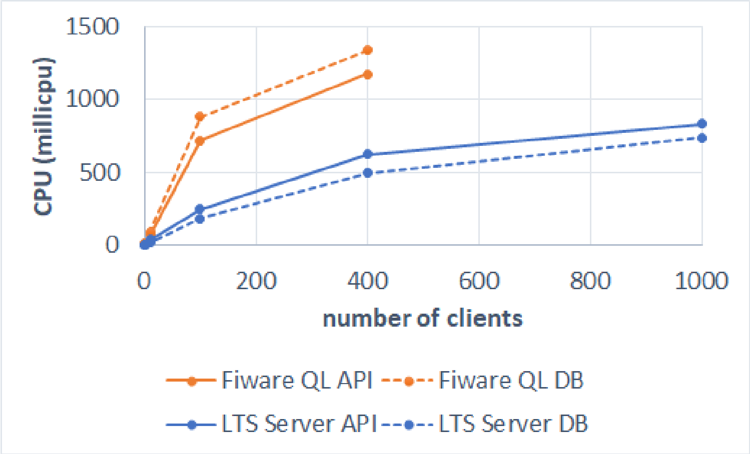


Fig. 5.6. Overview of the benchmark with CPU cost needed to publish the most recent observations.

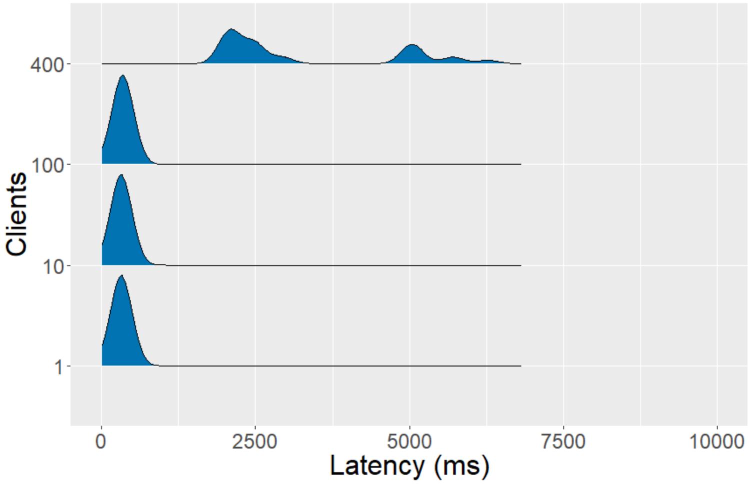


Fig. 5.7. Overview of the benchmark with the query response time (latency) of FIWARE Quantum Leap API, when publishing the most recent observations.

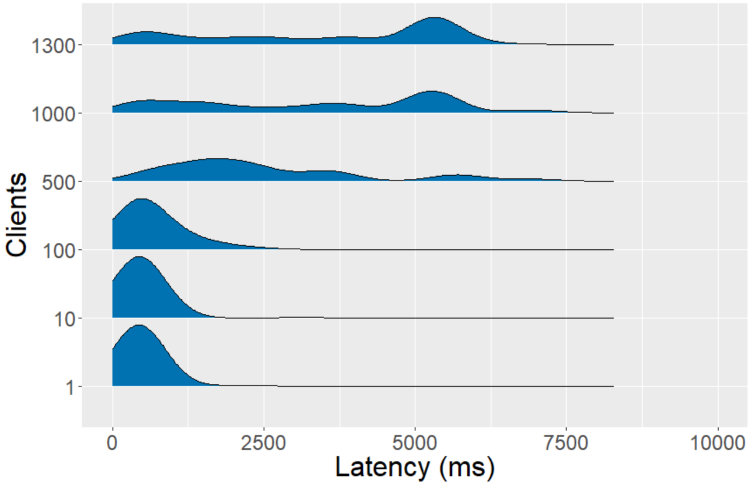


Fig. 5.8. Overview of the benchmark with the query response time (latency) of LTS Server, when publishing the most recent observations.

5.4.3.2 The absolute sensor values in a time interval that has not yet ended (b2)

This scenario benchmarks the absolute sensor values in a time interval (month) that has not yet ended. The CPU cost of the FIWARE QL API increases by a factor of twenty when scaling up to ten clients and doubled when scaling up from ten to one hundred clients (Figure 5.10). The memory usage of the FIWARE QL API remains stable (Figure 5.9). At a load of ten clients, the query response time of the FIWARE Quantum Leap AP remains below two seconds. From a hundred clients onwards, the query response time rises above ten seconds. The query response time at a load of four hundred clients raises up to twenty seconds, after which we get a timeout from the FIWARE Quantum Leap API (Figure 5.11). Figure 5.11 shows an overview of the benchmark with the query response time (latency) of FIWARE Quantum Leap API and figure 5.12 of the LTS Server. The CPU cost of the FIWARE QL API database increases by a factor of fifteen when scaling up to ten clients and doubled when scaling up from ten to one hundred clients (Figure 5.10). The CPU cost of the FIWARE QL API client is almost neglectable. As there is no client-side cache reuse, the bandwidth of the FIWARE QL API is 189.6KB, for an interval of one month.

The CPU cost of the LTS Server API, increases linearly with the number of clients (Fig. 5.10). From the second request, there is cache reuse of 30/31, as the last request cannot be cached. The CPU load on a thousand clients is lower than a load on ten clients in the case of the FIWARE Quantum Leap API. The load on the clients is 387 millicpu, compared to 2 millicpu in the case of the clients of the FIWARE QL API. The bandwidth at the LTS Server level is 542.2Kb per client, with a cold client-side cache. From the second query onwards, the bandwidth drops to 17.5Kb due to client caching.

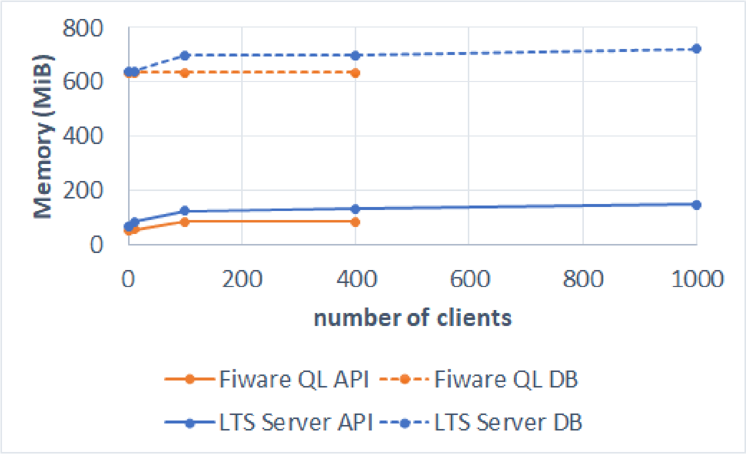


Fig. 5.9. Overview of the benchmark with the memory cost to publish the absolute sensor values
in a time interval that has not yet ended (b2).

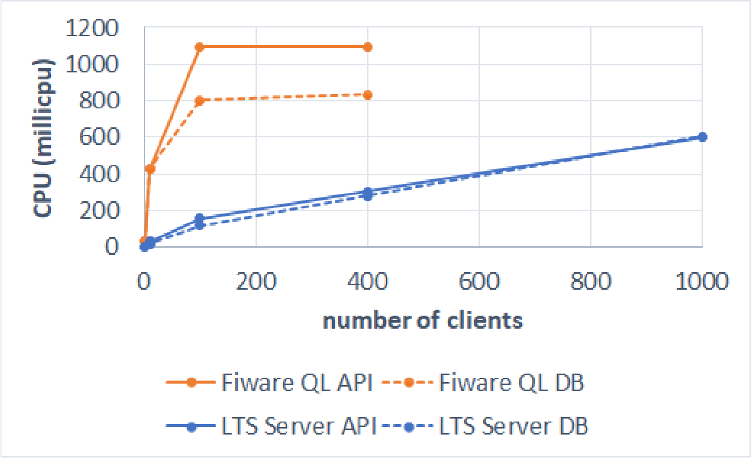


Fig. 5.10. Overview of the benchmark with the CPU cost to publish the absolute sensor values
in a time interval that has not yet ended (b2).

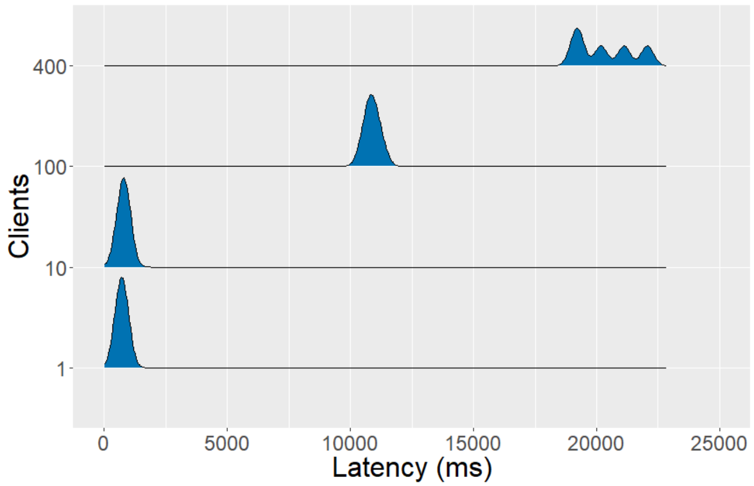


Fig. 5.11. Overview of the benchmark with the query response time (latency) of FIWARE Quantum Leap API, publishing the absolute sensor values in a time interval that has not yet ended (b2).

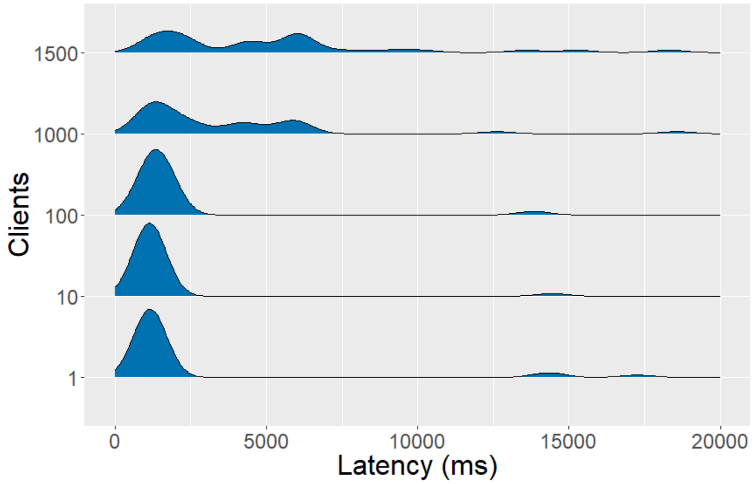


Fig. 5.12. Overview of the benchmark with the query response time (latency) of LTS Server, publishing the absolute sensor values in a time interval that has not yet ended (b2).

5.4.3.3 The absolute sensor values in a time interval that has ended (b3)

This scenario benchmarks absolute sensor values in a time interval (month) that has ended. The results of the FIWARE QL API are comparable with those of the absolute sensor values in a time interval that has not ended yet. From a hundred clients, the query response time of the FIWARE Quantum Leap API rises above ten seconds. The query response time at a load of four hundred clients raises to twenty seconds, after which we get a timeout (Figure 5.15). Figure 5.15 shows an overview of the benchmark with the query response time (latency) of FIWARE Quantum Leap API and Figure 5.16 of the LTS Server. As the time interval has ended, all 31 responses of the LTS Server API are fully cacheable. This allows to scale-up to 1000 clients without an extra CPU cost or increase of memory (Figure 5.14). The load on the clients is 401 millicpu, compared to 5 millicpu in the case of the clients of the FIWARE QL API. The bandwidth at the LTS Server level is 524.9Kb per client. From the second request, this drops to 17.5Kb.

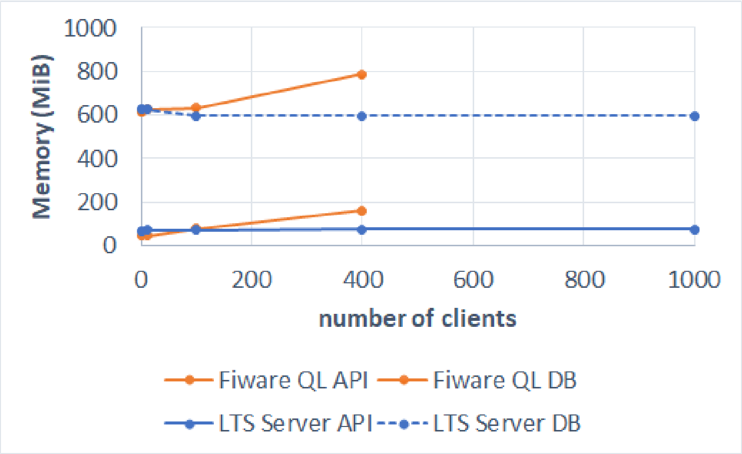


Fig. 5.13. Overview of the benchmark with the memory cost to publish the absolute sensor values in a time interval that has ended (b3).

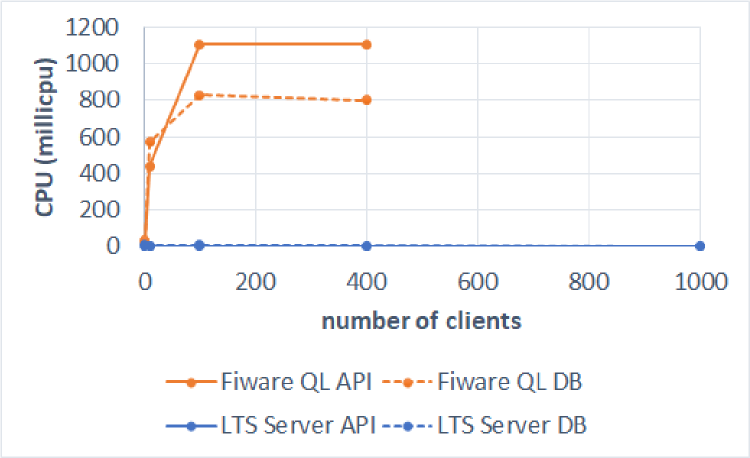


Fig. 5.14. Overview of the benchmark with the CPU cost to publish the absolute sensor values in a time interval that has ended (b3).

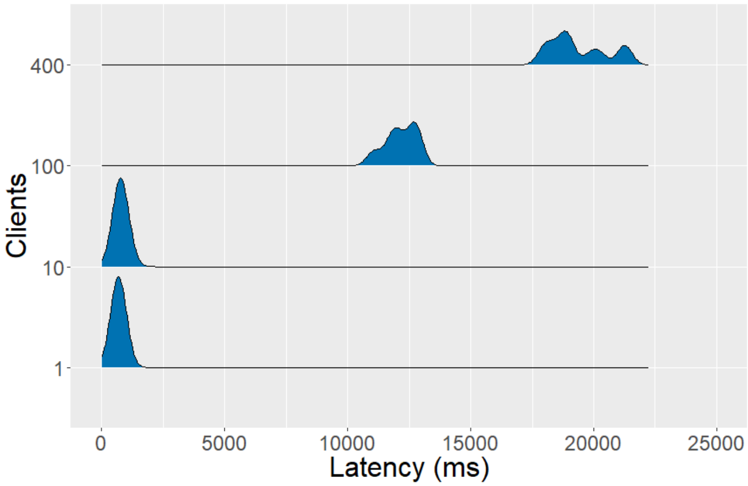


Fig. 5.15. Overview of the benchmark with the query response time (latency) of FIWARE Quantum Leap API, publishing the absolute sensor values in a time interval that has ended (b3).

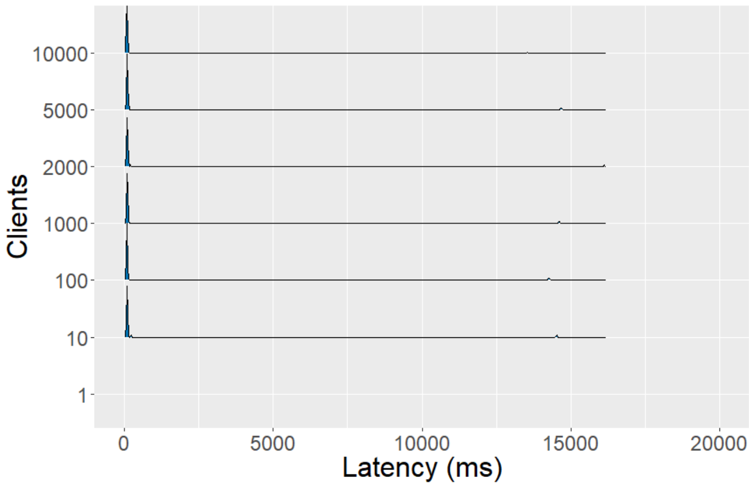


Fig. 5.16. Overview of the benchmark with the query response time (latency) of LTS Server, publishing the absolute sensor values in a time interval that has ended (b3).

5.4.3.4 The average sensor values in a time interval that has not yet ended (b4)

This scenario benchmarks the average sensor values in a time interval (hour) that has ended. The results of the FIWARE QL API are comparable with those of the absolute sensor values in a time interval that has not ended yet (Figure 5.18 and 5.19). From four hundred clients, the query response time of the FIWARE Quantum Leap API rises above ten seconds. The query response time at a load of four hundred clients raises to twenty seconds, after which we get a timeout (Figure 5.20). Figure 5.20 shows an overview of the benchmark with the query response time (latency) of FIWARE Quantum Leap API and Figure 5.21 of the LTS Server. As the time interval has ended, the responses of the LTS Server API are fully cacheable. This allows to scale-up to ten thousand clients with only a slight increase in the CPU cost and without an increase of memory (Figure 5.18).

The load on the clients of the LTS Server API is 367 millicpu, compared to 2 millicpu in the case of the clients of the FIWARE QL API. As the clients of the LTS Server API need to calculate the average, their load is significantly higher. The bandwidth at the server level is 524.9Kb per LTS Server API client — as the server responds with all the data from the last month — compared to 38,5KB for the FIWARE QL API that only responds the pre-processed average sensor value to the client. From the second request, the bandwidth of the LTS Server drops to 17.5Kb.

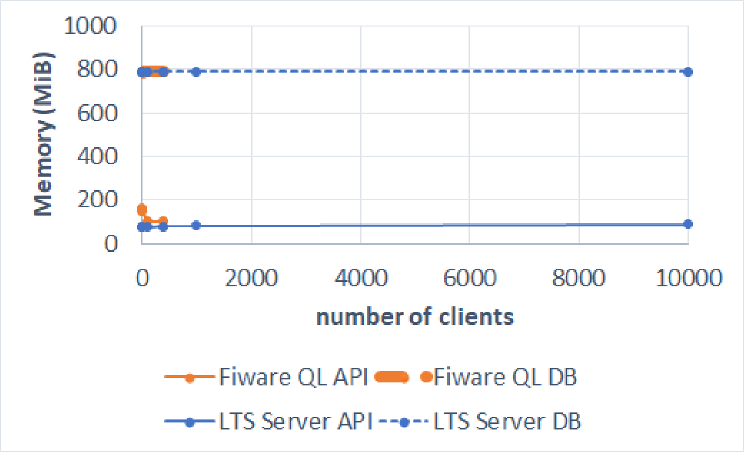


Fig. 5.17. Overview of the benchmark with the memory cost to publish the average sensor values — stretched to 10 000 clients — in a time interval (hour) that has ended (b4).

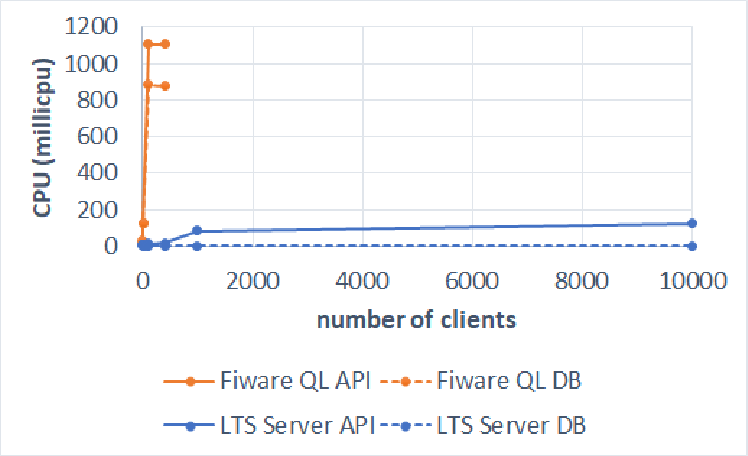


Fig. 5.18. Overview of the benchmark with CPU cost to publish the average sensor values — stretched to 10 000 clients — in a time interval (hour) that has ended (b4).

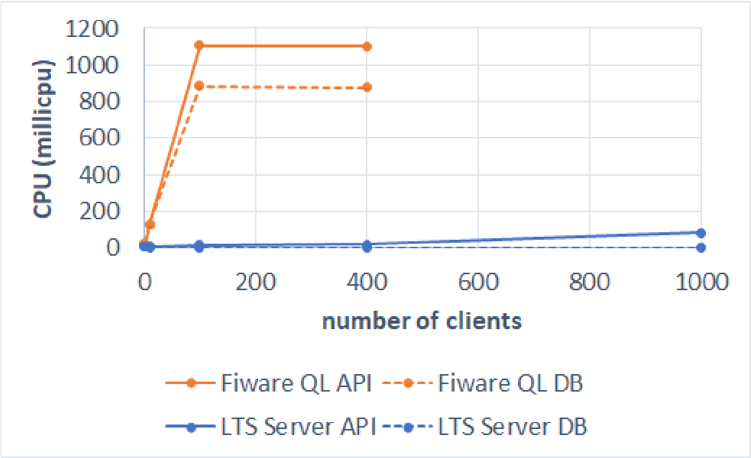


Fig. 5.19. Overview of the benchmark with the memory cost to publish the average sensor values in a time interval (hour) that has ended (b4).

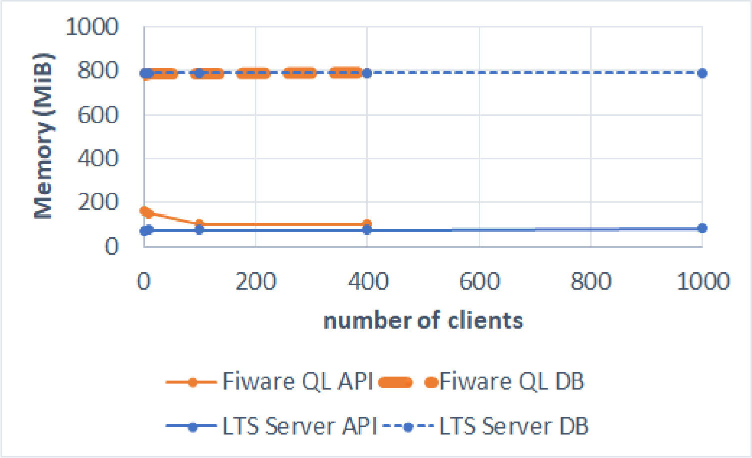


Fig. 5.20. Overview of the benchmark with the CPU cost to publish the average sensor values in a time interval (hour) that has ended (b4).

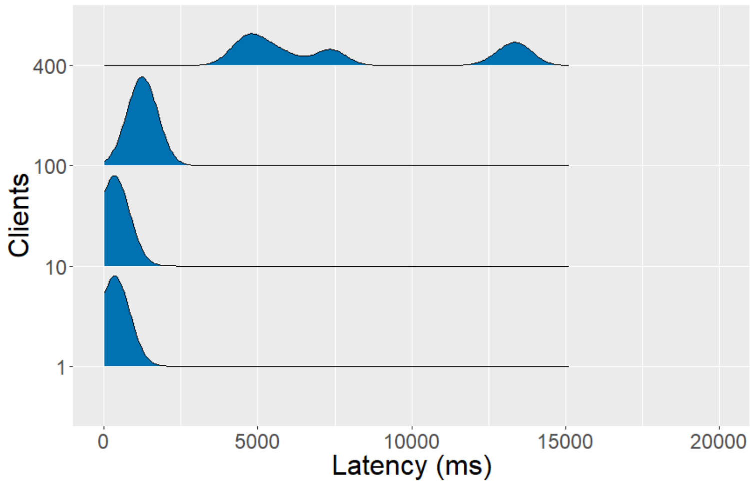


Fig. 5.21. Overview of the benchmark with the query response time (latency) of FIWARE Quantum Leap API, publishing the average sensor values in a time interval (hour) that has ended (b4)

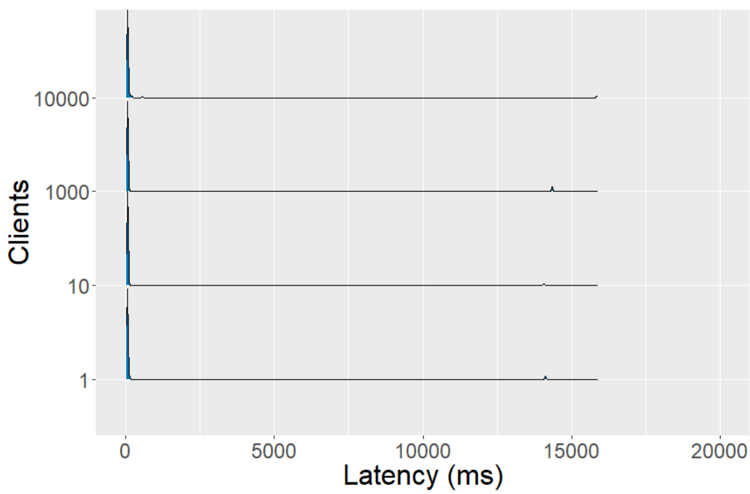


Fig. 5.22. Overview of the benchmark with the query response time (latency) of LTS Server, publishing the average sensor values in a time interval (hour) that has ended (b4)

5.5 Discussion and conclusion

In this chapter we focussed on **the first viewpoint of our research question** "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?". Although this chapter addresses multiple interoperability levels, the main contribution is technical interoperability. Technical interoperability covers interconnection of applications and infrastructures, including interface specifications that interconnect systems and services [26].

Government administrations as data providers make significant investments in collecting data, making it interoperable and publishing it for maximum reuse. Hence our research question addressed how data providers can develop a sustainable method for publishing and archiving open sensor data, in specific sensor data time series on air quality. In order to do this, we set-up a benchmarking experiment. We benchmarked the ubiquitous QuantumLeap API and a Linked Times Series API.

Our results show that the CPU use of the FIWARE QL API and underlying database increases linearly with the number of clients. Even with a load of only ten clients, the CPU use is a factor of four higher than the LTS Server API (b1, b2, b3, b4). From several hundreds of clients, the query response time of the FIWARE Quantum Leap API rises above ten seconds, after which we get a timeout (b1, b2, b3, b4). Scenario b4 illustrates the balancing efforts between the publisher and the re-user. The load on the clients of the FIWARE Quantum Leap API is 185 times higher, as the clients have to calculate the average sensor values. When a time interval has ended, the response of the LTS Server API is fully cacheable. This allows to scale-up to ten thousand clients with only a slight increase in the CPU cost and without an increase of memory (b4).

The benchmark showed that the LTS approach lowers the cost for publishing air-quality data and raises availability because of a better caching strategy. The results of the benchmark ascertain that — with an increasing number of clients — the LTS API has a lower CPU load than the FIWARE QuantumLeap API. Also, as the load of the LTS is constant for historic data, the cost becomes predictable, which is crucial for public bodies that need to determine their expenses beforehand. Building on the strength of the WWW — by using HTTP caching on the level of clients and servers — created not only a cost-benefit but also contributed to the stability of the air-quality endpoints. The FIWARE QuantumLeap API starts 'sputtering' at a load of 400 clients, in contrast with the LDF interface that still gets good results at 10,000 requests.

The bandwidth of the LDF endpoint is slightly higher, but the amount of resources a single request consumes — such as the load on the CPU and memory — is significantly lower, due to the cache reuse. The cache reuse is the ratio of items that are requested more than once from the cache instead of consuming server resources. The literature is consistent with our findings by limiting the server interface as it becomes more scalable due to Web caching [33,70]. The scenario that calculates the average sensor values in a time interval that has ended demonstrates that a part of the workload is shifted to the client but is still acceptable for the client. A non-measurable benefit is that the client can evaluate any given query on the client-side, without having to rely on server-side functionality other than downloading the right fragments. Next, the Open World Assumption (OWA) becomes applicable as more data can always be downloaded in order to get a more precise answer [58,60].

We discussed that the challenges related to linked open sensor data time series are not limited to the volume and velocity of the time series but also to their variety and researched the interoperability challenges, which are crucial when combining air quality data from different sources as well as linking them to other datasets such as traffic or weather data. We addressed the different interoperability levels, namely the legal, organisational, technical, and semantic level. Linked

Data facilitates interoperability on both a technical and semantic level. Context information such as temperature, humidity, and spatial information enriches air-quality data by reusing existing machine-readable RDF vocabularies. The principles of Linked Data make the data self-describing and machine-readable, which allows autonomous agents to reason on the sensor data. Linked Data builds upon the architecture of the WWW and uses typed links between data entities from disparate sources, described using the RDF.

Based on the above, we have three recommendations about archivability, indexing, and interoperability for future research. First, it is crucial to assure that time series are still accessible and usable for future generations, as they are valuable for research (e.g., on environmental changes). Therefore, a strategy needs to be defined that outlines which subsets of the data will be preserved in order to reduce the storage cost of the data in a digital archive. Second, research should consider a more dynamic method to fragment and index different types of time series, taking into account the available budget of the publisher. Finally, on the level of interoperability, future research should explore how to bridge between the NGSI — that redefines a knowledge representation in its own — and existing semantic assets.

Although our results are promising, there are limitations to our research. First, as the scenarios are limited, further validation of this method in a large variety of use cases and different types of sensor data time series is necessary to generalize our conclusions. Second, we should evaluate this method for real-time data streams. This research demonstrates the significant benefits of adopting the principles of Linked Data regarding air-quality time series as these principles not only provide interoperability towards external stakeholders but also foster a more sustainable and cost-effective architecture. As such, the LTS interface can serve as a valuable extension of the FIWARE stack. We expect that the insights from this chapter can speed-up the process of opening up sensor time-series data by public and private organisations. As such the contributions of our research, which builds upon the principles of Linked Open Data, are valuable for governments, organisations, and researchers that aim to publish air quality Sensor Data on a Web-Scale in a cost-efficient way.

5.6 References

[1] A. Abbas, G. Privat, Bridging Property Graphs and RDF for IoT Information Management., in: SSWS@ ISWC, 2018: pp. 77–92.

[2] T. Almarabeh, Y.K. Majdalawi, H. Mohammad, Cloud computing of e-government, (2016).

[3] Y.R. Baca Gómez, H. Estrada Esquivel, A. Martínez Rebollar, D. Villanueva Vásquez, A Novel Air Quality Monitoring Unit Using Cloudino and FIWARE Technologies, Math. Comput. Appl. 24 (2019) 15.

[4] M. Balazinska, A. Deshpande, M.J. Franklin, P.B. Gibbons, J. Gray, M. Hansen, M. Liebhold, S. Nath, A. Szalay, V. Tao, Data management in the worldwide sensor web, IEEE Pervasive Comput. 6 (2007) 30–40.

[5] P. Ballon, Smart Cities: Hoe technologie onze steden leefbaar houdt en slimmer maakt, Lannoo Meulenhoff-Belgium, 2016.

[6] P. Barnaghi, M. Presser, K. Moessner, Publishing linked sensor data, in: CEUR Work. Proc. Proc. 3rd Int. Work. Semant. Sens. Networks (SSN), Organised Conjunction with Int. Semant. Web Conf., 2010.

[7] T. Berners-Lee, R. Fielding, L. Masinter, Uniform Resource Identifier (URI): Generic Syntax. Request for Comments: 3986, (2005).

[8] C. Bizer, T. Heath, T. Berners-Lee, Linked data: The story so far, in: Semant. Serv. Interoperability Web Appl. Emerg. Concepts, IGI Global, 2011: pp. 205–227.

[9] bpost, Activity report 2017, 2017.

[10] L. Breslau, P. Cao, L. Fan, G. Phillips, S. Shenker, Web caching and Zipf-like distributions: Evidence and implications, in: IEEE INFOCOM'99. Conf. Comput. Commun. Proceedings. Eighteenth Annu. Jt. Conf. IEEE Comput. Commun. Soc. Futur. Is Now (Cat. No. 99CH36320), IEEE, 1999: pp. 126–134.

[11] R. Buyle, B. Van de Vyvere, Julian Andres Rojas Melendeza, D. Van Lancker, E. Vlassenroot, M. Van Compernolle, S. Lefever, P. Colpaert, P. Mechant, & E. Mannens, E. (2021). A sustainable open data platform for air quality data. *Data* (under review).

[12] M.J. Carey, M.J. Franklin, M. Livny, E.J. Shekita, Data caching tradeoffs in client-server DBMS architectures, in: Proc. 1991 ACM SIGMOD Int. Conf. Manag. Data, 1991: pp. 357–366.

[13] N. Castell, M. Viana, M.C. Minguillón, C. Guerreiro, X. Querol, Real-world application of new sensor technologies for air quality monitoring, ETC/ACM Tech. Pap. 16 (2013).

[14] H. Chourabi, T. Nam, S. Walker, J.R. Gil-Garcia, S. Mellouli, K. Nahon, T.A. Pardo, H.J. Scholl, Understanding smart cities: An integrative framework, in: 2012 45th Hawaii Int. Conf. Syst. Sci., IEEE, 2012: pp. 2289–2297.

[15] C. Coglianese, The transparency president? The Obama administration and open government, Governance. 22 (2009) 529–544.

[16] P. Colpaert, B. Abelshausen, J.A.R. Meléndez, H. Delva, R. Verborgh, Republishing OpenStreetMap's roads as Linked Routable Tiles, in: Eur. Semant. Web Conf., Springer, 2019: pp. 13–17.

[17] M. Van Compernolle, R. Buyle, E. Mannens, Z. Vanlishout, E. Vlassenroot, P. Mechant, "Technology readiness and acceptance model" as a predictor for the use intention of data standards in smart cities, Media Commun. 6 (2018) 127–139.

[18] M. Compton, P. Barnaghi, L. Bermudez, R. GarcíA-Castro, O. Corcho, S. Cox, J. Graybeal, M. Hauswirth, C. Henson, A. Herzog, The SSN ontology of the W3C semantic sensor network incubator group, J. Web Semant. 17 (2012) 25–32.

[19] P. Conradie, S. Choenni, On the barriers for local government releasing open data, Gov. Inf. Q. 31 (2014) S10–S17.

[20] ETSI, NGSI-LD: Context Information Management (CIM) and Application Programming Interface (API), 2019.

[21] European Central Bank, Payments statistics: 2018, 2019.

[22] European Commission, INSPIRE Maintenance and Implementation Group (MIG). Guidelines for the use of Observations & Measurements and Sensor Web Enablement-related standards in INSPIRE, 2008.

[23] European Commission, Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe, 2008.

[24] European Commission, European Interoperability Framework (EIF) for European public services: Annex 2, 2010.

[25] European Commission, Measuring air pollution with low-cost sensors. Thoughts on the quality of data measured by sensors, 2017.

[26] European Commission, New European Interoperability Framework (EIF), 2017.

[27] European Commission, European Interoperability Framework – Implementation Strategy, 2017.

[28] European Commission, Digital Single Market: EU negotiators agree on new rules for sharing of public sector data, 2019.

[29] European Environment Agency, Air Quality in Europe — 2018 report, 2018.

[30] L. Feigenbaum, G.T. Williams, K.G. Clark, E. Torres, sparql 1.1 protocol. Recommendation, World Wide Web Consortium, March. (2013).

[31] P. Fernández, J.M. Santana, S. Ortega, A. Trujillo, J.P. Suárez, C. Domínguez, J. Santana, A. Sánchez, SmartPort: A platform for sensor data monitoring in a seaport based on FIWARE, Sensors. 16 (2016) 417.

[32] R.T. Fielding, J. Gettys, J. Mogul, H. Frystyk, L. Masinter, P. Leach, T. Berners-Lee, Hypertext Transfer Protocol -- HTTP/1.1. Request for Comments, 1999.

[33] R.T. Fielding, R.N. Taylor, Architectural styles and the design of network-based software architectures, University of California, Irvine Irvine, 2000.

[34] R.T. Fielding, R.N. Taylor, Principled design of the modern Web architecture, ACM Trans. Internet Technol. 2 (2002) 115–150.

[35] G. Graefe, Query evaluation techniques for large databases, ACM Comput. Surv. 25 (1993) 73–169.

[36] J. Gubbi, R. Buyya, S. Marusic, M. Palaniswami, Internet of Things (IoT): A vision, architectural elements, and future directions, Futur. Gener. Comput. Syst. 29 (2013) 1645–1660.

[37] A. Haller, K. Janowicz, S. Cox, D. Le Phuoc, K. Taylor, M. Lefrançois, Semantic sensor network ontology, (2017).

[38] A. Haller, K. Janowicz, S.J.D. Cox, M. Lefrançois, K. Taylor, D. Le Phuoc, J. Lieberman, R. García-Castro, R. Atkinson, C. Stadler, The modular SSN ontology: A joint W3C and OGC standard specifying the semantics of sensors, observations, sampling, and actuation, Semant. Web. 10 (2019) 9–32.

[39] O. Hartig, C. Bizer, J.-C. Freytag, Executing SPARQL queries over the web of linked data, in: Int. Semant. Web Conf., Springer, 2009: pp. 293–309.

[40] D. Hasenfratz, T. Arn, I. de Concini, O. Saukh, L. Thiele, Health-optimal routing in urban areas, in: Proc. 14th Int. Conf. Inf. Process. Sens. Networks, 2015: pp. 398–399.

[41] J. Hendler, Data integration for heterogenous datasets, Big Data. 2 (2014) 205–215.

[42] K. Janowicz, A. Haller, S.J.D. Cox, D. Le Phuoc, M. Lefrançois, SOSA: A lightweight ontology for sensors, observations, samples, and actuators, J. Web Semant. 56 (2019) 1–10.

[43] P.A. Johnson, R. Sieber, T. Scassa, M. Stephens, P. Robinson, The cost (s) of geospatial open data, Trans. GIS. 21 (2017) 434–445.

[44] R. Keskisärkkä, E. Blomqvist, Event Object Boundaries in RDF Streams-A Position Paper, in: ISWC 2013 Work. Proc. OrdRing, 2013.

[45] G. Klyne, J.M. Carroll, Resource Description Framework (RDF): Concepts and Abstract Syntax, 2004.

[46] H. Kubicek, R. Cimander, H.J. Scholl, Layers of interoperability, in: Organ. Interoperability E-Government, Springer, 2011: pp. 85–96.

[47] M. Lanthaler, Hydra Core Vocabulary Specification, 2019.

[48] S. Lefever, P. Michiels, R. Van den Berg, An Open Smart City Vision and Architecture, 2018.

[49] F. Li, M. Vögler, M. Claeßens, S. Dustdar, Efficient and scalable IoT service delivery on cloud, in: 2013 IEEE Sixth Int. Conf. Cloud Comput., IEEE, 2013: pp. 740–747.

[50] J. Liu, Y. Li, M. Chen, W. Dong, D. Jin, Software-defined internet of things for smart urban sensing, IEEE Commun. Mag. 53 (2015) 55–63.

[51] L. Makra, P. Brimblecombe, Selections from the history of environmental pollution, with special attention to air pollution. Part 1., Int. J. Environ. Pollut. 22 (2004) 641–656.

[52] A. Mauri, J.-P. Calbimonte, D. Dell'Aglio, M. Balduini, M. Brambilla, E. Della Valle, K. Aberer, Triplewave: Spreading RDF streams on the web, in: Int. Semant. Web Conf., 2016: pp. 140–149.

[53] Y. Mehta, M.M.M. Pai, S. Mallissery, S. Singh, Cloud enabled air quality detection, analysis and prediction-a smart city application for smart health, in: 2016 3rd MEC Int. Conf. Big Data Smart City, IEEE, 2016: pp. 1–7.

[54] J.A.R. Melendez, G. Sedrakyan, P. Colpaert, M. Vander Sande, R. Verborgh, Supporting sustainable publishing and consuming of live Linked Time Series Streams, in: Eur. Semant. Web Conf., Springer, 2018: pp. 148–152.

[55] I. Mieck, Aerem corrumpere non licet, Luftverunreinigung Und Immissionsschutz Preußen Bis Zur Gewerbeordnung. (1869) 36–78.

[56] Nam, T.A. Pardo, Smart city as urban innovation: Focusing on management, policy, and context, in: Proc. 5th Int. Conf. Theory Pract. Electron. Gov., 2011: pp. 185–194.

[57] S. Nittel, Real-time sensor data streams, SIGSPATIAL Spec. 7 (2015) 22–28.

[58] P.F. Patel-Schneider, I. Horrocks, A comparison of two modelling paradigms in the semantic web, J. Web Semant. 5 (2007) 240–250.

[59] M. Penza, D. Suriano, M.G. Villani, L. Spinelle, M. Gerboles, Towards air quality indices in smart cities by calibrated low-cost sensors applied to networks, in: SENSORS, 2014 IEEE, IEEE, 2014: pp. 2012–2017.

[60] R. Reiter, On closed world databases. In "Logic and Databases", H. Gallaire and J. Minker eds, (1978).

[61] M. Vander Sande, Metadata and Control Features for Low-Cost Linked Data Publishing Infrastructures, Ghent University, 2018.

[62] X. Su, J. Riekki, J.K. Nurminen, J. Nieminen, M. Koskimies, Adding semantics to internet of things, Concurr. Comput. Pract. Exp. 27 (2015) 1844–1860.

[63] R. Taelman, P. Colpaert, R. Verborgh, E. Mannens, Multidimensional interfaces for selecting data within ordinal ranges, in: Proc. 7th Int. Work. Consum. Linked Data Co-Located with 15th Int. Semant. Web Conf. (ISWC 2015), 2016.

[64] R. Taelman, R. Tommasini, J. Van Herwegena, M. Vander Sandea, E. Della Valle, R. Verborgh, On the Semantics of Tpf-qs towards Publishing and Querying rdf Streams at Web-scale, Procedia Comput. Sci. 137 (2018) 43–54.

[65] M.D. Taylor, Low-cost air quality monitors: Modeling and characterization of sensor drift in optical particle counters, in: 2016 IEEE SENSORS, IEEE, 2016: pp. 1–3.

[66] I. Toma, E. Simperl, G. Hench, A joint roadmap for semantic technologies and the internet of things, in: Proc. Third STI Roadmapping Work. Crete, Greece, 2009: pp. 140–153.

[67] R. Tommasini, Y.A. Sedira, D. Dell'Aglio, M. Balduini, M.I. Ali, D. Le Phuoc, E. Della Valle, J.-P. Calbimonte, VoCaLS: Vocabulary and catalog of linked streams, in: Int. Semant. Web Conf., 2018: pp. 256–272.

[68] United Nations, 2018 revision of world urbanization prospects, United Nations New York, 2018.

[69] R. Verborgh, M. Vander Sande, P. Colpaert, S. Coppens, E. Mannens, R. Van de Walle, Web-Scale Querying through Linked Data Fragments., in: LDOW, Citeseer, 2014.

[70] R. Verborgh, M. Vander Sande, O. Hartig, J. Van Herwegen, L. De Vocht, B. De Meester, G. Haesendonck, P. Colpaert, Triple Pattern Fragments: a low-cost knowledge graph interface for the Web, J. Web Semant. 37 (2016) 184–206.

[71] vtvn, 'Zonnekaart' opnieuw online na teveel bezoekers op dag van lancering, Het Nieuwsbl. (2017).

[72] B. Van de Vyvere, P. Colpaert, E. Mannens, R. Verborgh, Open traffic lights: a strategy for publishing and preserving traffic lights data, in: Companion Proc. 2019 World Wide Web Conf., 2019: pp. 966–971.

[73] R.H. Weber, Legal interoperability as a tool for combatting fragmentation, (2014).

[74] World Health Organisation, Ambient air pollution: a global assessment of exposure and burden of disease, 2016.

[75] N. Zimmerman, A.A. Presto, S.P.N. Kumar, J. Gu, A. Hauryliuk, E.S. Robinson, A.L. Robinson, R. Subramanian, A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring., Atmos. Meas. Tech. 11 (2018).

[76] A. Zuiderwijk, M. Janssen, The negative effects of open government data-investigating the dark side of open data, in: Proc. 15th Annu. Int. Conf. Digit. Gov. Res., 2014: pp. 147–152.

CHAPTER 6. ORGANISATIONAL INTEROPERABILITY

Standards are always out of date.
That's what makes them standards.

— Alan Bennett.

This chapter scrutinises organisational interoperability by exploring how governmental processes can be streamlined by putting citizens truly in control of their data. We apply the decentralised Solid ecosystem to two high-impact public sector use cases. This chapter argues that Solid allows reshaping the relationship between citizens, their personal data, and the applications they use in the public and private sector. We detail how these processes can be streamlined by building upon existing Web standards and methods such as Linked Data and decentralisation. This chapter is based on the paper, 'Streamlining governmental processes by putting citizens in control of their personal data' [3].

6.1 Introduction

With the introduction of the General Data Protection Regulation (GDPR), the European Commission has provided a legal framework that aims to empower individuals in taking control of their personal information [11]. Such control is not necessarily a disadvantage to parties processing personal information: when used properly, GDPR can actually facilitate data flows that used to be much more complicated. GDPR, however, is mostly known for its complex legal effects on European companies dealing with large-scale personal data and may cost them significant resources in order to achieve and maintain legal compatibility. While international and multinational companies also have to respect GDPR rights for European data subjects, even when they do not have a physical European presence, several large players are —to put it lightly— slow with a correct adoption of GDPR. This has created a perverse reverse effect, where European companies that try to respect GDPR become less preferred as business partners, losing revenue to non-European companies that are more 'relaxed' with GDPR adoption [15].

Not all organisations that are subject to GDPR have questionable or malicious intent: some of them experience genuine difficulties in trying to adhere to the legal obligations. This is definitely the case for local, regional, and national governments, which need personal data to provide the services their citizens require. Governmental structures consist of multiple layers, and every layer consists of its agencies with their own data needs and processes that have grown historically. As a result, citizen data is spread across many places in many copies, leading to complex legal questions as well as numerous inconsistencies and repeated requests for data that is already present in other government administrations. These governments are a demanding party for a legally compliant technical solution to simplify all of their data needs.

The majority of data processes at the government level nowadays essentially aim to tackle the problem of how to move data as frictionless as possible from point A to B. Not only does this create a lot of technical challenges between the many different points, it also becomes a complex legal matter when a governmental 'data train' needs to pass by stations A, B, C, and D, where B and C are not legally allowed to see all of the data that A and D can. As such, complex processes exist to verify precisely what the access rights of B and C are, and to then reintegrate their results when pushing the data to D. A telling example is a low-emission zone (LEZ) in which certain vehicles are not allowed in a city centre, or only under certain conditions, because they emit too many harmful substances. In Flanders, a vehicle is linked to a natural person. When entering a LEZ, federal information linking the license plate to the owner is combined with regional data indicating whether a person has a disability. Finally the data is processed and the decision whether the vehicle is allowed is passed on to the city.

The Solid ecosystem [2, 16, 22] provides an answer by proposing a personal data pod for every citizen, such that all of their public and private data remains in one place. Instead of moving data between A and D, each of the agencies asks for permission to view a highly specific part of the data. That way, data does not have to be moved around, and GDPR compliance can be assessed automatically for every single data request. Control over personal data in our online and offline lives is a trending topic and therefore researched intensely [8, 17, 18, 20, 21, 23, 24, 27]. The key concept is that people can choose where they store their personal data, which build upon the principles of decentralisation. Blockchain is regularly referred to in this context as a solution for the management of personal data [8, 27]. Blockchain is a way for different parties who do not know each other to come to an agreement without the need for a referee or a trusted third party. This principle is essential, for example, for organising payments without a central bank or central manager, as the decentralised digital currency Bitcoin does [6]. Alas, Blockchains replicate data across many nodes. However, often, initiatives use Blockchain when this trusted third party is not required at all. If you have a central player or if the different parties trust each other, then you don't need a blockchain. Also, the immutability character of Blockchain, which implies that data cannot be deleted, might be a challenge in the context of Article 17 of the GDPR that gives people the right to erase their personal data [8, 9].

In this chapter, we explore the perspectives of control over personal data, and discuss two particular use cases that we have implemented using Solid. Solid provides a Web-based ecosystem that builds upon Open Standards and conventions [25]. According to Harrison, Pardo & Cook [13] an ecosystem is a metaphor often used to "convey a sense of the interdependent social systems of actors, organisations, material infrastructures, and symbolic resources that must be created in technology-enabled, information-intensive social systems" (p. 900). A telling example of digital ecosystems are Open Data ecosystems [26]. Open Data refers to the obligation of the government to make their non-privacy-sensitive and non-confidential data freely available on the Web [14]. The Open Data re-users depend on the data and metadata from the data providers, while the providers depend on the feedback of the re-users to increase the data quality [19, 26]. Albeit all the actors in the Open Data ecosystem are interdependent to develop their business efficiently and effectively, public administrations and policymakers are in the best position to bootstrap these open government ecosystems [13]. Zuiderwijk, Janssen & Davis state that the Open Data ecosystem challenges are related to "policy, licenses, technology, financing, organisation, culture, and legal frameworks and are influenced by ICT infrastructures" [26]. The challenges of Open Data ecosystems, which rewired the 'one-way street' into a 'bidirectional communication', could be paralleled to the challenges to put the citizen in control of their personal data as well [13, 19]. By applying the Solid ecosystem approach to two high-impact use cases, the Flemish Government aims to build up the skills and capacity to put the citizen back in control.

This chapter is further structured as follows. In the next section, we present the challenges that we aim to tackle. After that, we explain the basics around Solid in Section 6.3. Next, in Section 6.4, we discuss our approach for tackling the challenges using Solid, followed by a discussion of our implementation in Section 6.5. Finally, we conclude and present our lessons learned in Section 6.6.

6.2 Challenges

Local and regional governments in Flanders, the northern federated state of Belgium[127], aim to empower citizens in reusing their personal information online in different contexts such as public services, banking, health insurance, and telecom providers. Governments are often the custodian of authoritative personal data, such as a domicile address or medical information, which are administered by public administrations in various IS. Government administrations in Flanders share and reuse authoritative personal data between their various back-office applications to reduce the administrative burden for citizens [4], which is an implementation of the European 'once-only principle'[128]. However, public administrations are struggling to put the citizen in control.

A first challenge is that government administrations struggle to keep personal data such as email addresses, telephone numbers, or bank account numbers up-to-date. As some citizens rarely have contact with their government, personal data is often outdated in the various IS.

A second challenge concerns to allow citizens to reuse their data in a different context, such as a diploma when applying for a new job. The GDPR regulation 2016/679 states that "*In order to ensure that consent is freely given, consent should not provide a valid legal ground for the processing of personal data in a specific case where there is a*

[127] https://www.vlaanderen.be/en/discover-flanders

[128] https://ec.europa.eu/digital-single-market/en/news/ministerial-declaration-egovernment-tallinn-declaration

clear imbalance between the data subject and the controller, in particular where the controller is a public authority, and it is therefore unlikely that consent was freely given in all the circumstances of that specific situation." (European Commission, 2016, Article 43). To put it differently, the relationship between a government and a citizen is commonly considered as an imbalanced relationship, since the government wields more power than their citizens[129]. Therefore, a consent given by a citizen to reuse the authoritative data managed in government IS in the private sector, cannot be considered as freely given [9, 10]. Sharing data between government administrations in Europe is not based on a given consent but has a specific lawful basis.

Therefore, our main research question is: *how governmental processes can be streamlined by putting citizens in control of their authoritative personal data, within the context of the GDPR regulation*? This research question has two perspectives. On the one hand, how can citizens share their data with government administrations? On the other hand, how can citizens reuse their data stored in government IS in a different context?

This project evaluates how the decentralised principles of Solid [2, 20, 22, 25] can tackle these hurdles. Solid is an ecosystem that enables individuals to store data in their data pods. This gives users true control over their data, as they can choose where their data resides, and who can access it. The outcome, based on the principles of Linked Data and decentralisation, is valuable for putting the user back in control with respect to public administrations and private organisations.

[129] https://ec.europa.eu/info/law/law-topic/data-protection/reform/rights-citizens/how-my-personal-data-protected/can-my-employer-require-me-give-my-consent-use-my-personal-data_en#references

6.3 Solid

Solid [2] is a Web-based ecosystem that separates data from their applications, by providing people with their *personal data pod*, in which they can store data independently of the applications that they or others use to access that data. People can decide at a granular level which actors and applications can read from or write to specific parts of their data. The contrast with current application architectures is illustrated in Figure. 6.1. Instead of depending on a few applications that act as a gatekeeper of the data of large groups of people, the citizen is put in control of their personal data. Applications need to request access from the citizen in order to be able to operate on their data. Importantly, Solid is not an application or platform, but a protocol: a collection of Open Standards and conventions. It builds upon existing Web standards, including the Linked Data stack [2], which can be implemented by anyone.

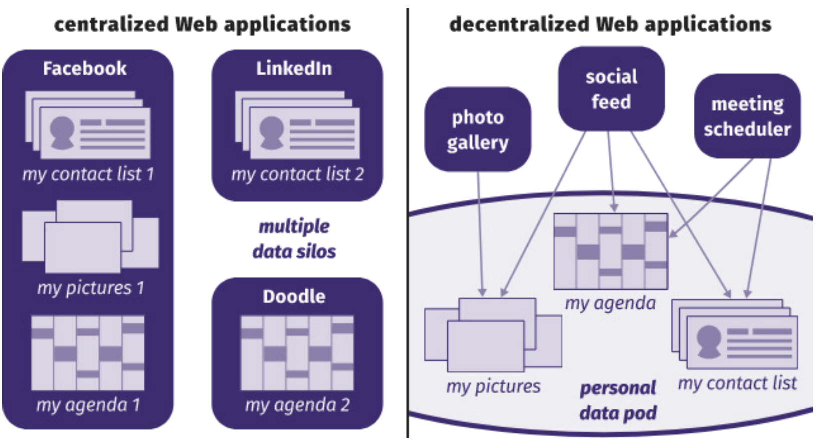


Fig. 6.1. Current applications are a combination of app and data. Thereby, the app becomes a centralisation point, as all interactions with that data have to go through the app. By introducing the concept of a personal data pod, Solid pushes data out of applications, such that the same data can be managed with different applications. This removes the dependency on a centralised application, as data can be stored independently in a location of the citizen's choice.

A data pod is a personal storage space that can exist anywhere on the Web, such as on your server, a shared community server, or a government-provided storage space. Within this data pod, the owner has full permissions regarding data creation, editing, and control management. The owner can decide to give specific permissions to other people, such as allowing family members to see their holiday pictures or allowing colleagues to read conference notes. Also, people, organisations, and applications can post a request to the public inbox of a pod to gain access to personal data. Within Solid, people have at least one data pod for themselves, but they can additionally have multiple other pods, for instance, for home data, work data, medical data, etc.

Whereas typical centralised applications require users to store their data within the application, Solid turns this around by making data personal and allows users to use any application on top of their data after granting explicit access. While simple applications work with just a single data pod, the real power of Solid becomes clear when applications *combine data* from multiple data pods, giving way to *decentralised applications*. For example, social network applications on Solid can store personal information such as posts, friends, comments, and likes in a personal data pod, while their visualisation will require combining data across different data pods. This solves two essential problems. First, data no longer needs to be *copied* in different applications, since applications will point to the single copy. Second, as a consequence thereof, synchronisation problems no longer occur: because there is only one copy of the data, applications can no longer have inconsistent versions of data.

Solid enables several capabilities that are typically missing in the current centralised web applications:

- **Independent identity**: people choose how they are identified and where their identity resides. In Solid, a personal identifier (WebID) is a URL[130].
- **Control over data**: people can grant and revoke fine-grained access permissions to specific parts of their data.
- **Choice of application**: the danger of vendor lock-in is avoided as data can flexibly be used by different applications.

For our purposes, Solid solves the aforementioned 'data train' problem, precisely because data does not move anymore between different government agencies. Instead, each government agency goes directly to the original source of the data, which is the data pod of the citizen. This addressed the problem of multiple copies and synchronisation, as well as the GDPR question of which agency has the right to access what data attributes of a citizen since each agency makes an individual request to the data pod. As such, the many processes focused on transporting data from one hop to the next, will be refocused on reading and writing data from a pod.

6.4 Approach exchanging personal information using Solid

In this section, we explain our approach for allowing citizens to share information with their government and vice versa using Solid. We first start by explaining the requirements of this approach. After that, we discuss two real-world scenarios that make use of this approach: (i) citizens sharing data, such as contact preferences (e.g., email address, telephone number) stored in a pod, and (ii) reusing authoritative government data in the private sector, such as diplomas, where the

[130] Solid uses the WebID-OIDC specification for authentication: https://github.com/solid/webid-oidc-spec

citizen keeps the diploma that has been digitally signed by the university and the government holds an indelible.

For our use cases, we assume that all citizens can be identified uniquely with a globally unique Uniform Resource Identifier (URI), referred to as a WebID[131]. This WebID points to a Linked Data document with more details about the citizen, in particular, a pointer to the personal data pod. Furthermore, we assume that all government departments and organisations have a WebID and data pod. An overview of the required components can be seen in Figure 6.2.

Typically, Solid data pods have a public inbox where anyone can post messages for the owner, where the messages can then only be read, modified, and removed by the owner. We assume this convention is met for all data pods, as we make use of this functionality for the communication between users.

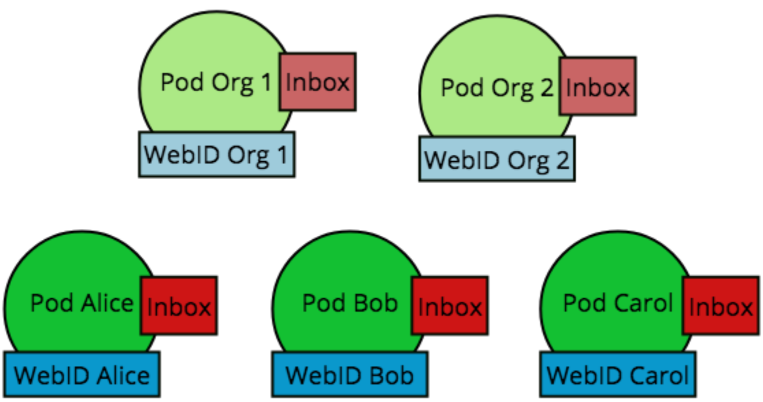


Fig. 6.2. The required components for our use cases. All governmental organisations (first row), all citizens (second row) have a data pod, WebID, and inbox.

131 https://www.w3.org/wiki/WebID

The Flemish government has developed a digital assistant that offers an integrated user experience when citizens interact with the different government administrations. A telling example is to provide citizens with notifications regarding the status of their public service, via a preferred channel. As the majority of the citizens have few digital interactions with their government, compared to interactions with the private sector, contact information and information about their preferences are often outdated. Therefore, the roles are swapped and the citizen's pod becomes the source for primary contact information and preferences. This use case addresses the first challenge and avoids that users have to keep their data up-to-date in the various portals of public and private organisations, which has an impact on the timeliness of the data.

We use an email address to illustrate this use case, which applies to any personal information.

- **Preconditions:** Citizen Alice (A) can be identified uniquely by her WebID. Also, A has a personal online data store (pod), hosted on a Solid Server (S). Likewise, organisation (O) has a WebID and a pod.
- **Use Case 1.1:** Share personal data. A authenticates to O, using secure delegated access. After successful authentication, A can grant O access to her email address by adding the WebID from O. O can read the email address from the pod after successful authentication. Extension: A can withdraw O the access to her email address.
- **Use Case 1.2:** Manage personal data. A authenticates to her pod, using secure delegated access. After successful authentication, A can add her email address to her pod via a user interface. Extension: A can modify or delete her email address.
- **Use Case 1.3:** Request access to personal data. O posts a request to the public inbox of A to gain access to the email address of A. After seeing this request, A grants O read access to her email address and send a notification to the public inbox of O. O receives the notification with a link to the original request. O can now read the email address from A.

Governments aim to empower citizens to reuse their personal information, stored in authoritative data sources on different governmental levels. Telling examples are sharing a diploma when applying for a new Job, as can be seen in Figure 6.3, or obtaining information about their income and government depths when applying for a loan. This use case evaluates a student that obtains a certificate from the university and addresses the second challenge. As a citizen cannot give the government consent to share their data with private partners, we put the user in control by storing the diploma in the citizens' pod. To put it differently, in the context of GDPR, the data subject becomes the controller of the data. This scenario indicates that Solid allows reshaping the relationship between citizens, their authoritative data and the applications they use in the public and private sector. If the citizen becomes an authentic source, legal agreements must be made to ensure that the authorities have easy access to the data. If the citizen refuses, the government can exercise this right as it does today in the tax context [1].

- **Preconditions:** Citizen Alice (A) can be identified uniquely by her WebID. Also, A has a pod, hosted on a Solid Server (S). Likewise, university (U) has a WebID and a pod. An employer (E) of A also has a WebID.
- **Use Case 2.1:** Registering as a student. A registers as a student at U, and has to provide her WebID. This will allow the university to send certificates after graduating.
- **Use Case 2.2:** Maintaining provenance until graduation. U maintains the whole provenance chain until the graduation of A. The provenance chain describes the history of a digital asset, in this case, a diploma, via a time-ordered sequence of provenance records. This includes all followed courses, grades, teachers, … This information is not publicly accessible, only A has read access to this.
- **Use Case 2.3:** Obtaining a certificate. A asks for a (summarised) copy of the certificate, so that she can share it with third parties. U will produce a summary of this certificate (not including the whole provenance

chain) and send this to the inbox of A’s data pod. This certificate is digitally signed by U using asymmetric encryption.

- **Use Case 2.4:** Sharing a diploma. Now that A has a copy of her diploma in her inbox, she can share it with anyone. For example, she can publish this on her data pod and give read access for her employer’s WebID.
- **Use Case 2.5:** Checking the validity of a diploma. If E wants to check if the diploma of A is valid, E has to check the signature of U on this diploma. E does this by extracting the signature from the diploma, determining the authority (U). This can be done using existing document signing mechanisms, such as XAdES [5].

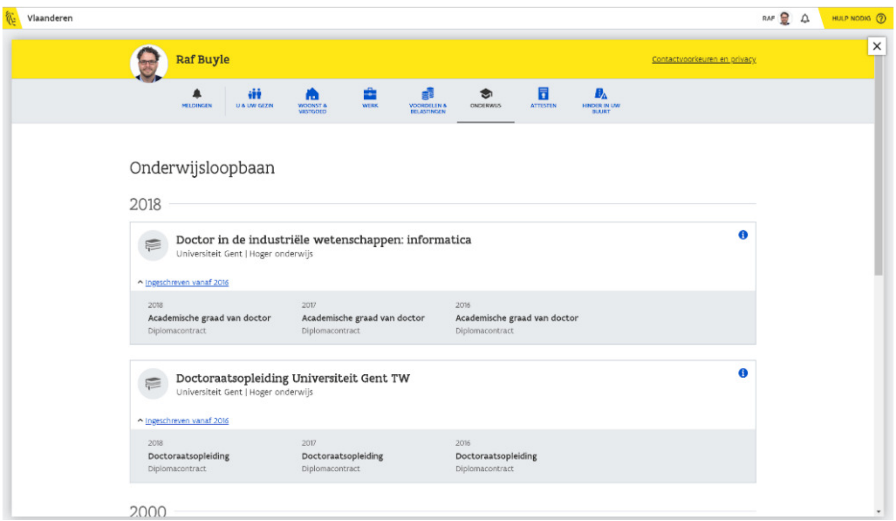


Fig. 6.3. Authoritative government data on diplomas, valuable for reuse in the private sector.

6.5 A digital assistant for Flemish citizens

In this section, we discuss the implementation of our approach into 'Mijn Burgerprofiel'[132] (My Citizen Profile), which is a smart digital assistant for Flemish citizens [4] with an overview of all their authentic information and status information of their interactions with the government. The authentication method of My Citizen Profile depends on whether the citizen is using services that process information under GDPR. The European security standard[133] 'electronic IDentification, Authentication and trust Services' (eIDAS) defines a substantial degree of confidence in the claimed or asserted identity of a person to substantially decrease the risk of misuse or alteration of the identity[134]. Users can access personal data via the My Citizen Profile by using their Belgian electronic identity card via a smart card reader or via their mobile phone, with a SIM card and their installed itsme® application[135].

As an example, we elaborate on the first use case that was discussed in the previous section, namely citizens sharing personal information (e.g., an email address). We leave the other use case as future work. As mentioned in Section 6.3, Solid detaches application from data. As such, the implementation of our approach requires two components: (i) storage for data pods, and (ii) an application for viewing and using relevant personal information. We discuss both components hereafter.

---

[132] https://overheid.vlaanderen.be/mijn-burgerprofiel

[133] https://www.eid.as/home/

[134] http://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32014R0910

[135] https://www.itsme.be/en/security

For our implementation, we make use of the Node Solid Server (NSS)[136] (version 5.0.1) to create and host data pods. If the user already has a pod, this can be used to share personal information. NSS implements the required specifications to allow users to register for a WebID and data pod, after which the server hosts this data pod and allows interaction using the Web Access Control specification[137]. NSS allowed us to create Solid pods for any citizens and governmental organisations. As such, the government provides data pods for all citizens by default. However, if citizens desire more control over their pod, they can choose to host a data pod themselves, for example by running NSS privately on their server.

In order to allow governmental organisations to request access to specific information of a citizen, or to view the actual information when access has been granted, we extended My Citizen Profile, where all Flemish citizens have a profile. Currently, this information is stored centrally within the databases of My Citizen Profile. For this work, we created a modified version of My Citizen Profile that instead stores information within the data pod of each citizen. The Flemish Government that hosts My Citizens Profile is a governmental organization, will also have one WebID, just like each citizen.

For our use case, we focus on storing the email address of a citizen. To achieve this, we implemented three components: a Solid linker, an email extractor, and an email visualizer. These components will be explained hereafter.

Within the profile settings of My Citizens Profile, we added a field where people can link their account with any Solid WebID, as can be seen in Figure 6.4. This involves logging in with any WebID via a pop-

---

[136] https://www.npmjs.com/package/solid-server

[137] https://github.com/solid/web-access-control-spec

up window. By default, each profile is linked with the default government provided WebID.

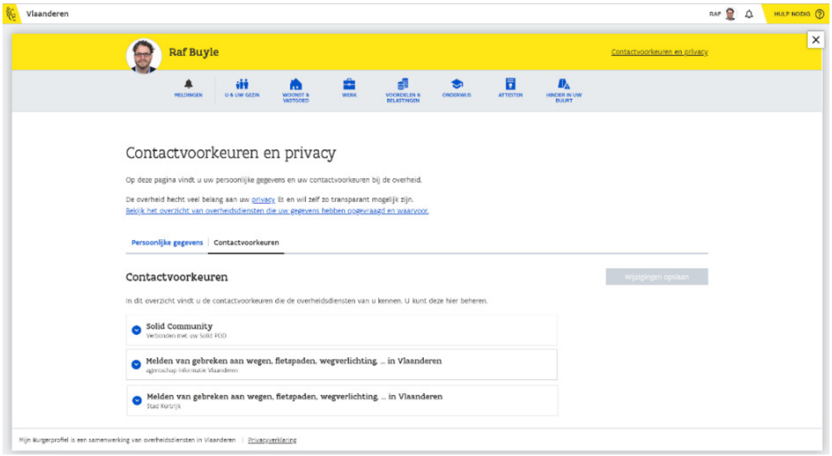


Fig. 6.4. The front-end design of the Digital Assistant, including the citizen's consent for reusing data from their personal data store.

If a citizen has a valid Solid WebID linked to its My Citizens Profile account, the application can attempt to extract its email address by following the links to the file in its data pod that contains an email address. Based on a WebID, the email extractor component can determine the URL through which the file is available in the user's data pod. With this URL, the extractor will perform an HTTP GET request, together with the authentication token of My Citizen Profile WebID.

If My Citizen Profile has been granted read access to this file by the citizen, the content of this file will be returned otherwise, an authorisation error will be returned by the data pod of the citizen. If no errors were encountered, the email extractor component will return the discovered email address.

On the personal My Citizen Profile overview page, a field is added that shows the email address of the user if this could be found. For this, the email extractor component is invoked based on the WebID that is linked to the current user. This information is always extracted on-the-fly, which means that this fact is never stored on any other location other than the citizen's data pod. This also means that when the citizen modifies the value, that My Citizen Profile, and any other authorised organizations, will be able to see the updated value immediately. This visualizer can be used in automated processes, such as sending reminders on, e.g., upcoming elections.

6.6 Discussion and conclusion

Although this chapter addresses multiple interoperability levels, the main contribution is on organisational interoperability. Organisational Interoperability refers to streamlined and aligned business processes across the different public sector administrations or organisations that act on behalf of the public sector. Organisational Interoperability is also crucial in other initiatives such as Smart Cities, where the aim is to combine vast amounts of (sensor) data for better decision making by creating a sustainable network of sensors and actuators, as discussed in Chapter 5.

In this research, we presented insights on the implementation of Solid in the region of Flanders. The Flemish government adapted the My Citizens Profile to be interoperable with the Solid ecosystem to put the citizen in control of their data. We addressed two compelling challenges, i.e., firstly that government administrations struggle to keep personal data-up to-date, and secondly to allow citizens to reuse their data stored in government IS in a different context. This initiative demonstrated that the Solid ecosystem provides an answer to the challenges by proposing personal data pod for every citizen, which enables them to share their data. Also, this method facilitates client side processing of personal data, which is not only crucial in terms of GDPR, but also relieves the application of this complexity.

New avenues for future research include investigating methods to keep the most recent version of a (summarised) copy of the authoritative data, such as a domicile address, in the users' data pod. This should ensure that the information that is shared by citizens with the private sector is always up-to-date. Another obvious extension to this research is to inform the user with the nature of the given consent to reuse data from their pod, including: the identity of the re-user, the purpose, the fact that data only will be used for automated decision-making, and/or information whether the consent is related to an international transfer of data [7]. This concept is referred to as 'informed consent' and could be implemented as a set of templates in combination with the users' preferences, which should be exchanged through a standardised vocabulary [12]. Also, all actions should be logged transparently in the pod, including access to data, data modifications, giving consent and revoking of the rights, comparable to expenses on our bank account [25]. This fine-grained and structured log can also be used to detect anomalies and data breaches by using machine learning algorithms. To complete, future research should certainly focus on the different challenges of open government ecosystems applied to the Solid ecosystem, more specific on policy, the role of the different actors and sustainable economic models.

Solid builds upon existing Web standards and methods such as Linked Data and decentralisation, therefore Solid can be seen as process innovation and organisational interoperability rather than technological innovation. As the Flemish My Citizen Profile also builds upon Web standards, including the Linked Data stack, integration with Solid pods was straightforward. We have used an email address to illustrate this case, but the intention is to broaden this to all personal data. The right as a citizen to have control over personal data could be paralleled with other basic needs. However, it is a challenge to ensure that people have at least one data pod. The Flemish Government provides a guaranteed, uninterrupted and minimal supply of electricity, gas and water for household use[138]. This principle could be extended by offering the citizens a free amount of data storage at a supplier of their choice. We expect that the insights from this Flemish Solid Pilot can speed up the process in public administrations and private organisations that face the same complexity when trying to put the user back in control.

6.7 References

[1] Belgisch Staatsblad. Wijzigingen met betrekking tot de onderzoeksmiddelen van de administratie. (2011). http://www.ejustice.just.fgov.be/cgi_loi/change_lg.pl?language=nl&la=N&table_name=wet&cn=2011041406 Accessed 9 July 2019

[2] Berners-Lee, T.; Verborgh, R. Welcome to Solid. (2019). https://rubenverborgh.github.io/Solid-DeSemWeb-2018/#title Accessed 9 July 2019

[3] Buyle, R., Taelman, R., Mostaert, K., Joris, G., Mannens, E., Verborgh, R., & Berners-Lee, T. (2019, November). Streamlining governmental processes by putting citizens in control of their personal data. In *International Conference on Electronic Governance and Open Society: Challenges in Eurasia* (pp. 346-359). Springer, Cham.

[4] Buyle, R., Van Compernolle, M., De Paepe, D., Scheerlinck, J., Mechant, P., Mannens, E., & Vanlishout, Z. Semantics in the wild: a digital assistant for Flemish citizens. In Proceedings of the 11th International Conference on Theory and Practice of Electronic Governance (pp. 1-6). ACM (2018). doi: https://doi.org/10.1145/3209415.3209421

[5] Cruellas, J. C., Karlinger, G., Pinkas, D., & Ross, J. Xml advanced electronic signatures (xades). W3C Recommendation (2003). http://www.w3.org/TR/XAdES Accessed 9 July 2019

[6] Decker, C., & Wattenhofer, R. (2013, September). Information propagation in the bitcoin network. In IEEE P2P 2013 Proceedings (pp. 1-10). IEEE.

[138] https://www.vlaanderen.be/vlaamse-overheid/persberichten/recht-op-minimumlevering-elektriciteit-gas-en-water

[7] de Montjoye, Y. A., Wang, S. S., Pentland, A., Anh, D. T. T., & Datta, A. On the Trusted Use of Large-Scale Personal Data. IEEE Data Eng. Bull., **35**(4), 5-8 (2012)

[8] Esposito, C., De Santis, A., Tortora, G., Chang, H., & Choo, K. K. R. (2018). Blockchain: A panacea for healthcare cloud-based data security and privacy?. IEEE Cloud Computing, **5**(1), 31-37.

[9] European Commission. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46. Official Journal of the European Union (OJ), **59**(1-88), 294 (2016)

[10] European Commission. Guidelines on Consent under Regulation 2016/679 Luxembourg: Publications Office (2018) https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=623051 Accessed 9 July 2019

[11] European Commission. It's your data – take control. Luxembourg: Publications Office (2018). https://ec.europa.eu/commission/sites/beta-political/files/data-protection-overview-citizens_en.pdf Accessed 9 July 2019

[12] Fatema, K., Hadziselimovic, E., Pandit, H. J., Debruyne, C., Lewis, D., & O'Sullivan, D. Compliance through Informed Consent: Semantic Based Consent Permission and Data Management Model. In Proceedings of the 5th Workshop on Society, Privacy and the Semantic Web - Policy and Technology (PrivOn2017), Vienna, Austria (2017)

[13] Harrison, T. M., Pardo, T. A., & Cook, M. Creating open government ecosystems: A research and development agenda. Future Internet, **4**(4), 900-928. MDPI AG (2012). doi: https://doi.org/10.3390/fi4040900

[14] Janssen, M., Charalabidis, Y., & Zuiderwijk, A. Benefits, adoption barriers and myths of open data and open government. Information systems management, **29**(4), 258-268 (2012). doi: https://doi.org/10.1080/10580530.2012.716740

[15] Jia, J., Jin, G. Z., & Wagman, L. The short-run effects of GDPR on technology venture investment (No. w25248). National Bureau of Economic Research (2018). doi: 10.3386/w25248

[16] Mansour, E., Sambra, A. V., Hawke, S., Zereba, M., Capadisli, S., Ghanem, A., ... & Berners-Lee, T. A demonstration of the solid platform for social web applications. In Proceedings of the 25th International Conference Companion on World Wide Web (pp. 223-226). International World Wide Web Conferences Steering Committee (2016). doi: https://doi.org/10.1145/2872518.2890529

[17] Mun, M., Hao, S., Mishra, N., Shilton, K., Burke, J., Estrin, D., ... & Govindan, R. Personal data vaults: a locus of control for personal data

streams. In Proceedings of the 6th International Conference on Emerging Networking Experiments and Technologies (p. 17). ACM (2010). doi: https://doi.org/10.1145/1921168.1921191

[18] Narayanan, A., Toubiana, V., Barocas, S., Nissenbaum, H., & Boneh, D. A critical look at decentralized personal data architectures. arXiv preprint arXiv:1202.4503 (2012)

[19] Pollock, R. (2011, March 11). Building the (Open) Data Ecosystem. http://blog.okfn.org/2011/03/31/building-the-open-data-ecosystem/ Accessed 9 July 2019

[20] Solid. Welcome to Solid (2019). https://solid.inrupt.com/ Accessed 9 July 2019

[21] Van Kleek, M., & OHara, K. The future of social is personal: The potential of the personal data store. In Social Collective Intelligence (pp. 125-158). Springer, Cham (2014)

[22] Verborgh, R. Ruben Verborgh on data & privacy. Imec Magazine (2019). https://www.imec-int.com/en/imec-magazine/imec-magazine-january-2019/back-to-the-future-how-we-will-regain-control-of-our-personal-data Accessed 9 July 2019

[23] Vescovi, M., Perentis, C., Leonardi, C., Lepri, B., & Moiso, C. My data store: toward user awareness and control on personal data. In Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication (pp. 179-182). ACM (2014). doi: http://dx.doi.org/10.1145/2638728.2638745

[24] Whitley, E. A. Informational privacy, consent and the “control” of personal data. Information security technical report, **14**(3), 154-159 (2009). doi: https://doi.org/10.1016/j.istr.2009.10.001

[25] Yeung, C. M. A., Liccardi, I., Lu, K., Seneviratne, O., & Berners-Lee, T. Decentralization: The future of online social networking. In W3C Workshop on the Future of Social Networking Position Papers (Vol. 2, pp. 2-7) (2009)

[26] Zuiderwijk, A., Janssen, M., & Davis, C. Innovation with open data: Essential elements of open data ecosystems. Information Polity, **19**(1, 2), 17-33 (2014)

[27] Zyskind, G., & Nathan, O. Decentralizing privacy: Using blockchain to protect personal data. In 2015 IEEE Security and Privacy Workshops (pp. 180-184). IEEE (2015). doi: 0.1109/SPW.2015.27

CHAPTER 7. LEGAL INTEROPERABILITY

Laws, are like sausages,
cease to inspire respect in proportion as we know how they are made.

—John Godfrey (1869).

This chapter discusses how to raise legal interoperability. This chapter examines a method to manage LBLOD. We argue that the method makes the legislation process more efficient, raises the quality of the decisions and lowers the barriers for reuse. This chapter is based on the paper 'Local Council Decisions as Linked Data' [2].

7.1 Introduction

Local councils are empowered by law, to make decisions on matters of importance to local communities. Decisions are made in formally constituted council meetings. In Flanders, local governments provide the decisions, or minutes, from these meetings to the Flemish

Agency for Domestic Governance (AGD) as unstructured data. These Council Decisions contain authentic and timely facts on, e.g., resignation of a local counsellor, the installation of a new one or the installation of a new traffic situation and their road signs. Local governments are the authoritative source for information, also available in authoritative registries, such as the registry of local counsellors or the Road Sign Database (RSD). In order to keep these registries up-to-date, local governments are obliged to update the information on local counsellors or road signs manually into a separate application provided the Flemish Government, yet the quality of the resulting register is suboptimal.

The EC defines *a base registry*[139] (BR) as a *trusted authentic* source of information under the control of an appointed public administration or organization *appointed* by the government. Maintaining a base registry comes with support and aligned processes at the level of the data providers, in this case, the local government. The RSD, which contains all road signs, their characteristics and road positions, alike the registry of local counsellors, did not live up to the expectations [6]. The Flemish Department for Mobility and Public Works created the database and inventoried the road signs. It then asked its 300 municipalities (for the municipal roads) to keep the database up-to-date. The municipalities, however, did not keep the database upto-date, as evidenced by, a.o., written question nr. 813 to minister Hilde Crevits in the Flemish Parliament (2013)[140]. The evaluation of RSD shows low scores on information, service, and system quality. The absence of net benefits will affect user satisfaction and the intention to use [1]. Local Council Decisions could provide valuable information however to such registries.

---

[139] https://ec.europa.eu/isa2/home_en

[140] https://www.vlaamsparlement.be/parlementaire-documenten/schriftelijke-vragen/870453

In this chapter, we study applying Linked Data technologies to harvest the data from the local council, as close as possible to the core processes, and publish it as Linked Data. The resulting dataset can provide information to the Base Registries. A subset of this information can also be re-used in line with the Decree on the re-use of PSI. Imagine a smart city where public decision-making is easy for all to follow using any digital channel.

7.2 Related work

OpenRaadsInformatie[141] publishes information from five local councils in the Netherlands as Open Data, as well as the OParl[142] project for local councils in Germany. Each of these projects use their own style of JSON API. The data from the municipalities is collected through APIs and by scraping websites and transformed to Linked Open Data. According to the Dutch project's evaluation [5], the lack of metadata at the source causes a direct impact on the cohesion between the different assets because they can't be interlinked. In this Proof of Concept, the data is linked at the source which allows enriching the data at an earlier stage. Next, the W3C Open Gov community group[143] is discussing and preparing an RDF ontology to describe, among others, people, organizations, events and proposals.

Finally, in Flanders, the interoperability programme of the Flemish Government, 'OSLO', focuses on the semantic level and extends the ISA Core Vocabularies to facilitate the integration of the Flemish base registries with one another and their implementation in business processes of both the public and private sector [4, 7].

---

141 https://www.vngrealisatie.nl/producten/open-raadsinformatie
142 https://oparl.org/
143 https://www.w3.org/community/opengov/

7.3 Implementation and demonstration

In the definition of the EC of a base registry, *trusted* means that, in this case, the local government is managing the Local Council Decisions conformant to best practices in all EIF-domains [3], more specific at the level of semantic and organizational interoperability and conformant to legal requirements in the Municipal Decree[144]. On a semantic level we defined a vocabulary for council decisions which will be adopted by OSLO. *Authentic* means that this is considered to be 'the' source of information which represents the correct status, and which is kept constantly up-to-date and is of highest possible quality. This is achieved by avoiding copying information manually into a separate form or application. When the registered data is part of the core processes of the local administrations and used in their IS, we expect this will improve the quality. *Appointed* means that the governing administration has a legal basis to collect and maintain the respective information.

We interviewed local governments on how they register and publish Local Council Decisions. We then organized three workshops which formulated the input for the Proof of Concept: two workshops were organized for creating a preliminary domain model, and one workshop was organized to create wireframes on how Local Council Decisions would be created and searched through in an ideal scenario. The domain concepts were formalized into two Linked Data vocabularies: one for the metadata and one for describing public mandates, formalized in https://lblod.github.io/vocabulary. The proof of concept consists of these components: an editor for local decisions, an HTML page publishing service responsible for URI dereferencing, a crawler for local decisions and two reuse examples on top of the harvested data.

---

[144] https://codex.vlaanderen.be/Portals/Codex/documenten/1013949.html

We introduce a virtual local government called VlaVirGem for which we can publish local decisions. The editor at lblod.github.io/editor is a proof of concept of such an editor, which reuses existing base registries. You can choose to fill out a certain template for decisions that often occur, such as the resignation of a local counsellor or the installation of a new one. When filling out the necessary fields, the editor will help you: for example, it will autocomplete people that are currently in office. You will then still be able to edit the official document, which contains more information such as links to legal background, context and motivation, and metadata. When you click the publish button, the decision is published as a plain HTML file on a file host. The URIs are created as hash-URIs from the document's URL.

A harvester is then set up using The DataTank[145], an open-source project to (re)publish data over HTTP. By configuring a rich snippets harvester, HTML files are parsed and some links are followed to discover the next to be parsed document. The extracted triples are republished for both the raw data as an overview of the mandates. This data is the start of two reuse demos at http://vlavirgem.pieter.pm: the first for generating an automatic list of mandates, and the second is a list of local decisions.

7.4 Conclusion

Although Local Council Decisions contain high quality information in the form of non-structured data, the information in the authoritative source for local mandates today does not. In order to reduce the workload to share this information (e.g., a newly appointed counsellor) with other governments or the private sector, the local decision can be published as a Linked Open Data document at the source.

[145] http://thedatatank.com/

The proof of concept shows that (i) an end-to-end approach, based on the developed Linked Data method, is feasible and that it can make the decision-making process more efficient: less manual work, governments may seek easier in regulation, and the Linked Data allows doing easy impact analysis when legislation is amended. (ii) We notice a quality gain in editing due to correct legal references (even referencing to decisions of their municipality) and the use of qualitative factual data (e.g., addresses linked to the Central Reference Address Database, a regional base registry). (iii) Finally, there are also efficiency gains in the publication of the decisions that are automatically published on the website of the local government, in the codex and without additional efforts suitable for reuse by third parties (Open Data). The insights created a political basis to build a base registry for Local Council Decisions in line with the best practices of this study. This project, funded by the Flemish Agency for Domestic Governance and Digital Flanders Agency (Flanders Radically Digital Programme), is a stepstone in the transition of the Flemish Government towards an information driven administration with simplified processes and better public services.

Although this chapter addresses multiple interoperability levels, the main contribution is legal interoperability. Legal Interoperability focusses on the various barriers in legislation when exchanging data across policy domains, different governmental levels or in a public-private context. We discuss how interoperability can be raised by embedding it in the legislative phase of local council decisions. Also, clauses in agreements between government administrations and software venders create hurdles to reuse the data outside the information system, which we addressed in Chapter 5.

These insights created a political basis to build a base registry for Local Council Decisions in line with the best practices of this study.

In 2018 Flemish municipalities became obliged[146] to publish their local decisions as Linked[147] Open Data. At the elections of January 2019, the majority of municipalities applied this method to register their local mandates. By the end of 2019, over a third of the software vendors for local governments adopted this method in their software.

7.5 References

[1] Benbasat, I., Barki, H.: Quo vadis, TAM? Journal of Association for Information Systems. vol 8. 3. 211-218 (2007). http://aisel.aisnet.org/jais/vol8/iss4/7

[2] Buyle, R., Colpaert, P., Van Compernolle, M., Mechant, P., Volders, V., Verborgh, R., & Mannens, E. (2016). Local Council Decisions as Linked Data: a proof of concept. In *15th International Semantic Web conference* (Vol. 1690). RWTH.

[3] European Commission, ISA: European Interoperability Framework (EIF)

[4] European public services: Annex 2. 21 (2010)

[5] European Commission, ISA: ISA handbook for using Core Vocabularies (November 2014). https://joinup.ec.europa.eu/site/core_vocabularies/Core_Vocabularies_user_handbook

[6] Kunzler,K: , Z.: Pilot Open Raadsinformatie Eindrapportage: 17 (2016). http://vng.nl/files/vng/20151112_eindrapportageopenraadsinformatiedefinitief.pdf

[7] Van Cauter, L.: Government�to�government information system failure in Flanders: an in�depth study. (2016). https://lirias.kuleuven.be/handle/123456789/533545

[8] Vanlishout, Z.: EU Core Vocabularies in Action: Linked Base Registries as a key enabler for eGovernment in Flanders. (2016). https://joinup.ec.europa.eu/sites/default/files/inline-files/SEMIC_Linked%20Base%20Registries%20as%20a%20key%20enabler%20in%20Flanders%20-%200_7.pdf

[146] https://beslissingenvlaamseregering.vlaanderen.be/document-view/5261ca28-dec1-11e9-aa72-0242c0a80002

[147] https://beslissingenvlaamseregering.vlaanderen.be/document-view/5261d284-dec1-11e9-aa72-0242c0a80002

CHAPTER 8
INTEROPERABILITY IN A HIGH-IMPACT PUBLIC SECTOR INTEGRATION PROJECT

Research is formalized curiosity.

It is poking and prying with a purpose.

— Zora Neale Hurston.

This chapter researches interoperability in a high-impact governmental integration project and outlines a method for raising interoperability between different IS and actors. We examine how semantic agreements are maintained and implemented end-to-end using the design principles of Linked Data. This chapter is based on the paper 'Semantics in the wild: a digital assistant for Flemish citizens' [4].

8.1 Introduction

Citizens expect an integrated customer experience from their government as they became accustomed to by electronic commerce services in the private sector [12]. Integrating public services from a citizen's point of view - even when these services are provided by different departments or authorities – is researched intensely and is often referred to as a 'one-stop government' [25, 22, 14, 23]. However, one-stop shop governments that integrate different services are scarce [20] and mostly stuck in vague visions [1]. The ambition of the Flemish regional government in Belgium is to digitize all interactions from public authorities to citizens and businesses by 2020 [24, 11]. The portals have several authentication methods, specific citizen profiles and a different 'feel' (layout and portal-flow). This entails that citizens have to follow-up and coordinate the public services on different portals which cause frustrations. Moreover, administrations often request information the government already has, which is in conflict with the 'once-only' principle[148]. Because of the autonomy of Flemish municipalities, the autonomy of the regional public sector agencies [18] and budget constraints, it is not feasible to rewire the entire ecosystem to a single portal. Therefore, via the programme 'Flanders Radically Digital'[149], the Flemish Information Agency is building a smart digital assistant, which will support citizens on the governmental portals of the regional and local administrations. The Smart Digital Assistant gives citizens an integrated customer experience, by providing a single-sign-on, a single profile with preferences and an overview of all interactions with the government, regardless of the portals citizens have used. On top of this, citizens have an insight into the information that the government is using in public services, which increases transparency. By integrating a smart component at the top of the government portals in Flanders, citizens have a recognizable entry point. The component, which

---

[148] https://joinup.ec.europa.eu/community/once-only-principle/home
[149] https://overheid.vlaanderen.be/informatie-vlaanderen/radicaal-digitaal

behaves similarly to a widget, provides citizens access to their personal information, all open transactions and notifications regarding public services (see Section 8.2). When consulting the status of a service, the citizen is provided with a link to a more specialized back-office application of the administration which handles the specific public service. In order to integrate the IS of the different administrations with the digital assistant, they need to be interoperable.

Interoperability is the ability of organizations to share information and knowledge, through the business processes they support, by means of the exchange of data between their ICT systems [9]. According to the EIF [7] multiple interoperability levels need to be addressed on the legal, organizational, semantic, and technical level. Because these levels assume a hierarchy in terms of maturity [13], the primary focus of this chapter is on the technical and semantic level.

This chapter researches how we can raise interoperability in an operational context, by supporting business analysts and developers in their complex design decisions to maintain semantic agreements. We focus on how information models can be aligned with existing standards and how to detect and resolve discrepancies. We outline a method for designing standardized programmable Web interfaces, which maintain the semantic agreements. As such, we address the following question: *'How to develop a scalable technique for raising and maintaining semantic and technical interoperability, within an operational public sector context?'.*

First we will provide an overview of the main concepts and outline the importance of interoperability. Next, Section 8.2 will describe the digital assistant in-depth, followed by a description of the critical success factors, challenges encountered and how they were addressed. In Section 8.3, we will discuss the method for raising and maintaining semantic and technical interoperability. After the short discussion in Section 8.3, this chapter ends with conclusions and future work.

8.2 Background

8.2.1 Strategy

Belgium is a federal state with three communities, three regions, and four language areas. The Federal State, the Communities and the Regions are at the top level and equal from a legal viewpoint[150]. Flanders is the northern federated state of Belgium[151] and an umbrella term for the Flemish Region and the Flemish Community. At bottom level are the municipalities, which are closest to the citizen. The different governmental levels are responsible for different policy domains, which causes fragmentation of the public services.

The Digital Assistant aims to facilitate a one-stop government in Flanders by creating an integrated user experience. Due to the autonomy of the various entities, the fragmentation of information and budget constraints, the Flemish government did not opt to develop a single entry point which integrates all services in a central portal (single window) but opted for the concept of a virtual window by integrating a smart assistant in the header of each portal, as illustrated in Fig. 8. These efforts are a first step in a strategy towards proactive public services where citizens do not have to take any action to receive a government service, often referred to as a no-stop-shop [19]. Qualitative research with citizens in Flanders in 2016, conducted in the context of the Digital Assistant, identified three main user requirements: (i) an overview of all interactions with the government, by means of status information and notifications, (ii) an insight into the information the government maintains and the ability to reuse this information, and (iii) personalized support. A non-functional

---

150 https://www.belgium.be/en/about_belgium/government/federale_staat/

151 https://www.vlaanderen.be/en/discover-flanders

requirement was the request to create a more uniform layout on all government portals.

Fig. 8.1: The front-end design of the Digital Assistant, including the citizen's preferences, access to personal information, active public services and notifications [16].

8.2.2 Building blocks

These findings led to the development and integration of following generic building blocks:

8.2.2.1 Single Sign-On (SSO): this building-block allows the citizen to authenticate once and be logged in to all government portals without further manual interaction [17]. The required authentication method depends on whether the citizen is using services that process privacy-sensitive information, the latter requiring substantial authentication. The European security standard[152] (eIDAS) defines a substantial degree of confidence in the claimed or asserted identity of a person to substantially decrease the risk of misuse or alteration of the identity[153].

8.2.2.2 Citizen profile: the citizen can manage preferences, including how he/she wants government administrations to interact regarding specific public services.

8.2.2.3 My Data: citizens can consult personal information that governments use in public services (once-only principle) and maintain in their authoritative information sources. This allows citizens to retrieve and consult information about their properties such as houses and land, learning certificates, and family situation.

8.2.2.4 Feedback loop: when citizens discover mistakes in the information they can give feedback, which is automatically dispatched to the responsible party.

8.2.2.5 Status information: the dashboard allows the user to consult the status of all interactions with the government. This overview contains also deep links to a more specialized back-office application of the administration which handles the specific public services.

8.2.2.6 Notifications: if the status of public service changes or a government administration wants to interact with a citizen, the notifications are sent via the preferred channel.

8.2.2.7 Contextual support: this building block brings the user in contact with the responsible government administration, including online help, chat with the helpdesk or support via telephone.

---

[152] https://www.eid.as/home/

[153] http://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32014R0910

8.3 Method

8.3.1 Critical success factors and challenges

In order to integrate the IS of the different administrations with the building blocks of the digital assistant, they need to be interoperable. The primary focus of this chapter is to raise interoperability on the technical and semantic level. The Flemish Government has an interoperability programme called OSLO [2]. OSLO brings together expertise from the public and private sector and delivers context-neutral vocabularies, in line with international information standards [3]. The specifications are published at data.vlaanderen.be[154]. Until today, the OSLO vocabularies were mainly applied to publish authoritative data sources, such as the base registry for addresses in Flanders. Since the semantics of the vocabulary terms are defined, the services implementing these vocabularies are self-describing. This supports exploration of information by automated agents and human users, which helps the latter by introducing a consistent lexicon across government administrations. In the case of the Digital Assistant, the integration services have to support use cases for the interaction with end-users on various platforms and integration with the back-office systems of the different government bodies. The services exchange both authoritative data and other information which merely supports the use cases and has no formal semantic agreements or schema. The requirements for the services are: (i) to make the authoritative data self-describing and (ii) to focus on ease-of-use in order to speed-up the adoption. by public administrations that focus on citizen-centric public services.

8.3.2 Towards machine-readable information

The data specification process in Flanders follows a transparent process. The semantic agreements are traceable and aligned to match

[154] http://data.vlaanderen.be/ns/

the different stakeholders: policy makers, domain experts, analysts and developers (see Fig. 8.).

8.3.2.1 Domain specialists: the semantic agreements are reached in open thematic working groups which consist of domain experts, both from the public and private sector. This approach is inspired by best practices of the EC [10]. The information-modelling follows a transparent process: all records of decisions[155], discussions[156] and models are publicly accessible, the latter is documented using the Unified Modelling Language™[157] (UML) [3] (see Fig. 8.: conceptual data model).

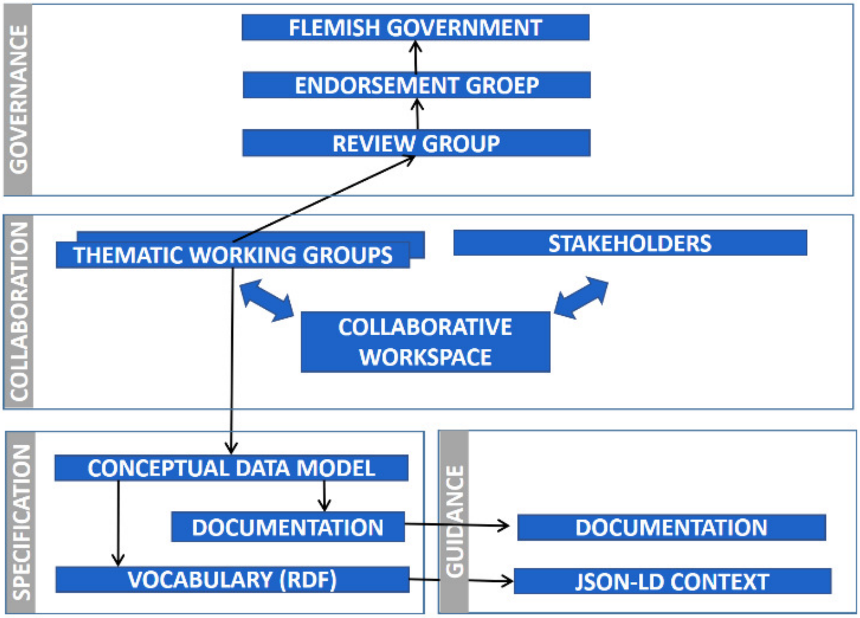


Fig. 8.2: An overview of how the semantic agreements are preserved and documented to match the different stakeholders.

---

155 https://informatievlaanderen.github.io/OSLO/
156 https://github.com/Informatievlaanderen/OSLO/issues
157 http://www.omg.org/spec/UML/

8.3.2.2 Analysts: because of its extensibility and for being a standard for data interchange on the Web, the Flemish Information Agency has chosen the RDF[158] as a data model and the principles of Linked Data for exchanging data. The UML information-model is mapped on properties of existing (international) vocabularies and transformed[159] to a RDF vocabulary (see Fig. 8.: vocabulary), which is the core of the formal data specification. This specification is then transformed into a formal specification, which adds additional constraints including mandatory properties and constraints on relations.

8.3.2.3 Policy makers: after a public review and a review by the OSLO review group, the formal specification is ratified by the endorsement group (see Fig. 8.: endorsement group). The endorsement group is empowered by a decree[160] and referred to as 'steering committee of Flemish Information and ICT-policy'. This means that these ratified formal specifications require mandatory implementation.

8.3.2.1 Developers: the Flemish Information Agency is building upon the principles of Linked Data, to allow data to be exposed and shared across different applications. Linked Data refers to a set of best practices for publishing and connecting structured data on the Web using international standards of the W3C" [26]. In line with the design principles[161] as asserted by Tim Berners-Lee in 2006, all the information objects are given a universally unique identifier which can be looked up via the Web, e.g., http://data.vlaanderen.be/id/adres/2179183 for an address. When a person or machine navigates to this identifier (URI) on the Web, standardized information is provided, using RDF as a data model. In addition, links to other useful datasets and resources are

---

158 https://www.w3.org/2001/sw/wiki/RDF
159 https://github.com/Informatievlaanderen/OSLO-EA-to-RDF
160 http://docs.vlaamsparlement.be/pfile?id=1213278
161 https://www.w3.org/DesignIssues/LinkedData.html

included. This strategy is already used for publishing authoritative data sources in Flanders including addresses[162] and organizations[163]. Although RDF is a simple data model, a lot of developers are overwhelmed by the (perceived) complexity [15]. Therefore, we have applied a developer friendly approach by using JSON-LD, which is based on Linked Data but also complies with the requirements of the digital assistant, namely, to make the authoritative data self-describing and to focus on ease-of-use in order to speed-up the adoption. JSON-LD is used by hundreds of millions of applications most often without the knowledge of the application users [15]. Another advantage is that this method allows combining self-describing data linked to an RDF schema with other information which has no formal semantic agreements. The information that is under the governance of OSLO is linked to the vocabularies using a JSON-LD context[164], which allows embedding the semantic agreements in JSON services. The identifiers of the governed information objects are dereferenceable and allow integrators to discover more about the authoritative reference data. The context file (see Fig. 8.: JSON-LD context), which bridges the semantics of the interface to the vocabulary is maintained and published by the authority. In Section 8.3.4 we will discuss the implementation of JSON-LD.

8.3.3 Unpacking the harmonization process

This section describes how the information models of the Digital Assistant are aligned with existing OSLO vocabularies and how they are inspired by the best practices of the Interoperability programme of the EC [8] and the W3C[165].

---

[162] http://data.vlaanderen.be/id/adres/2179183
[163] http://data.vlaanderen.be/id/organisatie/OVO002949
[164] https://www.w3.org/TR/json-ld-syntax/#the-context
[165] https://www.w3.org

8.3.3.1 Development of the conceptual data model. A first step in the development of semantic agreements is the use case modelling, which describes a specific usage of a system by one or more actors [6]. A second step is the design of a conceptual data model in UML, which is the most abstract form of a data model. It consists of UML Classes and their definitions, which represent things that exist in the real world (e.g., a person), and their associations and properties (e.g., a relation to family members).

8.3.3.2 Mapping of the information model to existing vocabularies. The goal of this step is to identify whether there already are existing qualitative vocabularies, in order to raise interoperability. In addition, the reuse of existing analyses lowers the development costs.

8.3.3.3 Detection of possible discrepancies. The goal of this step is to document the semantic design decisions and to prepare the design of the data model. Table 8.1, which is based on a method[166] of the interoperability programme of the EC, illustrates how one of the information models of the Digital Assistant is mapped to the OSLO vocabularies. The properties and associations are compared using the relations defined in the Simple Knowledge Organization System vocabulary[167] (SKOS). These relations are defined as (i) *closeMatch* indicates that the concepts can be used interchangeably across a wide range of applications, (ii) *relatedMatch* is used to link two concepts that are sufficiently similar that they can be used interchangeably in particular applications, (iii, iv) *broadMatch* and *narrowMatch* are used to state a hierarchical mapping link between two concepts and finally (v) *relatedMatch* is used to state an associative mapping link between two concepts.

---

[166] http://mapping.semic.eu/

[167] https://www.w3.org/2009/08/skos-reference/skos.html

Table 8.1: A subset of the vocabulary mapping, based on a method of the European ISA interoperability programme (terms are translated 168 to English).

Digital Assistant 'family relations'		SKOS Relation	OSLO Person vocabulary	
Class	Property		Class	Property
Person	name	Exact match	fullName	Person
Person	registry	Related match	citizenship	Person
Person	family members	Broad match	Has-Relation-With	Person
Person	Reference person	Narrow match	Family head	Person
Person	Administrator	No match	n.a.	n.a.

8.3.3.4 Resolving the discrepancies. If a property does not have a corresponding term in an existing vocabulary, we need to assign a new globally unique name. A good starting point for finding existing terms are the LOV[169]. It is important to evaluate whether the definition of the term matches the correct context. To ensure these names are unambiguous they are identified by Unique Resource Identifiers (URIs), generalized versions of a URLs used to locate web pages via a browser. By using HTTP URIs, they provide a link to a description online. If the one cannot find an existing term, one needs to define a new one; this process is often referred to as minting URIs [26].

[168] http://bit.ly/dig_assist_mapping_oslo
[169] http://lov.okfn.org/dataset/lov/

8.3.3.5 Development of the formal data specification

The vocabulary terms in the UML model are now mapped to the RDF vocabulary terms, as defined in Section 8.3.2.2 and 8.3.2.3. The UML model along with the mappings are then automatically transformed[170] into an RDF model [5]. In addition, documentation for developers and business-analysts is automatically generated from the RDF schema and enriched with contextual information.

8.3.4 Implementation

To comply both with the ease-of-use requirement and to make the authoritative data self-describing, we have based our approach on JSON-LD. JSON (JavaScript Object Notation) is a popular lightweight data-interchange format[171]. The disadvantage of JSON is that the services can only be documented in the human readable documentation, unlike XML[172] based services that can be annotated with machine-readable descriptions by using the XML schema language[173]. JSON-LD allows to make JSON documents self-descriptive and allows developers to work with Linked Data without the high entry barrier [15]. Being fully compliant with the classic JSON, the Flemish Information Agency has decided to create a blend: the objects that are under the governance of OSLO are JSON-LD enabled, whereas the other information, which merely supports the use cases, is in plain JSON integrated within the same document.

170 https://github.com/Informatievlaanderen/OSLO-EA-to-RDF
171 http://www.json.org/
172 https://www.w3.org/XML/
173 https://www.w3.org/XML/Schema

```
{
 "@context": {
   "Organisatie":"http://www.w3.org/ns/org#Organization",
   "voorkeursNaam": "http://www.w3.org/2004/02/skos/core#prefLabel",
   "alternatieveNaam":"http://www.w3.org/2004/02/skos/core#altLabel"
 },
 "@id": "http://data.vlaanderen.be/id/organisatie/OVO002949",
 "@type": "Organisatie",
 "voorkeursNaam": "Informatie Vlaanderen",
 "alternatieveNaam": "AIV"
}
```

Fig. 8.3: Code snippet of a JSON-LD object, which describes an organisation in the citizens portal.

When people communicate, the relation between linguistic expressions and what they express takes into account the context in which expressions are used and interpreted: 'the context of the conversation' [21]. A context in JSON-LD affords IS to communicate more efficiently by using shortcut terms, which can be compared as referring to the first name of a mutual friend, to communicate more quickly without losing accuracy[174]. The code snipped in Fig. 8. describes the *organisation* 'Digitaal Vlaanderen' in the citizens portal. As outlined in Section 8.3.2.4, dereferenceable URIs allow machines to browse the Web of data, as humans browse the Web of documents. The object *organisation* has a unique URI[175], which is governed by the Flemish Government[176]. The term *voorkeursNaam*, which is Dutch for *preferable name*, is associated[177] with the Object *organisation* which is the Dutch for an organization by using the JSON-LD context. The context maps the shortcut term *voorkeursNaam* to the URI '…/skos/core#prefLabel'[178], which provides a definition in line with international information standards. In the example in Fig. 8.3, the context is embedded in the JSON-document. In the services of the

[174] https://json-ld.org/spec/latest/json-ld/#the-context
[175] http://data.vlaanderen.be/id/organisatie/OVO002949
[176] https://overheid.vlaanderen.be/OSLO-URI-standaard
[177] https://json-ld.org/spec/latest/json-ld/#dfn-expanded-term-definition
[178] http://www.w3.org/2004/02/skos/core#prefLabel

Digital Assistant, the context is pointing to an external document[179] and generated automatically from the formal specification, for reasons of maintainability.

8.4 Conclusions

The aim to provide access to public services via a single-entry point is researched intensely, but governments still struggle to realize a one-stop government. The Smart Digital Assistant gives the citizen a unified customer experience without the need for rewiring all existing portals to a single channel. Although this approach is based on lean integration, interoperability is crucial. In this chapter, we proposed a method to raise interoperability on the technical and semantic level based on the architectural principles of Linked Data.

The method includes an implementation framework that describes how to make authoritative data self-describing. The semantic agreements are traceable and aligned to match the different stakeholders: policy makers, domain experts, analysts, and developers. We showed that the RDF can facilitate the semantic agreements and that JSON-LD allows developers to work with Linked Data without a high entry barrier. Our framework can be used by countries that face the complexity of integrating e-government portals. Simultaneously, our work can benefit e-government integration projects as it provides an end-to-end governance, as well as practical insights on the design of lightweight services. New avenues for research are to add machine-readable validating rules to the formal data specification. The set of conditions will be formalized using the Shapes Constraint Language[180] (SHACL), which will allow creating automated compliance tests. We will research if the set of conditions can be generated semi-automatically.

---

[179] http://data.vlaanderen.be/context/organisatie.jsonld
[180] https://www.w3.org/TR/shacl/

In this chapter we focussed on **the first viewpoint of our research question** “*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?*”. We analysed the Smart Digital Assistant, which provides a unified user experience for Flemish citizens when consuming public services on the various government portals. To realise this concept, the different information flows need to be harmonised and therefore interoperability is crucial. We presented an end-to-end method to raise technical and semantical interoperability, which builds on the principles of Linked Data.

The main contribution of this research is that we demonstrated that the design principles of the Semantic Web can facilitate interoperability within the public sector by adding context and useful links, using the RDF as a data model. Also, we demonstrated that decisions in relation to semantic agreements must be traceable, transparent, and consistent at all levels. Also, the format of the specifications and guidelines must be aligned to the different stakeholders to facilitate a levelled discussion.

8.5 References

[1] Bannister, F., Connolly, R. Forward to the past: Lessons for the future of e-government from the story so far. Information Polity 17.3, 4. pp. 211-226 (2012). DOI: http://dx.doi.org/10.3233/IP-2012-000282

[2] Buyle, R., et al. OSLO: Open Standards for Linked Organizations. Proceedings of the International Conference on Electronic Governance and Open Society: Challenges in Eurasia. ACM, pp. 126-134 (2016). http://dx.doi.org/10.1145/3014087.3014096

[3] Buyle, R. Towards interoperability in the public sector. The Semantic Web-ISWC 2017: 16th International Semantic Web Conference, ISWC 2017, Vienna, Austria, October 21-25, 2017. (accepted doctoral consortium paper – unpublished).

[4] Buyle, R., Van Compernolle, M., De Paepe, D., Scheerlinck, J., Mechant, P., Mannens, E., & Vanlishout, Z. (2018, April). Semantics in the wild: a digital assistant for Flemish citizens. In *Proceedings of the 11th*

International Conference on Theory and Practice of Electronic Governance (pp. 1-6).

[5] 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).

[6] Eriksson., et. Al. Business Modeling with UML. New York. pp. 1-12 (2000).

[7] European Commission, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. European Interoperability Framework Implementation strategy. http://eur-lex.europa.eu/resource.html?uri=cellar:2c2f2554-0faf-11e7-8a35-01aa75ed71a1.0017.02/DOC_1&format=PDF (accessed 26.08.17).

[8] European Commission, ISA. E-Government Core vocabularies Handbook. Publications Office of the European Union, Luxembourg. pp. 11-50 (2011).

[9] European Commission, ISA: European Interoperability Framework (EIF). Publications Office of the European Union, Luxembourg p. 5-37 (2017). http://dx.doi.org/10.2799/78681

[10] European Commission, Process and methodology for developing semantic agreements. p. 7 (2013).

[11] Homans., L. Vlaanderen Radicaal Digitaal 2015-2019. pp. 3-9 (2015).

[12] 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).

[13] Kubicek, H., Cimander, R. Three dimensions of organizational interoperability. European Journal of ePractice, 6, pp. 1-12 (2009).

[14] Kubicek, H., Hagen. M. One stop government in Europe: An overview. pp. 1-36 (2000).

[15] Lanthaler, Markus. ”Third Generation Web APIs.” pp. 48-111 (2014).

[16] Mostaert, K., Tegenbos, R. Vlaanderen Radicaal Digitaal – burgerloket. p. 24 (2017). http://bit.ly/VRD_2017_CPortal (accessed 26.08.17).

[17] Pashalidis, A., Mitchell, C. A taxonomy of single sign-on systems. In: Information security and privacy. Springer Berlin/Heidelberg, 2003. pp. 219-219.

[18] Roness, P., et al. Explaining autonomy of public sector agencies in Norway, Ireland and Flanders. EGPA Annual Conference, Study Group Governance of public organizations. pp. 4-59. (2007).

[19] Scholta, H., et al. From One-Stop to no-Stop-Shop: an e-Government stage model. In Proceedings of the 25th European Conference on

Information Systems (ECIS), Guimarães, Portugal, June 5-10, 2017 pp. 918-934.

[20] Schuppan, T., Koehl, S. One Stop Government: Stalled Vision or a Matter of Design?–Empirical Findings from Social Services in Germany. Proceedings of the 50th Hawaii International Conference on System Sciences. p. 2448 (2017). DOI: http://dx.doi.org/10.24251/HICSS.2017.296

[21] Stalnaker, R. On the representation of context. Journal of Logic, Language and Information 7.1. pp. 3-19 (1998).

[22] Tambouris, E. ”An integrated platform for realising online one-stop government: the eGOV project.” Database and Expert Systems Applications. Proceedings. 12th International Workshop on. IEEE, 2001.

[23] Tambouris, E., Archetypon, S., Wimmer, G. M. Online One-Stop Government: A Single Point of Access to. Electronic government strategies and implementation. pp. 115.(2004).

[24] Vlaamse Regering, Regeerakkoord van de Vlaamse Regering 2014-2019. p. 11 (2014).

[25] Wimmer, M. A. A European Perspective towards online one-stop government: the eGOV project. Electronic commerce research and applications 1.1: pp. 92-103 (2002). https://doi.org/10.1016/S1567-4223(02)00008-X

[26] Wood, D., Zaidman, M., Ruth, L., Hausenblas, M. Linked Data - Structured Data On The Web, Manning Publications Co., New York, 2014, p.4-235.

CHAPTER 9. CONCLUSION AND DISCUSSION

A Little Semantics Goes a Long Way.

— James Hendler[181].

This final chapter summarises the main findings of this dissertation and formulates an answer to the central research question. Additionally, it outlines the theoretical and practical contributions of my research. Finally, some limitations of the work in this dissertation and directions for future research are addressed.

[181] https://www.cs.rpi.edu/~hendler/LittleSemanticsWeb.html

9.1 Review of the research question and key findings

This dissertation aimed to research 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've researched this challenge both from the technical and political point of view in the context of e-government in the region of Flanders in Belgium. Our 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 within a public sector context. Our method allows datasets to be linked into a public sector knowledge graph governed by a public body. In Chapter 1, We've outlined the research questions and hypothesis. Chapter 2 sketched the concept and discussed the adoption criteria for Open Data standards. This dissertation answered the central research question across different chapters:

> **How can governments develop a scalable technique for raising and implementing semantic and technical interoperability, within an operational public sector context?**

This question had two perspectives. First, we focussed on the technical viewpoint:

> **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?**

In Chapter 4 we've presented the insights on the implementation of a Linked Base Registry for Addresses by unfolding the process followed towards raising semantic interoperability based on Linked Data principles. While implementing the address vocabulary, we stumbled on competing international semantic standards and difficult choices on how to extend them to fit the local context. To **enable business analysts and developers to maintain semantic agreements** and to cope with these challenges, it is crucial to have a transparent process and methodology for developing semantic agreements and a governance structure for making and institutionalising pivotal decisions in place. This can be realised through a policy framework for technical as well as domain-specific topics, alike to the OSLO[182] programme in Flanders. This policy also includes a URI standard for **persistent unambiguous identifiers** that supports government administrations by providing guidance that ensures that HTTP URIs are future proof[183]. The results show that the method of Linked Data can indeed increase semantic and technical interoperability and can lead to better adoption of data such as addresses in the public and private sector. **Chapter 5** delved into technical interoperability. We've researched what methods are suited for publishing Open Data time series in a sustainable, predictable and cost-effective way in the context of Smart Cities. Also, we analysed the REST architectural style that resembles the human Web, which builds upon hyperlinks, and a set of architectural constraints that facilitate architectural elasticity. The **uniform interface** simplifies the architecture and empowers software clients to evolve separately. Also, as all messages are self-descriptive and responses include links to possible actions, they can be **easily interpreted by clients**, making out-of-band documentation needless. Also, this chapter argues that the method of Linked Data not only raises technical interoperability but also lowers the publishing cost and raises the availability of the endpoints. This research demonstrated that the significant benefits of adopting the principles of Linked Data regarding air-quality time series

---

[182] https://data.vlaanderen.be/cms/Proces_en_methode_voor_de_erkenning_van_data standaarden_v1.0.pdf

[183] https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/document/uri-standard-guidelines-flemish-government

— as well as other fast and slow moving datasets — does not only provide interoperability towards external stakeholders but also foster a more sustainable and cost-effective architecture. Finally, more evidence was gathered by applying the process and method for raising interoperability to a high impact public sector integration project in Chapter 8. This chapter outlined a method for raising semantic interoperability between different IS and actors. We examined how semantic agreements are maintained and implemented end-to-end using the design principles of Linked Data. We analysed the Smart Digital Assistant, which provides a unified user experience for Flemish citizens when consuming public services on the various government portals. To realise this concept, the different information flows needed to be harmonised, and therefore, interoperability is crucial. We presented an end-to-end method to raise technical and semantical interoperability, which builds on the principles of Linked Data. The main contribution of this chapter is that we demonstrated that the design principles of the Semantic Web can facilitate interoperability within the public sector by adding context and useful links, using the RDF as a data model. We demonstrated that decisions in relation to semantic agreements must be traceable, transparent and consistent at all levels. Also, the format of the specifications and guidelines must be aligned to the different stakeholders to facilitate a levelled discussion.

Second, we've focussed on the political context:

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

In Chapter 3, we've outlined the process and method for developing Open Data standards. Also, we've addressed the interoperability hurdles at the different governmental levels and examine how the OSLO programme tackled these obstacles. We have determined that both the bottom-up and top-down approach were vital to create the

necessary political support at the different governmental levels. The bottom-up approach, where different government levels and private partners are collaborating on interoperability, is crucial for **building consensus**. The transparent data specification process assures a level playing field. The process is aligned with the principles[184] of international standardisation bodies, i.e., due process, broad consensus, transparency, balance, and openness. This bottom-up process is combined with formal top-down governance. The Steering Committee 'Flemish Information and ICT-policy', which is empowered by a decree[185], is responsible for the governance. When the governance was formalised, the development of new data standards and the adoption of standards by public and private IS increased swiftly. The acceleration of the adoption of data standards shows the importance of formal governance for the trust of the various stakeholders. Finally, more evidence was gathered by applying the results to a high impact public sector integration project in **Chapter 8,** which was discussed in the previous section, demonstrated that decisions in relation to semantic agreements must be traceable, transparent and consistent at all levels to assure a level playing field.

Our results show that it is possible to reach semantic agreements and overcome the political hurdles within an operational public sector context by using a meet-in-the-middle approach. Throughout this dissertation we demonstrated that the design principles of the Semantic Web can facilitate interoperability within the public sector by adding context and useful links, using the RDF as a data model. Also, we demonstrated that 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

---

[184] https://open-stand.org/about-us/principles/
[185] http://docs.vlaamsparlement.be/pfile?id=1213278

stakeholders (e.g., technical, business, and policy) to facilitate a levelled discussion.

9.2 Practical contributions

9.2.1 Impact

In this section, we discuss the practical impact and outcome of the different research topics which are addressed throughout this dissertation. The process and method for raising semantic and technical interoperability within an operational public sector context were embraced by the 'Steering Committee of Flemish Information and ICT-policy' in 2018[186]. This government embraced governance raised the development and adoption of data standards in the region of Flanders. At the time of writing this dissertation, the number of ratified data standards by the Flemish Government has grown to more than ninety[187]. In 2015 the focus of the interoperability programme was at the Core Vocabularies including persons, addresses, buildings, public services, objects in the public domain, and legislation. The early adopters were the Digital Flanders Agency and The Agency of Local Regional Governance. In 2019, we noticed a shift towards business domains, which affirms the trend towards adoption of data standards in the mainstream business domains including tourism, a digital twin for mobility infrastructure, MaaS, cultural heritage, tourism, air quality, water quality, and the subsoil. The community has grown to over four hundred authors from the public sector, private sector, and academia collaborating on the development of data standards[188]. In 2019 the process and method were adopted on the Belgian interfederal level[189]

[186] https://data.vlaanderen.be/standaarden/erkende-standaarden/proces-methode-ontwikkeling-standaarden/proces-methode-ontwikkeling.html

[187] https://data.vlaanderen.be/standaarden/

[188] https://data.vlaanderen.be/standaarden/

[189] https://github.com/belgif/review/blob/master/Process/201906-ICEG%20-%20process%20and%20method.docx

(the federal level, the communities, regions, and language areas in Belgium) too and on top of that the European ISA programme researched the potential to apply this method on the European level. Also, the OSLO semantic interoperability programme has structural funding that is a combination of a grant from the Flemish government and co-funding from various government administrations. The estimated budget of the OSLO interoperability programme in 2021 will exceed 1500 person-days, not including the work of the vast community that cocreates the standards in the various business domains.

The high-impact project, which we discussed in **Chapter 8**, examines a method to publish LBLOD. The LBLOD programma has won several awards for its innovative approach including the Agoria 2019 Open Data e-Gov Award[190], 'Gouden Byte' award[191] and the 2020 Publica Technology Award[192]. At the time of building this prototype in 2016, we researched the possible impact on the reuse of data and the legislative processes and the opportunities for reuse. In his 2019-2023 policy letter, the Vice minister-president of the Flemish Government and Flemish Minister for Local and Provincial Government announced[193] LBLOD is the linking-pin between on the one hand local governments and on the other hand the regional government and the private sector. All local governments are obliged to publish their local decisions in line with the OSLO standards via an Open Source Linked Data Editor provided by the Flemish Agency for Local and Provincial Government (ABB) or via an OSLO compliant software of an ICT service provider. By the end of 2019, 30% of the local governments were using third-party software that is OSLO compliant. The method makes the

[190] https://www.agoria.be/nl/Modernisering-burgerlijke-stand-wint-e-gov-Award-2019

[191] https://lokaalbestuur.vlaanderen.be/nieuws/tool-gelinkt-notuleren-wint-gouden-byte-2018

[192] https://publica-brussels.com/winnaars-publica-awards-2020/

[193] https://publicaties.vlaanderen.be/download-file/32209

legislation process more efficient, raises the quality of the decisions and lowers the barriers for reuse. During the installation of the new councils, 130 municipalities saved 67 days when publishing 8055 new public mandates as Linked Data [12]. In 2021, ABB will focus on new policy domains, including mobility. Local Council decisions will provide data to both the RSD as well as private initiatives including route planners. As more municipalities join in and new policy types become available, the Flemish Agency for Local and Provincial Government expect this profit to increase in the future. Taking into account 10.000 decisions related to traffic regulation, local governments save four hours to create the local council decision by providing context information in their RSD. This saves the administration 1.250.000 EUR each year [13]. Also, this lowers the barriers and costs for private re-users too.

In the 2019 September Declaration[194] — the annual government policy statement of Flanders — the Minister-President of Flanders, Jan Jambon, stated that "*the strength lies in unlocking digital data, both in the government and in the private sector*" and that "*citizens and businesses must gain more control over their own data*" (Jambon, 2020, p. 24) [6]. In the context of the September Declaration and the COVID-19 recovery plan of the Flemish government, there is an extensive focus on digitization. The digital transformation in Flanders includes two high-impact projects: an IoT sensor data broker and self-sovereign identity for Flemish Citizens. The IoT sensor data platform, alike **Chapter 6**, will focus on a sustainable, predictable, and cost-effective way to publish data in the context of Smart Cities. The project will build upon the method of Linked Data to raise interoperability. Also, the broker embraces the method of LDF to lower the publishing cost and raise the availability of the endpoints. The self-sovereign identity builds upon the insights of My Citizens Profile in **Chapter 8** and Solid in

---

194 https://publicaties.vlaanderen.be/download-file/38745

Chapter 6. This decentralised approach based on the Solid ecosystem and Linked Data will allow reshaping the relationship between citizens, their personal data, and the applications they use in the public and private sector. During the first co-creation sessions to bootstrap the ecosystem, promising use cases were identified in the domains of mobility, health, media, and the government.

9.2.2 Implications for practice

In this section, we discuss how practitioners can apply our findings in the field. Our approach to raising interoperability combines the process to reach technical and semantic agreements by broad consensus and an end-to-end method based on the principles of Linked Data to maintain the semantic agreements within an operational public sector context. This can be applied in four steps (Fig. 9.1):

1. **Set up a formal governance** by anchoring the standardisation process at an existing governance body or initiating a new governance body. This is crucial for the trust of the various stakeholders and impacts the adoption of data standards.
2. **Agree on a transparent process** to reach semantic and technical agreements. The process outlines the roles of the different actors and specifies how consensus can be reached among stakeholders. Reference implementations of this process are applied and documented in Flanders[195] and on the Belgian interfederal level[196].
3. **Install an end-to-end method** based on the principles of Linked Data. This implies that all records of decisions, discussions and models are publicly accessible; the latter is documented using a formal language based on RDF. The method must include an implementation framework that ensures semantic agreements are

[195] https://data.vlaanderen.be/standaarden/erkende-standaarden/proces-methode-ontwikkeling-standaarden/proces-methode-ontwikkeling.html (Dutch)

[196] https://github.com/belgif/review/blob/master/Process/201906-ICEG%20-%20process%20and%20method.docx (English)

traceable and aligned to match the different stakeholders: policy makers, domain experts, analysts, and developers. Reference implementations of this process are applied and documented in Flanders[197] and on the Belgian interfederal level[198].

4. **Cocreate data standards**: the semantic agreements are reached in open thematic working groups, consisting of domain experts from the public sector, private sector and academia. These working groups follow the process and method within a formal governance framework.

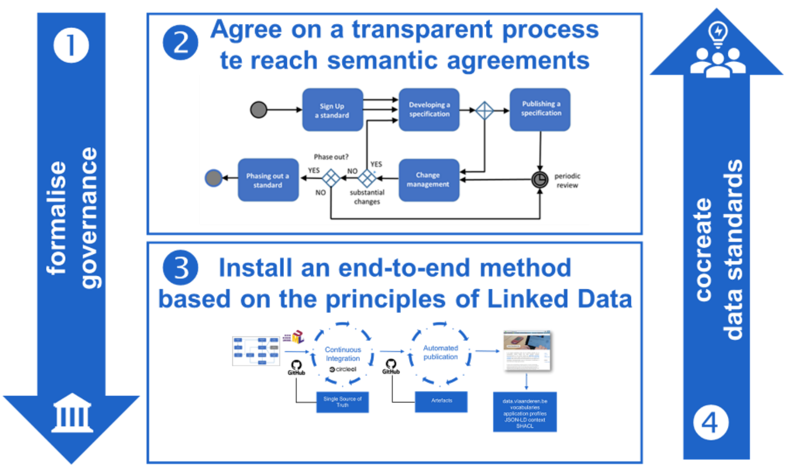


Fig. 9.1: raising interoperability in the public sector.

[197] https://data.vlaanderen.be/standaarden/erkende-standaarden/proces-methode-ontwikkeling-standaarden/proces-methode-ontwikkeling.html (Dutch)

[198] https://github.com/belgif/review/blob/master/Process/201906-ICEG%20-%20process%20and%20method.docx (English)

9.3 Reflections

In contrast to the rest of this dissertation, I will write this section in the first person, as it includes my reflections.

First, during my research, I had a privileged position to participate in the developments of standardization and digital transformation in the public sector in Flanders. I was responsible for the OSLO standardization programme from the early start in 2012 by the Flemish ICT organization. In 2015, when the Flemish Government became OSLO's lead, I continued to play this role. Since I started my PhD in 2016, I have combined the role of researcher with my standardisation activities in the public sector in Flanders. As a privileged observer, analyst, and critic, I gathered my data via action research. This refers to the fact that I was involved as co-practitioner in the setting under study and combined theory and practice [7, 11]. Action research is an established research methodology that is often applied in the social sciences and medical sciences since the 1950s [2]. According to Avison et al. action research is "*an iterative process involving researchers and practitioners acting together on a particular cycle of activities, including problem diagnosis, action intervention, and reflective learning*" (p. 94) [1]. Action research involves some ethical and professional challenges for the researcher-practitioner. On the one hand, as I knew all the internal decisions and politics, on an ethical level, it was important to have informed consent from the government administration [2]. In all the papers I've written as a first author, the participating Flemish Government administration was co-authoring. On the other hand, on a professional level, it's important to keep the focus on the research goals and obligations to develop general knowledge [2]. The stunning coaching by my research groups IDLab and MICT, peer-reviewed papers and my broad international research network ensured to keep the focus on my obligations to develop general knowledge. Regardless of these challenges, action research provided an answer to the needs for relevant research in the fast changing domain of IS research, in this case within an operational public sector context [2].

Second, I would like to focus on a limitation. I've conducted my research for raising semantic and technical interoperability in the Flemish public sector in Belgium. Standardisation is a multistakeholder activity, in this context with coördination between different government administrations, the private sector, and academia within a specific policy framework. The governance model in Flanders that finds its roots in geospatial e-services and standards can be characterised as a mix of hierarchical and network governance [3]. Network coordination has an important impact on addressing these complex problems [9]. In 2019, the process and method was adopted on the federal level, the communities, regions, and language areas in Belgium[199] too and furthermore successfully applied to the standardisation of URIs[200]. As this interfederal initiative — which operates within a different policy framework — was successful, we have a first indication that this process and method can be applied in different public sector contexts. Although Belgium is often referred to as an ideal test market and a bridgehead to the common market, further research is needed to evaluate what factors influence the adoption of this process and method and if we can generalise our conclusions to other regions and countries [10].

9.4 Directions for future research

The TRAM-model reveals that innovativeness is an important influencer for the use intention of data standards. We expect that other parameters which are not included in the model might have an impact on the use intention such as organisational factors and potential network effects because data standardisation is a multistakeholder

[199] https://github.com/belgif/review/blob/master/Process/201906-ICEG%20-%20process%20and%20method.docx

[200] https://github.com/belgif/thematic/blob/master/URI/iceg_uri_standard.md

activity as well (e.g., coordination between agencies and the context of policy frameworks). Therefore future research is needed to explore the effects of network governance in order to speed up the adoption of Open Standards to raise interoperability in complex ecosystems. We suggest researching the impact of organisational impediments (e.g., lack of support from top management) and economic impediments [8]. Therefore, in our future research, we wish to research and develop a model that extends the TRAM model with the model of Provan & Kenis that includes the structural characteristics of the Network Governance Forms [9].

We want to evaluate the parameters that have an impact on the use intention of data standards on the federal level, the communities, regions, and language areas in Belgium.

When the lower levels of interoperability are in place, organizational and legal interoperability can be realized. Because organizational interoperability is about processes and, therefore, closer to end-users, it seems useful to involve end-users - such as citizens - in the co-creation process.

We outlined an implementation framework that describes how to make authoritative data self-describing. The semantic agreements are traceable and aligned to match the different stakeholders: policymakers, domain experts, analysts, and developers. We showed that the RDF can facilitate semantic agreements. The set of conditions are formalized using theSHACL, which allows for automated compliance tests too, e.g., evaluating if an information system is compliant with an application profile such as *a local council decision* or an *address* is straightforward. Also, there are guidelines[201] to create a new vocabulary or application profile. When — as an implementing or tendering party — combining different vocabularies that are not in the

---

[201] https://github.com/Informatievlaanderen/OSLO-handleiding/blob/master/Modellering/OSLO-Modelleringsregels.pdf

same governance, it can be ambiguous to define if an information system is compliant. Therefore further research is needed to develop an unambiguous set of rules and maturity levels which allow adequate audit and control records [4]. At the time of writing this dissertation, a working group was started-up to define a compliance framework for data standards in the region of Flanders.

The output variables that are used to evaluate interoperability — including cost reduction, quality of service delivery, technological benefits and improved efficiency — mainly apply to NPM. As society is shifting towards a digital era, in our future research we would like to evaluate the impact of interoperability by using criteria that align on DEG. Possible directions to measure the impact include network simplification, increased transparency for the citizen, no-touch public services and new applications and business-models.

Throughout this dissertation, we debated the implementation of public services, including when analysing the Smart Digital Assistant, which provides a unified user experience for Flemish citizens when consuming public services. In Flanders, we have defined an application profile for public services — such as requesting a subsidy — which extends the European definition of public service. Although we have a standardised model for defining the information — such as the criterion to be entitled to receive this subsidy and the evidence that has to be delivered — the implementation of the algorithms is untransparent. Given an average mix of 250 business application for each local government that support public services, and the fact that criteria of these algorithms are defined in local, regional, and federal legislation, algorithms and software are often outdated. Also, on September 16th 2020, von der Leyen stated in the State of the Union at the European Parliament Plenary that "*We want a set of rules that puts people at the centre. Algorithms must not be a black box and there must be clear rules if something goes wrong. The Commission will propose a law to this effect next year*" (p. 7) [14]. Additionally, there is a link with the GDPR (Articles 13-15) that states data subjects have the right to have insights of how their data is processed by automated or artificially intelligent IS, however the feasibility of this right is often contested in the literature [5, 15]. The cities Amsterdam and Helsinki already publish a registry of algorithms that are used in public services[202]. In the innovation project FAST[203] we explored a method 'OSLO-STEPS' which allows creating modular and machine-readable algorithms. This decentralised approach allows creating an algorithm that is based on decentralised rules that originate from the local, regional, and federal level. As machines can reason on these algorithms, public services can be generated automatically, which will ensure that the IS will be more up-to-date and transparent. Further research is needed to evaluate how we can apply and extend our process and method to develop transparent, decentralised algorithms and turn this dream into reality! Quod erat demonstradum.

9.5 References

[1] Avison, D. E., Lau, F., Myers, M. D., & Nielsen, P. A. (1999). Action research. *Communications of the ACM*, 42(1), 94-97. https://dl.acm.org/doi/pdf/10.1145/291469.291479

[2] Baskerville, R. L. (1999). Investigating information systems with action research. *Communications of the association for information systems*, 2(1), 19. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.982.6795&rep=rep1&type=pdf

[3] Chantillon, M., Crompvoets, J., & Peristeras, V. (2017). The governance landscape of geospatial e-services—The Belgian case. ISPRS International *Journal of Geo-Information*, 6(9), 282-307.

[4] Delbaere, M., & Ferreira, R. (2007). Addressing the data aspects of compliance with industry models. *IBM Systems Journal*, *46*(2), 319-334.

---

202 https://algoritmeregister.amsterdam.nl/
203 https://www.imec-int.com/en/what-we-offer/research-portfolio/fast

[5] European Commission. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46. Official Journal of the European Union (OJ), 59(1-88), 294 (2016).

[6] Jambon, J., *Septemberverklaring van de Vlaamse Regering 2020*. 2020. p. 24. https://www.vlaanderen.be/publicaties/septemberverklaring-van-de-vlaamse-regering-2020

[7] Lewin, K. (1946). Action research and minority problems. Journal of social issues. 2,4, 34-46.

[8] Ouma, F. K. (2014, May). Impediments to interagency statistical information sharing amongst government agencies in Uganda: A G2G adoption. In *IST-Africa Conference Proceedings,* 2014 (pp. 1-11). IEEE.

[9] Provan, K. G., & Kenis, P. (2008). Modes of network governance: Structure, management, and effectiveness. Journal of public administration research and theory, 18(2), 229-252.

[10] Stridsberg, A. B. (1962). Launching Your Product in Belgium. *Journal of Marketing*, *26*(1), 12-18.

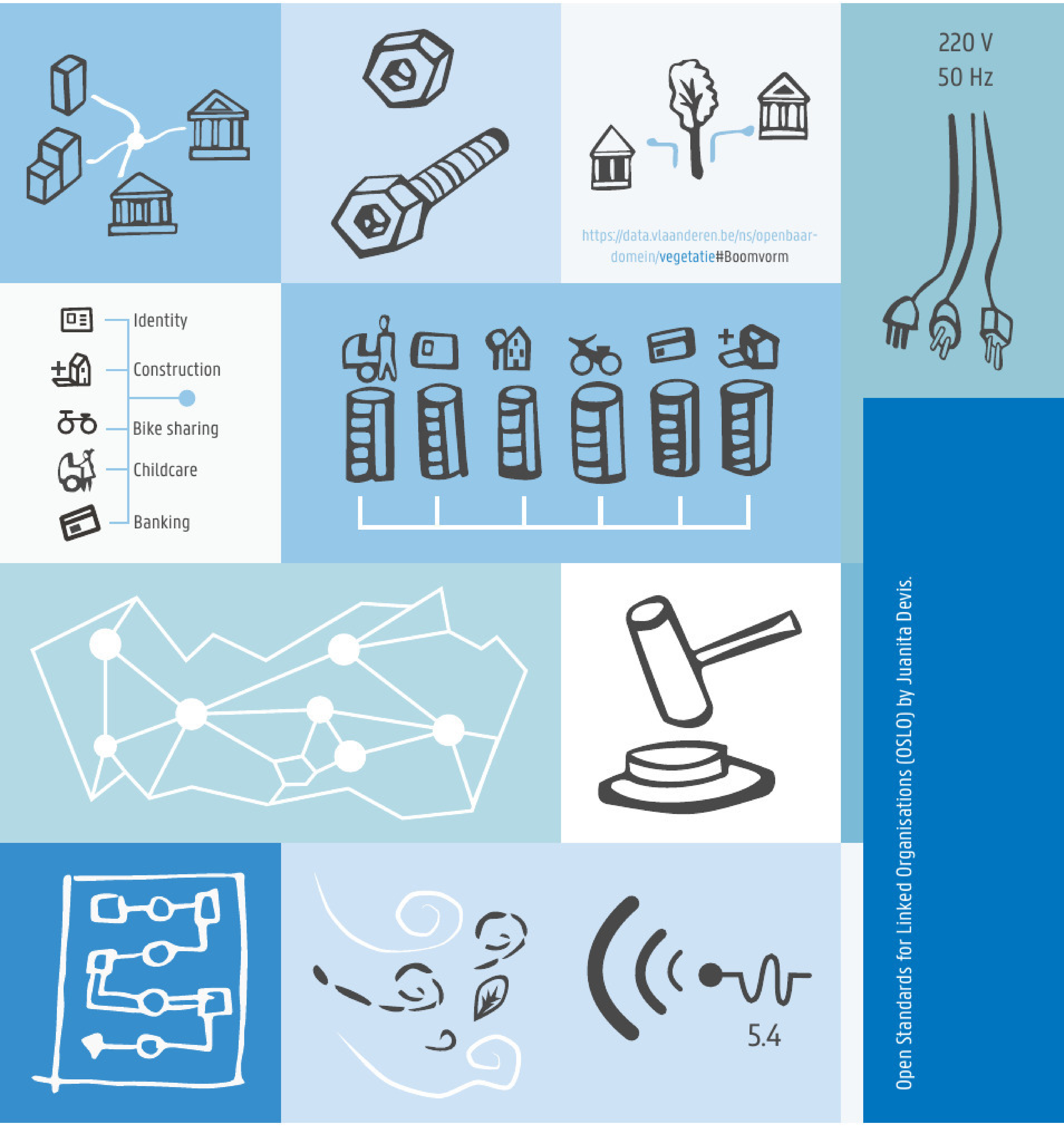
[11] Susman, G. I., & Evered, R. D. (1978). An assessment of the scientific merits of action research. *Administrative science quarterly*, 582-603. https://www.jstor.org/stable/pdf/2392581.pdf

[12] *Volders, V. (2020).* Vlaanderen Radicaal Digitaal II: Lokale Besluiten als Gelinkte Open Data (LBLOD). Unpublished manuscript, Flemish Agency for Local and Provincial Government, Brussels, Belgium.

[13] Volders, V., Buyle, R. (2020). *Semantics at your fingertips How Belgian municipalities create Linked Open data without additional effort.* [PowerPoint slides]. Policy making after shock is digital-ready – Are you ready? https://lokaalbestuur.vlaanderen.be/nieuws/lblod-op-government-after-shock-17-november-2020

[14] von der Leyden, U., *State of the Union Address by President von der Leyen at the European Parliament Plenary*. 2020. p. 7. https://ec.europa.eu/commission/presscorner/detail/ov/SPEECH_20_1655

[15] Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the general data protection regulation. International Data Privacy Law, 7(2), 76-99.



Open Standards for Linked Organisations (OSLO) by Juanita Devis.