FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

Transparenz von Gemeindefinanzen in Österreich durch Linked Open Budget Data

Eine Analyse des Stands der Technik und ein Ausblick für zukünftige Anwendungen

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Wirtschaftsinformatik

eingereicht von

Paul Blasl
Matrikelnummer 0926926

an der
Fakultät für Informatik der Technischen Universität Wien

Betreuer: Univ.-Prof. Mag. Dr. Michael GETZNER
Ass.-Prof. Dipl.-Ing. Dr. Johann BRÖTHALER
Fachbereich Finanzwissenschaft und Infrastrukturpolitik (IFIP)
Department für Raumplanung

Wien, 26.11.2017

Paul BLASL

Michael GETZNER

FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

Transparency of local government finances in Austria through Linked Open Budget Data

An analysis of the state-of-the-art and a prospect for future applications

DIPLOMA THESIS

submitted in fulfilment of the requirements for the degree of

Diplom-Ingenieur

as part of the Master Programme in

Business Informatics

submitted by

Paul Blasl
student number 0926926

at the
Faculty of Informatics, Vienna University of Technology

Advisor: Univ.-Prof. Mag. Dr. Michael GETZNER
Ass.-Prof. Dipl.-Ing. Dr. Johann BRÖTHALER
Centre of Public Finance and Infrastructure Policy (IFIP)
Department of Spatial Planning

Vienna, 26.11.2017

Paul BLASL

Michael GETZNER

ERKLÄRUNG ZUR VERFASSUNG DIESER ARBEIT

PAUL BLASL
KRIEHUBERGASSE 9, 1050 WIEN

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, 26.11.2017 ______________________________

Paul BLASL

DANKSAGUNG

Mein besonderer Dank gilt in erster Linie meinen Eltern, die mir mein Studium und damit diese Arbeit überhaupt erst ermöglicht haben, die mich in all meinen Entscheidungen unterstützen und die Voraussetzungen für das Leben schufen, welches ich heute führen kann.

An dieser Stelle möchte ich auch meinen Freunden herzlich danken, die immer für mich da waren, mir außerordentliche, persönliche Unterstützung zukommen ließen und damit wesentlich zum Gelingen dieser Arbeit beigetragen haben.

Und zuletzt möchte ich auch Herrn Dr. Johann Bröthaler danken, der mir für fachliche als auch inhaltliche Verbesserungen zur Seite stand und nicht zuletzt den gesamten Werdegang dieser Arbeit betreute. Zudem verschaffte er mir die Möglichkeit eine Diplomarbeit in einem für mich außerordentlich interessanten Fachgebiet zu schreiben und gleichzeitig über ein modernes, technisches Konzept eine Verbindung in die Welt der Informatik zu schaffen.

KURZFASSUNG

Durch die technischen Entwicklungen und Fortschritte im Bereich der Informationstechnologien sind die heutigen Möglichkeiten als auch die Notwendigkeit von Transparenz im Bereich der öffentlichen Finanzen größer denn je. Das Ziel, Regierungsbudgets in einer nachvollziehbaren, verständlichen, transparenten und offenen Weise zu veröffentlichen, verursacht nicht nur Anforderungen an die öffentlichen Haushaltsdaten an und für sich, sondern auch an die dazugehörigen Budgetprozesse und -systeme sowieso die zugrundeliegenden technischen und rechtlichen Konzepte. Im Zuge dieser Diplomarbeit wird der diesbezügliche aktuelle Stand der Technik, der Forschung und der international bereits bewährten Methoden auf Open und Linked Data ausgeweitet, um diese gemeinsam auf Gemeindefinanzen in Österreich anzuwenden. Damit soll ein neues Niveau in Bezug auf Transparenz und fachlicher Praxis – verglichen mit dem derzeitigen Stand hierzulande – erreicht werden.

Dementsprechend, wurde ein beispielhafter Ansatz mit Hilfe von neuesten Technologien auf dem Gebiet der Wirtschaftsinformatik konzeptioniert, implementiert und veröffentlicht. Dieser Ansatz verschafft Gemeindefinanzen – durch die Integration und Verwendung der Grundlagen einer übergreifenden Vision eines semantischen Webs – selbst definierte Semantik als auch eigenständige Bedeutung und beseitigt damit bestehende Probleme bei der Veröffentlichung als auch der Verbreitung solcher Daten. Schließlich wurde auch eine qualitative Bewertung dieses Ansatzes durchgeführt, womit sichergestellt wurde, dass das Ergebnis üblichen Standards und Regelwerken entspricht, die sich für solch eine Anwendung als relevant erweisen.

Nicht zuletzt, analysiert und zeigt die vorliegende Arbeit den aktuellen Stand von bereits veröffentlichten Budgetdaten international, aber auch insbesondere in Österreich. Dafür werden alle notwendigen theoretischen Grundlagen in einer umfassenden Art und Weise erörtert, wobei jüngste, relevante Entwicklungen und Literatur beachtet wurden, damit eine aktuelle Perspektive auf die Themen und die zugrundeliegenden Konzepte von Gemeindefinanzen, Linked und Open Data als Ausgangspunkt für den Rest dieser Arbeit gewonnen werden kann.

ABSTRACT

Not until the recent technical developments and progresses with respect to information technologies, the possibilities but also the needs to gain budgetary transparency over governmental finances are greater than ever. The goal to publish information about budgets of governmental authorities in a reasonable, comprehensive, transparent and open way induces challenges not only with respect to the data itself but also to the embedded budgetary processes and systems as well as underlying technological and legal concepts. In the course of this Diploma Thesis, relevant state-of-the-art approaches and international best practices are expanded towards Open and Linked Data for an application on local government finances in Austria. Thereby, a new level of best practise and transparency – compared to the status quo – is created in this part of the world.

Accordingly, an exemplary approach was conceptualized, implemented and made publicly available with latest technologies in this field of Business Informatics. This approach enhances data on local government finances with self-defined semantics and meaning by incorporating and using the basic foundations of the overall Semantic Web vision, and thereby, overcomes present issues regarding the publication and distribution of such data. In the end, a qualitative assessment has been performed as well, in order to ensure that the result complies with common standards and guidelines relevant for such an application.

Last but not least, the present work also shows and analyses the current state of published budgetary data internationally and, in particular, in Austria. Therefore, also all the necessary theoretical principles are introduced in a comprehensive manner while carefully considering recent developments and relevant literature, so that an up-to-date perspective can be gained over the topics and underlying concepts of local government finances, Linked and Open Data – as a basic foundation for the rest of this work.

CONTENTS

LIST OF ABBREVIATIONS

In the course of this work a variety of different relevant abbreviations is defined and used. The most common ones are listed in the table below for convenient look up – including a reference to the respective definition within this work.

Abbr.	Full term	Defined on
CC	Creative Commons	p.52
CSV	Comma-Separated Values	p.29
EB	Extraordinary Budget	p.17
ESA	European System of Accounts	p.4
IBP	International Budget Partnership	p.64
LOBD	Linked Open Budget Data	p.62
LOD	Linked Open Data	p.58
OB	Ordinary Budget	p.17
OBD	Open Budget Data	p.57
OGD	Open Government Data	p.53
OKI	Open Knowledge International	p.48
ÖStP	Österreichischer Stabilitätspakt	p.1
RDF	Resource Description Framework	p.34
SPARQL	SPARQL Protocol And RDF Query Language	p.40
URI	Uniform Resource Identifier	p.31
URL	Uniform Resource Locator	p.31
VRV	Voranschlags- und Rechnungsabschlussverordnung	p.9
W3C	World Wide Web Consortium	p.30
XML	eXtensible Markup Language	p.32

CHAPTER 1
INTRODUCTION

1.1 PROBLEM DEFINITION

Most kinds of communities around the world are organized in some way or another. Usually, these communities developed some sort of governmental authorities, which try to coordinate them to achieve more than every single individual could do on its own. In order to do so, those governments collect money – also known as “taxes” – from their community to achieve those goals. Nowadays, in most parts of the western world, where citizens are allowed to participate in elections of their own government, those governments are facing the dilemma to gain and maintain trust from their community, that they are using the collected taxes for the benefit of all in the most efficient way while the community itself cannot be fully aware of all implications of each single action by the government. Moreover, most governments consist of various different, hierarchically organized bodies, which themselves need to be informed mutually about the activities made by the others and the respective cash flows incoming and outgoing as well as in between them. The idea of opening up governmental data and especially budgetary data of the public sector was born among other things due to this demand for more transparency of governmental activities. Not until the recent technical developments and progresses with respect to information technologies, the possibilities but also the needs to gain this budgetary transparency of governmental finances are greater than ever.

With respect to Austria, since 2005 the directive of the European Parliament and Council on the re-use of public sector information (PSI directive) [1] and since 2012 the Austrian Stability Pact or ÖSTERREICHISCHER STABILITÄTSPAKT (ÖStP 2012) [2] are some of the basic regulations, which try to enforce more transparency and openness for public finances. However, the goal to publish information about budgets of governmental authorities in a reasonable, comprehensive, transparent and open way induces challenges not only with respect to the data itself but also to the embedded budgetary processes and systems. Now, it is insufficient that a municipal budget is just displayed at the town hall. Although, some governmental authorities were already publishing budget data on the Internet, it is mandatory to do this especially in an open way from now on. This so-called “Open Budget Data” does (at least) consist of reusable and machine-readable files. So, PDF files or even proprietary formats like Microsoft Excel spreadsheets are not sufficient anymore.

While the act of publishing data in this way seems to be quite simple, the real challenges arise when it comes to the content itself. While the total expenditure of a town or even a country can be determined quite simple, more sophisticated questions like “How much money does a city spend on local public transport?” are often unanswerable. The problem becomes worse if someone tries to compare the expenditures for local public transport in different municipalities all over Austria or even across the European Union. This is due to the fact that complex budgetary data is heterogeneous organized and published and therefore additionally not really comparable. Although, there exist common standards like UN’s “Classification of the Functions of Government” (COFOG) [3] to classify government’s expenditures, the allocation in those classifications (if appropriate and detailed ones exist anyway) differ from country to country or even from city to city within Austria itself. Besides that, most classifications are not that fine-grained to analyse detailed questions like the one about local public transport anyway. Furthermore, cash flows within the public sector can be quite complex themselves. The Austrian Fiscal Equalization is an extraordinary example for this. In

April 2016, it became publicly known that in only three out of eight Federal States (except Vienna) shares of common taxes were calculated correctly over years, because local authorities could not understand the calculations themselves, so nobody noticed the resulting errors [4, p. 82]. Moreover, the budgetary data of different bodies of the public sector is published (if anyway) on different platforms in different formats under different licences – from the official "Government Finance Statistics" provided by Austrian's statistical office over separate governmental Open Data platforms through to private initiatives. There is not only no common and open data integration between those sources, but also often relevant parts cannot be compared in a semantically correct way as no appropriate metadata is provided to each single dataset to set it in the right context.

As the requirements for real Open Budget Data are manifold and the need for transparency is evident, researches and projects in recent history tried different approaches with the given improving technology. However, issues still persist and are awaiting new and more appropriate solutions.

1.2 AIM OF THIS WORK

The aim of this work is to provide an exemplary approach – on the basis of current state-of-the-art and best practice – for data integration and interconnection of local government finances and the respective developments with respect to Open Budget Data in Austria. This approach should provide solutions to the current, mentioned issues especially concerning transparency and openness. In order to do so, relevant requirements will be worked out and analysed, international best practice and state-of-the-art will be adopted and applied on budgetary data from Austrian municipalities and last but not least, the result will be evaluated according to common standards and guidelines, which will be narrowed down for the given application to reach a new level of best practice focused on the peculiarities of Austria's governmental budget structure.

So, in short, following general questions shall be addressed:

→ Does the current status of Open Budget Data in Austria increase transparency of local government finances?
→ How do worldwide best practices regarding Open Budget Data look like? How can transparency be improved or ensured with that? Is such an approach applicable to local government finances in Austria?

As seen later in this very Diploma Thesis, so-called "Linked Data" as an appropriate technology seems to provide the means to solve big issues of current approaches and is already in use in state-of-the-art applications of governmental Open Budget Data internationally. Linked Data itself is a method of publishing structured data. This data in turn can be interlinked and become more useful through subsequent semantic queries. If the principles of Open Data are applied, this approach is called "Linked Open Data". So, in the course of this work, a method should be found, how this very technology can be applied to governmental efforts towards Open Budget Data in Austria. So, following more detailed question in addition to the ones above shall be addressed too:

→ How can the principles of Linked Open Data be applied on local government finances in Austria?

To limit the scope of this Diploma Thesis in an appropriate way, the elaborated approach will focus on local government finances of municipalities in particular – the governmental levels of Federal States and National State in Austria will be not directly addressed; although, the principles may be applied to them with ease in future works. Furthermore, the mentioned approach will concentrate on the data model, format, creation and its retrieval; not on its actual end-user-friendly representation. There exist already various papers, which focus on topics related to the latter point, and furthermore, it would be – in the author's view – out of scope for this Diploma Thesis as well as business informatics' primary applications in general.

To sum it up, the primary scientific contribution of this Diploma Thesis shall be an exemplary approach – with all the implications and details already explained – that will combine the principles of "Open Data" and "Linked Data" with budgetary data of municipalities in Austria, in order to provide the means for future applications on a new, more transparent, coherent, reasonable, comprehensive and comparable level of implementation to better fulfil overall requirements by current society, governmental institutions and law.

1.3 METHODOLOGY & APPROACH

The methodology to reach these expected results consists in a multi-staged approach:

→ In the first part of this work, theoretical principles relevant for this work will be worked out and analysed. To be more exact, at first the public sector itself, its definition and composition, and Austrian's local government finances in particular, their structure and realization as well as their legal basis are investigated. Already established flows of information regarding budgetary data in Austria will be addressed too at this point. Afterwards, the terms of "Linked Data" and "Open Data" including their derivates "Open Government Data", "Open Budget Data" and "Linked Open Budget Data" are elaborated and current respective general standards and guidelines will be laid out.
All of this background information is established by extensive literature review and by scanning recent scientific papers relevant for the respective topics to gain an up-to-date perspective on them.

→ In the second chapter, the current status of Open Budget Data in Austria will be ascertained and analysed. In parallel, also international as well as implementations across the European Union in particular will be presented and compared with the status quo in Austria. Especially, applications of Linked Open Budget Data will be of interest at this point. As a consequence, an assessment will take place with the goal to work out common best practices and state-of-the-art implementations on the basis of general standards and guidelines found in the first part as well as general relevant requirements.

→ With best practices and state-of-the-art implementations at hand, an exemplary approach for transparent Linked Open Budget Data in Austria was constructed as part of the third and last part of this Diploma Thesis. Based on predefined goals (according to the stated standards and guidelines as well as relevant requirements) a concept and subsequently, the prototype implementation will be presented, concentrating on the data model, format, creation and its retrieval; not on its actual end-user-friendly representation. This prototype focuses on exemplary (already open) municipal budgets with time series data. The underlying concept as well as its prototype implementation itself will be assessed with the predefined goals to measure the accuracy of fit of the created approach and model in the last stage of this part.

At the very end of this work, a short recap over the results will be made including the current state-of-the-art as well as the outlined exemplary approach and possible goals for Open Budget Data especially in Austria. Additionally, open questions and issues regarding the discussed topic shall be addressed too.

CHAPTER 2
THEORETICAL PRINCIPLES

As the topic of governmental budgetary data, its structure as well as its distribution is quite diverse and complex, basic principles and general macroeconomic terms & definitions will be explained within the following chapter. Additionally, technical basics, standards and latest developments about data disclosure and distribution shall be outlined. On the one hand, the reading of subsequent chapters should be simplified as a result. On the other hand, a more sophisticated and differentiated approach for dealing with the topic of this work itself shall be made possible for the reader.

In the **first** part of this chapter, the composition of the public sector, its complexity and also its boundaries will be outlined. This discussion will not be reduced to the eponymous local governments or municipalities, respectively, but will also include the public sector as a whole for a holistic view on the vast complexity and extend of the subsequently discussed topic. The **second** part will address the budgets of governmental authorities themselves, their structure, development as well as their publication and distribution. While this discussion will be focused on the peculiarities of Austria's municipal budget structure, also relevant legal basics about the fiscal management in Austria in general shall be laid out. The **third** part of this chapter will deal with the second essential topic relevant for this work: Linked Data. In this part, technical definitions, principles and mechanics of this interesting field of Informatics necessary for understanding subsequent chapters will be explained. In the **last** part of this chapter, Open Data, derived terms, various relevant principles, standards & guidelines shall be reviewed – addressing the third thematic pillar of this Diploma Thesis in the end.

2.1 THE INSTITUTIONAL FRAMEWORK – THE PUBLIC SECTOR

To get a better understanding on the limitations of budgets of different governmental authorities as well as municipal budgets in particular – what is included and what is not –, the definition of the public sector and its subsectors has to be explained at first.

The primary source for this definition used in this work as well as in the European Union is the EUROPEAN SYSTEM OF ACCOUNTS (ESA). It is a system of national and regional accounts as an internationally comparable framework for describing the total economy of individual communities, different regions, countries, groups of countries and their relations with other economies [5, p. 1]. Its most recent version was officially adopted and released as a regulation [6] by the European Parliament and the council of the European Union in 2013 (as ESA 2010). According to this regulation, since September 2014 all member states of the European Union are obligated to use this revised framework and to conduct all data deliveries to the European Union's statistical office EUROSTAT in accordance with the ESA 2010. This version replaced its predecessor, which was published in 1996 [7] – the EUROPEAN SYSTEM OF ACCOUNTS 1995 (ESA 95). These systems derive from the original EUROPEAN SYSTEM OF INTEGRATED ECONOMIC ACCOUNTS published in 1970 with a second, slightly modified edition appearing in 1978. The ESA 2010 (as well as its predecessors) is fully consistent with the United Nation's SYSTEM OF NATIONAL ACCOUNTS (SNA), which contains worldwide guidelines on national accounting. However, the ESA incorporates certain adaptations for usage in the European Union as well as for its special data needs [8, p. 186].

2.1.1 The composition of the Public Sector

The ESA defines so-called "institutional units", which made up total domestic economies, and thereby provides the means to measure accurately economic activities and to differentiate different parts of economies form each other:

> *"The economy of a country is a system whereby institutions and people interact through exchanges and transfers of goods, services and means of payment (e.g. money) for the production and consumption of goods and services.*
> *In the economy, the units interacting are economic entities that are capable of owning assets, incurring liabilities and engaging in economic activities and in transactions with other entities. They are known as institutional units."* [5, p. 25]

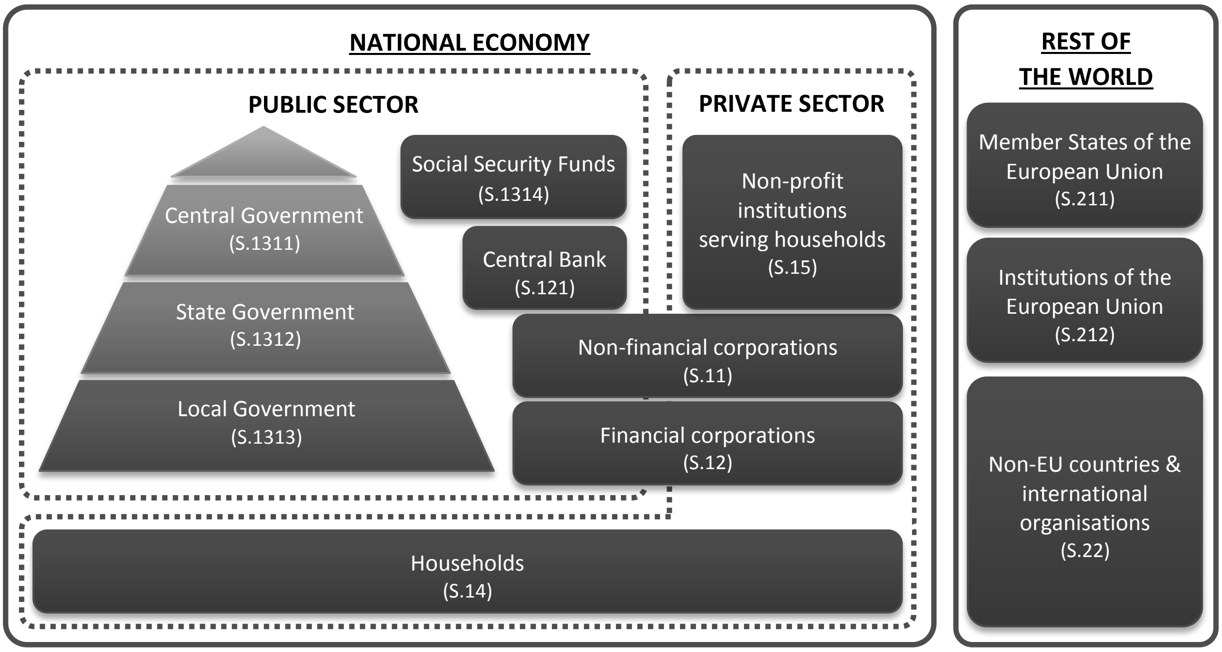
In macroeconomic analysis the actions of each institutional unit individually are not of primary interest, naturally. Instead, the aggregated activities of similar institutions are considered and therefore, these units are categorized by the ESA into groups – so-called "institutional sectors", which can be divided into subsectors themselves in most cases [5, p. 31]. The five primary, mutually exclusive institutional sectors (with their respective sector codes) defined by the ESA [5, p. 12] are:

→ non-financial corporations (S.11)
→ financial corporations (S.12)
→ general government (S.13)
→ households (S.14)
→ non-profit institutions serving households (S.15)

These five institutional sectors in conjunction with their corresponding units make up total domestic economies. With that, the ESA enables (relatively) fine-grained and standardised accounting for each individual sector, its sub-sector as well as the total economy. Additionally, non-resident units are defined, which can interact with these five institutional sectors and their respective units. These non-resident units are grouped into a sixth institutional sector: the rest of the world (S.2).

The most important institutional (sub-)sectors relevant for this work are depicted in **FIGURE 1** with

FIGURE 1 – Overview of institutional (sub-)sectors defined by the European System of Accounts



The codes within the brackets represent the sector codes according to the ESA 2010.
Source: European System of Accounts 2010 [5, p. 31 ff.], own illustration

all major groupings for those sectors. In particular, the public and the private sector as the major groupings of a domestic or national economy, respectively, and their differentiation are of most interest for this part of this work, naturally.

Of course, the ESA provides sophisticated means to distinguish both categories. **TABLE 1** sets out the two general criteria, how the ESA's methodology allocates institutional units to certain (sub-)sectors.

TABLE 1 – General criteria for the allocation of institutional units into ESA's sectors of economy

Criteria	Controlled by government (public sector)	Privately controlled (private sector)
Non-market output	General government	Non-profit institutions serving households
Market output	Public corporations	Private corporations

Source: European System of Accounts 2010 [5, p. 8], own illustration

As depicted, on the one hand, it distinguishes, whether a unit is controlled by the government or not, so whether it belongs to the public or private sector, respectively: The government controls a single unit by definition, if it owns "more than half the voting shares or otherwise controlling more than half the shareholders' voting power" [5, p. 33] – in other words, if it is able to determine general corporate policy and has major influence on all economic activity of an institutional unit (e.g. by appointing a new management). This form of control is called "dominant influence by virtue of ownership" or – in short – simply "dominant ownership". According to the ESA 2010, it is explicitly not necessary for this that the government holds the voting share or the voting power directly[1]. This would be called "indirect interest" or "indirect control".

On the other hand, the ESA differentiates between market and non-market producer. This is done primarily via the so-called "50% criterion"[2]. As per ESA 2010, a market producer has to cover at least 50 percent of its production costs[3] by its sales[4] over multiple years [5, p. 57]. If the institutional unit is a non-market producer, which is controlled by the government, it is generally considered as part of the general government sector. Market producers in turn are classified as non-financial and financial corporations [5, p. 60]. So, for example, the central bank of a country is seen as a market producer and financial corporation controlled by the government and is therefore part of the public corporations sector as well as the general public sector (cf. **FIGURE 1**).

With that, a clear differentiation of the public sector should be possible: The public sector consist out of institutional units which are mainly controlled by the government. It includes both non-market producers as well as market producers. So, even if a corporation is financially independent of the government, it is considered as part of the public sector, if the government controls it – although it is

[1] If corporation A is mainly owned by corporation B (by – at least – the half of A's voting shares), a third corporation C, which owns more than half of the voting shares for corporations B, also controls corporation A indirectly by definition.

[2] Additionally to this quantitative criterion, the ESA 2010 includes also qualitative criteria for differentiation of institutional units between market and non-market producer. However, this primarily affects units, which make most sales only with the government.

[3] For the purpose of this criterion, the ESA 2010 defines [5, p. 59] production costs as "equal to the sum of intermediate consumption, compensation of employees, consumption of fixed capital, other taxes on production plus costs of capital. Other subsidies on production are not deducted." The costs of capital (which can be dedicatedly approximated by the net actual interest payments) were added explicitly in this definition for the 2010's revision of the ESA.

[4] With respect to this definition by the ESA 2010 [5, p. 59], sales means turnover excluding taxes on products while including "all payments made by general government or the institutions of the Union and granted to any kind of producer in this type of activity" – this corresponds to the definition of output at basic prices.

categorized into a different sub-sector than the general government. However, as interested readers might already suspect, in reality the described distinction is quite complex and boundaries become rapidly blurred. Having that said, these definitions also give a good overview about the vast scale of the public sector in general, which will be explicitly shown for Austria in the following, as well as about all players within and outside of an economy, which can interact with each other, from the perspective of each individual institutional unit.

2.1.2 The Public Sector in Austria

The Finance Statistic Regulation or GEBARUNGSSTATISTIK-VERORDNUNG (GebStat-VO)[5] in its 2014's revision obliges the Austrian's statistical office STATISTIK AUSTRIA to publish a list of all institutional units, which are associated with the public sector by the definition of the ESA 2010, annually not later than March 31st [9]. Therefore, liabilities, operating result, headcount, if applicable, shareholding of the government (in percent) and for all public corporations reporting liabilities higher than 0.01% of the GDP (for Austria, this means about 30 Mio. Euro), time series data about the 50% criterion[6] are gathered from all primary institutional units of the general government[7], the units they are controlling as well as all respective controlled sub-units [10, p. 3].

The mentioned report about all institutional units allocated to the public sector for 2016 is shown in aggregated form in **TABLE 2**. Austrian's capital Vienna – fulfilling both roles of state as well as local government – and its controlled units are assigned to the latter sector (S.1313).

TABLE 2 – The number of institutional units of the public sector in Austria by March 2016

Subsectors / controlling subsector	S.11 2)	S.12 3)	S.13 4)	Total
Central Government (Bund)	239	10	257	**506**
State Government (Länder)	402	22	319	**743**
Local Government (Gemeinden) 1)	2,147	6	4,302	**6,455**
Social Security Funds	5	2	60	**67**
Total	**2,793**	**40**	**4,938**	**7,771**

1) Numbers by municipalities (incl. Vienna) as of January 2016
2) S.11 = Non-financial corporations (those ones counted among the public sector)
3) S.12 = Financial corporations (those ones counted among the public sector)
4) S.13 = Institutional units allocated to the general government (or public non-market producers)

Source: Statistik Austria [11], own illustration

The data in **TABLE 2** is presented like this: 506 units known as part of the public sector are controlled by the central government in Austria by March 2016. 239 of them are non-financial corporations (e.g. the public service broadcaster ORF and its divisions), 10 financial corporations (e.g. the central bank) and 257 institutional units controlled by the *central* government are seen as part of the *general* government (e.g. the statistical office STATISTIK AUSTRIA itself). Furthermore, Austrian's government controls 2,793 non-financial corporations, which are in fact market-producers.

Additionally, the data in **TABLE 2** shows that 6,455 of 7,771 or 83% of all public institutional units are controlled by the local governments. Or in other words, 6,455 units of the public sector in Austria are the 2,100 Austrian municipalities themselves, or numerous corporations controlled by them. This emphasises the complex structure of the public sector, especially in the local government sector – even in a small country like Austria. This complexity will be addressed by this work specifically.

---

[5] More information on Austria's Finance Statistic Regulation will be given in chapter 2.2.6.
[6] The so-called "50% criterion" differentiates market producers form non-market producers (see also chapter 2.1.1).
[7] These are with their respective sector codes according to the ESA 2010 (cf. **FIGURE 1**): the central government (S.1311), state governments (S.1312), local governments (S.1313) and social security funds (S.1314).

2.2 THE MUNICIPAL BUDGET IN AUSTRIA

The public fiscal management is a necessary theoretical foundation to approach and understand the topic elaborated by this Diploma Thesis. Therefore, public budgets, their structure as well as basic regulations and general principles related to fiscal management in Austria will be addressed in the following sub-chapters. To limit the scope in an appropriate way, the fiscal management of municipalities in particular is of primary interest – the specifics on the governmental levels of Federal States and National State in Austria won't be directly addressed; although, the same (or similar) principles apply also to those parts of the public sector.

2.2.1 Principles of Fiscal Management of Local Governments

In their role as an institutional unit, local governments have their own individual fiscal management by definition. Although the ESA declares local governments in general, their country-specific constitution is made up by national law. In Austria, local governments in form of municipalities are defined by the Federal Constitutional Law or BUNDES-VERFASSUNGSGESETZ (B-VG). According to the articles 115 and 116 B-VG [12] municipalities are seen as local congregations that are self-governed and autonomous economic bodies. Those articles also state that every municipality is part of a federal state and all territory in turn is part of some municipality.[8] With that, the distinction and definition of local governments in Austria is quite precise and clear. However, with that the Austrian regulations imply also uniform municipalities [13, p. 58]: Deduced from the B-VG, all municipalities are organized basically in the same way – regardless whether they are industrial cities or agricultural villages. Obviously, this is not true for all local governments, like – for example – Vienna, which is fulfilling both roles of state as well as local government, or other cities with their own statute[9], which are treated legally differently in some aspects. So, this definition is also called "the fiction of uniform municipalities". Nevertheless, this definition causes all basic regulations including overall principles of fiscal management to apply to all municipalities equally – with all the resulting advantages and disadvantages.

With that in mind, the B-VG constitutes in Article 13 basic principles for fiscal management in Austria: Paragraph 1 states that all taxation regulations are issued by the central and the respective state governments based on an own constitutional regulation, the Fiscal Constitutional Law or FINANZ-VERFASSUNGSGESETZ (F-VG 1948), which lays the foundation for all subsequent laws that treat fiscal management of local governments in particular. Additionally, Paragraph 2 states that the central, state and local governments together shall ensure with their fiscal management the balance of the total economy as well as sustainable budgets.

As with most definitions of the B-VG, it does not give any quantitative criteria for those aims. However, the actual assessment of budgets on this foundation and their comparison with these basic principles is up to audit by different institutions, eventually – this part of fiscal management will be treated in more detail later in sub-chapter 2.2.2.

Anyway, those qualitative aims are extended even further in the Federal Constitutional Law later in Article 119a, paragraph 2, B-VG. Through this regulation for auditing municipal budgets by Federal States in general, more detailed principles can be deduced; they are: thrift ("Sparsamkeit"), efficiency ("Wirtschaftlichkeit") and expediency ("Zweckmäßigkeit"). As these terms are recurring principles in

---

[8] This means, that there are *no* unincorporated areas in Austria by law – in contrast to other countries (e.g. Germany).

[9] Municipalities with a population over 20,000 *can* request their own statue according to Article 116, paragraph 3, B-VG. Currently, there are 15 of such cities in Austria: Eisenstadt, Graz, Innsbruck, Klagenfurt am Wörthersee, Krems an der Donau, Linz, Rust, Salzburg, St. Pölten, Steyr, Villach, Waidhofen an der Ybbs, Wels, Wien and Wiener Neustadt.

various Austrian laws dealing with financial management, they are explained in the following in the context of fiscal management [14, p. 50]:

→ *Thrift* means that the use of funds should be restricted to the absolute minimum.
The question is whether a matter is actually necessary.
→ *Efficiency* means that something is reached for the intended purpose with the best possible outcome on the condition of minimum use of funds.
The question is whether a matter is worth the money.
→ *Expediency* means that a matter can be used for the intended purpose.
The question is whether a matter brings the planned benefits.

Although, these principles seem to be quite obvious and reasonable (especially, considering public finances build up on taxes payed by all of us), they are often violated nevertheless, resulting in partially desperate financial situations, where governmental bodies reach the edge of bankruptcy. Therefore, those principles are a central part for auditing governmental budgets.

On top of these basic principles and regulations regarding fiscal management, there exists a broad range of budgetary laws or laws relating to public finance in Austrian's legislation. One of the most important ones of those regulations for this Diploma Thesis is the Budgeting and Accounts regulation or VORANSCHLAGS- UND RECHNUNGSABSCHLUSSVERORDNUNG (VRV), which is based on §16, paragraph 1 F-VG 1948 [15]. The latter one states that the Federal Minister of Finance (in accordance with the Austrian Court of Audit) can govern budgets and statements of account for the sake of consistency. Therefore, the VRV regulates in detail a uniform structure for budgets and accounts for all territorial authorities (or primarily Federal States as well as municipalities) from a functional and economical point of view, which facilitates a holistic, fiscal and economic analysis of such reports. The VRV itself goes back to the beginnings of the Second Austrian Republic. On 9th July 1949, directives on the creation of budgets and statements of account for Federal States, municipalities and associations of local authorities were issued by the Federal Ministry of Finance, which were first applied for the fiscal year 1950 [16, p. 7]. Those directives as a voluntary agreement were replaced by the first version of the VRV as a mandatory law in 1974. Since then, this regulation is negotiated between the Association of Towns and Municipalities as well as all other governmental bodies and subsequently, was enacted in identical form by the Federal Ministry of Finance multiple times – law on order, so to say [14, p. 62] – with their most recent implementations called "VRV 1997" and "VRV 2015" [17].

Additionally to the general principles for fiscal management by the B-VG, the VRV is aware of more specific budgetary principles. These are in aggregated form [16, pp. 9-15]:

→ *Completeness* (including accuracy and truth)
→ *Temporal and factual commitment* (of target and actual values)
→ *Annuality* (including maturity and prior approval)
→ *Publicity* (in all parts of the budget cycle)

Due to their importance, these terms shall be discussed in more detail in the following at this point.

Completeness

All settlements (all revenues as well as expenditures) must be budgeted and accounted completely, unabridged and unbalanced with their total gross value. Additionally, they must be budgeted and afterwards accounted individually. As a consequence, nothing can be done without preliminary planning of each single matter. This policy includes also an obligation for "non-declaration". So, if there was e.g. no build up in reserves, it is not allowed to simply omit this position in a budget with a note that there is nothing to declare – it has to be present anyway, reported with a declared zero.

Temporal and factual commitment

According to the principle of temporal and factual commitment, the government is bound to the budget in two forms:

→ Usage of funds only in the given period (as *temporal commitment*)
→ Usage of funds only for the given matter (as *<u>qualitative</u> factual commitment*) at the given amount (as *<u>quantitative</u> factual commitment*)

Although there exist exceptions[10] for these commitments, funds cannot be used in another period as budgeted or for other matters, if there is no resolution that does not state something else.

Annuality

All governmental authorities have to budget their estimated revenues and expenditures for each single calendar year prior to its actual beginning. For each of these periods also statements of account have to be issued afterwards. So, the fiscal year according to the VRV corresponds to the calendar year. All revenues and expenditures mature not later than the 31st December must be included in the same fiscal year. Maturity in this context covers all claims and liabilities resulting from deliveries and services, which were performed by no later than the end of the respective fiscal year.

Publicity

The principle of publicity states that all stages in the budget cycle of a governmental authority should happen in public in a democratic community. Hence, public budgets and statements of accounts have to be approved in public sessions of the respective governmental body and be on display for the public, subsequently. The most recent amendment of the VRV, the VRV 2015 [17], even states that budgets and statements of accounts must be published accessibly in the Internet (Article 6, paragraph 10 & Article 15, paragraph 4).

Especially, the latter principle in a different form will be very important to this Diploma Thesis later on. In chapter 2.4.3, the theoretical principles about the disclosure of governmental budget data specifically will be discussed in detail, accordingly.

Although even more principles can be deduced from the VRV and similar laws, the discussed ones should give an appropriate overview of the foundations of fiscal management of municipalities in Austria for the rest of this Diploma Thesis. In any case, all of these basic budgetary principles in Austrian's legislation emerged over time and have established in fundamental law since then. They provide not only the means for budgeting and accounting in a proper way, but also for efficient, suitable and impact oriented budgetary governance [16, p. 9].

Last but not least, it should be pointed out that the so-called "generally accepted accounting principles"[11] in the private sector apply also to public finances in general [18, p. 82]. However, as they are equivalent in some parts to those principles already discussed and further details do not really concern the rest of this Diploma Thesis, there will be no further discussion at this point.

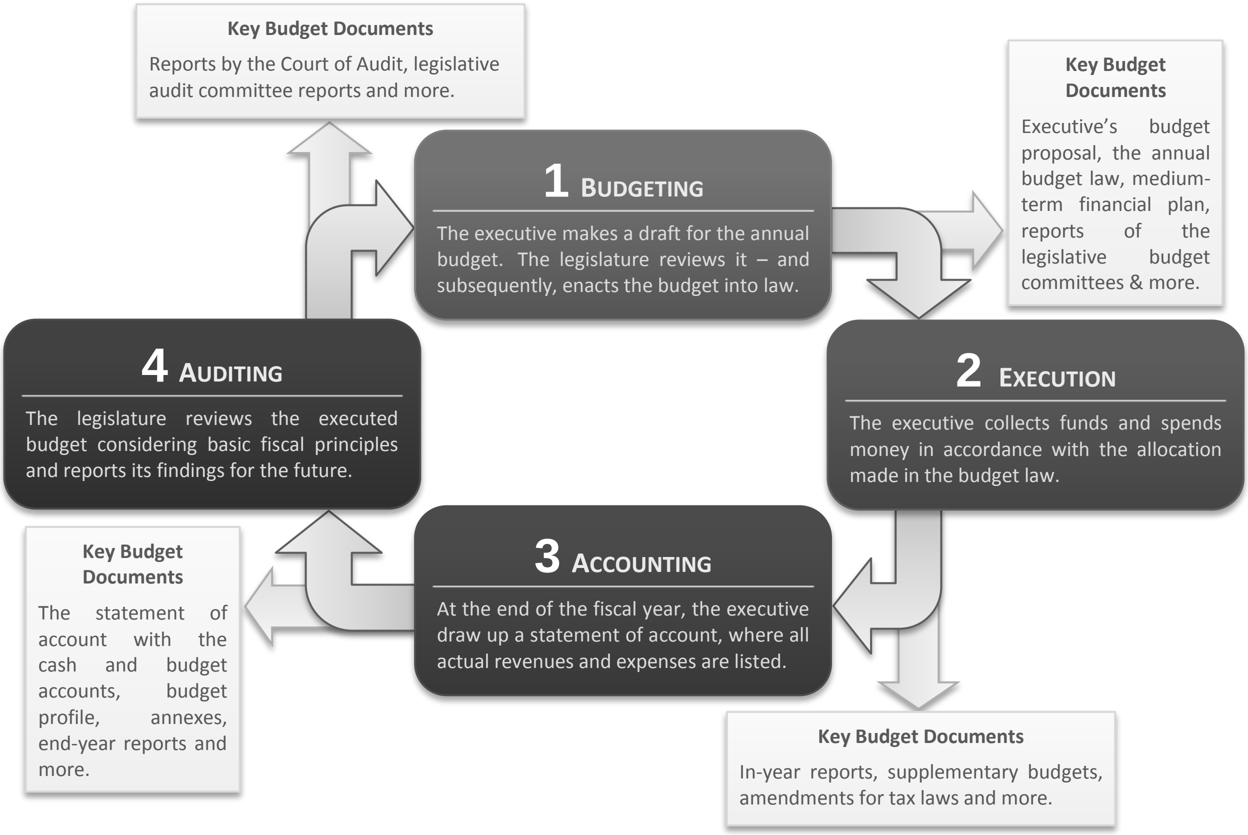
---

[10] The VRV states explicit possibilities for exceptions to the principle of temporal and factual commitment e.g. with loan transfer or excess, or supplementary budgets. However, they are *always* subject to authorisation by the local government.
[11] These principles are codified in the Austrian Commercial Code or UNTERNEHMENSGESETZBUCH (UGB), §190 ff..

2.2.2 The budget cycle in general

The process of budgeting and accounting for governmental authorities built on the fundamental principles discussed in the previous sub-chapter 2.2.1 is extensively specified and regulated by Austrian's VRV and related laws. However, this overall budget process is not specific to municipalities or to Austria's governmental authorities in general. In fact, it can be found in most countries and their respective territorial authorities. For a better understanding of this process, often the model of a so-called "budget cycle" is used [16, pp. 18-24]. A depiction of this model with short explanations and the most important outputs of each single step can be found in **FIGURE 2**.

FIGURE 2 – The different steps of the general budget cycle in the public administration



Source: VRV [17], Klug [16, pp. 18-24], Lucke et al. [19, p. 2], own illustration

This budget cycle simplifies the overall process in a reasonable way, on the one hand, but explicitly shows the constant repetition of budgetary activities over multiple fiscal years, on the other hand. This is of the utmost importance: All process steps are repeated in a similar way every fiscal period, not only for the sake of transparency and comparability, but also to prevent any (possibly large-scale) errors from happening in the first place by providing a comprehensive guideline for the overall fiscal management of public bodies under general accepted and applicable principles.

Although, the single steps of this process are not of primary interest for this work in particular and are not discussed in detail subsequently, some of their artefacts or outcomes are important documents with data, which will be in the limelight in the course of the rest of this Diploma Thesis. As these documents as well as their respective data and their specifics – regulated for municipalities by the VRV – are important for understanding subsequent parts of this work, the following discussion will concentrate on the budgetary documents and regulations as part of the general budget cycle for municipalities in particular.

Budgets

A budget is a report about a financial plan, where scheduled projects are depicted with all of their financial consequences. The budget is usually made in the first step of the budget cycle and prior to the applying fiscal period (cf. FIGURE 2). These and more principles (as discussed in chapter 2.2.1) are written down for municipalities in the VRV. Accordingly, the budget for the next financial year has to be drafted by the local government and after that, enacted into law by the local council in advance. Therefore, it is a legal binding basis for the execution of the budgeted revenues and expenditures of the local administration [20, p. 8].

The planned revenues are understood as minimum levels and the expenses as maximum amounts. The actual payment and intake orders are made with respective documentations and in accordance with the two-man rule [16, p. 19].

Medium-term financial plan

Besides the annual budget, all municipalities are obliged to make a medium-term financial plan since 2012[12] according to Article 15 ÖStP 2012 [2]. Under the terms of its second Annex, such a plan includes (among other general fiscal key figures) all expected revenues and expenditures for at least five fiscal years. With that, weaknesses and shortcomings of financial planning that solely relies on annual budget planning should be overcome [21, p. 29]. A unique fiscal plan for each single year leads inevitably to a temporal as well as factual isolated point of view for matters and it is no suitable instrument for taking into account medium- or long-term effects. In fact, it misleads the public administration to see revenues, expenditures as well as investments as isolated events ignoring possible consequences. Therefore, the mandatory medium-term financial plan has the task to coordinate and inform optimally about investments and their impacts in factual, temporal and financial respects. This should insure the optimal use of limited funds (cf. with the principles of thrift, efficiency and expediency in chapter 2.2.1).

Statements of account

A statement of account – under the terms of the VRV – is a report about the realisation of a financial plan, whether and to what extent a municipality has performed its duties in accordance with the preliminary budget [14, p. 68]. All accounting must be covered by this respective budget. Therefore, all accounts are listed with their target values (originating in principle from the respective budget) as well as their actual values (originating from the actual accounting). So in fact, a statement of account illustrates the finances dually. Additionally, all principles applying to budgets also apply to statements of account.

The local administration has to make this report immediately, but in most cases no later than three months after the end of each fiscal year – the actual timing is established by the respective municipal code of each Austrian Federal State. A statement of account consists of the cash and budget accounts as well as multiple supplements and annexes – like asset and liabilities listings, but the most important one by far is the budget profile. The details of the structure of a typical municipal statement of account (as well as a budget in general) will be examined in chapter 2.2.4.

Last but not least, this annual statement of account with a report of the audit committee has to be reviewed by the local council in a public session each year.

---

[12] The respective regulation became applicable for all municipalities, the Federal States (except Salzburg) and the central government on 1st January 2012. Later, on 1st January 2013 this regulation became also applicable for Salzburg [138].

Budgetary audit

After the execution and respective accounting of a fiscal year within a municipality, the respective budget and statement of account is up to audit by the superior legislature and executive. For municipalities in Austria, the right to do so is established by the B-VG in multiple articles.

Article 119a, paragraph 1, B-VG states that, in general, the central as well as the state government practise supervisory legislation for all municipalities; paragraph 2 establishes the right for budgetary audit according to the principles of thrift, efficiency and expediency by the Federal States. The results shall be communicated to the local government and the mayor has to inform the supervisory authority about all actions based on this report within three months. Furthermore, according to paragraph 4, the respective supervisory authority has each and any right to audit all municipal matters; the local government has the obligation to give all necessary information anytime.

Additionally, Article 127a, B-VG establishes the right for the Court of Audit to monitor the fiscal management of local governments and associated institutional units. However, the Court of Audit can carry out such checks on its own initiative only for bigger municipalities. For a long time, it was only authorized to do so for municipalities in Austria with a population of at least 20,000 citizens. Since 2011, the Court of Audit is permitted to audit on its own behalf all municipalities now with only a population of at least 10,000 citizens due to an amendment of the B-VG [22]. This extended its responsibility from 24 to 71 municipalities [23, p. 45] – or to about 37.6% of the total budget volume of all Austrian municipalities. Furthermore, the Court of Audit is also able to become active for municipalities with a population under 10,000 citizens upon request of the respective Federal State's legislature or executive according to Article 127a, paragraph 7 & 8, B-VG. However, such a request can only be made two times a year and only with sufficient justification. Paragraph 5 states again that the results of this audit shall be communicated to the local government and the mayor has to inform the Court of Audit about all actions based on this report within three months. The Court of Audit shall report in turn the results of this overall process to the federal as well as central government and make them public afterwards (Article 127a, paragraph 6, B-VG).

But what if a local government ignores all principles and concepts discussed so far and the audit delivers a damning indictment, e.g. over years? What are the consequences of the violation of the budget cycle in this case? First of all, it is important to state out that the budget is a municipal law, for instance like a short-stay parking regulation, and municipalities as such are self-governed and (especially) autonomous economic bodies by law (see chapter 2.2.1). Therefore, "nobody" can *punish* the local government due to the violation of its own regulations. However, as the local government is democratically legitimised, the superior legislature allows the public to vote their local government out of office, which should be the primary penalty for all persons in charge [14, p. 76].

However, the respective municipal codes offer the possibility to appoint a government commissioner in case of a "lack of power to act". Article 119a, paragraph 7, B-VG establishes the foundation to do so. With that in mind, the supervisory authorities or the Federal States and their governments, respectively, are able to dissolve a local council and suspend the mayor from office. According to the jurisdiction of the Austrian Supreme Administrative Court [24], the *assumption*, that the local government is incapable of acting, is sufficient for this step. On top of this, those circumstances have *not* to be proven by the supervisory authorities. After that, the government commissioner takes over all *necessary* official duties from the local government. Such steps have been performed a few times in the past. For example, the local government of the Lower Austrian municipality Deutsch-Wagram has been removed from office this way in 2009 due the repeated violation of the budget cycle [25].

2.2.3 Definitions & differentiations for municipal budgets

Throughout the general budget cycle, there are a lot of budgetary interdependencies by definition. However, elements reoccurring in the annual budget as well as the respective statement of account have their own interdependencies and some overall concepts and definitions have to be differentiated to understand such reports over the whole budget cycle in their entirety. Therefore, the most important ones of these concepts for Austrian municipalities shall be discussed in the following. Later in this Diploma Thesis, these concepts will be addressed regularly.

Sources & usage for municipal finances

Before looking into the actual structure of municipal budgets, the sources and the usage for local government finances shall be explained very briefly. First of all, municipalities finance themselves via various streams of revenues, which can be grouped into four bigger categories [26, pp. 41-53]:

- → *shares of common taxes* (levied by the central government)
- → *own taxes* (taken by the local government itself)
- → *fees and charges* (for municipal services, e.g. utility fees like for waste & water disposal)
- → *intergovernmental transfers* (from other public bodies and levels of government)

Regardless from where we are in Austria, shares of common taxes make up the biggest part of all municipal revenues – overall, with more than 30%, seen nationally [26, p. 38]. These shares originate – as their name might suggest – from the biggest common taxes taken by the central government; above all: Value Added Tax (VAT), Income Tax, Wage Tax, Corporate Income Tax and Mineral Oil Tax. These centrally generated tax revenues are distributed afterwards to Federal States and municipalities. The legislative basis for this distribution is established in the F-VG 1948 (see also chapter 2.2.1) and the subsequent Fiscal Equalization Law or FINANZAUSGLEICHSGESETZ[13] (FAG).

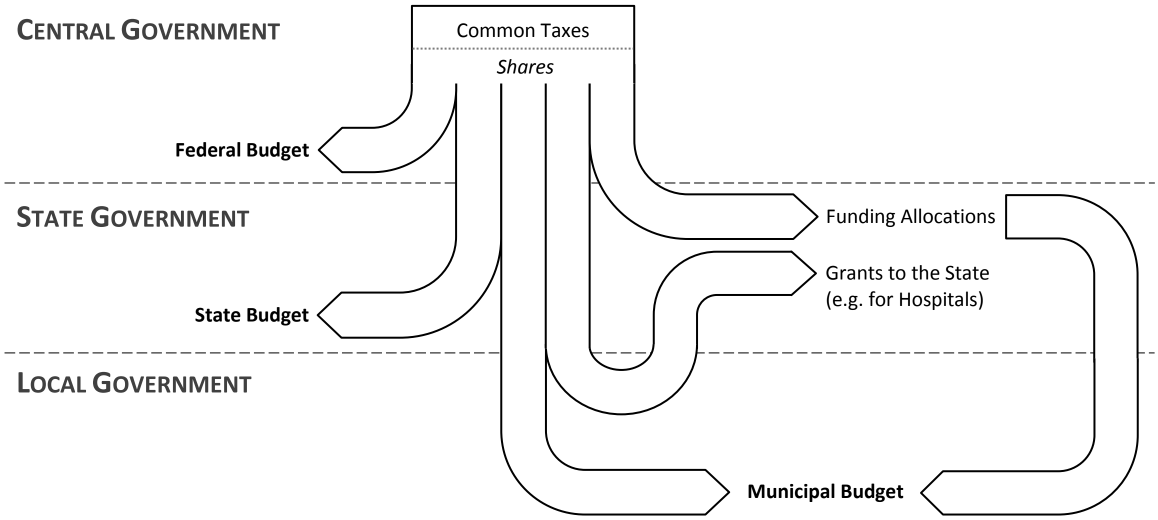
As the very details of these respective laws and this distribution and are quite complex[14] and are also not of primary interest for this work, only a short overview [14, pp. 54-60] shall be given in order to understand certain elements of budgetary documents recurring in this Diploma Thesis: The tax revenues taken by the central government – whose amount is of course dependent on economic trends – are distributed to Federal States and to municipalities in particular via the number of citizens, which have their primary residence within the respective community (secondary residences are *not* included). The more people live in an area, the more revenues it will get. It is important to note, that the general share of common taxes a municipal receives is *not* depending on actual performance or so. However, these “shares of common taxes” are handled for all municipalities (except Vienna) by their respective Federal State, who withholds a certain amount of this money[15] for later allocation. These “funding allocations” are made for those municipalities, which communicate demand for specific matters. Besides that, some of this money received by municipalities as a share of common taxes has to be transferred to the State Government right away for fulfilling duties in the name of the respective local government – these transfers are called “grants” to the State Government. The best-known examples for such a transfer are grants to the

[13] The FAG only applies for a certain period of time; since 1989, its validity is limited to four years. However, if no new FAG is enacted, the old one – according to current legislation – still applies until the legislature passes a new law. The current version of this law – the FAG 2017 – was enacted in December 2016 and applies until the end of 2021 (§31, FAG 2017 [27]).

[14] In fact, the FAG is one of the most complex fiscal laws in Austria, if not the most complex one. The Fiscal Equalization in Austria has evolved over decades and is burdened with a lot of background stories, which are necessary to know and to make any sense and get any context out of the pure text of this law. Attempts to revise the law and its regulations in a more comprehensible way failed in the past. So, today only a few experts understand the regulations to their full extend and need to answer specific questions by public bodies, how to proceed in certain situations [14, pp. 53-54].

[15] According to Article 12, paragraph 1, FAG 2017 this is 12.8% of all funds for municipalities calculated per Federal State.

FIGURE 3 – General abstract flow of tax revenues by the central government according to the Fiscal Equalization Law



Source: FAG 2017 [27], Parrer [14, p. 54], own illustration

respective state for its hospitals. The reason behind this is – as the whole Fiscal Equalization – very historical and quite complex too.[16] An overall summary of this distribution of common taxes is depicted in **FIGURE 3**.

Nevertheless, the remaining municipal revenues are taken by the local government itself. Besides some taxes (e.g. municipal and property taxes) that have to be paid directly to the municipality by all citizens or companies equally, the local government also collect fees and charges for actual services (e.g. utility fees, short-term parking fees). In this sense, the superior legislative allows the local government to request some sort of retail price for certain duties, which it has to fulfil anyway – however, there has to be some legal legitimisation and also some local regulation, which allows it to do so. Also, the amount of money that can be claimed this way is normally limited by superior laws. Besides that, the local government can generate revenues from normal economic activities. The municipality can provide services – like a cooperation – and sell it to the people on the free market. These fees through usage of optional municipal facilities and services (e.g. admission or entry fees to the public swimming pool, sports field and so on) have to be distinguished from all other charges and fees due to legal legitimisation.

Municipalities know also various other, but rather insignificant sources of income. Most relevant are usually transfers of capital from other public bodies (e.g. for special projects by the European Union or to other municipals for special cooperation or syndicate services). Excluded so far, were also revenues from capital reserves and borrowings – so, financial assets – as well as sales of movable and immovable property. However, as those assets have to be built up or paid through the other sources of income, they are (or should be) *no primary* source of revenue – rather a *secondary* and *temporary* one – and have been excluded on purpose so far. However, financial flows between the total budget and such assets will be discussed later in this chapter in more detail.

Last but not least, all of these revenues are used for all kinds of municipal matters. Most of them are in the (assigned) “sphere of activity” of a municipality in a way or the local government is simply the cost bearer for certain governmental activities[17] as direct actor anyway. Most of the duties a local

[16] For example, there is a general grant to the state governments since 1948, which was introduced as a compensation for the foundation of the first (!) republic 1918 (e.g. property taxes were a state matter at this time) [14, p. 60].

[17] The Austrian Constitutional Court has declared that certain costs (as personnel and material costs) due to assigned activities have to be paid by the public body that acts (e.g. the municipal for overtime or running costs due to general elections) [14, pp. 58-59].

government has to perform are regulated in some way or another. However, in general these functions and respective expenses can be categorized into two types [18, pp. 100-101]:

→ *compulsory expenditures* and
→ *discretionary expenditures*

The first ones result from legal regulations and obligations (e.g. education spending, infrastructure costs or housing subsidies), the others out of the self-conception of a municipality, which tries to provide optional services for the benefit of the community by its own choice (e.g. public swimming pools or libraries). There are no encompassing or even always valid lists which state out what municipal expenditures are actually compulsory or discretionary [14, p. 59]; however, relevant literature[18] tries to make approximate classifications.

However, municipal expenses can be also distinguished by their type. This view – in contrast to the functional view discussed so far – can be seen as an economical classification. Common types of expenses for municipalities are, in decreasing order of their general magnitude [26, pp. 39&54-68]:

→ operating and administrative expenditures
→ personnel costs (including salaries for elected bodies and pensions)
→ transfer payments (to other public bodies)
→ investments in movable and immovable property
→ debt service (including interest rates)

Obviously, there is some sort of overlapping concept between these two discussed classifications of expenditures. So, also (sub-)types of costs can be distinguished into compulsory or discretionary. This overlap as well as this dual view on expenditures – functional versus economical – (that can be also applied in a similar way to revenues) will be of interest in chapter 2.2.4.

Differentiation of municipal revenues & expenditures

All these revenues and expenditures – in general, every single cash movement in a municipality, respectively – has to be accounted accordingly. However, certain matters, which do not affect the municipal budget as such, have to be treated differently. These matters are sorted out from the annual budget. This is called "budget ineffective accounting", in contrast to "budget affecting accounting", which covers usual activities of a local government. Budget ineffective matters are revenues and expenditures that are not finally taken by a public body; instead, they will be settled with third parties. These include [16, pp. 23-24]:

→ *Transitory Funds* (revenues taken for third parties, e.g. VAT)
→ *Advancing Funds* (expenditures covered by third parties, e.g. input tax)

As such matters do not affect the preliminary budget by definition, they belong also to the so-called "transitory accounting" and need not be budgeted in the first place [14, p. 63]. However, the budget ineffective accounting must be handled as soon as possible – within the same fiscal year, in the best case – and appears in statements of account as actual values within summary accounts – without target values, of course. Also, matters, whose correct budget account is not known (yet), can be accounted as budget ineffective temporarily [16, p. 23].

[18] In works by Parrer [14, pp. 119-122] or Hiebler [18, pp. 100-101] such classification are made referencing regulated cost approaches. These approaches as functional classes for budget items will be addressed in more detail in chapter 2.2.4.

Furthermore, the budget affecting accounting – so, all revenues & expenditures, which are taken by a local government eventually – is differentiated into ordinary and extraordinary budgets [21, p. 10]:

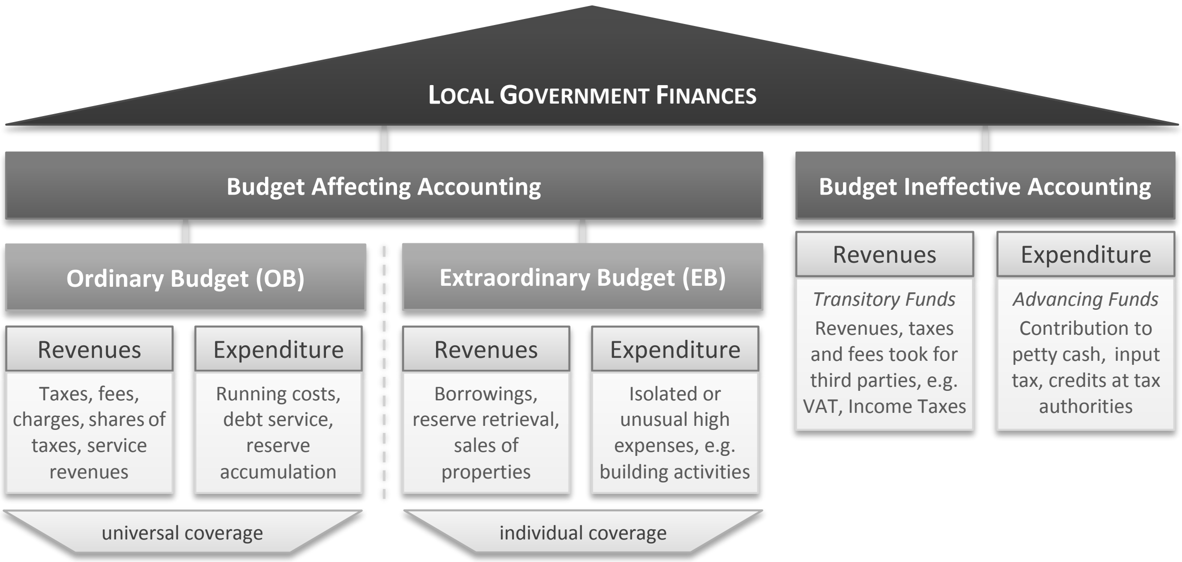
→ The *Ordinary Budget (OB)* covers all revenues and expenditures resulting from regular economic management of a public body.

→ Within the *Extraordinary Budget (EB)* expenditures are only included, if they are seen as isolated or as unusual high in the respective matter. Such expenditures may be budgeted, if and only if there are specific extraordinary revenues that cover them.

As a consequence, the so-called "universal coverage principle" applies to the ordinary part of the public budget. So, typically, all regular expenditures (e.g. running costs) have to be covered by regular revenues (e.g. municipal taxes) – without earmarking. In contrast to that, every single extraordinary expenditure has to be covered – in its entirety or partially – by a dedicated earmarked extraordinary revenue (e.g. borrowings). This is called the "individual coverage principle". The VRV states explicitly that directly accounted cash flows between OB and EB are *not* allowed. So, the key principle of this budgetary concept is that regular expenditures must *not* be covered by extraordinary revenues. This complies with a strict fiscal ordering principle, which has proven itself in practice [16, p. 14].

In **FIGURE 4** the just described structure of local government finances is depicted collectively in its entirety with some extra exemplary revenues and expenditures within the respective categories.

FIGURE 4 – The budgetary structure of local government finances including exemplary revenues and expenditures



Source: Parrer [14, pp. 64-65], Enzinger et al. [21, pp. 9-11], own illustration

The transfer of capital from other public bodies to municipal budgets has been somewhat excluded in this explanation on purpose, as that would have gone beyond the scope of this sub-chapter. However, some aspects regarding this matter will be addressed later on, if necessary. At this point, it is only worth mentioning, that transfer of capital from other public bodies can be accounted as revenue or expenditure to the OB as well as the EB – depending on its source and usage.

Possible financial flows within municipal budgets

It is important to note at this point that according to the VRV financial flows from the OB as well as the EB – so the global budget as a whole – to financial assets *are* allowed – opposed to direct cash flows in between them. In fact, the VRV states unambiguously that financial assets should be established by the OB, while the EB should be covered by those financial assets [14, p. 65].

Under normal circumstances, interest rates, debts and reserves are paid from the OB. Naturally, this should only happen to such extend as it does not jeopardize the payment of regular expenditures and as long all expenditures together are covered by regular revenues. The EB is covered by revenues from those financial assets as well as by sale of movable and immovable property. The latter case or "possibility" is not only quite uncommon, but also substantially endangering for the total budget of a municipality, if performed over a long time.

Last but not least, it is important to note the described concepts in this sub-chapter run like a golden threat through the whole budget cycle – beginning from the budget over the statement of account to the budgetary audit.

2.2.4 The structure of municipal budgets

Besides the overall concepts and general principles relevant for municipal financial management described so far, the actual structure of accounting in local governments in all its details has been excluded somewhat until now. Despite the current revision of the VRV induces significant changes to the structure of municipal budgets overall – these changes will be discussed in the next sub-chapter 2.2.5 – this overall structure is based on a single-entry bookkeeping system down to the present day. This system is also called "cameralistics"[19]. The term of cameral accounting is often used synonymously as a term for bookkeeping in public administration in general. The VRV as a legal basis for that and local fiscal management in Austria in general is not referring to its accounting structure directly as being cameral – only in minor annotations. However, the VRV still implies it in its entirety unambiguously.

Cameral accounting as a form of bookkeeping only compares the revenues with the expenditures in a fiscal plan – the budget – over a respective fiscal year [14, p. 78]. A difference in the balance is compensated with the accumulation or retrieval of financial assets (debts & reserves), respectively – as discussed in the last sub-chapter 2.2.3. As a consequence, the total results of all revenues and expenditures are always equivalent and offset each other. Therefore, the cameralistics are also called "cash accounting" [18, p. 81] or simply "revenue- & expenditure-accounting" [21, p. 7].

With that, the cameralistics in its core shows only the complete view on all financial *changes* in a fiscal period, but gives no overview of the current stock of financial assets [14, p. 80]. That's why, for most municipalities there is no (complete) statement of assets available. And therefore, also no accounted depreciation or economical profit and loss statement is possible. However, the VRV still makes several annexes compulsory, which can give some answers to these matters, and its recent revision tries to overcome this issue in its entirety. Nevertheless, in its core the VRV still incorporates cameralistics and sticks to it. Additionally, the Federal States themselves can make their own (stricter) regulations on municipal budgeting, but, as a superior national law, the VRV provides the common basis and overrules possible loopholes in state law.

In order to get a holistic overview on the cameral accounting in municipal budgets, the subsequent discussion in this sub-chapter will follow a bottom-up approach: starting from individual accounting entries, their classification via general groupings through to the summarizing parts of annual municipal budgets and statements of account, the structure of fiscal management for local governments will be discussed in detail. So, processes later in this Diploma Thesis should be easier to understand for the reader.

[19] from the Latin word „camera"; meaning literally: room, vault; in the sense here: princely treasure chest [14, p. 77].

The structure on its most detailed level: accounting entries

According to the VRV, all payment and intake orders shall be accounted as target values on their respective appropriate accounts. These debit positions show all claims for revenues and liabilities for expenditures. So, a revenue is accounted at the time at which a claim is emerging; an expenditure – on the other hand – as a result of an emerging liability. If a payment is received or made due to a claim or liability, it is accounted as actual value on the very same account. Thereby, a claim or liability is settled in the terms of cameral accounting. In this way, the separation of powers is also established on the accounting level [18, p. 81] – with orders on the one hand and executions on the other.

The way, how every payment and intake is accounted, is crucial for this Diploma Thesis. Each revenue and expenditure is associated not only to a single account with its target and actual value. In fact, it is classified in three ways according to the VRV. As discussed in chapter 2.2.3, each cash flow can be classified in a functional as well as in an economic way. Additionally, it can be classified in a fiscal way too. VRV's overall classification for revenues and expenditures is summarized in **TABLE 3**.

TABLE 3 – Classification of revenues and expenditures in municipal budgets according to VRV

First Classification (fiscal)	Second Classification (functional)	Third Classification (economic)
Budgetary Indicator (Haushaltshinweis)	Approach (Ansatz)	Account (Posten/Konto)
1	**123456**	**123456**

Source: VRV [17], Enzinger et al. [21, p. 11], own illustration

In this sense, the classification for each and every cash flow is structured like this:

- → The first digit describes the overall *fiscal* classification called HAUSHALTSHINWEIS or – within this work – "budgetary indicator". It reveals whether the respective cash flow is a revenue or expenditure, belongs to the ordinary or extraordinary budget or is accounted as budget ineffective (cf. the differentiation of municipal revenues & expenditures in chapter 2.2.3). The respective possible indicator numbers for municipalities according to the VRV are listed in **TABLE 4**.
- → The second group classifies the respective cash flow in a *functional* way, which means, that it describes for which kind of matter or "job" the respective cash flow has taken place [14, p. 70]. In the terms of the VRV, this is called HAUSHALTSANSATZ or – within this Diploma Thesis – "approach". The first three decadal divided digits are completely regulated, where the first decade defines so-called "groups", the first and second decades "sections" and all three mandatory decades together "subsections".

TABLE 4 – *Fiscal* sub-classification of revenues and expenditures in municipal budgets according to VRV

Budgetary Indicators (Haushaltshinweise)		
Indicator	Cash Flow Classification	Fiscal Classification
1	Expenditure	Ordinary Budget
2	Revenue	Ordinary Budget
5	Expenditure	Extraordinary Budget
6	Revenue	Extraordinary Budget
9	Expenditure	Budget Ineffective Accounting
0	Revenue	Budget Ineffective Accounting

Source: VRV [17], Enzinger et al. [21, p. 12], own illustration

TABLE 5 – Structure of *functional* sub-classification of revenues and expenditures in municipal budgets according to VRV

Approaches (Ansätze)			
	Group	Section	Subsection
0	Representative Bodies & General Administration	each **0** to **9**	each **0** to **9**
1	Public Order and Security		
2	Teaching, Education, Sport and Science		
3	Arts, Culture and Cults		
4	Social Welfare and Housing Subsidies		
5	Health		
6	Road- and Waterworks & Transport		
7	Economic Promotion		
8	Services		
9	Financial Management		

Source: VRV [17], Enzinger et al. [21, p. 12], own illustration

The VRV allows also possible sub-classifications under those regulated and compulsorily given subsections that can be made freely by each public body individually in an optional fourth and fifth decade. The also optional sixth decade is regulated again by the VRV and serves for an additional *fiscal* (sub-)classification. If this classification is used, but not the optional classification in the fourth or fifth decade, the latter must be given with "0" anyway.

The structure of such approaches within the mandatory first three decades with a listing of all groups according to the VRV is shown in **TABLE 5** above.

→ Finally, the third and last group classifies the respective revenues or expenditures in an *economic* point of view. For municipalities, this classification is made up by at least three also decadal divided digits, where the first decade defines so-called "classes", the first and second decades "subclasses" and all three mandatory decades together (again) "groups". The VRV again allows further possible sub-classifications with three additional decades at the maximum; all of them can be freely chosen by each municipality individually.
All in all, these economic classifications are called HAUSHALTSPOSTEN or -KONTEN in the terms of

TABLE 6 – Structure of *economic* sub-classification of revenues and expenditures in municipal budgets according to VRV

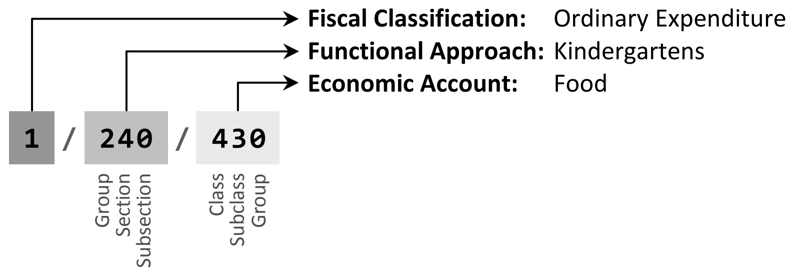
Accounts (Posten/Konten)			
	Class	Subclass	Group
0	Facilities	each **0** to **9**	each **0** to **9**
1	Inventories		
2	Cash, Claims, active accrued Revenues and Reserves		
3	Liabilities, Deferred Credits to Revenues		
4	Durable and Non-Durable Goods & Merchandise Usage		
5	Personnel Payments		
6 **7**	Other Operating and Administrative Expenditures		
8	Recurring Revenues		
9	Equity and Final Accounts		

Source: VRV [17], Enzinger et al. [21, p. 13], own illustration

the VRV or – within this work – “accounts”. They state, which kind of revenue or expenditure is meant – taking into account business as well as economic aspects [21, p. 13].
The structure of such accounts within the first three compulsory decades with a listing of all classes according to the VRV is shown in **TABLE 6** on the previous page.

Together, these classifications constitute so-called HAUSHALTSSTELLEN or “budgetary accounts”, respectively, [16, p. 17] and accompany each transaction from its order in the annual budget up until to the listing in the respective statement of account more than a year later [14, p. 70]. An example for such a budgetary account according to the VRV – omitting all optional parts – is given in **FIGURE 5**. The individual classifications can be separated “appropriately” according to the VRV; in most cases dots or slashes are used, like in the depicted example.

FIGURE 5 – Exemplary budgetary classification according to VRV: expenditures for food in kindergartens



Source: VRV [17], own illustration

The described structure for the classification of each and every individual accounting entry or cash flow, respectively, in municipal budgets was established with the first revision of the VRV in 1974 [28, p. 10]. The register of all accounts followed the uniform scheme of accounts by the ÖPWZ (Austrian Centre for Productivity and Profitability) [16, p. 17], which is also used by industrial companies in Austria. However, VRV’s register of accounts was never adjusted to new revisions of the uniform scheme of accounts down to the present day [28, p. 10]. Instead, adaptations were made separately due to the necessities in the course of Austria’s joining to the European Economic and Monetary Union as well as similar new statistic requirements. So, by today, the VRV is aware of nearly each 400 approaches and accounts for classification of municipal cash flows.

Public corporations in municipal budgets

As seen in chapter 2.1, a proper differentiation between the general government and other institutional sectors is made already at the level of the European Union in a supranational way. This differentiation beats its way also down to municipal budgets in Austria and their structure in order to provide a proper overview about the finances of a local government as such – and not of their associated institutional units.

Therefore, all public corporations but also all business enterprises of a municipality in general are distinguished from the remaining governmental budget. The reason for this is that their separate budgets – including all of their debts, but also reserves – should *not* be reflected in the budget of the municipality itself – except for possible withdrawals of profits or subsidies granted by the local council. Such corporations need to fulfil three requirements according to the VRV [20, p. 48]:

- → The institutional unit needs to be a public corporation in the sense of the ESA (see also chapter 2.1.1). So, it needs to be a market producer, which should be able to cover at least 50% of its production costs on its own or by its sales, respectively.
- → The corporation needs to keep a full set of own accounts.
- → The corporation needs to have also an own management, which shall have the freedom of choice regarding all operational matters.

If these requirements are met, all cash flows in a municipal budget regarding those public corporations are classified, and hence, accounted within the approach sections 85-89; although, their activities do not need to have something in common with approach group 8 “services” altogether [14, p. 74]. If a corporation does not meet the “50% criterion” for whatever reason, it is budgeted separated likewise within the approach sections 81-83 as a public or “business-like” institution in contrast to a corporation. Thereby, it is accounted separately from the rest of the municipal budget in fact, but it still belongs to the general government in the books and respective summaries.

Examples for public corporations in a municipality can be institutions providing water supply and disposal or waste collection, but also a cinema, tavern, funeral undertaker, forestry or a public utility company in general can be counted as such a business enterprise. On the contrary, public swimming pools or a street cleaning company can be examples of public institutions.

General summarizing elements of municipal budgets

All accounted cash flows are summarized in one way or another in the annual budget or the respective statement of account of a municipal in order to get an overview of the local government finances in a simple way. Therefore, the VRV regulates the structure of those budgetary documents in detail.

First of all, individual revenues and expenditures regarding a specific budgetary account are only listed in different aggregated forms as a sum of all cash flows accounted to the respective budgetary account – not individually – in municipal budgets and respective statements of account. In this sense, for budgeting as well as accounting each and every budgetary account is always listed with their respective target and actual values (in case of statements of account). Additionally, there are two main superior aggregations present in all budgetary documents according to the VRV:

- → The main overview of all revenues and expenditures grouped by all ten approach groups (first decade of the classification) for the OB as well as the EB.
- → The budget profile with its very own aggregation systematic, which provides the means for evaluation of the municipal budget in an economically way [21, p. 16].

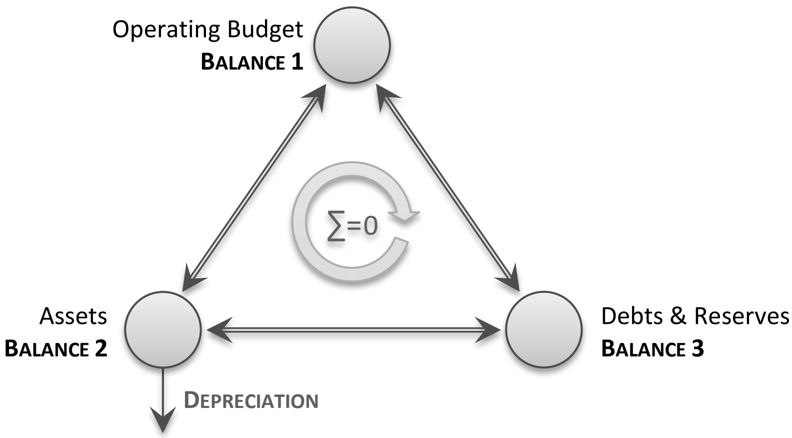
Given its significance, the budget profile will be discussed subsequently in more detail. On the top of this, the VRV also requires a cash balance for the statement of account and several annexes – also for budgets. Some of them – like listings for personnel costs or transfers to and from other public bodies – aggregate the mentioned cash flows in their own way. Others in turn – like the liabilities listing – give information that go beyond the cameralistics core or the cash flow-point of view. However, as a complete explanation of these elements would go far beyond the scope of this work and are also not of its primary interest, there will be no further discussion at this point.

The relational triangle of fiscal balances (or the budget profile)

To understand why the budget profile in particular is of such utmost importance in the municipal budgetary documents, one needs to know how finance management works.

A total budget has to be balanced at the end of each fiscal period; *no* residual values must remain.[20] A deficit has to be financed somehow (e.g. through making debts); surpluses have to be used too (e.g. for building up reserves). As a consequence, all accounting is interrelated to each other in the form of a relational triangle as depicted in FIGURE 6. The operating budget (balance 1), owned assets (balance 2) and all debts & reserves (balance 3) are correlated; a change in one of them has direct consequences to the others. Why? Simply because – as already mentioned in the beginning – the aggregated sum of all balances (the total budget) has to be equal zero.

FIGURE 6 – Relational triangle of fiscal balances in municipal budgets



Source: Köfel [29, pp. 91-93], Parrer [14, p. 81], own illustration

As real assets are losing value or depreciate, respectively, and have to be replaced at some point in time, eventually, there is a natural suction pressure on this relational triangle originating from the asset balance. If no assets should be forfeited, the operating budget including debts and reserves (sum of balance 1 & 2) has to be generating a surplus for reinvestments in assets.

In an ideal situation with healthy finances, a positive operating budget (balance 1) will provide the means to build up assets (negative balance 2) and reserves (negative balance 3). Those reserves also serve reinvestments in assets at a later point in time.[21] So, in case the operating budget is negative in one single fiscal period – for whatever reason, the balance can be reached and even investments can still be made. But what happens if the operating budget (balance 1) is negative over a longer period? Both balances 2 & 3 will become positive, which means, that assets are sold and debts are made to finance the operating budget. Obviously, this trend cannot be maintained forever and will likely result in bankruptcy in the end [14, p. 83].

[20] Annual transfers for settlements are excluded from this explanation, as they have to be accounted properly as well.
[21] This corresponds to the possible financial flows in municipal budgets discussed in sub-chapter 2.2.3.

The budget profile in accordance with the VRV displays this relational triangle in exactly the same way (even in more detail) through aggregation of all cash flows of different accounts. The depicted TABLE 7 below shows a consolidated overview of such a budget profile in its entirety.

TABLE 7 – Consolidated overview of a complete budget profile according to the VRV

№	±	Description	Sum OB + EB	thereof A85-89[1)]	Sum w/o A85-89
I.		**PROFILE**			
19	+	Operating Revenues	*(1)*		
29	–	Operating Expenditures	*(2)*		
91	**=**	**BALANCE 1: Operating Result**			*(3)*
39	+	Revenues from Asset Accounting	*(1)*		
49	–	Expenditures from Asset Accounting	*(2)*		
92	**=**	**BALANCE 2: Asset Account Result w/o Financial Trans.**			*(3)*
59	+	Revenues from Financial Transactions	*(1)*		
69	–	Expenditures from Financial Transactions	*(2)*		
93	**=**	**BALANCE 3: Financial Account Result**			
94	**=**	**Annual Result w/o final settlement[2)]**		*(4)*	
II.		**DEDUCTION OF THE FISCAL BALANCE**			
70	±	Annual Result w/o A85-89 & Financial Transactions	*(3)*		
71	±	Clearing of Annual Result of A85-89	*(4)*		
95	**=**	**Fiscal Balance ("Maastricht Result")**			
III.		**OVERVIEW TOTAL BUDGET**			
79	+	Total Revenue	*(1)*[3)]		
89	–	Total Expenditure	*(2)*[3)]		
99	**=**	**Administrative Annual Result**	0		

1) Sum of all cash movements classified as such ones for approach sections 85 to 89
2) Sum of Balance 1, 2 and 3
3) Total sums including annual transfers and transfers between OB & EB

Source: VRV [17], Enzinger et al. [21, p. 22], own illustration

All balances correspond to the ones already discussed and depicted in FIGURE 6. All depicted sums for revenues and expenditures are made up by multiple sub-aggregations not shown. For easier orientation in every aggregation or line, respectively, regulated index numbers are shown leftmost. So, the described balances have always the indices 91, 92 and 93. The VRV regulates specifically which accounts (in most cases subclasses and groups) have to be aggregated for each index. On the other hand, the budget profile lists three columns. In the first column the sum of all target values for the respective line is shown. The second column lists the part of this sum accounted for public corporations or approach sections 85 to 89, correspondingly. The third and last column aggregates only the sum associated with the general government (so, without any public corporation), eventually. The values for the OB and the EB are always aggregated and not displayed separately.

At the end of the actual budget profile always follows the deduction of two results:

- → the fiscal balance or "Maastricht Result" and
- → the administrative annual result.

The "Maastricht Result" is originating from regulations of the European Union and allows in its form only public corporations to balance their total budget including the third balance or finance themselves via debts or reserve retrieval over a fiscal year, respectively. For the general government the latter are left out of consideration on purpose. The reason is that the legislature wants to visualise possible debt financing [14, pp. 192-193]. Therefore, the "Maastricht Result" shows the amount of equity financing by the local government and is one of the most important values in a municipal budget, because of that.

Last but not least, the administrative annual result at the very end of the budget profile shows that all balances are indeed equal zero according to the discussed relational triangle of fiscal balances. Therefore, annual transfers and – interestingly – transfers between the OB and the EB are settled beforehand. As discussed in the previous chapter 2.2.3, the latter should not happen, actually. Nevertheless, the VRV offers this loophole for municipalities. After all, the administrative annual result must equal zero anyway.

2.2.5 The reformation of Local Fiscal Management

As mentioned in sub-chapter 2.2.1, the VRV has been revised several times since its first version in 1974. Over the years, several amendments of the VRV became applicable. However, the most recent one, the VRV 2015 [17], was issued in October 2015 and can be considered as one of the biggest changes to Austrian budget law in recent history [30, p. 20].

After the budget law reform of the central government in 2013, the aim is to unify the budget law for all other public bodies [31, p. 45] – federal states as well as all municipalities – in Austria. In order to do so, elements of the double-entry bookkeeping system are introduced in existing frameworks like the VRV. According to this logic, a three-component calculation is established and therefore, now all budgets and statements of account will illustrate three main aggregations [31, p. 45]:

- → A *profit and loss account*, which lists all in- and outflows of resources – independent of the actual point in time where the corresponding payment took place. So, this calculation shows all *actual* values accounted in a municipal budget.
- → A *finance statement* includes all payments (revenues & expenditure) made in a fiscal period and gives an overview about the overall financial liquidity. So, this calculation shows all *target* values accounted in a municipal budget.
- → An *asset statement*, which also can be considered as a "balance" in a business point of view, opposes all assets and liabilities with each other.

The real novelty to the VRV is the last aggregation. Thereby, the VRV overcomes one of the biggest issues of the incorporated cameralistics – the only cash flow-centred point of view on budgets. Now, all municipalities have to value all owned assets in their entirety and need to account deprecation explicitly. For any part of the budget, the asset base is not ignored anymore. Therefore, the VRV 2015 provides in Annex 7 an encompassing list of economic lifetimes for possible assets.

Additionally, the explicit separation of OB and EB is dropped by the VRV 2015. One of the major reasons behind this might be that there was criticism [28, p. 9] that the separation as such was vaguely defined and allocations to OB or EB were inconsistent, anyway. So, according to this argument, the overall comprehensibility and comparability was not given or, at least, hampered.

Last but not least, more than ten new additional supplements and annexes are required in budgets and statements of account with the VRV 2015 – resulting in a total of 35 instead of 23 possible additions to the core of the budgetary document [32, p. 11].

However, as complete asset valuations and also associated adjustments to the overall accounting and respective IT systems can be quite an overwhelming assignment considering that there are about 2,100 municipalities all over Austria, broad transition periods were provided too. So, according to

Article 40, paragraph 2, VRV 2015, for all municipalities, which are audited by the Court of Audit[22], the VRV 2015 applies by no later than the fiscal year 2019; for all other municipalities it takes effect for the fiscal year 2020, at the latest. Hence, from 2021 onwards all budgets and statements of account will comply with these new regulations.

In spite of that, the consequences on the actual accounting system in its details are quite negligible. The approach and account classifications discussed so far are more or less left untouched[23] and the accounting system itself too. Not even the mentioned budget profile has been revised or updated in the VRV 2015 to a greater extend.[24]

With respect to this Diploma Thesis, most of the work and the explanations are referring to common concepts of the VRV 2015 and its predecessor, the VRV 1997, if applicable. As the entire recent municipal budget data published down to the present day and most of it published until 2020 will correspond with the VRV 1997, this work will stick to the VRV 1997 in most parts, anyway. As the underlying accounting system remains more or less unchanged, minor changes mentioned can be incorporated by future works easily. Possible preceding as well as subsequent amendments to the VRV 2017 are expectable in any case.

2.2.6 Information flows of Budget Data

Aside from overall concepts, the actual content of a municipal budget and its structure, it is necessary to understand the respective information flows of this budget data to get a holistic overview about this topic in the sense of this work. Especially, the flows of budget data between the public bodies on different governmental levels and the interchange with external entities should be addressed in the following. The actual usage of budgetary data outside the public sector and its implications will be of major interest later in this Diploma Thesis – beginning in sub-chapter 2.4.3.

The basis for such information flows in Austria, but also worldwide, is primarily the official so-called "Government Finance Statistics". These finance statistics provide information and overview about the financial management of all public bodies nationally. Elements of respective reports are listings of (mostly) aggregated revenues, expenditures and assets as well as debt and personnel levels of individual public bodies like municipalities, which serve as statistical overview for economic and fiscal needs, delivery obligations as part of the European Union as well as a foundation for calculations on the total economy according to the ESA [33, p. 36].

In Austria, the legal basis for such statistical surveys is provided by the Federal Statistics Law or BUNDESSTATISTIKGESETZ (BStatG 2000). As the actual finance statistic is not regulated by this superior law, according to Article 4, paragraph 4, BStatG 2000 [34] the GebStat-VO 2014 – as discussed in chapter 2.1.2 – takes care of its details. The most recent revision of this regulation became applicable at the beginning of 2014 (Article 12, GebStat-VO 2014) and was not only more extensive than its predecessor from 2002, but also introduced new report obligations for municipalities [35]. So, some reports have to be made now quarterly (also for associated institutional units) and also have to include contingent liabilities and shareholdings of the local government.

In order to establish a regular interchange of budgetary data between the different bodies of the public sector (and their heterogeneous information systems), common data formats have to be used, obviously. Austrian's statistical office STATISTIK AUSTRIA specifies these data formats after

[22] So, currently all municipalities with a population over 10,000 citizens – see also chapter 2.2.2.
[23] Only additions are made: two (!) approach classifications in sections 02 & 04 and about 30 account groups.
[24] Besides the explicit statement of revenues and expenses through equalization payments for financial derivatives (indices 58 & 68), only minor adjustments in the association of account groups have been made for municipal budget profiles.

TABLE 8 – Record types according to GHD and respective statistical surveys by the STATISTIK AUSTRIA

№	Record Type	Record contents (exemplary)	yearly reportable	quarterly reportable
01	**General data**	Information about the municipality, its respective contact, the subsequent content, creation and resolution date	yes	yes
02	**Budgetary accounts**	Preceding year's statement of account, current budget & current medium-term financial plan for all budgetary accounts	yes	yes
03	**Debts & liabilities**	Levels at beginning and end of reporting period, intakes and payments, creditor, credit period and interest rate	yes	yes
04	**Financial assets and claims**	Levels at beginning and end of reporting period, intakes and payments, debtors, period and interest rate	yes	yes
05	**Other assets**	Levels at beginning and end of reporting period, all ins and outs, respective asset categories	yes	no
06	**Personnel**	Number of staff employed, employment relationships and hours worked for each personnel category & cost approach	yes	no
07	**Other finance statistical figures**	Municipal tax rates and payments for each public-private partnership	yes	no
08	**Shareholdings in public corporations**	Information about the corporation, the ownership structure, equity, liabilities, operating result, headcount	yes	no
09	**Contingent liabilities**	Levels at beginning and end of reporting period including medium-term forecasts, all ins and outs, creditor, duration	yes	no
91	**Stability Pact figures**	Additional personnel data according to the ÖStP, e.g. average pension amount and retirement age	yes	no

Source: Statistik Austria [36], own illustration

consultation of Federal Ministry of Finance and the affected institutional units according to Article 5, paragraph 4, GebStat-VO 2014. The respective data format for Austrian municipalities is specified as municipal budget data or GEMEINDEHAUSHALTSDATEN interface (GHD). Such a national data format or interface – as it is called by STATISTIK AUSTRIA – exists now for more than 15 years [33, p. 36] in Austria. In fact, it consists out of block formatted textual data in ASCII or representation according to ISO 8859-1 with a fixed record length of 400 characters [36, p. 3]. The content of this data which has to be transferred via GHD as well as its requested frequency is listed in **TABLE 8**. A similar data format is specified for Federal States as LANDESHAUSHALTSDATEN interface (LHD) by the STATISTIK AUSTRIA [37].

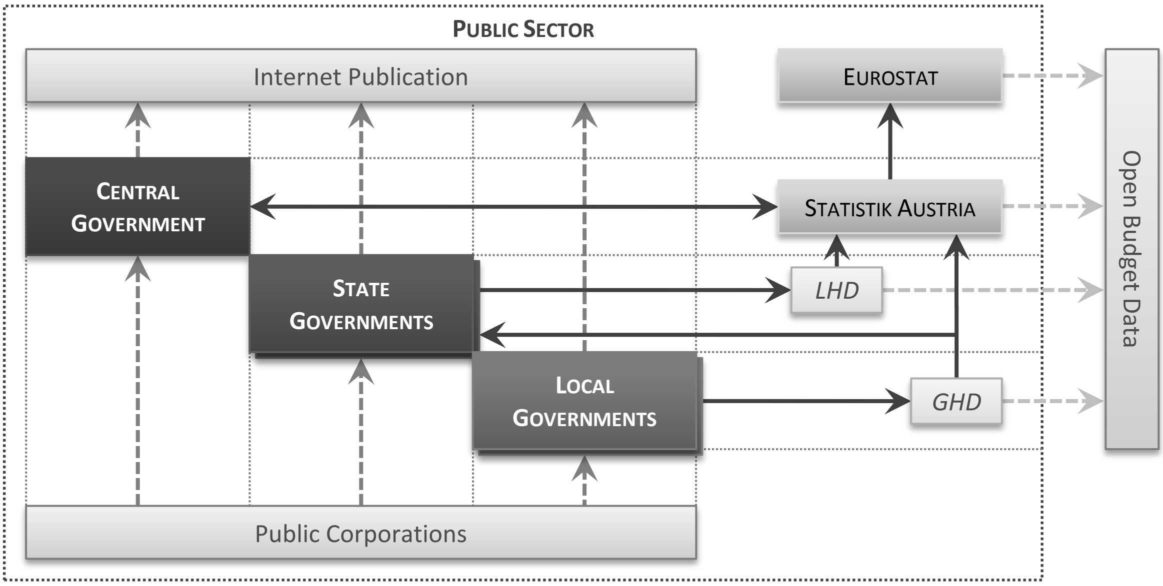
According to Article 5, paragraph 4, GebStat-VO 2014, the actual data transfer has to be performed via electronical means. However, it does not state which medium should be used technically. In fact, TXT-files are used, which are transmitted via email, uploaded on websites or on file servers via FTP. The actual flow of this data looks like this for the main parts of the general government [33, p. 37]:

- → The *central government* transmits its budget data directly to the STATISTIK AUSTRIA. Most probably, STATISTIK AUSTRIA has direct access to the central government's budget information system.
- → All *state governments* provide their budgetary data according to LHD and transmit it directly to the STATISTIK AUSTRIA.
- → Each and every municipal *local government* in Austria provides its budgetary data according to the discussed format of GHD. In contrast to the other public bodies of the general government, this data has to be transmitted and gathered first by the respective responsible authority of the State Government, which in turn sends all of this budgetary data collectively to the STATISTIK AUSTRIA for each Federal State.

The STATISTIK AUSTRIA itself has to transmit aggregated information according to the ESA 2010 derived from all of its collected budgetary data regularly to European Union's statistical office EUROSTAT as part of delivery obligations due to Austria's membership in the European Union as well as European Economic and Monetary Union. Furthermore, the STATISTIK AUSTRIA shall open to the public all (or rather most) statistical data “in some suitable form” according to Article 19, paragraph 1, BStatG 2000. However, especially the collected budgetary data is available publicly only in aggregated form through the systems of the STATISTIK AUSTRIA, but no individual data or account entries. This is due to Article 19, paragraph 2 & 3, BStatG 2000, which forbids the publication of personal-related data without prior approval of all affected parties – which is not given for fiscal data of public bodies, not surprisingly.

Apart from that, each public body – Federal States, municipalities as well as the central government itself – can (and some of them do, actually) publish its budgets as well as the respective statements of account on the Internet to the public by itself – also sometimes in an “open”, reusable and machine-readable way. More on that will be addressed later in several subsequent chapters of this work.

FIGURE 7 – Information flows of Budget Data in Austria



Source: Bröthaler [33, p. 35], own illustration

An overview of these flows of information or budgetary data, respectively, in Austria is depicted in FIGURE 7 above. Please note, that this is only an abstract overview and real-life information flows are much more complex and also different flows not shown can appear under special circumstances.

2.3 THE TECHNICAL FRAMEWORK – LINKED DATA

After getting in touch with the subject's domain of this Diploma Thesis, the necessary technical background for the subsequent work shall be addressed in the following. As described in chapter 1.2, so-called "Linked Data" should provide those technical means to implement an exemplary approach for Austrian municipal budget data, like it has been done by international state-of-the-art and current best practices in similar ways. Such examples will be of primary interest later in chapter 3 of this work; the mentioned exemplary approach in chapter 4. But right now, Linked Data as a term and its concepts shall be addressed in more detail.

In general, down to the present day most of the data distributed via the World Wide Web – today's most influencing and most used information source – is "just" a *Web of Documents* or, to be more precise, Hypertext [38, p. 23]. Most of all websites we know consist out of this "hypertext", along with the most popular one: Wikipedia. The term "hypertext" itself describes a text, which structures information objects web-like[25] through cross-references ("hyperlinks") [39, p. 2]. To declare such texts on computers in a suitable way, the Hypertext Markup Language (HTML) was invented by Tim Berners-Lee during the early 90s [39, p. 7].

However, as such hypertexts concentrate on human-readable, eponymous *texts* instead of actual data, this data is still often published in *Documents on the Web* (instead of in a *Web of Data*) – like Austrian budgetary data in PDF files or Microsoft Excel spreadsheets (see chapter 3.1.3). Linked Data should provide such a *Web of Data* in addition to the current *Web of Documents* [40]. While the actual technical mechanics and the respective foundation for this web shall be explained later in this very chapter, the actual aim of Linked Data is to overcome issues by former data storage & distribution techniques [38, p. 22]: Today, in most cases, data is associated and, thereby, incorporated by specific applications. Metadata and data schema information are incorporated as such by those applications, instead of being separated from the actual application logic. A database or an export of its data is built specifically for certain applications, usually. For example, column headings in the ubiquitous comma-separated-value format (CSV) are not really immediately useful, as the actual meaning of those textual designators is ambiguous and can differ from document to document. Furthermore, while looking at such a CSV file, humans *may* be able to derive some value from those column headings; a program, which is not explicitly built for this respective data schema, is *not* necessarily able to do so. That's why in the end, *Documents on the Web* including their incorporated data are *not* (easily) reusable or even machine-readable for third-party programs. In contrast to that, Linked Data – composing a *Web of Data* – *is*. Therefore, Linked Data becomes increasingly important when it comes to distributing data in an open way (see also chapter 2.4.4).

2.3.1 The Semantic Web

To understand Linked Data in its entirety, it is necessary to understand the related term of the "Semantic Web" as well. The Semantic Web – as a well-known term nowadays – lays the foundation for Linked Data and describes the overall concept of this field in information science.

Both concepts were initially coined by famous Tim Berners-Lee. While Linked Data was described by him in a seminal design note [41] in 2006, the concept of a Semantic Web can be traced back to the beginnings of the World Wide Web [42, p. 17] and a "roadmap" for a Semantic Web was outlined by Berners-Lee already in 1998 [43]. The Semantic Web can be understood as the effort to give the

---

[25] Hence, the term "World Wide Web" (as a "web" of "hypertext documents"), which was also the eponym for the first browser for such hypertexts, invented by Tim Berners-Lee himself [39, p. 7].

information (and the data) on the World Wide Web self-defined semantics and meaning. The idea is explicitly *not* to replace all current websites (and/or the data on them), but to enrich or extend them in a way, so that (especially automated) information retrieval is simplified. This was also expressed by Tim Berners-Lee himself in an article, which was published in SCIENTIFIC AMERICAN in 2001:

> *"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." [44, p. 35]*

Although, Berners-Lee described specific theoretical and technical concepts for such a Semantic Web quite early in his roadmap, it is important to understand that the primary meaning behind this term should be understood rather as an overall vision than an actual technical instrument. Berners-Lee emphasizes this in one of his books in the year 2000:

> *"The first step is putting data on the Web in a form that machines can naturally understand, or converting it to that form. This creates what I call a* Semantic Web *– a web of data that can be processed directly or indirectly by machines." [45, p. 191]*

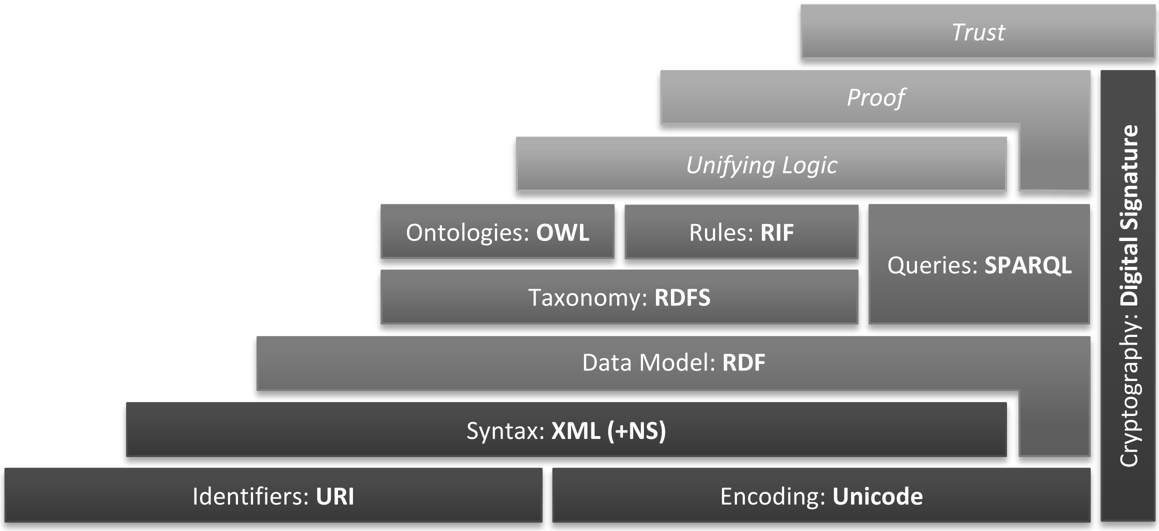
Therefore, the Semantic Web has to be seen as a goal or the target for this vision, while Linked Data – seizing this idea – is a (possible) content of this Semantic Web and provides the technical means to reach this goal [42, p. 17]. Additionally, Linked Data omits certain parts of the Semantic Web-vision by definition. While the Semantic Web also incorporates higher & more sophisticated concepts and goals – like representation of knowledge and logic, as we will see soon –, Linked Data concentrates primarily on publishing and linking structured data. This simplification – just like the one done with Hypertext on the Web in the first place – shall lower entry barriers for data providers possibly publishing Linked Data [46, p. 69].

However, soon in the development of the – at that time still young – World Wide Web, it became clear that certain standardizations have to be made by a central and independent authority preventing the Web from breaking apart and to ensure interoperability. For this reason, the WORLD WIDE WEB CONSORTIUM (W3C) was founded by Tim Berners-Lee in October 1994 [47]. Many of the technological standards that will be mentioned in this chapter building up the foundation for Linked Data were introduced and are still managed by the W3C.

Also, the Semantic Web itself is made up by such technological standards originating from the W3C. It consists out of several conceptual layers introduced by Berners-Lee in his roadmap for it in 1998 [43]. Each layer basing upon its underlying counterpart can incorporate different of these standardized technologies. This layer model illustrating the conceptual architecture of the Semantic Web is nowadays known as SEMANTIC WEB CAKE, SEMANTIC WEB LAYER CAKE or SEMANTIC WEB STACK. An illustration of the original layer model was presented for the first time by Berners-Lee on a scientific conference in Washington D.C. in December 2000 [48]. Since then, the model was revised and distributed in serval versions by the W3C. The most recent one [49] originates from 2007. Based on this version – with additions from recent relevant literature – **FIGURE 8** illustrates the SEMATIC WEB LAYER CAKE including exemplary technologies for the different conceptual layers, which – for the sake of simplicity – match those used for Linked Data as discussed later.

At this point, the structure of this layer model shall be explained in order to give an overview of the environment Linked Data is settled. In general, the SEMATIC WEB LAYER CAKE can be distinguished into basic technologies, data modelling & description means and overlying concepts, which are partially yet to be specified – this differentiation is represented in **FIGURE 8** by the colouring (from dark to light). The actual contents of those groups of layers are discussed in the following.

FIGURE 8 – An up-to-date view on the SEMANTIC WEB LAYER CAKE



Source: Berners-Lee [48], Bratt [49], Obitko [50], own illustration

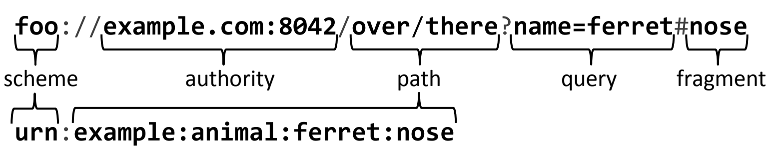
Hypertext Web Technologies

The lowest layers of the SEMANTIC WEB CAKE – depicted in dark grey in **FIGURE 8** – define and incorporate standard technologies originating from the *Web of Documents* or *Hypertext*. Most of them are now used for decades since the beginning of the World Wide Web and have been evolved under the supervision and/or standardization of the W3C to the present day. These technologies built – without a change – the foundation for the Semantic Web.

The very basic technical basis within this part is the digital *encoding* of all data. In recent years, an international standard called UNICODE has proven to be a common ground for all software development around the world. While providing the means to encode technically more than one million characters on the basis of fundamental ASCII encoding, its most recent version 9.0 contains currently 128,172 characters [52, p. 3] from languages all over the world. The long-term goal is to digitally encode every single meaningful character or text element from all known writing systems with one single, universal standard [52, p. 4]. As preceding legacy character encoding systems were inconsistent as well as incomplete, interoperability of systems with data from all around the world was not guaranteed. The UNICODE standard should resolve this issue as being also compliant to ISO/IEC 10646 [52, p. 902].

Besides that, data need not only be encoded but also addressed and *identified* in some suitable way. This is done via so-called "Uniform Resource Identifiers" (URIs). Originally, those identifiers could be separated into two groups [51, p. 7]: Uniform Resource Locators (URLs) and Uniform Resource Names (URNs). While URLs provide the means to locate certain resources (e.g. Hypertext documents) by describing their primary access mechanism (e.g. its network "location" via HTTP), URNs are intended to actually "name" resources in a unique and persistent way (i.e. even if a resource becomes unavailable). Incorporating both concepts of URLs and URNs, with URIs this separation no longer

FIGURE 9 – Standardized structure and elements of "Uniform Resource Identifiers" (URI) including examples



Source: Berners-Lee et al. [51, p. 16], own illustration exists. Thereby, the standard for URIs provides a simple, but also flexible way to address and identify resources or data. So, URIs are also used to digitally address and make statements of all kind of resources within the Semantic Web – from things like places or individual people through to abstract concepts like hierarchical ordering. The standardized structure of URIs in general (including its individual elements and two examples) is depicted in **FIGURE 9**.

On top of these very basic technological concepts and standards, it has to be also defined how the actual data is structured; in other words, how the *syntax* of the data looks like. This is done in most parts within the Semantic Web through the so-called "eXtensible Markup Language" (XML) and respective namespaces (NS). Such eponymous mark-ups are used to provide additional information for a stream of data within this data – including the data's storage layout and hierarchical organized, logical structure [53, p. 1]. As a kind of meta-language, XML provides also the means to impose constraints on this layout and its structure. Thereby, it becomes machine-readable additionally. However, XML can be seen as an alphabet, which does *not* inherently provide any vocabularies or semantics whatsoever, as it has no build-in mechanism to convey such information automatically [44, p. 37]. The meaning of individual data is still up to individual applications, while the actual processing is well-defined. HTML, which builds the foundation the current World Wide Web, can be seen as syntactical sibling[26] of XML.

Last but not least, the integrity of the data is of the essence. If data can be manipulated freely by third parties or its origin cannot be determined for sure, its value and usage is limited, misleading or even dangerous. Therefore, *cryptography* can be used to ensure the integrity of any data stream. For example, data can be equipped with digital signatures providing the means to verify its origin and content [43]. While this is already quite common in the current World Wide Web, respective standards for the Semantic Web are still to be defined. However, the concept of the Semantic Web and its Layer Cake – as depicted in **FIGURE 8** – assumes cryptography as a fundamental technical column for its final vision and the goal for a *Web of Data*.

Semantic Web Technologies

The intermediate layers of the SEMANTIC WEB CAKE – depicted in mid-grey in **FIGURE 8** – consist out of technologies and concepts, which are inherent for building Semantic Web applications. There already exist supervised standards by the W3C for all of them. As the details of those standards are also an integral part of Linked Data, they will be discussed subsequently in the separate chapter 2.3.2. For now, only a short overview of the elements of this group of layers with respect to the Sematic Web as such shall be given.

In order to give data self-defined semantics and meaning, several things have to be considered: First of all, a standardized *core data model* for all applications using the Semantic Web has to be in place for information interchange. A popular common standard for this is the so-called "Resource Description Framework" (RDF). Based on statements in a form of so-called triples, it enables to represent information about resources in the form of graph. RDF is based on all underlying layers of the SEMANTIC WEB CAKE: URIs to reference other resources and concepts, XML for serialization (in transmissions and/or storage files) and UNICODE for standard fundamental digital encoding of XML data. On top of this, RDF Schema (RDFS) provides the means to create vocabularies based on the RDF. Thereby, hierarchies of resource classes and properties can be established – so-called "*taxonomies*". So, in fact RDFS is a meta-language, which can be used to describe an RDF

[26] Only XHTML does truly extend XML and is conform to its standards. Nevertheless, its development has been stopped in favour of HTML's version 5. However, both HTML as well as XML are derivates of the so-called "Standard Generalized Markup Language" (SGML), which is in turn standardized by ISO 8879 [53, p. 1] – hence, their designation as "sibling".

(sub-)language or data model, respectively, that can be used to represent certain data via RDF. However, as capabilities of the RDFS are somewhat limited, two elements are still missing to really providing the means to create real semantic data. On the one hand, so-called "*ontologies*", which are based on description logic and introduce missing descriptive elements (like cardinalities or equality), bring the power for automated reasoning to the Semantic Web. Therefore, while extending RFDS, the "Web Ontology Language" (OWL) was created by the W3C as a standard to describe such full-blown ontologies. On the other hand, a rather new concept was introduced to the SEMANTIC WEB CAKE: *rules*. They are providing the means that go beyond the constructs available from taxonomies and ontologies. As exemplary standard, the Rule Interchange Format (RIF) was introduced in 2010 [54] by the W3C. Since such rules are not explicitly required for Linked Data in particular, the rest of this work will not go into details with this part of the Semantic Web.

Last but not least, as it is necessary to retrieve semantic data for respective applications in a proper way as well as storing it, also some form of *querying language* is necessary as a part of the SEMANTIC WEB CAKE. To query any RDF-based data (including statements involving RDFS and OWL), "SPARQL Protocol and RDF Query Language" (SPARQL[27]) was introduced by the W3C. SPARQL seem to be related to other query languages for relational databases like SQL; however, there are important differences in some parts. If nothing else, the fact, that is designed uniquely and exclusively for querying triplestores (consisting of RDF triples), should be the most outstanding feature of SPARQL.

High-Level Technologies for the Semantic Web

Apart from those technologies seen so far, the SEMANTIC WEB CAKE incorporates even more concepts at the top – depicted in light grey in **FIGURE 8**. As already mentioned, respective standards in this group of technological layers are still to be defined. However, they shall be introduced in a nutshell at this point anyway in order to provide a holistic overview of the components of the Semantic Web.

The lowermost layer within this group is also the only one, which is already in use in relevant works by now. Although no common standards exist, such applications, which are classified as *unifying logic* in the sense of the Semantic Web, do. Based on the data described and stored under the terms of the underlying layers (so, e.g. as RDF graph in a specific ontology), it is possible to deduce (automatically) new information at this point. This is called "reasoning" and is one of the most astonishing benefits that the Semantic Web can provide. Mentioned applications, incorporating specific algorithms (so-called "(semantic) reasoners"), implement and support exactly this very step.

Additionally, it is necessary to ensure reasonable results in itself. So, all reasoning steps shall be traced and the results – including those from simple queries – should be verified (somehow).[28] Conceptually, these processes are incorporated by the *proof* layer of the SEMANTIC WEB CAKE. However, there are still no applications, let alone, any standards implementing this concept.

Last but not least, *trust* shall be built for all data and results from queries and deductions by verifying its contents and identify its source through the means of cryptography. As already mentioned, technologies and application implementing this layer in the sense of the Semantic Web are still unrealized to this day too.

On top of all those layers of the SEMANTIC WEB CAKE, applications with user interfaces can be built truly forming a Sematic Web as envisioned in the first place.

[27] SPARQL is a recursive acronym; so, an acronym which refers to itself.

[28] That means verification outside the "narrow" boundaries of constraints usually applied in the current World Wide Web [43]; so, *no trivial* verification e.g. done by a webserver which is granting access to certain documents only to certain users (in this case, some data on the webserver itself says that requests originating from somebody, who knows a certain combination of a username and a password, have to be granted – which is already the end of verification chain).

2.3.2 Technical mechanics behind Linked Data

As we have seen in chapter 2.3.1, there are several technologies composing Linked Data as a mean to reach a Semantic Web. However, they need more explanation for further usage in this work. Therefore, especially the standards depicted in the intermediate layers of **FIGURE 8** (except RIF for rules), which compose the technical core concepts for Linked Data, shall be discussed in the following.

Resource Description Framework – RDF

The "Resource Description Framework" (RDF) was originally designed by the W3C for describing metadata [55] in a vendor-neutral and operating system-independent way in order to provide interoperability between electronic applications through machine-readable information. A first public draft for this standard was published in August 1997; the final W3C recommendation at the beginning of 1999 [55] incorporating the RDF data model itself as well as a XML serialization syntax. The current version of this framework was released in February 2014 as RDF 1.1 [56].

By now, RDF is considered as a formal language standard with well-defined inherent semantic for the description of structured information. In contrast to XML, it isn't about the syntactically correct representation of documents as a data *format*, but a data *model* for organizing information and relations:

> *"RDF is a datamodel for objects ("resources") and relations between them, provides a simple semantics for this datamodel, and these datamodels can be represented in an XML syntax." [57]*

With that – and as we will see later on, RDF resembles characteristics shown by classic data modelling approaches like UML class diagrams or Entity-Relationship-Models. However, there are also quite essential differences, which are required to provide actual semantic Linked Data.

The idea behind RDF itself is to make "statements" about so-called "resources". Each such RDF statement consists of a *subject*, a *predicate* and an *object*: the *subject* names the resource being described, the *predicate* what property of the resource is going to be described and last but not least, the *object* holds the value of this property itself. This combination is known as RDF "triple". For example, the sentence "the sky is blue" can be expressed as an RDF statement or triple – with "the sky" as subject, "is" as predicate and "blue" as the object. It is important to note, that RDF names especially subjects and predicates not arbitrarily and freely, but uses URIs instead for referencing things in a globally unique and consistent way. Such things can be resources or real-world objects, respectively, like a place or a book, but also abstract concepts, like relationships between those resources. URIs are used to *reference* those actually meant things in an abstract, but also machine-readable way. Additionally, an object can be a literal (like "blue" in "the sky is blue"), but also a URI itself, referencing other resources (like "Peter" in "Paul knows Peter", assuming that "Peter" is a

TABLE 9 – Tabular representation of exemplary RDF triples

Subject *(Resource Identifier)*	**Predicate** *(Property Name)*	**Object** *(Property Value)*
faust	title	*Faust. Eine Tragödie.*
faust	creator	*Johann Wolfgang von Goethe*
thisDiplomaThesis	title	*Linked Open Budget Data in Austria*
thisDiplomaThesis	creator	*Paul Blasl*

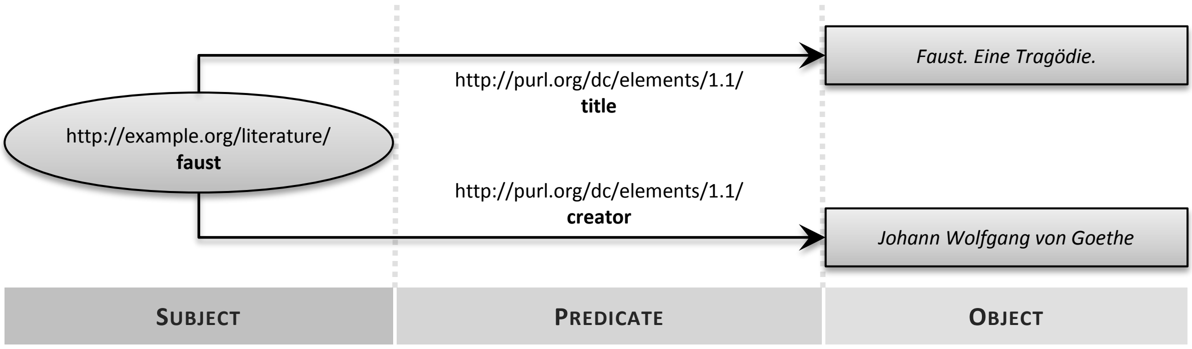
Note, that full URIs for subjects and predicates are omitted in this example for the sake of clarity and briefness.
Source: Own illustration

resource which is described further in additional statements). Examples for such RDF triples are listed in tabular form within **TABLE 9**. Using this structure and concept, almost everything can be described via RDF.

All in all, such RDF triples form intrinsically a labelled, unidirectional multigraph [56]. In other words, RDF's subjects, predicates and objects can be represented graphically; with subjects and objects as nodes and predicates as directed edges – all of them labelled accordingly with URIs and/or literals. Such an RDF graph can have multiple, distinct edges between two nodes, whereas these two nodes do *not* necessarily have to be distinct as well.[29] The later type of graph is also called a "multigraph".

By convention, resources are represented in illustrated RDF graphs as ellipses, in contrast to literals, which are normally depicted as rectangles. So, they are easily distinguishable. The example in **FIGURE 10** illustrates an RDF graph for the first two triples listed in **TABLE 9**.

FIGURE 10 – Graphical representation of exemplary RDF triples



Source: Own illustration

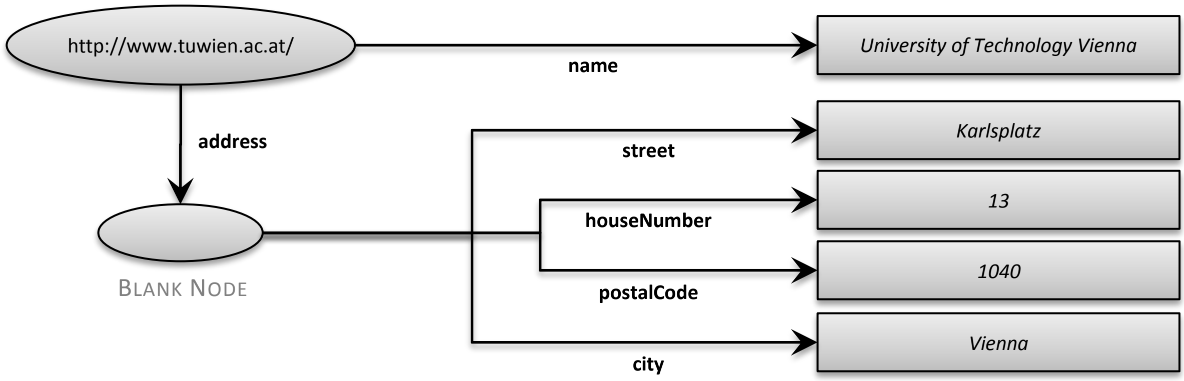
This very graph states for example, that "Johann Wolfgang von Goethe" is the creator[30] of a certain literary work referenced by the URI "`http://example.org/literature/faust`"[31]. As "Johann Wolfgang von Goethe" is modelled as literal in this example, he cannot be described further, because literals can never be inserted as a subject (but URIs can).

However, literals themselves are not necessarily "just" strings according to the RDF standard. They can be (and implicitly are always) denoted to a certain data type and/or spoken language. Those data types (such as strings, numbers and dates) are always compatible to those from XML schema, even if not defined as XML's built-in types. The way how actually this "annotation" of a data type or a language is done, is up to certain implementations of RDF, as described later on in this chapter. Nevertheless, most implementations also support so-called "simple literals" without any denoted data type and language (as in the examples in this very chapter). According to the RDF standard, in this case "`http://www.w3.org/2001/XMLSchema#string`" should be assumed as data type – without any statement of the used language.

Yet one conceptual element is missing for discussion of RDF. This specific framework is also aware of a mechanism used for grouping several statements about one single subject. This mechanism is called "blank node" or "anonymous resource". In fact, a statement's object can be a resource or blank node without any URI or literal at all (i.e. it is blank). This blank node in turn can be referenced as subject for multiple subsequent statements. Blank nodes can be looked upon as a grouping tool,

[29] This means in turn that a subject is able to make a statement about itself; being subject and object in the same RDF triple (like in "Paul knows himself" or – to be more precise – "Paul knows Paul").
[30] For this predicate (similar to the predicate referencing a "`title`"), in this example the standardized URI "`http://purl.org/dc/elements/1.1/creator`" from the "Dublin Core" project's metadata element set [139] is used.
[31] Referenced URIs for RDF resources need not be dereferenceable, in the sense that, if this URI is used as URL in a web browser, a valid web page would be returned. In general, in RDF any URI can be used as long as it is unique and consistent.

FIGURE 11 – Graphical representation of an RDF graph with a blank node



Note that, full exemplary URIs for all predicates are omitted in this example for the sake of clarity and briefness.
Source: Own illustration

possibly making a series of RDF statements clearer to comprehend. However, the W3C's standardization of RDF makes no reference to any internal structure of blank nodes or how they should be represented in possible implementations. As blank nodes are transitive (with no actual sematic content within themselves) and can be simply omitted (via merging the respective statements) in an extreme case, this happens to be no real problem whatsoever, actually; although, in certain realized implementations with their help more complex statements are possible then without them. However, this is *non-standardized* RDF-behaviour. Additionally, the implementation and usage of blank nodes may entail their own downsides and difficulties [58].

An example of an RDF graph with such a blank node is depicted in **FIGURE 11** above. In this example, all address related statements are made by the means of this blank node (referenced through the predicate "`address`"); hence, grouping those attributes in a meaningful way.

To summarize the main elements of RDF's data model in short: A subject may either be a URI reference or a blank node. A predicate is always a URI reference. And an object can be a URI reference, a literal as well as a blank node itself. All of them together form a RDF triple or statement.

As initially stated, RDF is a data model – not a data format. Therefore, means are needed to store RDF triples in some appropriate way for later retrieval. Hence, RDF graphs can be serialized textually. The first RDF serialization standard was introduced by the W3C as part of the first specification [55] of the RDF data model itself in 1999. Based inherently on the XML format, it is called "RDF/XML". An exemplary RDF serialization with this very syntax for the listed statements in **TABLE 9** is shown in **LISTING 1** below. As RDF relies heavily on XML concepts (like namespaces) as well as underlying concepts (like URIs), the usage of this format seemed to be just logical; nevertheless, as it became obvious the years after its specification, the "RDF/XML" syntax never became really popular. There

LISTING 1 – Exemplary RDF statements in serialized representation ("RDF/XML" syntax)

```
<rdf:RDF xmlns:rdf ="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:dc11="http://purl.org/dc/elements/1.1/">

    <rdf:Description rdf:about="http://example.org/literature/faust">
        <dc11:title>Faust. Eine Tragödie.</dc11:title>
        <dc11:creator>Johann Wolfgang von Goethe</dc11:creator>
    </rdf:Description>

    <rdf:Description rdf:about="http://example.org/literature/thisDiplomaThesis">
        <dc11:title>Linked Open Budget Data in Austria</dc11:title>
        <dc11:creator>Paul Blasl</dc11:creator>
    </rdf:Description>

</rdf:RDF>
```

Source: Own creation

exist several reasons for that. First and foremost, one reason is the potential complexity and the difficulty of processing and using this syntax at all, but if nothing else, it was also the complex structure for such simple statements with (always) only three elements in XML, which makes them difficult to comprehend for one at first glance due to XML overhead [59, p. 27] (cf. **LISTING 1**).

This is why new serialization standards for RDF were invented. By removing the XML overhead altogether, at first the "Notation 3" (N3) syntax was introduced (again) by Tim Berners-Lee himself in 2005 as a personal project ("with his director hat off") [60]. As the name suggests, this syntax was especially designed for RDF triples, omitting syntactical overhead as much as possible. Back then, Berners-Lee described it as a teaching language "basically equivalent to RDF in its XML syntax, but easier to scribble when getting started" [60].

However, N3 never became an official standard by the W3C or any other institution [59, p. 28]. Because full N3 syntax has an expressive power that goes potentially much beyond the RDF standard itself, in 2007 [61] Dave Beckett began to create the so-called "Terse RDF Triple Language" – usually better known as "Turtle" – as a subset of N3, which can only serialize valid RDF graphs (and no more). Like N3, "Turtle" does not rely on XML and is generally recognized as being more readable and easier to edit manually than its XML counterpart. **LISTING 2** below shows the serialized RDF graph from **LISTING 1** or **TABLE 9**, respectively, in "Turtle" syntax.

LISTING 2 – Exemplary RDF statements in serialized representation ("Turtle" syntax)

```
@prefix lit:  <http://example.org/literature/> .
@prefix dc:   <http://purl.org/dc/elements/1.1/> .

lit:faust               dc:title    "Faust. Eine Tragödie." ;
                        dc:creator  "Johann Wolfgang von Goethe" .
lit:thisDiplomaThesis   dc:title    "Linked Open Budget Data in Austria" ;
                        dc:creator  "Paul Blasl" .
```

Source: Own creation

Furthermore, string literals can be denoted in RDF to a certain language, as mentioned earlier. Additionally, literals can be declared as being of a certain data type. Below **LISTING 3** shows a serialized listing in "Turtle" syntax including some examples of such easily-made "annotations".

LISTING 3 – Exemplary RDF serialized statements ("Turtle" syntax) with denoted language and data types

```
@prefix lit:  <http://example.org/literature/> .
@prefix dc:   <http://purl.org/dc/elements/1.1/> .
@prefix xsd:  <http://www.w3.org/2001/XMLSchema#> .

lit:faust   dc:title    "Faust. Eine Tragödie."@de ;
            dc:creator  "Johann Wolfgang von Goethe" ;
            dc:date     "1829-01-19"^^xsd:date .
```

Source: Own creation

The tag "@de" means that the string "`Faust. Eine Tragödie.`" is written German language; "`^^xsd:date`" that the string "`1829-01-19`" is representing a chronological date.

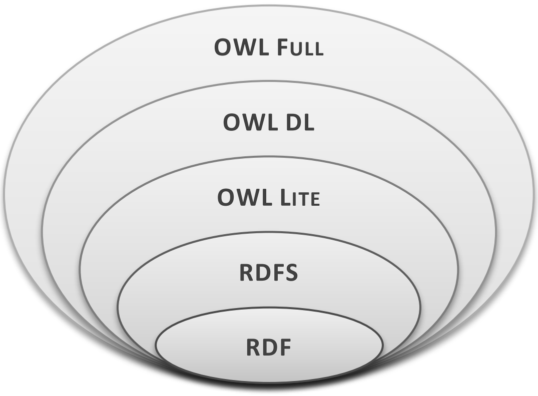
However, to this day still a lot of applications use the "RDF/XML" syntax, just because it was specified years earlier. It can be expected that other more suitable serialization syntaxes will replace "RDF/XML" in the years to come. Despite its name, in particular "Turtle" is moving the quickest in the race for popularity [59, p. 28] and is used already by more and more RDF applications. As a natural consequence, it has been standardized by the W3C as a recommendation [62] for RDF 1.1 in the year 2014.

Taxonomies & Ontologies with RDFS & OWL

Although RDF provides a model and syntax to describe resources, it does not give any indication about the actual meaning of the statements made. Furthermore, one can make up new URIs for all the resources and properties in one's own RDF triples, if one don't find existing ones; however, these leads to a potential high redundancy of (*what should be* consistent and unique) URIs.

For this reason, for Linked Data taxonomies and ontologies are created, building a formal naming and definition of some resources' types, their properties and the interconnections between entities of those types for various domains. This leads to vocabularies, hierarchies and relationships for a domain's resources, which can be automatically processed by applications – due to ontological defined meaning and semantic. In order to create such ontologies as well as taxonomies, appropriate definition languages are needed. For Linked Data, the most used ones are built on each other with RDF providing the fundamental basis – as depicted in **FIGURE 12** below.

FIGURE 12 – Hierarchy and semantic scale of RDF and respective meta-languages



Source: McGuinness & van Harmelen [57], own illustration

With the "Resource Description Framework Schema" (RDFS), a first draft of such a meta-language was published in April 1998 [63] – the same time around RDF's specification was underway; the final W3C recommendation at the beginning of 2004 with a recent revision as "RDF Schema 1.1" in the year 2014 [64].

In general, RDFS provides the means to create simple classifications – as well as hierarchies of them – for building taxonomies and to declare basic relationships in between them. **LISTING 4** below shows such an exemplary taxonomy in RDFS syntax. Based on this schema, statements in RDF can be made and due to the declared taxonomy, it can implicitly deduced – for example – that statements valid for "Pets" also apply to "Dogs" (being a subclass) as well as that everything that "eats"

LISTING 4 – An exemplary taxonomy written in RDFS

```
<rdf:RDF xmlns:rdf ="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">

    <rdfs:Class rdf:ID="Pet"/>

    <rdfs:Class rdf:ID="Dog">
        <rdfs:subClassOf rdf:resource="#Pet"/>
    </rdfs:Class>

    <rdfs:Class rdf:ID="Feed"/>

    <rdf:Property rdf:ID="eats">
        <rdfs:domain rdf:resource="#Pet"/>
        <rdfs:range rdf:resource="#Feed"/>
    </rdf:Property>

</rdf:RDF>
```

Source: Own creation

something is an “`Animal`”, although this has not to be explicitly declared in those statements. It is important to note, that exactly this very concept of deducing implicit statements (as they may become far more complex than in the mentioned example) trough taxonomies (as well as ontologies later on) is one of the major features in overlying technological layers in SEMANTIC WEB CAKE (especially for reasoners). Furthermore, one may notice that the mentioned RDFS example in **LISTING 4** is written in XML again. In contrast to RDF itself and its respective statements, it is still quite common to declare taxonomies and ontologies in XML, although they could be expressed also in syntaxes like “Turtle” anyway (because all meta-languages are full RDF, too). One reason may be that ontologies need *not* to be human-readable in most parts, anyway, as they are made for information interchange between applications primarily (e.g. for verification of a domain’s statements and respective deductions).

However, RDFS is not sufficient to declare complex ontologies with more complex relations and respective restrictions. For example, it is impossible to declare via RDFS that a dog cannot be a cat at the same time (although being an animal anyway) or that every animal has to eat at least one feed (to prevent starvation). This is called disjointness or a cardinality constraint, respectively. Both concepts – as well as others – are not able to be modelled in RDFS. Hence, RDFS is used as description language for less complex applications or ontologies, respectively, like the “Dublin Core” project’s metadata element set[32], where such, more complex concepts are not necessary.

Therefore, the so-called “Web Ontology Language” (OWL) was introduced to overcome these restrictions of RDFS. Like RDFS, the W3C was the institution to standardize yet another semantic meta-language as a recommendation. A first draft of its features was published in July 2002 [65]; the final release took place (interestingly) at the very same day as for RDFS’ recommendation in February 2004 [57]. In 2009, a revised version of the ontology language as introduced as “OWL 2” by the W3C, which was again revised in a second edition in late 2012 [66].

OWL is based on and compatible to RDFS[33] with the same syntactical foundation of XML. However, it “adds more vocabulary for describing properties and classes: among others, relations between classes (e.g. disjointness), cardinality (e.g. "exactly one"), equality, richer typing of properties, characteristics of properties (e.g. symmetry), and enumerated classes” [57].

For different use cases as well as different computational and semantic requirements, three increasingly expressive sublanguages of OWL were designed from the very beginning [57]:

→ *OWL Lite* was introduced as easily implementable description language with lower formal complexity as its sisters. Mainly for building taxonomies with simple constraints, it should be easier to provide tool support for OWL Lite than for its more expressive relatives.

→ *OWL DL* – named due to its correspondence to description logic – provides maximum expressiveness while retaining computational completeness (all conclusions are guaranteed to be computable) and decidability (all computations will finish in finite time). Therefore, while it includes all OWL language constructs, they can only be used under certain restrictions.

→ *OWL Full* is designed to be fully equivalent to semantic expressiveness of RDFS – omitting all restrictions of OWL DL. However, with this freedom no computational guarantees can be assured.

Each of these sublanguages of OWL is a superset or extension of its simpler predecessor. Hence, every valid ontology written with OWL Lite is a valid ontology as of OWL Full, while the reverse is not

[32] This ontology [139] was used in the first part of this chapter for description of literary relationships (like for the predicates “`title`” & “`creator`”).
[33] OWL even uses some of its features, like the predicates declaring taxonomies – e.g. everything depicted in **LISTING 4**.

(necessarily) true. Furthermore, it is important to understand that – expect for OWL Full – OWL is covering only a restricted set of concepts and possibilities of RDF or RDFS. In other words, every statement made by the means of RDF or RDFS can be expressed as valid ontology in the sense of OWL Full; however, not necessarily of OWL DL or even OWL Lite. Nevertheless, this self-curtailment was made intentionally, because statements made with ontologies using all features of OWL Full may become undecidable: existing reasoners may not be able to make all possible deductions – or a complete reasoning as such. Additionally, implementations of such reasoner supporting OWL Full do not exist to this date and it is unlikely that any reasoning software will be able to support complete reasoning for every feature of OWL Full in the future [57].

As OWL ontologies are quite complex and its language constructs are manifold, no example is given at this point, as in the course of this work an existing ontology based on OWL will be used anyway.

However, when creating a new ontology for a specific domain, it is obvious to choose the correct description language – RDFS or any of OWL's sublanguages; especially, if the subsequently emerging Linked Data should support reasoners and, thereby, the full potential of the Semantic Web.

SPARQL Protocol and RDF Query Language – SPARQL

As mentioned in the previous sub-chapter 2.3.1, some form of querying language is necessary in addition to the standards discussed so far for retrieval of already stored sematic Linked Data in respective applications. By now, the "SPARQL Protocol and RDF Query Language" (SPARQL) has become prevalent for querying sets of RDF statements. Again, the W3C was the main driver behind this standard's development and its standardization: A first draft was published in late 2004 [67] – just after the final releases for RDFS and OWL; the one of SPARQL itself as a recommendation happened in early 2008 [68]. Therefore, SPARQL can be considered as the youngest standard among the Linked Data technologies. In spring 2013, SPARQL was revised as version 1.1 [69].

Although SPARQL resembles certain characteristics of querying languages used for relational databases (like SQL), it is designed to operate on data that is organized in a directed, labelled graph – more specifically, RDF's data model – as sets of "subject-predicate-object" triples. So in fact, SPARQL has more in common with querying languages for "document-key-value" stores (like NoSQL) than with others.

Most kinds of SPARQL queries describe one or multiple (if necessary, optional) triple patterns, which are also known as "basic graph pattern" and are like RDF statements or triples, respectively, except that each of the subjects, predicates or objects may be a variable. If a basic graph pattern matches a subgraph of a total set of RDF statements (so, if this subgraph may be substituted for the variables and the resulting RDF graph is equivalent to the original subgraph), a result matching the actual query is found. All matching results are called "solution sequence". As a matter of fact, there may be zero, one or multiple solutions to each query. Due to the nature of the query's graphical represented underlying data, this list of solutions is possibly unordered and not all included values may be of the same data type. It is important to note that SPARQL is orientated on properties (of the queried RDF graph and its included triples) rather than on actual classes – this concept can be compared to the one called "Duck Typing"[34] in object-oriented programming languages.

Other than that, SPARQL provides a full set of analytic query operations such as for joining or sorting data. Additionally, it supports different kinds of queries for retrieving raw value data (`SELECT` query),

[34] When using "Duck Typing" the actual type of an object is not described by its class, but rather by its attributes and methods. This term originates from the so-called "Duck Test" which describes abductive reasoning. It is usually described with the following expression: "If it looks like a duck, swims like a duck and quacks like a duck, then it probably is a duck."

subgraphs as valid RDF graphs on their own (CONSTRUCT query) as well as querying simple truth statements (ASK query) and descriptions (in RDF) of data found (DESCRIBE query). One of the major features of the most recent version 1.1 of SPARQL, is the possibility to also alter – add, replace and delete – underlying datasets or RDF statements and subgraphs through respective SPARQL queries.

In **LISTING 5** below, an exemplary SPARQL query is shown, which is designed to query the data used in previous examples in this very sub-chapter (like **TABLE 9** & **LISTING 1**). Although this example is deliberately kept simple, it includes multiple characteristic language elements of SPARQL.

LISTING 5 – An exemplary SPARQL query with several typical elements

```
PREFIX dc: <http://purl.org/dc/elements/1.1/>

SELECT ?uri ?creator ?title
WHERE
{
    ?uri    dc:creator  ?creator ;
            dc:title    ?title .
    FILTER (?creator = "Johann Wolfgang von Goethe")
}
ORDER BY ASC(?creator)
LIMIT 1
```

Source: Own creation

Described in natural language, the query in **LISTING 5** should find a subgraph with RDF triples containing the “`creator`” and “`title`” predicate with exactly the same subject (`?uri`); the object within the triple containing “`creator`” as predicate should represent the literal string “Johann Wolfgang von Goethe”. Certain elements are selected as result (`?uri`, `?creator` & `?title`). These results as raw data for a query on the RDF data from **LISTING 2** would look like this:

FIGURE 13 – An exemplary result for a SPARQL query

```
-----------------------------------------------------------------------------------------------
| uri                                     | creator                      | title                |
===============================================================================================
| <http://example.org/literature/faust>   | "Johann Wolfgang von Goethe" | "Faust. Eine Tragödie." |
-----------------------------------------------------------------------------------------------
```

Source: Own illustration

One unanswered question is *where* the RDF data is actually stored, when it is queried via SPARQL. Of course, small sets of RDF triples can be easily stored in serialized form (like in the discussed “Turtle” syntax) and be queried from there by modern SPARQL implementations – so-called “processors”. Nevertheless, if you want to handle big amount of data in form of triples, it is necessary to store it in a special database also known as “triplestore”. Like relational database management systems, such “triplestores” index data and decide which data to load into memory and when; however, they are especially optimized for the storage and retrieval of eponymous triples. Both commercial and Open Source “triplestores” are available nowadays. All of them providing certain starting points or interfaces for querying and retrieving results from underlying datasets; these are commonly called “SPARQL endpoints”.

With SPARQL, it is also possible to distribute a query to multiple such endpoints at the same time – joining results from various data sources and domains; a procedure known as “federated query”. This feature was introduced with SPARQL 1.1 too. As data can be equipped with self-defined semantics and meaning via RDF, SPARQL is predestined to combine potentially incoherent data sources; hence, its federated queries provide the means to do so.

2.3.3 The composition of Linked Data

While being part of the holistic Semantic Web vision, Linked Data provides possible technical means to reach this vision as mentioned in the previous sub-chapter 2.3.1. Thereby, the concept of Linked Data itself concentrates basically on publishing semantical annotated and interlinked data:

> *"Linked Data refers to data published on the Web in such a way that it is machine-readable, its meaning is explicitly defined, it is linked to other external data sets, and can in turn be linked to from external data sets." [42, p. 2]*

With that, higher level goals and concepts of the Semantic Web (like the upper parts of the SEMANTIC WEB CAKE) are neglected with Linked Data, as a consequence. However, this is done – as already stated – intentionally for the sake of simplicity for eventual data providers. For the purpose of this Diploma Thesis, exactly this simplification is required, as – later on – the focus will be on municipal budgets' data model, format, creation and its retrieval, but *not* on (possible) subsequent processes.

Nevertheless, various things have to be considered before creating such data. For this reason, in his seminal design note from 2006, where he coined the term of Linked Data, Tim Berners-Lee already laid out four rules – nowadays known as principles – for creating and publishing Linked Data [41]:

1. Use URIs to name and identify things in a unique and consistent way.
2. Use HTTP URIs to provide the means to look up those names (making them *dereferenceable*[36]).
3. Provide useful information, using open standards (like RDF), when a URI is looked up.
4. Refer and interlink to other things with URIs, so that, people can discover more things.

The interested reader may notice that these four principles can be easily met with the technologies described so far. However, as these basic principle also suggest, Linked Data is more than an arbitrary accumulation of various technologies; instead, it is a concept for publishing data in a specific way. Berners-Lee himself stated out that violating the specifically designed principles would *not* break *anything*. However, the respective data provider would miss a chance for interlinking data for later reuse. So, he called the principles also "expectations of behaviour" [41] for Linked Data.

With respect to the principles themselves, they are serving multiple purposes: First of all, they ensure user- as well as machine-readability (rule 2 & 3). One as the other should obtain useful information through dereferenceable unique and consistent names (URIs) using the existing World Wide Web – HTML for the human beings, RDF (building on well-defined taxonomies and ontologies via RDFS and OWL) for machines. Other than that, it shall be ensured (rule 4) that the overall vision of a *Web of Data* – in addition to a *Web of Documents* – can be reached at some point.

URI Design Issues

Despite that the technologies discussed so far provide the technical foundation to create data with self-defined semantics and meaning, one can easily see that the identifiers for the data themselves, the URIs, are of primary interest when it comes to creating appropriate Linked Data. Their usage in combination with – according to Berners-Lee's four principles for Linked Data – the "expectation", that they are dereferenceable on the World Wide Web using an everyday web browser, is anything but simple. As it is a substantial part of this work to design and create Linked Data, this issue shall be addressed further in the following.

[35] This is the Answer to the Ultimate Question of Life, the Universe, and Everything. (DON'T PANIC.)

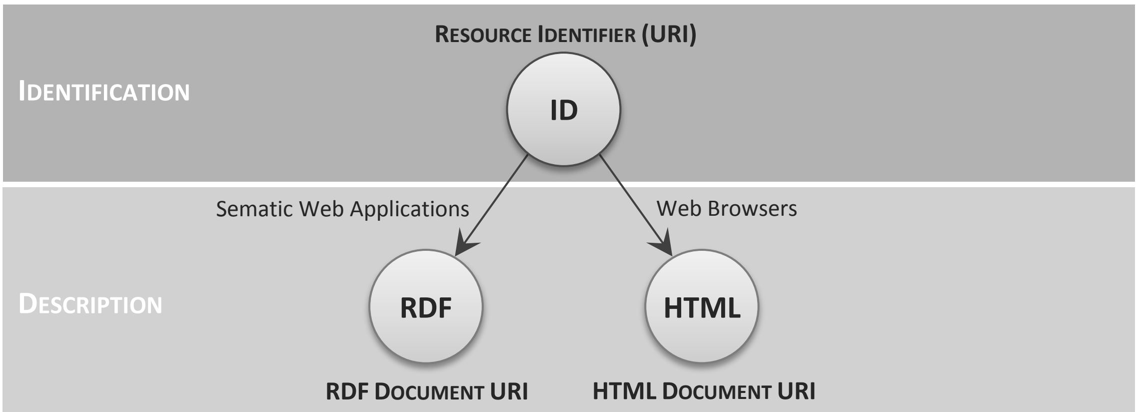
[36] While URIs can "reference" nearly anything (like a real-world thing or just a web document), they can be "dereferenced" for retrieving the actual object (e.g. the web document itself via a web browser). Hence, the verb *dereferencing* describes the process to obtain something, at which a pointer (the URI) points; if something is *dereferenceable*, it is possible to do so.

First and foremost, the W3C itself suggests [70] that developers should *only* create URIs for Internet domains they also control. If URIs are used, which one does *not* control, in most cases it will obviously become impossible to comply with Berners-Lee's four principles and actually provide dereferenceable data for those URIs as well. And, if nothing else, one cannot commit to the domain's permanence, when a third party possibly abandons it after some time.

Furthermore and more importantly, as URIs are used to reference absolutely anything in the sense of Linked Data and not just web documents (e.g. real-world objects like a place or a book, but also abstract concepts like relationships between those objects, or even non-existing things like a unicorn), it is important to know what a dereferenceable URI actually identifies or what it refers to. According to the second and third rule of Berners-Lee's four principles [41], it is expected that a URI used in Linked Data should be dereferenceable via HTTP and a document or data obtained in this way shall provide – both human- and machine-readable – information, which describes the resource that is identified by this very URI. However, if the URI *identifying* a real-world object is equal to the URI *describing* the very same object (for conformity with the four principles), this leads to a major issue of this URI itself, which should be unambiguously identifying the object and nothing more. For example, if "`http://www.example.com/id/alice`" identifies a person called "Alice", then it would be *absolutely* inappropriate that a server being addressed via HTTP for dereferencing this URI responds with a HTTP status "`200 OK`" as the server cannot deliver Alice herself, but only a description or data about her, respectively.

Obviously, the requirements of URIs and by Berners-Lee's four principles or Linked Data itself conflict with each other at this point. A technical solution is needed, so that the URI identifying a resource can be separated from the URI(s) describing and containing data for the very same resource – while providing a "link" to the latter for dereferencing the URI via HTTP. The resulting desired relationship between a resource and its describing documents is depicted in **FIGURE 14** below.

FIGURE 14 – The desired relationship between Linked Data URIs identifying and describing the same resource



Source: Ayers & Völkel [71], own illustration

Although, the describing document (and its respective URI) can be the same for both human ones as well as machines[37], more often they will be served in separate documents – providing more flexibility regarding data representation and encoding. This fact is also reflected by **FIGURE 14**.

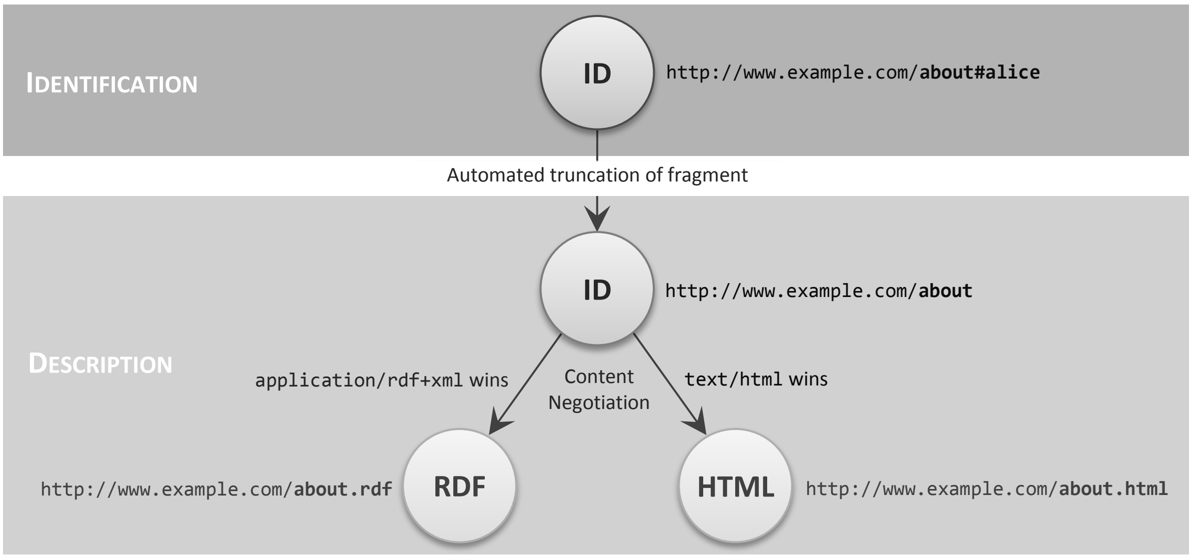
In a note for implementers of RDF [71], the W3C proposes two technical solutions to this very problem: implementing identifying URIs as "Hash URIs" or "303 URIs". Both have their own advantages and disadvantages; and both shall be discussed in detail at this point.

[37] Semantical annotated (meta-)data can be also embedded into HTML using standards like "RDFa" or "microformats", from which RDF data can be extracted. However, as these technologies are not used within this work, there will be no further discussion at this point.

The first approach using *Hash URIs* for identifying non-document resources makes use of HTTP and the structure of URIs themselves. A "fragment identifier" added to a URI (and separated by a hash sign "#" from the rest of a URI) allows indirect identification of a secondary resource while referencing a primary resource and giving additional identification information. This fragment – as depicted also in **FIGURE 9** in chapter 2.3.1 – has to be stripped off prior to dereferencing a URI by the client, and thus the identification of the secondary resource (within the primary resource) is also solely managed by this very client [51, p. 25]; the server *must not* be informed about this fragment [72, p. 45].

In the sense of Linked Data, that means, if a client dereferences a URI identifying a real-world object using a *Hash URI*, it has to remove the URI fragment before requesting the resource. Therefore, this object cannot be retrieved directly, and with that, does not necessarily identify a web document. So, in this way non-document resources can be identified without creating ambiguity.

FIGURE 15 – Using URI fragments for separation of URIs used for identification and description of a resource



Source: Ayers & Völkel [71], own illustration

Above, **FIGURE 15** illustrates this procedure including a process called "content negotiation", which is used to automatically and appropriately deliver human-readable data (in this example, HTML) or machine-readable data (RDF) to the client. This process is specified in detail in the specification of HTTP [72, p. 37 ff.]. At this point, it is only worth mentioning, that the client can specify via the "`Accept`" HTTP header field response media types that are acceptable for it (like HTML or RDF). It is also possible to define multiple data type including an (optional) weight for an indication of preferences. However, of course, the server has to be appropriately configured to support this content negotiation as intended. This process is also possible for other data types used in the World Wide Web or Linked Data (in addition to, or without RDF or HTML).

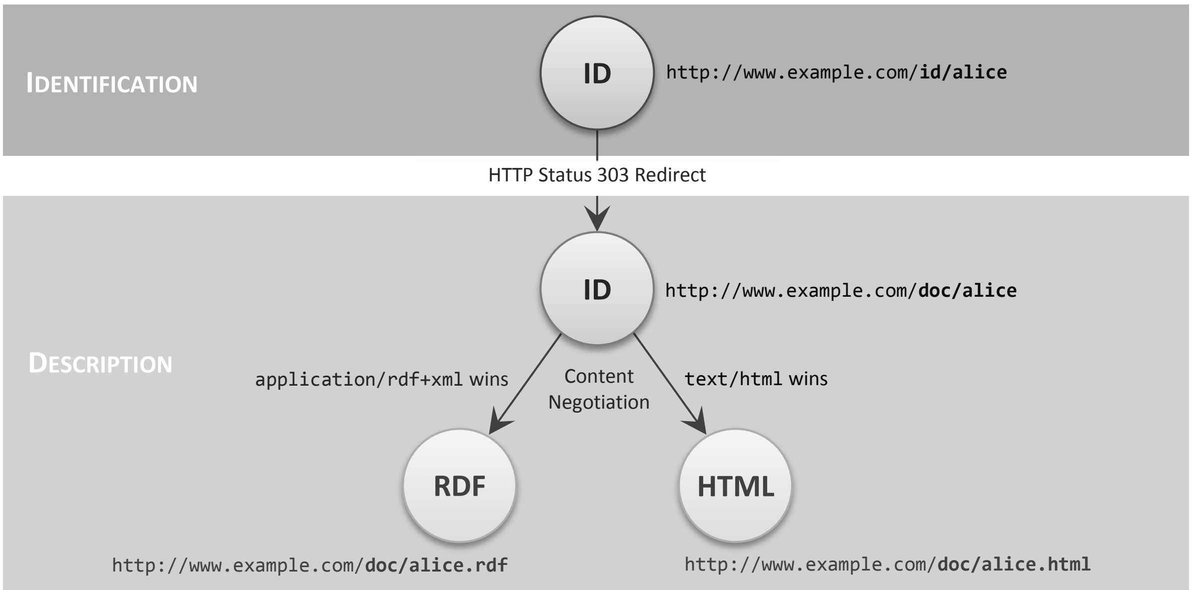
Anyway, **FIGURE 15** defines "`http://www.example.com/about#alice`" as URI identifying Alice and "`http://www.example.com/about`" providing a document with information about her – and probably others – in various (automatically negotiated) data formats. The stripped fragment can be subsequently used by the client to find this very information within the returned document.

Other than that, the second approach using *303 URIs* takes the server up on its promise. In this case, the server responds to requests relating to identifying URIs with the HTTP Status Code "303". In general, this "`303 See Other`" status code indicates that the server wants to redirect the client to a different resource, which is given as another URI in the "`Location`" HTTP header field and shall give an indirect response to the original request. Usually, this procedure is used after a HTTP POST request for redirecting the client to a representation of the just posted data. With respect to Linked Data, the specification of HTTP clearly gives also the possibility to use this procedure for identifying URIs too:

> *"A 303 response to a GET request indicates that the origin server does not have a representation of the target resource that can be transferred by the server over HTTP. However, the `Location` field value refers to a resource that is `descriptive` of the target resource, such that making a retrieval request on that other resource might result in a representation that is useful to recipients without implying that it represents the original target resource." [72, p. 57]*

So, for example, as the server cannot deliver Alice as a person in a suitable representation after a GET request for the URI "`http://www.example.com/id/alice`", which identifies her in person merely, the server can redirect the client to "a resource that is *descriptive* of the target resource" through responding with HTTP Status Code "303" to the original request. If subsequently the same content negotiation upon the request of the new URI is made, this mechanism called using *303 URIs* for Linked Data – as depicted in **FIGURE 16** – is equivalent to the one using *Hash URIs* in its result.

FIGURE 16 – Using HTTP Status 303 Redirects for separation of URIs used for identification and description of a resource



Source: Ayers & Völkel [71], own illustration

Similar to **FIGURE 15**, **FIGURE 16** defines "`http://www.example.com/id/alice`" as URI identifying Alice and "`http://www.example.com/doc/alice`" providing a document with information about her – and this time, only her – again in various (automatically negotiated) data formats. If the content of these documents in those representations (like the depicted RDF and HTML document) about the original resource differ substantially, the server should – with content negotiation right at the identifying URI – immediately redirect the client to separate descriptive URIs (e.g. for a machine-readable document to "`http://www.example.com/doc/alice`" and for the human-readable one to "`http://www.example.com/people/alice`") making content negotiation there obsolete. In this case, those URIs refer not two versions of the same document, but different documents altogether [71]. This procedure holds true also for *Hash URIs*, if (and only if) the descriptive documents differ substantially.

Which one of the two discussed approaches is more suitable for Linked Data, depends on the specific application. While *Hash URIs* have the advantage of reduced communication overhead through serving multiple resources via one single URI and request, which in turn reduces access latency, this advantage comes with (possibly) high data loads as multiple descriptions of other resources are delivered in conjunction with the desired one, because they are in the same document. On the other hand, *303 URIs* are more flexible because the redirection for each identifying URI can be configured separately on the server. Additionally, this policy can be changed easily and centrally later on. However, a client looking up several resources may trigger a lot of redirects. Hence, a high number of those redirects may cause in turn higher latency and server load as well. Of course, providing a SPARQL endpoint for automated data retrievals can overcome this issue – but only to a certain extent, as dereferencing of such URI will still take place by humans (especially for popular domains).

Apart from this, URIs used for Linked Data in general shall be designed ensuring certain characteristics in mind as a note for implementers of RDF by the W3C states clearly [71]:

→ *Simplicity* comes with the usage of short, mnemonic URIs. Although there is no (relevant) technical limitation when it comes to length of URIs, the structure of URIs should be simple and unambiguous.

→ *Stability* means that a URI – once being established – should remain this way as long as possible. As technologies may change in the future, implementation-specific parts (like file extensions, e.g. `.rdf`, `.html` or `.php`) should be omitted for client usage. Also, versioning (with alphanumeric identifiers like `v1`, etc.) is not recommended. Identifying URIs should be unique, consistent and able to live forever. Such URIs are also called "Neutral URIs" [70].

→ *Manageability* is curial as URIs controlled by one shall be issued also in way that one can manage them in a convenient way. For example, keeping all identifying URIs on a separate sub-domain (e.g. "`id.example.com`") can simplify server migrations in the future.

With that in mind, the four principles of Linked Data as well as overall usability and convenience of used URIs, that reference and identify domain-specific resources, can be ensured easily by respective applications.

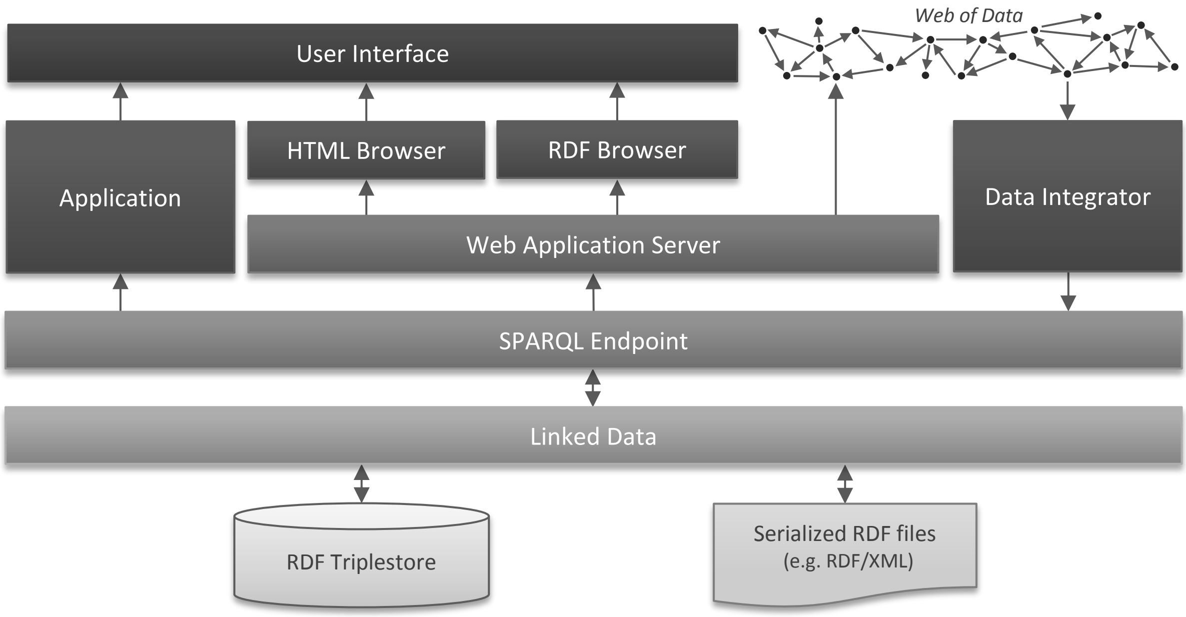
Software Architecture of Linked Data Applications

So far, the various elements of Linked Data and their interaction with each other have been discussed in detail. However, one last thing has to be elaborated before moving on: the composition of all those elements into a coherent architecture found in relevant best practices and scientific works that can be used in this specific work later on. As the interested reader might have noticed, a lot of the interrelationships of the discussed technologies and standards have been laid out already. Nevertheless, up to this point *no* overall picture has been given. As this is crucial for the rest of this Diploma Thesis and definition of its scope, such an overview shall be presented now in a nutshell.

The overall architecture of (possible) best-practice applications for Linked Data has been of interest in a lot of relevant literature so far (like in [73], [70] or [46]). In addition, the SEMANTIC WEB CAKE gives an overview for the overall picture and the technical possibilities. However, these works on those topics only provide a fragmentary, complicated, or (often & in some parts) overly detailed and fine-grained architectural discussion to some extent. For this reason, a general overview of typical software architecture for Linked Data shall be created at this point, which can be used later on.

First of all, it is important to note that such an application can follow two strategies: publishing or consuming data. There already exist several applications in this field – also in particular regarding governmental, and specifically, for budgetary data as seen in chapter 3.1 – implementing one or even

FIGURE 17 – Typical overall abstract architecture of applications for Linked Data



Source: Own illustration

both strategies. Furthermore, an application, that actually publishes Linked Data, can do this for two purposes: on the one hand, for the good of all and a *Web of Data* via an open interface (cf. chapter 2.4.4); and on the other hand, for the sake of its own application and respective individual representations through a stand-alone user interface (or multiple ones). Note, that the latter part is optional: as long the published information is useful, no independent application is necessary, essentially. The foundation of all this is an independent storage of Linked Data hosted by the application itself.

A graphical representation of this overview is given in **FIGURE 17**, where RDF triplestores and/or serialized data (in syntaxes like "RDF/XML" or "Turtle") are the backbone for a SPARQL endpoint providing data creation, edit and retrieval services for saved Linked Data. An overlying independent application or webserver provides the means for a user interface based on this foundation (like HTML and RDF browsers or more sophisticated, independent applications). Additionally, the webserver can publish new Linked Data to the *Web of Data* for further third-party usage. From this origin, also data can be obtained (obviously) for own usage (via the means of a special so-called "data integrator") and deducing new information, which can be published in turn in some case.

Of course, not all parts of the elements shown in **FIGURE 17** are necessary for a typical Linked Data application (e.g. an application can only *publish* data without *consuming* any or vice versa). However, it gives a short overview of the general software architecture of typical Linked Data applications.

Anyway, there are always two sides of the same coin. As the reader might have noticed, this very chapter concentrated on the detailed *publication* of Linked Data for the most part. But Linked Data is about *publishing* *as well as* *consuming* data in a specific way. There would be also a lot more to tell about the latter part, but this would go far beyond the scope of this very Diploma Thesis. At this point, it is only worth mentioning, that as complicated as creating Linked Data can get, also consuming this data can be. Nevertheless, Linked Data provides manifold possibilities like the Semantic Web vision itself.

2.4 DISCLOSURE OF GOVERNMENTAL BUDGET DATA

Last but not least, the disclosure of digital data in general but also with respect to governmental, including budgetary data shall be addressed in the final part of the second chapter of this Diploma Thesis. With that, the complete theoretical foundation for the rest of this work and subsequent usage will have been elaborated after that.

The mentioned disclosure of data or its result, respectively, is commonly known as "Open Data" – eponymous for the title of this work. In the following sub-chapter, this term as well as its derivates (including – most importantly – "Linked Open Budget Data") is defined in detail. Additionally, relevant principles, standards and guidelines, which will be also necessary later on, are laid out. Furthermore, also the economical but also general benefits of "Open Data" in certain parts will be presented on the basis of relevant scientific literature.

2.4.1 Open Data

The human culture is shaped by the desire to enhance existing things. Everyone builds upon those things created by others. In the digital world, this is now easier than ever: Digital data can be retrieved, used, modified and shared again with minimal effort – not only by a certain group of people, instead (theoretically) by *everyone* – as physical barriers (and distances) are neither here nor there. However, there are laws – basically all around the world – which (can) guarantee the rights of one, who has created something in the first place – the so-called "author". These author's rights are also known as "copyright" in the English-speaking world. Additionally, the data itself can be encoded and distributed in such a way, that subsequent usage by others can be hampered, or even be impossible due to its structure. The concept of "Open Data" tries to introduce standards and guidelines in order to overcome these issues – among others – through elimination of the last barriers of data exchange in today's society for the sake of all and enhancement of all our knowledge in particular.

The idea of Open Data's concept is not new. It originates for digital contents from the idea of "Open Source" [74, p. 6], which relates to software, whose source code is freely available for everyone. The reason for creating such Open Source software in the first place was that developers did not want to start from scratch each time implementing a new application. Instead, there was a certain need to build upon source code by others – if possible without any (commercial or legal) hurdles. The success of this Open Source-principle is based mostly on the fact that it does explicitly allow also commercial re-use [74, p. 8]. With Open Data, the footsteps of Open Source shall be followed for digital data in general, possibly achieving similar popularity as its forefather.

While the general concept of Open Data in particular seem to be quite obvious at first glance, an explicit definition is not that easy. In fact, there is no generally accepted or standardized definition for Open Data. In relevant literature different kind of definitions are made – most of them are vague with a lot of room for interpretation. However, there are institutions like OPEN KNOWLEDGE INTERNATIONAL (OKI) [38] that provide a quite holistic and sophisticated definition regarding this matter. As a global non-profit organisation focusing on *open* knowledge, its OPEN DEFINITION is a well-known policy to define "open" with respect to digital content and data. This definition is used and followed within the rest of this very work too. Therefore, it shall be presented at this point in a nutshell.

[38] OKI was formerly known as "Open Knowledge Foundation" (OKF) and renamed in May 2016 [140].

The Definition of Openness

OKI's Open Definition is about works published in specific conditions (e.g. a data format) under certain circumstances (e.g. licences). A "work" – according to this definition – is a piece of knowledge being transferred. A "licence" in this sense is a legal condition under which a work is published and provided to others. A term often raised in this context is "public domain", which "denotes the absence of copyright and similar restrictions, whether by default or waiver of all such conditions" [75] – regardless of the actual regional legal possibilities as seen later on.

An "*open* work" must show following characteristics according to this Open Definition [75]:

→ The work must be in an open *legal state* – either dedicated to public domain or provided under an "open license" (see below).
→ The *access* to a work must be provided in a sense that it is provided always in its entirety and at no more than a reasonable one-time reproduction cost. For this reason, the work *should* be available through the Internet without charge; however, there is *no* obligation for that.
→ The work must be published in a machine-readable *structure*, in which the individual elements of the work can be easily accessed and modified.
→ The *format* of this work has to be open in way so that no restriction – monetary or otherwise – is placed upon its usage. Hence, it can be fully processed with at least one Open Source tool.

In addition, an "open licence" in the sense of the first bullet point of the list above must satisfy following (required) conditions[39] according to the Open Definition [75]:

→ The licence must allow *free use* of the licenced work.
→ The licence must allow *redistribution* of the licenced work (in its original or a derived form).
→ The licence must allow *modification* of the licenced work (including its redistribution).
→ The licence must allow *separation* of parts of the licenced work (incl. their distribution).
→ The licence must allow *compilation* of the licenced work with others (without placing restrictions on each of those other works).
→ The licence must allow *no discrimination* against any person or group.
→ The licence must allow *propagation* without the need to agree to any additional legal terms.
→ The licence must allow *application* of the licenced work for any purpose.
→ The licence must allow *usage at no charge* – without any fee, royalty or other compensation.

All in all, these extensive definitions of characteristics of openness can be summarized in the following statement also made by OKI itself [75]:

> *"Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness)." [75]*

The first draft of this Open Definition was released by OKI in August 2005; the first official release took place one year later as Version 1.0. The most recent revision of the definition was published lately in November 2015 [75].

As seen so far, the definition of Open Data strongly implies legal as well as structural considerations. The later (with open access to data in an open format) has been discussed extensively with Linked Data in the previous chapter 2.3. However, the legal aspect has been somewhat omitted up to this point. With an overview of relevant Austrian laws as well as available open licences in the sense of the Open Definition above, the definition of Open Data shall be rounded off in the following.

---

[39] OKI's Open Definition knows also acceptable restrictions to these conditions, like the requirement for attribution of creators, integrity, share-alike (when distributions of the work must remain under the same or similar license) and others.

Legal aspects of Openness

Within Austria's legislation, the most relevant regulations regarding Open Data in particular are the ones addressing copyright as well as data protection. Although, various other laws and regulations exist – on national level as well as by the European Union – which may be relevant for Open Data, this work concentrates its respective discussion on the national ones in order to stay within the scope of this Diploma Thesis.

First of all, one's rights as an author in particular are established by national law in most parts. For those laws the principle of territoriality holds [76]; so, for example: Austrian legislation on copyright applies to Austrian territory; German legislation on the same matter only to German territory, and so on. Hence, this copyright can only be enforced by the means of the respective national law, where legal protection is claimed by an author (so-called "lex loci protectionis"). In turn, this means, the law of the country, where something was created or registered, can*not* be applied in general. So, in case one's rights as an author are violated, only the copyright of the respective country is applied, where infringement took place – regardless where the author and his creation are originating. Although, there does not exist any global obligation for it, the principle of "lex loci protectionis" is generally accepted as the prevailing rule for choosing law in the field of copyright law – at least as far as it concerns the existence, validity, scope and duration of those regulations [77, p. 57].

As this work concentrates on the disclosure of budgetary data of *Austrian* municipals primarily for use in *Austria*, only *Austria's* legislation on copyrights shall be addressed extensively in the following. Apart from that, a uniform level of protection can be assumed – at least within the borders of the European Union – due to the progressing harmonisation of national regulations [76]. In Austria itself, the respective regulations are established by the national Copyright Law or URHEBERRECHTSGESETZ (UrhG). It originates from the period of Austria's first republic and was issued for the first time in 1936. The most recent revision of the UrhG was enacted as applicable law in October 2015.

The UrhG itself has its own specific definition for works that are protected by this national copyright. According to Article 1, paragraph 1 UrhG [78], works are unique, intellectual creations in the fields of literature, musical arts, fine arts or cinematics. The respective requirements on uniqueness and intellectuality are rather low. So, with that, nearly everything that differs from the ordinary is probably protected by this law [76]. Besides computer software in general (which are classified by the UrhG as works of literature), raw digital data – in the form of databases – is also affected by the UrhG: databases are treated as compilations worth protecting through Article 6 and 40f, respectively.

But what's the point of copyright with respect to Open Data? In short, the UrhG explicitly states that the author is entitled *solely* to use, modify, reproduce and distribute his own work (Articles 14, 15, 16, 18a & 21). Therefore, the legislative's default mode (in Austria, but also in most European countries) is "all rights reserved". So, if an author states nothing else, his work can*not* be used by *anyone*. That obviously contrasts with the definition of compiled Open Data as declared in the OPEN DEFINITION as described before. However, the UrhG provides the possibility to decidedly permit the use of a copyrighted work ("Werknutzungsbewilligung" – Article 24, paragraph 1). This permission of an author acts relatively and enables others (besides the author) in possession of this permission to use, modify, reproduce and distribute the work within the limits the respective permission grants. Such permissions correspond with licences in the sense of the OPEN DEFINITION. Additionally, the UrhG knows an exclusive permission for usage a copy righted work, which acts absolutely ("Werknutzungsrecht" – Article 24, paragraph 2) – however, this permission would *not* comply with the definition of an open licence due to its (exclusive) nature. Other than that, the UrhG does not allow the *complete* transfer or release of an author's copyrights (Article 19, paragraph 2). So, *no* dedication of a work to public domain would be legally effective – at least, in Austria; in contrast to the USA, where this is possible. Nevertheless, a work can become part of public domain in Austria: In general, this happens 70 years after the author's death (Article 60 & 61), which is obviously no viable option for Open Data. Hence, certain open licences are necessary for publishing Open Data in the first place – otherwise, it would be just copyrighted, compiled data nobody is legally allowed to use.

On top of this, Open Data is not only restricted due to copyright, but also through data protection laws. The protection in this sense is related to *personal* data only. Data containing personal details (or details, which enable to draw conclusions about a person indirectly) can*not* be generally open in the sense of the OPEN DEFINITION. Based on a directive[40] of the European Union, this is guaranteed by the Data Protection Act or DATENSCHUTZGESETZ (DSG) as a fundamental right in Austria. Its currently applicable version dates back to the year 2000 with various revisions to this date [79].

The term "data" is specifically defined as "personal data" by the DSG 2000 in its Article 4, item 1. Personal data in this sense is information about any natural person, community or legal entity, which allows identifying the respective individual or makes it identifiable indirectly, at least. For example, this is the name or age of a person, but also his or her social security number, email address or telephone number. Additionally, also "sensitive data" as such is defined in Article 4, item 2, as information especially worth protecting (like about someone's political views, religious affiliation or sex life). Although, processing of such data (personal as well as sensitive data) is allowed under specific circumstances and with explicit consent of the affected individual (Article 1, paragraph 2), in most cases it cannot be part of Open Data; if nothing else, due to the fact as personal data shall *not* be generally open and freely accessible as well as (re-)distributable according to the DSG 2000.

The new GENERAL DATA PROTECTION REGULATION by the European Union [80], which becomes directly applicable (in contrast to the current directive) in 2018, won't change anything regarding Open Data or the statements at this point, because data protection will be strengthened rather than loosened.

Open Licencing

In order to deal with national copyrights in a proper way, various standardised open licences exist. These licences naturally do *not* conflict with respective national law. Instead, most of them are built upon it and adopted regional versions for specific copyright laws are created to ensure legal force. Indeed, without any copyright in its today's form *no* Open Data could be published as respective licencing operates upon regulations about copyright [74, p. 8], rather than ignoring it.

Such open licences especially for Open Data or databases, respectively, were designed by OKI's project "Open Data Commons" (ODC) [81]. Since its beginning in December 2007, ODC has released three different open licences, which are also compliant with the OPEN DEFINITION:

→ the "Public Domain Dedication and Licence" (PDDL)
→ the "Attribution Licence" (ODC-by) and
→ the "Open Database Licence" (ODC-ODbL).

While the first one tries to dedicate a database to the public domain, the second only requires the attribution of authorship in redistributed works and the latter additionally warrants that such works are published under the same licence (ODC-ODbL). As already mentioned with respect to UrhG, an explicit dedication to public domain is not legally effective within Austria's legislation. However, for this particular case, the PDDL (as well as similar open licences) provide "fall-back terms", which declare a waiver of all copyrights to the maximum extent possible by national law.

---

[40] The (not yet uniform) data protection rights within the European Union are based on the Directive 95/46/EG of European Parliament and Council. It was issued in October 1995 and had to be implemented as national law three years later [141].

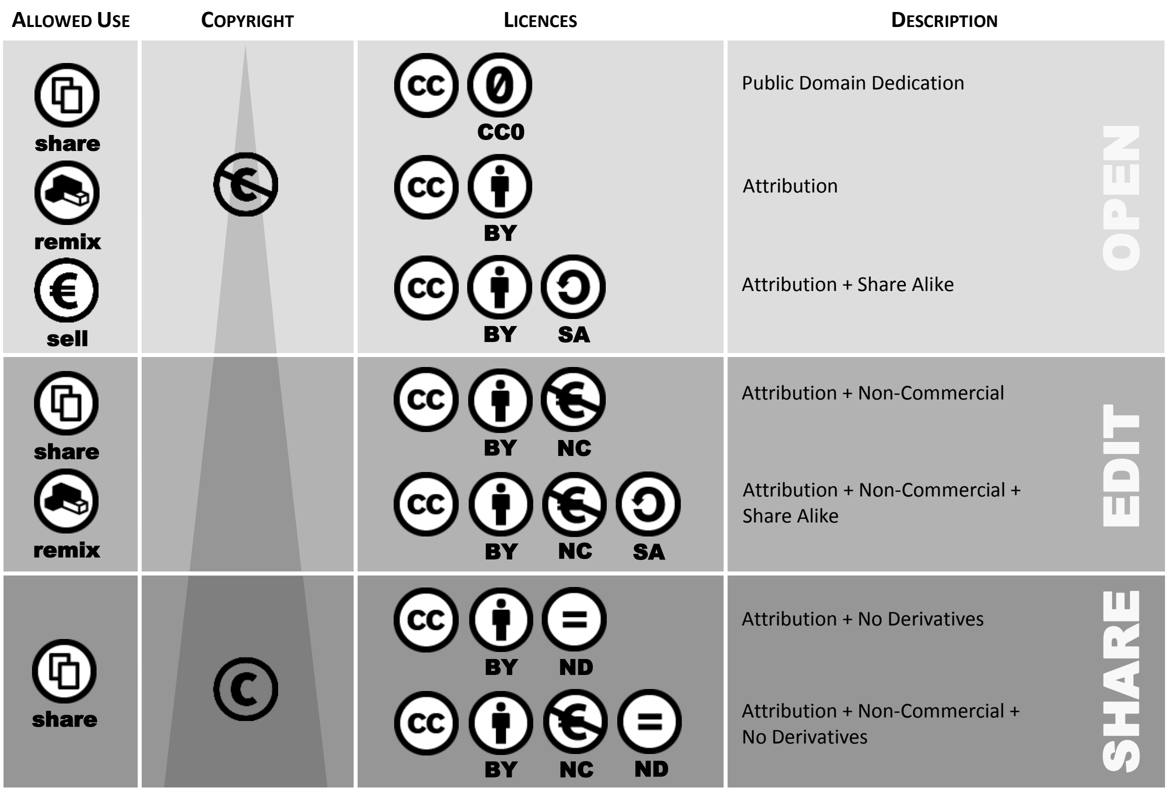
More popular – especially with respect to more content-related applications – are the every man's licences of Creative Commons (CC), a non-profit organization dedicated also to free exchange of knowledge like OKI. It was founded as an initiative at Stanford University by the American legal scholar Lawrence Lessig in 2001 [74, p. 8]. To this day, the initiative's goal is to provide comprehensible licences – also for legal laymen. These licences are offered for free and in a modular way, so that authors can inform others that their copyrighted work can be used freely – under specific circumstances. Creative Commons itself states that until 2017 there were over 1.2 billion works published under the conditions of one of CC's licences [82].

The CC licences themselves consist of a selection of four different legal modules, which apply certain restrictions on the permission of otherwise free usage of copyrighted work:

- → *Attribution* requires an attribution to the actual author in redistributed works.
- → *Share Alike* requires redistribution of subsequent works under identical legal conditions.
- → *Non-Commercial* forbids commercial redistribution, like with monetary compensation.
- → *No Derivatives* requires the original work to remain unchanged in redistributed works.

All possible combinations of these modules as currently available licences of Creative Commons are depicted in **Figure 18** below. As only the first two modules listed above are compliant with the Open Definition of OKI, Open Data can only be published under the CC-BY and the CC-BY-SA licence. They resemble OKI's ODC-by and ODC-ODbL licences, respectively. Additionally, an equivalent licence to OKI's PDDL is also available by CC: the CC0 – applying none of the restrictions of the above modules. These three licences – shown in the upper part of **Figure 18** – are as the only open ones by CC.

Figure 18 – The spectrum of Creative Commons Licences from Public Domain (top) towards all rights reserved (bottom)



Source: Creative Commons [83], Kimpel [74, p. 2], own illustration.

Each of these licences is provided in three versions: the *common deed* as simple summary of the respective licence, the *legal code* serves for legal communication as a strict and complete juridical version, and the *digital code* provides machine-readable information for the licence in question.

2.4.2 Open Government Data

In case the concept of Open Data is applied to governmental data, the respective result is called "Open Government Data" (OGD). This term describes data about the public sector that is released in an open way by the government or the respective administration in the interest of the general public *without* each and any restriction of free use or distribution [84, p. 6]. With respect to general interests and laws, obviously personal data (see also chapter 2.4.1) as well as classified or infrastructure-critical data is explicitly excluded from the (possible) set of OGD.

Principles for Open Government Data

Most of the current developments in the area of OGD are based on the Open Government initiative started in January 2009 by the US administration of Barack Obama right after his inauguration. The goals were to strengthen democracy, public trust as well as to promote efficiency and effectiveness of the government. In order to do so, three general principles for an open government were established and released as a presidential memorandum[41] for the heads of all US executive departments and agencies [85]:

→ *Transparency* shall insure accountability of the government. The public should be provided with information about the tasks performed by their government. The free availability of this information is an essential basis for transparency.

→ *Participation* shall provide the government with collective information and expertise as well as dispersed knowledge. With that, the government's effectiveness and the quality of its decisions should be improved. Furthermore, the public should be involved directly into the process of decision making.

→ *Collaboration* shall actively engage the public in the work of their government. Therefore, new tools, methods and systems should improve cooperation not only between the government and the private sector, but also within and between all levels of the government itself.

These principles were adopted multiple times internationally for national implementations of OGD. Also, in Austria these principles were incorporated literally in the official framework for OGD platforms by the Federal Chancellery in 2013 [86, p. 3].

Before the Open Government initiative of the US government started and right after the OPEN DEFINITION was published by OKI for the first time, in October 2007 30 US open government advocates met in Sebastopol, California, how OGD should be published by the government after its previous publications were usually inconsistent and incomplete. In August 2010, the Sunlight Foundation, yet another non-profit organisation dedicated to free exchange of knowledge, but especially of information by the US government, adopted and extended the list of eight principles defined in Sebastopol to ten. These guidelines greatly resemble the definition of Open Data by the OPEN DEFINITION, interestingly. For this reason, they won't be discussed here (again) in full detail, but they are in short:

→ *Completeness* (of published data; nothing should be omitted including descriptive metadata)
→ *Primacy* (data should published in its collected and stored form and *not* be aggregated)
→ *Timeliness* (of the publication of OGD in order to maximize utility)
→ *Ease of physical and electronic access* (via open formats and APIs for bulk access)
→ *Machine-readability* (of data files for easier processing including respective documentation)
→ *Non-discrimination* (data should be accessible *without* registration or identification)

[41] Presidential memorandums as well as executive orders in the US administration have the force of law and order all or specific government agencies to do something, or to start a regulatory process.

- → *Use of commonly owned standards* (which should be open and free to use as data format)
- → *Licensing* (*no* restrictions should be imposed on the use of OGD)
- → *Permanence* (data made available online, should remain online)
- → *Usage costs* (*no* costs should be involved accessing OGD as this limits access again)

These principles are often addressed by relevant literature with respect to OGD and are widely considered as general guidelines for this topic [38, p. 11]. Again, Austria's governmental framework for OGD platforms incorporated them as well [86, pp. 5-7].

Legal aspects of Open Government Data

Corresponding to Open Data in general, there exists of course also a legal framework for OGD in particular. Because the European Union considered the usage of information and data provided by the public sector as an economic potential, which should strengthen the competitiveness of the economy of the European Union [87, p. 43], quite early, already in 2003 a Directive on the re-use of Public Sector Information (PSI directive) was issued [1]. This directive has become applicable through national implementations by no later than July 2005 (Article 12, PSI directive). Basically, *all* documents of the public sector should become (digitally) accessible and usable – explicitly for any (commercial or non-commercial) purpose – as long as copyrights of third parties or general interests are not violated. Overall, the PSI directive was quite anticipatory as it incorporated principles later found also in policies for Open Data in general, like the Open Definition: free use (Article 3), electronic access & data formats (Article 5), transparency (Article 7), licencing (Article 8) and non-discrimination & non-exclusiveness (Articles 10 & 11). Additionally, a uniform maximum deadline of 20 business days for requests on public sector information (Article 4, paragraph 2) and harmonized tariff principles for such publications (Article 6) were defined. Although, the latter point conflicts with the requirement of usage at no charge of Open Data or the Open Definition in particular, it introduced a price limit all over Europe. This limit is equivalent to the respective marginal costs for reproduction, provisioning and distribution of the publication. Provided that the fee cannot be justified properly, in most cases, this means that this data is released (practically) for free anyway [88].

In June 2013, the PSI directive was revised by the European Union to adopt for incorporating cultural heritage information (like from libraries, museums and archives in general), as well as new developments with respect to Open Data in general and OGD in particular. So for example, the definition of "open formats" and respective standards as a requirement for publication of public sector information was included into this amendment. Respective implementations into national law should have become applicable by no later than July 2015.

However, it has to be stated clearly that the PSI directive does *not oblige* public institutions to publish or allow the usage for their documents to this day. The directive only regulates the framework under which the publication and reuse of public sector information has to happen.

In Austria, the national implementation of or the Federal Act for the PSI directive, respectively – called Informationsweiterverwendungsgesetz (IWG) – was published not until November 2005; its revision due to the amendment of the European PSI directive in July 2015 [89]. Due to the situation of legal competence in Austria, also laws on the level of all Federal States were necessary to implement the directive to its full extend also for the public bodies of states as well as municipalities [87, p. 44]. The process was completed in September 2007 [90] with Salzburg being the last Federal State to implement the PSI directive in its legislation. Overall, these implementations in Austrian law resemble the PSI directive in most parts almost literally.

On top of this, in Austrian legislation especially the UrhG has to be revisited when it comes to OGD. The reason for that is its Article 7, paragraph 1. It says that all laws, regulations, official decrees, announcements and decisions as well as works made exclusively or primarily for official use shall not enjoy copyright protection. This means that they are dedicated to public domain by law. However, Article 7 UrhG only includes works of literature (including computer software), but not compilations like databases for compiled Open Data. Furthermore, officially made maps are also excluded from this clause (Article 7, paragraph 2). Therefore, OGD publications made by public bodies in Austria are explicitly licenced anyway. Accordingly, most of them (and all of them on official OGD platforms) are licenced with CC's attribution licence (CC-BY). Additionally, this is established (including a decidedly reference to the CC-BY licence) in Austria's governmental framework for OGD platforms [86, p. 11].

Benefits of Open Government Data

At this point, some questions may arise: Why is OGD needed? Why and wherefore do someone want to have this governmental data? And what is the added value by doing so? In the following, qualitative as well as quantitative answers shall be given to these legitimate questions in a nutshell.

First of all, a lot of relevant (also scientific) literature for OGD is addressing the mentioned questions nowadays. This Diploma Thesis will refer to a German study from 2010 by Lucke and Geiger summarizing the qualitative part. Most of its discussion orientates on the general principles for Open Data and OGD as already discussed. However, the core findings can be broken down to seven qualitative benefits of OGD for society and its administration [84, pp. 10-14]:

→ *Opening of Government and Administration*
Open Data originating from the public administration enables each citizen (theoretically) to use this data to inform oneself about the activities of one's government. Additionally, this OGD can be used by the public to monitor and supervise these activities. In the best case, this works as positive non-discriminating public relations work since the government is operating on (legal) behalf of its own citizens as well as it is financed by taxes paid by them.

→ *Reuse & Recycling*
Citizens themselves can choose from the overall dataset provided by their government, which data is of interest for certain interest groups. So, the value of gathered data can be maximized.

→ *Transparency*
OGD can make the activities of the public administration more transparent, because decisions, subsequent activities and potential consequences are made visible and comprehensible. Hence, public acceptance for governmental activities may rise and accusations of corruption and mismanagement can be contradicted easily by referring to respective and transparent OGD. Additionally, actual incidences can be detected and treated faster. Releasing OGD to the public demonstrates self-assurance and willingness to handle criticism. Furthermore, benchmarking and respective federal competition is enabled through OGD.

→ *Participation & Collaboration*
Through participation (involvement) and collaboration (integration) of the public into decision making and the subsequent governmental activities triggered by OGD, entry barriers to politics, related interest and engagement are lowered. The resulting collective consultation and governance can reduce political apathy and lower consulting fees by third parties.

→ *Better Governance*
Political and administrative activities can be more effective and efficient as distribution of information is easier and faster, and a wide variety of interpretations and feedback can be gathered due to the publication of OGD. In addition, some applications and services need not to be implemented by the government itself as citizens are willing do this on their own.

→ *Open Innovation*
Through re-usage and distribution of OGD, a high potential for innovation and economic development is created. The public and eventual developers need not express their demands; they can just implement on their own and provide their findings and results to others. Additionally, the potential of yet unused data may be unfold.

→ *Promotion of economic development*
Private companies, universities and researchers – like any others – can use and process OGD to create new evaluations, applications, products, services or even business models. As a consequence, the attractiveness as a business location, and if nothing else, the quality of life for citizens through new services can be improved.

Besides that, the benefit of OGD has been tried to be estimated also quantitatively. As a digital agenda for the European Union – titled "turning government data into gold" – was published by the European Commission in December 2011 [88], it referred to a 40 Billion Euro "boost" to the Union's economy each year that OGD could provide. Since then, this prominent figure was recited frequently by relevant literature as well as by media [91, p. 13]. Various scientific studies[42] are underlying this and similar figures. The one given by the European Commission is originating from a study by Graham Vickery, who investigated this area on behalf of the European Commission in 2011 [92]. In fact, he summarized other preceding studies, thereby giving a holistic, quantitative overview and analysis for the economic impact of OGD. Thus, Vickery estimated the *average* *direct* economic potential of OGD within the European Union with 40 Billion Euro by the year 2009 – while 35.3 Billion and 47.1 Billion Euro were the minimum and maximum values for this averaged estimate [92, p. 38]. In an optimistic case, where governmental data is opened up incorporating all principles for Open Data supported by decent infrastructure, the aggregated *direct* and *indirect* economic benefits for the economy of the European Union could have been in the order of 200 Billion Euro (1.7% of the Union's GDP) by 2008 according to Vickery [92, p. 37].

With respect to Austria, one can project this figures spatially as well as temporally. While such a complete calculation[43] is given within the Annex (see p. 113), the notable results should be given at this point: The figures originating from Vickery's study and projected temporally would result in a potential *direct* economic impact of OGD from 43 through to nearly 76 Billion Euro for all members of the European Union by 2016. Broken down to Austria, this would give a national *direct* impact in the order of 1 to 1.8 Billion Euro for the year 2016. Furthermore, the overall *direct* and *indirect* effects for the economy of the European Union in 2016 would yield in potential economic benefit from 227 through to 344 Billion Euro, or 5.3 to 7 Billion Euro for Austria, respectively.

And that's not all. The study made by Vickery reveals possible reduction of costs due to OGD as well. Estimated efficiency gains in existing operations[44] all over Europe could be nearly 10 Billion Euro worth [92, p. 4]. In contrast, the possible governmental revenues for data from the public administration in the whole European Union are considered *relatively* low (compared to the potential overall economic impact of OGD): 1.4 to 3.4 Billion Euro [92, p. 25] could be gained through licencing and associated activities of data suppliers.

---

[42] A review and summary of the most prominent studies, which address the economic impact of OGD quantitatively, is given by Huber et al. in a study by the Vienna University of Economics and Business from 2013 [91, pp. 39-47].

[43] The calculation is using a temporal projection via nominal growth of the European Union's GDP (as a pessimistic forecast) as well as through the estimated market growth for governmental data by 7% [92, p. 36] taking into account the maximum and minimum values of Vickery's study. This results in the estimated (somewhat wide) intervals given by this work for the year 2016. The spatial projection is calculated through time series data of Austria's share of the European Union's GDP.

[44] These potential efficiency gains can be achieved through improving accessibility of information necessary for obligatory environmental impact assessments, for research & development and more rapid and comprehensive access to information for the general public according to Vickery [92, p. 4].

Whatever the exact potential economic impact of OGD and the amount of associated savings & revenues may be, a magnitude in the upper single-digit billion range for Austria and even in the three-digit billion range for the whole European Union is more than remarkable. With that, the reasoning behind opening up governmental data should have become reasonable by now. Nevertheless, it should be mentioned at this point that most benefit lies within geo-spatial data [91, p. 47] far ahead from any other field of subject – including budgetary data.

2.4.3 Open Budget Data

The concept of so-called "Open Budget Data" (OBD) originates basically from Open Data and Open Government Data. With respect to the principles behind OGD in general, OBD in particular is about a common understanding of opening up budgetary data of the public sector and its institutional units.

Data of the public budget is called OBD, if it is released by the government and public administration without each and any restriction of free use or distribution in the interests of the general public. Primarily, this data consists out of financial plans and budgets as well as budgetary reports like statements of account and similar documents without personal data, which are provided in an open, reusable and machine-readable data format [19, p. 6]. The granularity of OBD published in this way is still a topic of ongoing discussion in related literature and expert groups. However, according to that, the budgetary raw data released as OBD does not necessarily represent each and every single governmental account entry to the deepest level. This would be bulk data, e.g. for Austrian municipalities of about two million accounts per year [33, p. 34]. Thus, the aggregation level should be chosen via a common, coherent approach for the various public bodies and their sweeping sizes (e.g. all aggregated budgetary accounts for every fiscal period).

The foundations for the provisioning of such OBD are (existing) budgeting systems of the governmental authorities' or all other institutional units' (of the public sector) internal accounting. Although, these systems can be considered as modern, but also heterogeneous nationwide [33, p. 36], in most cases standardized data interfaces exist for the purposes of official finance statistics for decades by now – for example, like in Austria the GHD and LHD interface (see chapter 2.2.6).

In general, OBD can improve budgetary transparency of public bodies beyond the scope of such ordinary finance statistics. If governmental authorities implement and provide interfaces for OBD with respective interoperability, budgeting systems cannot be only interconnected more easily, but also up to date analyses – in an extreme case – on a daily basis are made possible. This enables impact-oriented control by the government as well as the public itself in a timely manner [19, p. 6]. Additionally, disadvantages of analogue data formats like media disruption and long shipping times can be overcome with OBD.

However, to this date still reservations with respect to OBD – especially by politicians – persist. So, no standardized, enforceable and in the end also punishable obligations (including respective timing) for OBD are in place all over Europe or in Austria in particular. In fact, there are concerns that apart from additional costs for publishing this data and respective adaptations as well as increased administrative expenditures because of budgetary misinterpretations, there is no added value other than enhanced transparency [19, p. 6]. Anyway, the former discussion of benefits of OGD should provide a more sophisticated and differentiated approach regarding this matter.

A lot of regulations and theoretic principles related to OBD and respective information flows in Austria were already touched briefly in chapter 2.2. A more detailed discussion about the current status in Austria regarding OBD – also with respect to the topic of this Diploma Thesis – will take place in chapter 3.1.3 later on.

2.4.4 Linked Open Data

While the idea of Open Data (see chapter 2.4.1) is an indirect descendant of the Social Web, the idea of Linked Data (see chapter 2.3) is built upon the concept of the Semantic Web [38, p. 25]. Accordingly, "Linked Open Data" (LOD) combines the two principles of Open Data and the technologies and means collected under the term Linked Data as described by Tim Berners-Lee in his already discussed seminal design note from 2006:

> *Linked Open Data (LOD) is Linked Data which is released under an open licence, which does not impede its reuse for free. [41]*

With that, Linked Open Data describes – at its best – structured, highly interconnected and syntactic interoperable information or datasets, respectively, which are distributed over various data sources and repositories (like SPARQL endpoints) within the same or different organisation and can be used by everyone without any restrictions. As a consequence, LOD is an important mechanism for information management and integration; hence, in turn, it becomes more and more important in state-of-the-art implementations of data interfaces and storages for Internet-based services and applications [38, pp. 17-19].

In this way, data is not just published somewhere *on* the World Wide Web, but instead *in* the Web by using LOD; this way, the goal to achieve a real & infinite *Web of Data* – additionally, to a *Web of Documents* – moves a little bit closer (see chapter 2.3). The possible result is also known as "Global Giant Graph" [73, p. 29] according to the technical background of Linked Data, using RDF graphs.

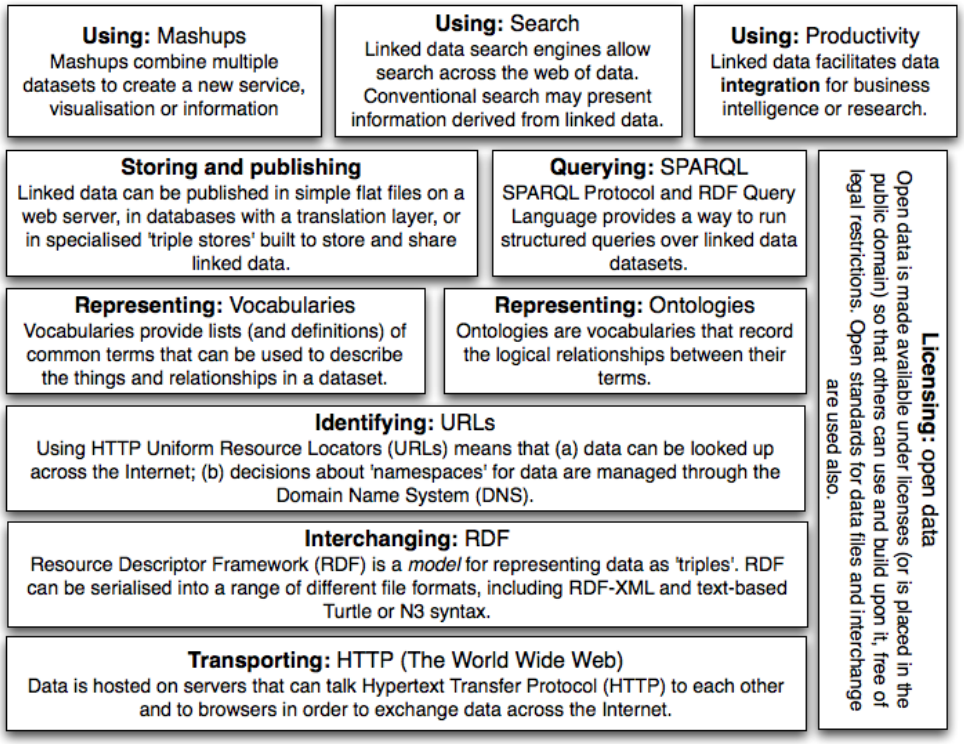
Despite the perception, that the transition from "just" documents on the Internet to real LOD might seem to include a structural break, and with that, being a big step, the concept of interconnected & open data can be reached more easily than maybe expected. Representing that this process is more a progressive, stepwise transition rather than an encompassing adoption in one go, Tim Berners-Lee (again) introduced a so-called "5-Star-Scheme" at the "Gov 2.0 Expo" in Washington D.C. in 2010 [38, p. 17] – four years after he coined the term of "Linked Data" in the first place. As depicted in **TABLE 10**, it classifies published datasets from protected, not reusable content (without any star) through to real LOD composing an open *Web of Data* (gaining a total of five stars). Thereby, authors as well as consumers can rate and flag available LOD in a common and consistent way.

TABLE 10 – The six classifications for Linked Open Data according to the Tim Berners-Lee's 5-Star-Scheme

The 5-Star-Scheme for Linked Open Data (by Tim Berners-Lee)			
Classification	Name	Description	Examples
WD	WEB DATA	Data in whatever format on the Internet *without* an open licence	DRM protected music files
★ OL	OPEN LICENCE	Data in whatever format on the Internet *with* an open licence	Images or PDF files published with CC-BY or similar licences
★★ OL RE	RE-USEABLE	Structured, reusable & machine-readable data on the Internet *with* an open licence	Excel spreadsheets
★★★ OL RE OF	OPEN FORMAT	Structured data in a non-proprietary open format on the Internet *with* an open licence	TXT or CSV files
★★★★ OL RE OF URI	URIS FOR IDENTIFICATION	Data with URIs to identify globally contained concepts and objects with all from above	RDF data possibly using SPARQL endpoints
★★★★★ OL RE OF URI LD	LINKED DATA	Data with links to other data to provide context including everything from above	Linked Open Data cloud

Source: Berners-Lee [41], Cyganiak [93], own illustration

FIGURE 19 – An overview of the elements composing a so-called “Linked Open Data Stack”



Source: Davies [94]

The technical implementation of LOD itself relies heavily on the concepts and standards based on the idea of a Semantic Web as discussed in chapter 2.3. However, as these concepts leave a margin for the actual implementation and some of them – envisioning the Semantic Web to its full extend – overshoot the aim of LOD in particular, there is no encompassing definition of what *is* and what *is not* covered by the term “Linked Open Data”. Likewise, currently published data classified as LOD to some extend or even completely (e.g. with 5 stars from the 5-Star-Scheme) is often available in serval formats – sometimes in parallel.

In order to give some guidance, Tim Davies created a concrete stack of technologies commonly considered as relevant technical components for LOD. This stack is depicted in its original form in **FIGURE 19**. Resembling prominent elements from the SEMANTIC WEB CAKE (see **FIGURE 8** on p. 31), it explains and concentrates on existing implementations and solutions for LOD. Hence, it shows technologies and their relation with each other required for building LOD.

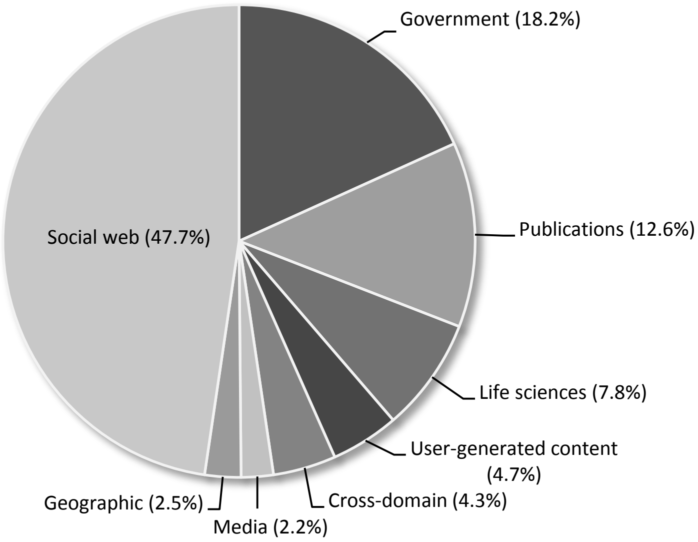
As most of its elements were already explained in some form in chapter 2.3 or are at least self-explanatory, there will be no full discussion of this “Linked Open Data Stack” at this point. However, it is still worth mentioning that a separate and independent licencing layer is laid out for obvious reasons (especially, for bringing the *Openness* to Linked *Open* Data). Furthermore, three prominent benefits or use cases for LOD are shown at the top of the stack in **FIGURE 19**: mashups, search and productivity. While the latter is quite reasonable and was discussed already (like for OGD in chapter 2.4.2), the other two may need a little bit of explanation.

So-called “mashups” as well as search demonstrate both the far-reaching benefits of LOD. On the one hand, mashups describe new information and media content generated seamlessly through (re-)combination of already existing data. Famously used by the web search engine Google, for example, maps can be enriched with satellite images, traffic information and pictures displayed from several sources for the displayed location. On the other hand, LOD can enable search engines (like Google, again) to semantically process user requests and (theoretically) providing results with much higher precision. For example, a search for "beetle" would be able to distinguish results for the animal from the car named after it – as it does Google already nowadays. However, this again may extend the approach of LOD much further to the Semantic Web as such, because working solutions for the layer containing *unifying logic* within the SEMANTIC WEB CAKE (like reasoners) and overlying concepts (see also chapter 2.3.1 and **FIGURE 8**) may be needed when using LOD for these matters.

In any case, LOD as depicted in **FIGURE 19** is the foundation of all kinds of such and similar use cases, because respective applications are able to provide more accurate and up-to-date (open) data in an easier and faster (automated) way as a result to each user request.

Already published LOD like this originates from the "Link*ing* Open Data Project" (LOD-P) founded in January 2007 [38, p. 40]. Supported by the W3C's Semantic Web Education and Outreach Group, it can be seen as a continuing community effort, which identifies existing datasets complying with the OPEN DEFINITION (including respective open licences), converts them to RDF according to Linked Data principles and publishes them again on the Internet [42, p. 5]. With that, the LOD-P can be considered as the initial origin and the beginning of LOD as well as its shaping *Web of Data*.

FIGURE 20 – Shares of various domains within the Linked Open Data cloud



Source: Schmachtenberg et al. [95], own illustration

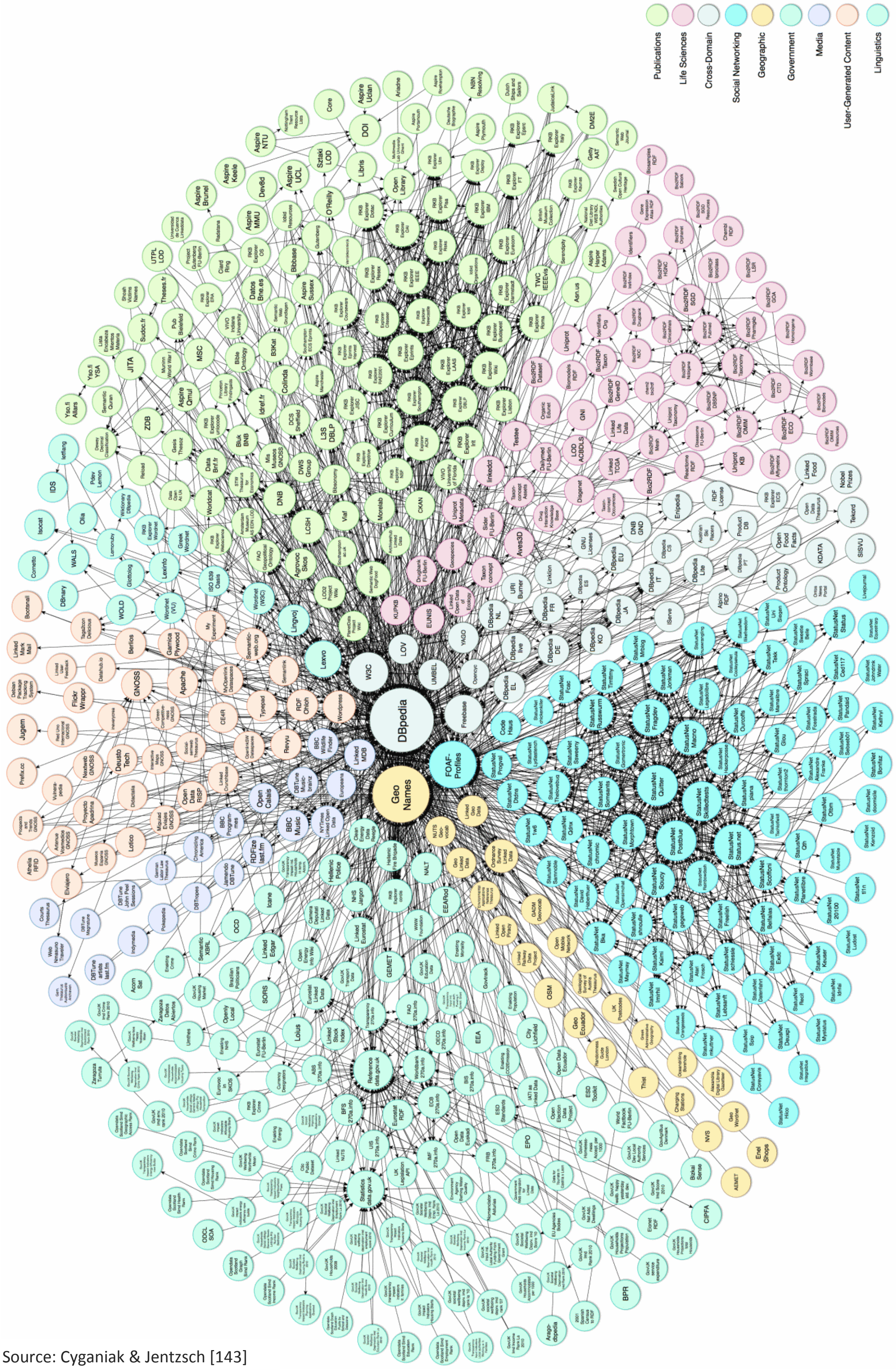
Since this project started, this *Web of Data* has grown remarkably and now includes datasets produced by various organisations, data providers or private initiatives, like media (e.g. the BBC or the NEW YORK TIMES) or governments (e.g. USA or UK) [38, p. 40]. Measuring the amount of this published LOD and its structure is obviously challenging. However, numerous studies – including some from the community of the LOD-P itself – addressed this question through techniques like automated crawling. Based on 1091 known datasets[45] back then, a study from the University of Manheim performed in 2011 investigated the structure of the Linked Open Data "cloud" at that time [95]. The shares of the contained thematic domains are depicted in **FIGURE 20** above. As shown, LOD from the social web prevails with nearly 47.7% of all datasets, followed by governmental data (18.2%) and scientific publications (12.6%). Obviously, (Linked) Open Data from the public administration played a bigger role already at this time.

The actual range and scale of all known published LOD is regularly mapped by the community of the LOD-P, or to be more precise by Richard Cyganiak (National University of Ireland, Galway) and Anja Jentzsch (Free University of Berlin). The status of this map by August 2014 is depicted in **FIGURE 21**.

---

[45] Datasets in this sense describe an arbitrary amount of data (in terms of RDF triples) published by one single organisation.

FIGURE 21 – The Linked Open Data cloud diagram as of August 2014



Source: Cyganiak & Jentzsch [143]

The illustrated Linked Open Data cloud diagram in **FIGURE 21** consists of nodes representing each and every known dataset published as LOD by 2014 and directed edges the RDF links between them. Furthermore, the diagram is differenced by colour into thematic groups like in the mentioned study from the University of Manheim. Obviously, the distribution of the number of datasets between the various domains has changed between 2011 (as depicted in **FIGURE 20**) and 2014. This is due to the increasing expansion rate of the LOD cloud. As of 2011, it constituted of over 31 Billion RDF triples [38, p. 40] in 1.091 datasets [95, p. 4]; by 2016 it has grown to over 149 Billion RDF triples in 2973 datasets [96], which nearly corresponds to an increase by a factor of five in its volume in not less than five years.

Additionally, the Linked Open Data could diagram in **FIGURE 21** shows that certain datasets serve as giant hubs interlinking huge amount of data. Above all “DBpedia”, a community effort to extract and publish structured information from Wikipedia (like from its right-handed “infoboxes”), “GeoNames”, providing RDF information of millions of geographical locations worldwide, and “FOAF” (an acronym of “friend of a friend”) profiles, which are datasets describing persons, their properties and relationships, are the most notable and popular linking hubs in the current *Web of Data*.

After all, LOD describes a principle of extending the core technologies of Linked Data towards the vision of the Semantic Web – while not (yet) forming a Sematic Web as described in chapter 2.3.1 to its full extend.

2.4.5 Linked Open Budget Data

Last but not least, “Linked Open Budget Data” (LOBD) describes the combination of the two concepts of “Open Budget Data” (see chapter 2.4.3) and “Linked Open Data” (see chapter 2.4.4) in the end. Hence, all data of the public budget that is released by the government and public administration without each and any restriction of free use or distribution in the interests of the general public as Linked Data is called “Linked Open Budget Data”.

This relatively new method of interlinking the published governmental budget data makes it possible to use this data beyond organisational and institutional borders – like for fiscal (meta-)statistics, analyses, comparisons, maps and publications. Through this combination of budgetary data, – for example – known or potential relationships and dependencies can be easily deduced, shown and understood by everybody – and not only by the institutional units of public sector itself within their hierarchical structured (and closed) information flows (as seen in chapter 2.2.6). Added values emerge – like for OGD in general – when this LOBD from different sources – as innovative recombined data – creates new insights [19, p. 7].

Despite the possibilities, actual full implementations of governmental LOBD are still rare. However, some nascent examples are appearing on the scene in recent years. The purpose of the following major chapter is not only to unveil and describe such examples, but also analyse and assess their status according to the overall theoretical principles and guidelines outlined in this preceding chapter. After this scientific analysis, the exemplary application of eponymous LOBD on municipalities in Austria will be the subsequent and last task of this Diploma Thesis, in the end. The theoretical principles presented so far as well as the upcoming analysis of the state-of-the-art will be the basic foundation for reaching the primary scientific contribution of this work.

CHAPTER 3
OPEN BUDGETS IN PRACTICE

In order to fulfil mentioned overall requirements by current society, governmental institutions and laws, existing implementations of OBD or LOBD, respectively, can be already found all over the world. A comparison of these implementations, especially with their current status for municipalities in Austria, in addition to their quality, extent as well as their compliance to the presented standards and guidelines will be of interest in the upcoming chapter. The aim of this will be to ascertain the current status of OBD implementations for public finances (especially with respect to Austria) and find current best practises in this field of information science.

So, in the **first** part of this chapter, current implementations of OBD, related literature and existing surveys will be presented as a consequence. The discussion will cover examples on the international level as well as on the level of the European Union. The status in this field in Austria – considering especially eponymous municipalities – will span over the last section of this very part. The following, **second** part of this chapter will assess the found implementations according to the standards and guidelines from the previous chapter as well as general relevant requirements. Furthermore, these implementations will be compared with each other. In doing so, common best practices and state-of-the-art implementations shall be worked out. Later on, this will be the basic foundation for the last major chapter of this Diploma Thesis. The study and assessment in the following part of this Diploma Thesis took place in August 2017 – the results should be considered accordingly.

3.1 CURRENT IMPLEMENTATIONS OF OPEN BUDGET DATA

In following, chosen examples for implementations and publications on governmental OBD and LOBD in particular will be presented. The creators of such data sources can be governmental bodies themselves as well as private initiatives, which gather and publish budgetary data with prior approval of the responsible governmental authority. Overall, the respective data of these examples can refer to all different parts of the public sector and all governmental levels. The reason for such a more extensive approach on this part is that *all* available and suitable examples for existing OBD shall be analysed in order to work out the current status of governmental OBD and respective common best practices in the end. However, for Austria in particular the extended discussion will focus obviously on municipalities and their status on OBD, anyway.

3.1.1 International

International initiatives on OBD reach far back before the concepts about digital “openness”, related laws for public authorities (see chapter 2.4.1) or even Linked Data (chapter 2.3) became public. Since then, unsurprisingly some private organisations came up that try to survey and publish information on the current status of Open Data regarding governmental budgets, as well as some that actually provide coherent platforms to release and distribute such information on a supranational level. In addition, there are also notable examples on OBD created and maintained by governmental bodies themselves as a matter of fact. In the following, the most outstanding and famous examples of such international implementations as well as surveys on OBD shall be presented in a nutshell.

International surveys on Open Budget Data

When it comes to actual implementations of OBD and their assessment, the question may arise, how an implementation's status and its compliance to general underlying principles like publicity, openness and transparency can be measured in an objective and sophisticated way. Of course, various methodologies exist regarding this matter so far – mostly consisting out of questionnaires or standardized, practical reviews, or a combination of both. Internationally seen, the most notable ones in this field are the OPEN DATA INDEX and the OPEN BUDGET INDEX. As their name suggest, the former one measures Open Data from several fields in the public sector (including public finances) whereas the latter one concentrates on budgetary data only. However, both of them focus solely on budgets of the central governments – neglecting the levels of Federal States and municipalities altogether. Nevertheless, in order to get an impression of the current status of latest OBD implementations and their measurement, both will be introduced in a nutshell subsequently.

First of all, the OPEN DATA INDEX [97] is created and published by OKI's Open Government Data Working Group [87, p. 11] since 2013 on a yearly basis for 94 different countries all over the world in 15 thematic areas[46] as well as an aggregated summarized grading for national "openness". This data gathered by experts in the field of Open Data has the aim to provide an overview on the respective situation worldwide [87, p. 11].

In practice, Austria's results in the OPEN DATA INDEX for 2016 were somewhat devastating: With respect to budgetary data for public finances (by the central government) Austria occupied the 76th place at the rear end of the ranking (among countries like Iran & Kenya). This is because the data was published by the Federal Ministry of Finance in non-open formats (PDF & Microsoft Excel spreadsheets), in various separated files and had no dedicated open licence, although it was still publicity available, up-to-date and obtainable free of charge. On top of this, the OPEN DATA INDEX showed that there was no data available for governmental spending on a detailed transactional level for Austria; instead, only high-level aggregated numbers were published. Therefore, as the level of detail for this data was not met according to the survey's requirements, Austria – at the end of the respective ranking – has been seen as completely non-transparent in this field of governmental data among astonishing 82 out of the evaluated 94 countries.

In general, according to the 2016's OPEN DATA INDEX a total of 32 countries (e.g. Germany & the UK) were ranked in the first place for completely transparent governmental budgets and only two for transparent governmental spending, namely Colombia and Greece. Overall, Taiwan ahead of Australia and the UK were on the top of the leader board in the aggregated OPEN DATA INDEX for the best OGD at that time.

On the other hand, the more focused OPEN BUDGET INDEX exists since 2006 and was developed by the INTERNATIONAL BUDGET PARTNERSHIP (IBP), a private initiative formed in 1997 by the CENTER ON BUDGET AND POLICY PRIORITIES in Washington D.C. [19, p. 5] which should "promote transparent and inclusive government budget processes as a mean to improve governance and service delivery in the developing world" with the ultimate aim "to ensure that public resources are used more effectively to fight poverty and promote equitable and sustainable development in countries around the world" [98].

As a mean to reach this goal, the OPEN BUDGET INDEX is published typically every two years as part of the so-called OPEN BUDGET SURVEY by the IBP and compares national budget transparency and

[46] Besides the governmental budgets and spending (on detailed transactional level) these are: national statistics, data on governmental procurement, laws and their drafts, administrative boundaries, maps and geo information including land ownership information, data on water & air quality and weather forecasts, company registers and general election results.

accountability around the world. This is done through assessment of the three components of a "well-functioning budget ecosystem" according to the IBP [99, pp. 11-12]:

→ *Transparency* through public availability of information and data on public finances
→ *Participation* with possible opportunities for citizens to take part in the budget process
→ *Oversight* by reliable and strong controlling authorities and institutions like the legislature and an overseeing national Court of Audit.

The quality and status of each component is evaluated through an actual scientific methodology [99, pp. 61-63]: a questionnaire – consisting out of 140 multiple-choice-questions for the latest survey in 2015 – is completed by researchers – like from academic institutions or civil society organizations – typically based in the respective country. The responses to each question are reviewed and rated on a scale from 0 (standard not met) to 100 (standard fully met). Subsequently, the results are simply averaged; the (carefully considered) amount of questions in each individual thematic area[47] results in its desired weighting within the overall rating – and, with that, targeting the thematic goal of the IBP: to be able to reveal undesirable developments, possible improvements and prospects in a sophisticated and comprehensive way.

The OPEN BUDGET INDEX in particular is a partial result of this assessment. It concentrates on budgetary transparency through public availability of eight budgetary key documents [99, p. 21]: the pre-budget statement, the executive's budget proposal, the enacted budget, citizens' budgets, and in-year, mid-year & end-year reports as well as audit reports. Hence, the OPEN BUDGET INDEX for a specific country reflects the public availability of these documents including their comprehensiveness and timeliness [99, p. 62]. Corresponding to its significance, there are 109 respective questions related to the OPEN BUDGET INDEX out of 140 in 2015's overall survey.

Although, IBP's OPEN BUDGET SURVEY considers itself the largest assessment on governmental budget ecosystems, covering 102 countries in its latest report in 2015 [99, p. 12], Austria was not (yet) included. As its next report scheduled for the end of 2017 will incorporate – with 115 – 13 more countries [100], there is a slight chance this will change in the near future.

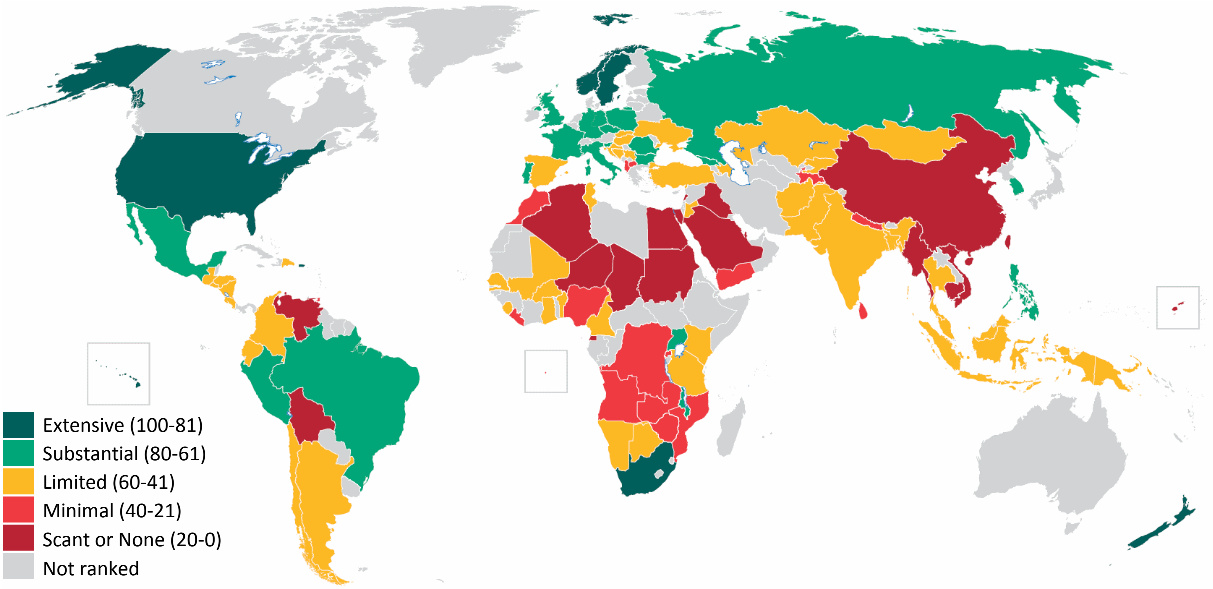
Furthermore, the survey does *not* necessarily require digital, machine-readable distribution of budgetary documents – also hard copies were still acceptable upon its latest release, reflecting that in certain parts of the world Internet access is limited – and thereby is also *not* compliant to the OPEN DEFINITION (see chapter 2.4.1) by OKI. However, even then – in 2015 – merely four percent [99, p. 24] of published, evaluated budgetary documents were released as hard copy *only* and not on websites of governmental authorities. Interestingly, only about 17 percent [99, p. 24] of published documents on such websites were machine-readable. The upcoming 2017's survey will consider only online publication as transparent [100], eventually.

All in all, the 2015's OPEN BUDGET INDEX reveals that governments typically fail to publish OBD and the majority of the world's population cannot access sufficient[48] budget information [99, p. 16]: 78 out of 102 evaluated countries – a large majority – merely provide insufficient[48] budgetary documents and data; 68 percent of the world's population is living in these countries. No wonder that the averaged overall OPEN BUDGET INDEX is 45 out of 100.

[47] Indeed, the survey is adopted regularly according to developments regarding transparency as well as from the experience of past surveys. Hence, the questionnaire was revised with the last version of the OPEN BUDGET INDEX in 2015, to incorporate a shifted (more accurate) importance of other budgetary documents in comparison with the Executive's Budget Proposal. However, the overall basic methodology remained unchanged.

[48] The OPEN BUDGET SURVEY considers [99, p. 16] budget transparency as "sufficient", if (and only if) the respective OPEN BUDGET INDEX is above the score of 60 for one, individual country. Correspondingly, if this score is less or equal than 60, budget transparency within the respective country is considered as "insufficient".

FIGURE 22 – The OPEN BUDGET INDEX with its worldwide grouped results regarding budget transparency from 2015



Source: International Budget Partnership [99, pp. 17-18]

In 2015, the overall ranking of the OPEN BUDGET INDEX [99, p. 7] was led by New Zealand (with a score of 88 out of 100), followed by Sweden (87), South Africa (86), Norway (84) and the USA (81), while Saudi Arabia and Qatar came in last (in a tie, with a score of 0). The worldwide performance of all surveyed countries is depicted in **FIGURE 22** above; a full list of the overall ranking with all scores and their underlying data from the respective questionnaires is given in the Annex (see p. 114).

Qualitatively spoken, five countries missed to publish their budget altogether; 29 did not release *any* end-year report (including a statement of account like in Austria) [99, p. 21]. Only 16 out of the 102 countries published all three evaluated key documents in the budgetary process [99, p. 21]. Surprisingly, from the documents found out not to be publicly available, more than half of them were not produced at all by the respective public administration [99, p. 21].

Additionally, there are several countries, which show a highly volatile level of budget transparency [99, pp. 35-36]: For example, Ghana has not published any end-year report in 2006, prepared one for internal use only in 2008, made one publicly available in 2010, published any again in 2012 and has released one to the public again in 2015. Overall, this unfavourable phenomenon is observable in every tenth country surveyed and obviously prevents constant budget monitoring and analyses by private initiatives or researchers.

One of the key findings of the OPEN BUDGET SURVEY is that willingness by the government is the most influential factor for budget transparency and the actual implementation of OBD [99, p. 20]: Most of the countries at the rear end of the ranking have weak democratic institutions or are governed by – as considered to be – autocratic regimes, which do not (want to) provide any budgetary information (or as little as possible). Not surprisingly, these countries are also those ones which tend to be perceived as being the most corrupt. Additionally, they perform worse on the Human Development Index. Oil- and gas-depended, undemocratic countries eventually conclude this end of the OPEN BUDGET INDEX (Chad, Equatorial Guinea, Iraq, Qatar, Saudi Arabia, Sudan and Venezuela).

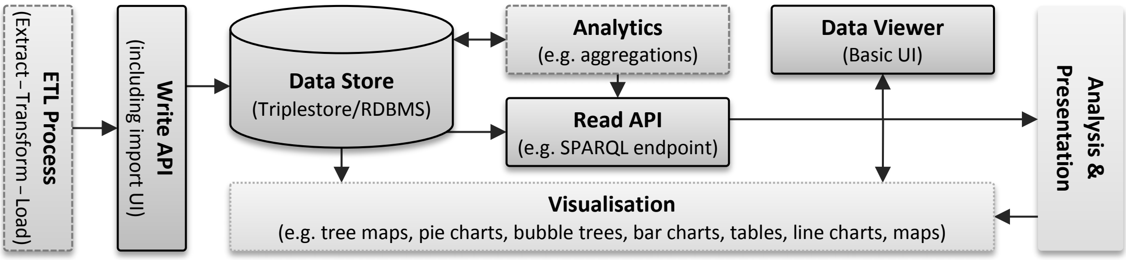
In the end, if insufficient budget information is given, the public (and if it is not produced at all, also the government) is left behind unable to fully understand or monitor how much public finances are raised and how they are spent – and according to the OPEN BUDGET SURVEY, the majority of world's population has to live actually with this situation in their own country.

Private Open Budget Platforms

In contrast to the governmental distribution of OBD, there are also private initiatives, which address this issue as well. One of the biggest platforms for OBD is provided by OKI's OPENSPENDING project [101], which can be seen not only as a simple database, but also includes various specifications and tools for standardized processing and visualisation of the incorporated budget data. Besides that, it is an international, community-driven project (especially with respect to the data collection) with the aim to create the world's largest and highest quality database on public finances in a comprehensive and open way [102].

Started in 2011 in its today's form, by August 2017 OPENSPENDING incorporated nearly 2,200 datasets with over 45 million records from 76 countries [101] – including Austria[49]. These datasets consist of budgetary (aggregated) data as well as data of individual financial transactions [103, p. 96]. With OPENSPENDING NEXT a new next-generation version of the platform was launched in mid-2017. Thereby, the project wanted to address various issues with its previous, original version as well as to return to its original philosophy of *centralising* data storage and *decentralising* visualization of this data [102] – analogous to OPENSTREETMAP, an open geographic map of the world, which most people do not encounter on the project's website, but on various other sites, where it is embedded in some form or another. The abstract structure and range of OPENSPENDING and similar platforms in this form with respect to overall data processing is depicted in **FIGURE 23** below.

FIGURE 23 – The abstract structure of Open Budget Platforms and their data flows using the example of OPENSPENDING



The border strength of each individual part depicts its respective importance to OPENSPENDING's current range of functions.
Source: Own illustration based on Pollock et al. [104, p. 20]

Several international as well as regional projects derived from the original OPENSPENDING project. One of the most notable seen internationally – especially with respect to this very Diploma Thesis – emerged from a scientific work at the University of Leipzig [103] in 2014 and had the goal to transform OPENSPENDING's OBD to LOBD; hence, its name: LINKEDSPENDING. The current webpage [105] of LINKEDSPENDING provides a user interface as well as a full-blown SPARQL endpoint based on ONTOWIKI – an open-source software supporting Linked Data and its distribution. Although this derived project seems to be abandoned by now and no new data was published since 2014, all of OPENSPENDING's data back then was transformed resulting in more than 900 datasets with 5 million so-called observations comprising nearly 110 million RDF triples, which are today available [105].

Besides the somewhat out-dated data and its scale, the method chosen for transforming is particularly interesting for further use in this Diploma Thesis: With a classic ETL[50] approach, the contents of OPENSPENDING's database were transformed automatically into an RDF based data cube through a specifically implemented application. Especially, the used RDF ontology for this data cube – standardised based on a W3C recommendation – will be of more interest later in the last chapter of this work. The full approach of creating LINKEDSPENDING is discussed in detail in [103].

[49] In August 2017, there were only 2 Austrian datasets including more 60,000 records [101] on the OPENSPENDING platform.
[50] ETL stands for "extraction", "transformation" and "load": extraction form data sources, transformation into the targeted data format and load to publish it on the target database.

Governmental Open Budget Data by the United States of America

Due to its global importance, its outstanding performance in surveys about governmental OBD and its historic contribution to the principles of Open Government Data in general (see chapter 2.4.2), the public finances released as OBD by the governmental bodies in the USA shall be presented as an international example of such an (governmental) implementation outside of the European Union.

In accordance with the already discussed presidential memorandum of the US administration (see also chapter 2.4.2) and related laws, most notably the FEDERAL FUNDING ACCOUNTABILITY AND TRANSPARENCY ACT of 2006 (FFATA) and the DIGITAL ACCOUNTABILITY AND TRANSPARENCY ACT of 2014 (DATA Act) [106, p. 1], budgetary transparency of the US public finances made relatively good progresses in comparison to other countries in recent years and its today's level can be considered as quite remarkable. Consequently, this development is also reflected by surveys like the OPEN DATA INDEX as well as the OPEN BUDGET INDEX and by now, the USA is at the upper end of both rankings.[51]

The data on the budgets as well as the respective statements of account of the central government is published by the US administration or, to be more precise, by the Office of Management and Budget of the White House digitally every year since 1996 [107]. These publications are dedicated to public domain or released in terms of CREATIVE COMMONS' attribution licence (CC-BY) – depending on their origin. In addition, the data itself is provided as processed and maintained time series data on revenues and expenditures from 1962 onwards (as historic actual values) including the budget for the current year as well as a medium-term financial forecast or plan, respectively, for the next five years. The data's level of detail is on individual (aggregated) account level differentiated by the various US agencies and their bureaus. Provided data formats[52] include open and machine-readable CSV files. A more detailed, overall description of this published data can be found in [108].

Besides that, US efforts on transparency including the levels required by laws like FFATA and the DATA Act go as far as releasing data on actual transaction data about governmental spending. Following the requirements induced by FFATA in 2006, the website "USAspending.gov" launched in December 2007 and provides fine-grained data on prime recipient transactions of governmental funding submitted by the US agencies making federal contract, grant, loan, and other financial assistance awards [109]. The data is published in anonymised way on the basic level of individual account records. It is again available in the open, machine-readable CSV data format and also via a proprietary webserver endpoint. However, the underlying licence cannot be considered as open according to OKI's OPEN DEFINITION.[53] Announced for release in late 2017, the new version and major overhaul of the platform incorporating recent legislative's changes to relevant law (especially the DATA Act) will not change anything about this according to that what could be seen so far.

Similar platforms like "USAspending.gov" with similar level regarding openness are also available on the level of Federal States in the USA. For example, the Texan Comptroller of Public Accounts provides comparable data as the central government [110]. Furthermore, regarding transactional data, this part of Texas' federal government publishes even *every single* transaction made by public authorities – more than one million records every month, all obtainable in large CSV files for free.

However, to this date there does not seem to be an actual implementation on LOBD for US public finances – not by the public authorities, at least. Nevertheless, this should not belittle the overall accomplishments in the field of OBD by the US administration and a respective local semantic implementation seems to be simply a matter of time.

---

[51] The USA got the 11th place on the overall OPEN DATA INDEX in 2016 and the 5th place on the 2015's OPEN BUDGET INDEX.
[52] Almost all budgetary publications since 1996 comprise PDF versions, Microsoft Excel spreadsheets as well as CSV files.
[53] Bulk access or commercial use of the data is forbidden among other restrictions – see also respective comments in [97].

3.1.2 European Union

In the following, this work will narrow down its focus for the presentation of exemplary OBD implementations regionally to the area of the European Union and the European Union itself. Still, the discussion will primarily address the publication of data on public finances at other governmental levels than the one of municipalities. Nevertheless, it will give an appropriate overview about the status of OBD within the European Union as well as current state-of-the-art and best practices in this field of information science in Europe.

Implementation of Open Budget Data by the European Union itself

As a joint venture by multiple European countries, standing above everything with respect to public authorities, the European Union is obviously the biggest governmental body in this part of the world. The European Union's budget itself has a volume of nearly 144 Billion Euro in the year 2016 [111]. It should be actually obvious that respective transparency is of the utmost importance – especially, in recent years, for the sake of public's perception and, if nothing else, the public inner communication and audit. It would be only logical for the European Union and its somewhat governing Commission to take the lead in opening up its own finances with state-of-the-art OBD. However, this is not necessarily the case as seen subsequently.

Obviously realising this (and similar) deficits regarding OGD on the level of the European Union, the European Commission released its decision on the reuse of Commission documents in December 2011 [112] – eight years after the PSI directive was issued by the same supranational institution (see chapter 2.4.2). With that, interestingly all documents by the European Commission *shall* be released under their own, somewhat vague "licence", so to say – instead of being simply published as public domain or with one of the available and already existing open licences, for example. Although, this decision requires the publication of these documents without any restriction regarding reuse – also allowing also commercial use (Article 6), "where possible and appropriate" in a machine-readable data formats (Article 8) and that being free of charge "in principle" (Article 9), the incorporated terms and exceptions are so substantial and also in large parts obscure that this decision cannot be implicitly applied to all publications made by the European Commission in general. Consequently, the latter cannot be generally considered as openly licenced in the sense of OKI's OPEN DEFINITION as well. Nevertheless, this decision appears to be the major catalyser for recent developments in this field on the level of the European Union, apart from intending decidedly the implementation of a single point of access to its structured data (Article 5) – more on that later.

According to its decision, the budget of the European Union published by the European Commission as its executive branch is released since the fiscal year 2013 in machine-readable XML and CSV data formats [111] – inexplicably, without any values on planned *revenues* (in contrast to the provided data on expenditures) and their breakdown. The latter ones are contained in respective budgetary documents from the European Commission available as PDF files only. However, these files and therefore, respective budget data is available back to the fiscal year 1958. For terms on reuse and similar conditions, the respective publications refer to the mentioned decision by the European Commission. The same applies to actual values on the budgets or statements of account, respectively – no such machine-readable data is published by the European Commission, only PDF documents.

Furthermore, following the mentioned Article 5 of its own decision, the European Commission set up the "European Union Open Data Portal" (EU ODP) as a single point of access to its structured data in 2012 [113]. However, especially with respect to budgetary data, the EU ODP does not provide *any* additional value or information. In fact, (exactly) the same machine-readable data on the European Union's budgets as available on the website of the European Commission itself can be obtained here as well. Its only benefit of an implemented, state-of-the-art SPARQL endpoint is wasted as the EU ODP only provides (more or less meaningless) metadata for the provided XML and CSV files by its means – the actual data and contents within these files cannot be queried through this endpoint and additionally, the metadata itself does not describe the dataset sufficiently in a machine-readable way. Interestingly, there exists a full-blown RDF ontology specifically for the European Union's budget on the same platform [114], although obviously unused for an actual application on the EU ODP to this day. The overall shortcomings of the EU ODP regarding OBD can be observed quantitatively via its usage: all datasets together[54] were downloaded only less than 100 times from the EU ODP since their publication until August 2017.

The deficits of European Union's implementation on its own OBD is also criticized by recent literature [115], which points out the lack of standardisation, but also of the consolidation and completion of available budgetary data as well as existing inconsistencies and missing metadata. The reluctance for cooperation and involvement in this field by the European Commission itself was also revealed. Hence, the collection, analysis and understanding of budgetary data of the European Union can be considered as "a complex task which hampers EU Budget transparency" [115, p. 27].

Governmental Open Budget Data by the United Kingdom

Although the European community failed somehow to publish OBD on the level of the European Union down to the present day, there are some outstanding examples among its member states, which may prove to be of major interest for this very specific Diploma Thesis. One of the two subsequently examined countries is the United Kingdom as one of the paragons according to the recent OPEN DATA INDEX (see chapter 3.1.1).

Responsible for respective publications, but also for the development and execution of the central government's public finance and economic policies in this country is the British government department also known as "HM Treasury"[55] – the British equivalent to the Austrian Federal Ministry of Finance. All respective content released by the Treasury – as well as all other parts of the UK's government – is licenced explicitly with the so-called "Open Government Licence" (OGL) [116], a legal framework introduced by the UK's National Archives in September 2010. The terms of OGL are compatible and equivalent to the ones of CC-BY and ODC-by (see chapter 2.4.1). Therefore, OGL is also compliant to OKI's OPEN DEFINITION and, as a consequence, all of the subsequently discussed publications of HM Treasury.

Besides regular budgetary documents, that shall be discussed later on, HM Treasury is releasing nowadays huge amounts of the contents of its own accounting database in regular intervals. Starting in June 2010, the "Combined Online Information System" (COINS) provided the basis for first extensive data releases [117]. By September 2013, COINS was replaced by the "Online System for Central Accounting and Reporting" (OSCAR) and provided even more extensive data to the public since then [118]. The reason for this system change – including a conceptually caused structural break in time series data – was to create a more user-friendly system that provides key management information and data for the government itself as well as public reporting [119]. Therefore, the quality of data these systems hold was improved by simplifying data structures. Additionally, data validation and control processes were strengthened. To be precise, COINS as well as its successor OSCAR can be considered as consolidation systems and databases rather than actual accounting applications [120]. They do not hold the details of individual financial transactions. In fact, they are used to collect financial data from all across the public sector including the individual governmental

[54] These are four budgetary datasets on expenses by the European Union for the fiscal years 2013 to 2016.
[55] "HM Treasury" is a shortcut for "Her Majesty's Treasury" – sometimes more informally referred to as "the Treasury".

departments. This is done in order to support centralised fiscal management, the preparation of estimates for future budgets and respective discussions, the production of public expenditure statistics like UK's "Whole of Government Accounts" or "Public Expenditure Statistical Analyses" and to meet data requirements of the regional Office for National Statistics.

In order to be able to publish consistent historical time series data all UK's governmental departments are required to actively maintain this data within OSCAR for up to nine years: five historic (or outturn) years, the current year and up to three future (or plan) years depending on the timing of the latest spending review [120].

The financial data is modelled within COINS and OSCAR as a hyper-dimensional data cube (more on that later in chapter 4.1): Each record can be identified uniquely by several "facts" (e.g. the data type – like target or actual values, the point in time to which the data relates, etc.) and is associated with three hierarchical organized, main dimensions [121, pp. 4-5]:

- → The *organizational* dimension identifies the responsible governmental department.
- → The *accounts* dimension names the economic nature of each record.
- → The *segment* dimension records the function associated with the data, similar to a account.

In this way, the UK's data model for accounting of public finances resembles features of the Austrian VRV (accounts and approaches, see also chapter 2.2.4), albeit being conceptually more sophisticated.

Although the respectively released data originating from COINS and OSCAR concerns again – like in the USA – governmental spending and expenditure without respective revenues, the provided information is detailed and complies with modern OBD standards: the data is provided – for free and openly licenced – in a machine-readable CSV format and consists of detailed forecasts and outturns – or values on budgets and statements of account – published annually after final balancing of each fiscal year and during each of them on a quarterly basis as in-year reports [118]. The respective released amount of data (and its level of detail) is quite astonishing: Relating to the annual report for the fiscal year 2016, 2.4 Gigabyte of CSV data with more than 1.5 million records on governmental spending are available as extract from OSCAR's database.

This amount of data is far too large for loading into widely available analysis tools such as Microsoft Excel[56]. Moreover, the extracted data from OSCAR is not self-explaining at all. In fact, OSCAR is a highly technical information source, which requires detailed knowledge of the used encoding schemes and all attributes in conjunction with the financial background to make any sense of the raw data. These limitations and semantic circumstances were also realised by HM Treasury during the first release of data extracts from COINS and, as a consequence, already in 2010 – the same year the first data release took place – the data was also released as Linked Open Data [117] resulting in the first known LOBD implementation of this kind by a governmental authority in the world. The detailed conversion process of COINS data into an RDF graph was described by Ian Dickinson on behalf of UK's public administration in [120]. In a nutshell, for this process detailed conceptual considerations were made in advance: A clear and comprehensible URI scheme for referencing the converted data was developed and an already existing RDF ontology was used to model the data in a semantically correct way. Interestingly, this ontology, although it was still in development at this time, is identical to the one used for LINKEDSPENDING four years later (see chapter 3.1.1). As already mentioned, this ontology will be also of major interest later in this Diploma Thesis.

However, where there is light, there is also shadow: Even if the UK is the country with the most advanced publication on OBD and even LOBD with respect to governmental expenditures, the budget

[56] For example, Microsoft Excel's limit to the size of its spreadsheets is situated at about 1 million rows [144].

FIGURE 24 – The "Daily Bread" calculation of UK's governmental spending per day by "Where does my money go?"

Source: Open Knowledge International [123]

itself as well as respective statements of accounts are not published in a machine-readable format. Only PDF versions of all budgets since 1992 are available on the website of HM Treasury [122].

Nevertheless, UK's budget – and due to respective extensive releases of COINS and OSCAR, especially regarding governmental spending – has been a reference for various leading visualisations in the field of governmental OBD. One of the most notable and frequently mentioned projects of this kind is called "Where does my money go?" based on data published via the OBD platform OPENSPENDING. This website-based visualisation [123] was developed by OKI after an idea originating from as early as 2007 and enabled the general public to consume the information of UK's OBD in an easy and comprehensive way. Thereby, "Where does my money go?" set an example for various similar projects succeeding this one. The tree-based structure of its visualisation was reused as well as the proportional calculation of the usage of each one's taxes as depicted in **FIGURE 24**.

Governmental Open Budget Data in Germany

The second example on governmental OBD of a European country that shall be presented in this very chapter is as Austria's great neighbour – with respective influence – Germany. Additionally, the Federal Republic is ranked – likewise as the UK – at the first place in the subcategory for OBD in 2016's OPEN DATA INDEX. This means Germany publishes its central government's budget in full compliance with OKI's OPEN DEFINITION. Enough to have another, very briefly look at an exemplary implementation of OBD.

Since 2012, Germany is publishing its central government's budget in a machine-readable way on a website of its Federal Ministry of Finance [124] and a general OGD platform [125]. Nowadays, this data on Germany's budgets is available as Microsoft Excel spreadsheets as well as CSV and XML files – besides the commonly used PDF versions. In Germany, these publications are dedicated to public domain, which is achieved by referencing their Copyright Law that is more generous regarding governmental data collections compared to Austria's respective legislation (see chapter 2.4.2).

However, respective statements of account with actual values on Germany's governmental revenues and expenditures are not available in an open way – only as PDF versions by the Federal Ministry of Finance. Additionally, a release of data about public finances by the means of LOBD still does not exist in Germany as well.

3.1.3 Austria

Last but not least, the implementation of OBD for public finances in Austria shall be investigated in the following sub-chapter. In short, recent developments regarding Austrian legislation concerning OBD, which were not yet discussed in chapter 2, as well as the current overall status of OBD publications by public and private institutional units on municipal budgets and statements of account in particular will be of interest subsequently.

Recent legal developments regarding Open Budget Data in Austria

As mentioned in chapter 2.4.2, since 2003 or 2005, respectively, the PSI directive by the European Union as well as the IWG as its regional implementation regulate the framework under which the publication and reuse of public sector information and, as a consequence, also OBD has to happen. However, there was *no* obligation for governmental authorities to do so after all. Moreover, the national UrhG places governmental publications of budgetary data (as compilations like databases for compiled data) under general copyright law by default. Hence, the implementations regarding OBD for public finances are still in their early stages of development as seen later on.

Nevertheless, in recent years there seems to appear a glimmer of hope: With the ÖStP from the year 2012 central, state and local governments have agreed upon – among others – an actually ground-breaking regulation on governmental OBD by mutual consent – Article 12, paragraph 1, ÖStP:

> *"Die Haushaltsbeschlüsse der Länder und der Gemeinden sind in rechtlich verbindlicher Form zu fassen und öffentlich kundzumachen. Bund, Länder und Gemeinden haben ihren jeweiligen Rechnungsvoranschlag und Rechnungsabschluss inklusive aller Beilagen zeitnahe an die Beschlussfassung in einer Form im Internet zur Verfügung zu stellen, die eine weitere Verwendung ermöglicht (z.B. downloadbar, keine Images oder PDF)." [2]*

This means that Austrian public bodies are in principle obliged to publish their budgets as well as their statements of account in an open, reusable and machine-readable way since 2012 or 2013[57], respectively. Furthermore, according to ÖStP's Article 12, paragraph 4, each and every "relevant" institutional unit allocated to the public sector in accordance with the ESA as well as all governmental funds shall be revealed and their finances made public this way.

Respective legal implementations of this regulation of the ÖStP 2012 can be found in corresponding municipal codes of the Austrian Federal States. In respective equally worded amendments Lower Austria's municipal code[58] has adapted the related regulation of the ÖStP at the beginning of 2014, Carinthia's[59] in February 2015. Other Federal States in Austria still did *not* make respective amendments to this date.

Although, in general these regulations do not mention the concept of "openness" strictly speaking (which leaves the possibility to interpret them more liberal), the requirement on potential further reuse does induce it, if things get thought through to the end. Also, the terms of the PSI directive and the national IWG are applying to such publications as well. So indirectly, appropriate (open) licencing and similar features for real Open Data can be expected this way. A more direct national definition or, at least, a respective explicit reference within the ÖStP itself and future, succeeding regulations would be preferable, anyway.

---

[57] As mentioned in chapter 2.2.2, the ÖStP became applicable for all municipalities, Federal States (except Salzburg) and the central government on 1st January 2012. Later, on 1st January 2013 the regulation became also applicable for Salzburg [138].

[58] See Lower Austria's municipal code's Article 73, paragraph 5 for budgets; Article 84 for statements of account [145].

[59] See Carinthia's municipal code's Article 86, paragraph 9; Article 90, paragraph 2 for statements of account [146].

Governmental Open Budget Data in Austria

Currently, publications on governmental OBD in Austria are primarily made in three different ways:

- → through national open data platforms like "data.gv.at"
- → on official websites of the respective public body itself
- → via platforms provided by private initiatives

While the latter option will be of separate interest in the subsequent section of this chapter, the first one is meant to be as a "single point of contact" for citizens as well as meta-platforms comprising all different scattered (open) data catalogues in Austria. Thereby, OGD shall be centralized available for automatic as well as manual queries. Responsible for development and maintenance of "data.gv.at" is the Federal Chancellery in conjunction with the City of Vienna. It has been online now since April 2012 [126]. In October 2014, even a similar (pilot) platform for LOD as a centralized catalogue for respective data repositories in Austria was launched [127]. However, beside 18 "basic datasets" that were released with the platform itself and which provide only metadata for the actual documents (none includes OBD), no more datasets have been published since then and the project seem to have been somehow abandoned since its launch in October 2014, as no activity is observable ever since. Besides that, as the second listed option above indicates, individual public bodies can of course choose to make publications on their own websites and open data platforms (if any exists) – without any enrolment to a central platform like "data.gv.at", however, these publications are hardly (or, at least, only with respective efforts on data collection) trackable and, subsequently, usable.

In any case, the real situation of implementation on governmental OBD in Austria via these means presents itself quite differently: Although applicable legal regulations and respective decisions of various public authorities may intend something different, and also respective infrastructure for public OGD distribution is obviously available, there is still *no* nationwide implementation of municipal OBD in Austria. The central government's budget makes no exception.

The latter is available on the website of the Austrian Federal Ministry of Finance [128] since the fiscal year 2012 – however, only as PDF files with supplemental Microsoft Excel spreadsheets. Additionally, these publications are not licenced in an open way, but instead – even more critically – explicitly all rights on these documents are reserved by a respective copyright notice. The same applies to respective statements of account for Austrian public finances, but they are actually published as PDF versions *only* – with some data available directly on the website of the Federal Ministry of Finance. It is no wonder that the central government has not published *anything* regarding public finances on the national open data platform "data.gv.at" to this date as well. So, in fact, there has not been *any* real OBD publication whatsoever by Austria's central government down to the present day.

With respect to Austrian municipalities, the situation looks a little bit more differentiated: indeed there are respective OBD publications available. However, they still remain within narrow limits up until now: only *eight* datasets on each budgets and statements of account for municipal finances are available on the governmental platform "data.gv.at". Included are respective datasets for the three biggest cities in Austria: Vienna, Graz and Linz. Most notable is the one of the City of Graz offering apparently the most detailed OBD on its finances since 2009 – with more than 8,000 records or individual aggregations of accounts and cost approaches per dataset and fiscal year. In comparison, Vienna's budget is available as time series data since 2001 – however, with only about 2,700 records per fiscal year and no recent publications for 2016, 2017 & 2018. On top of this, published data on its statements of account consists only of highly aggregated sums (ten on revenues and seven on expenditures), anyway. So, the small amount of published OBD on governmental platforms is obviously varying in structure, quality and even quantity. An improvement remains to be seen.

Private distribution of Open Budget Data in Austria

Obviously, the direct distribution of OBD of Austrian public bodies is not (yet) expedient and effective. Therefore, the only alternative are private initiatives. The most successful by far is the web-based platform "OffenerHaushalt.at" by the so-called "Zentrum für Verwaltungsforschung" (KDZ) [129], a non-profit society originally founded by the Austrian Association of Towns and Municipalities which focuses on fundamental work about various economic topics in the field of activities of the public sector.

Since November 2013, local governments can publish data on their budgets, respective statements of accounts, liabilities and debts on "OffenerHaushalt.at" for free in an open way – in compliance with recent regulations like the ÖStP 2012. In addition, mayors can simply "unlock" their municipality on the platform and approve to the (open) publication of already present datasets originating from Austria's statistical office STATISTIK AUSTRIA for the years from 2001 onwards. Interestingly, these datasets consists only of *target* values originating from *statements of accounts* of respective years. According to the KDZ, values on budgets are not obtainable through STATISTIK AUSTRIA; instead, this data has to be uploaded by the municipal authorities themselves. However, most data on this platform consists actually of the data from STATISTIK AUSTRIA only, but it can be considered as *sufficiently* detailed as it contains all revenues and expenditures on the level of – according to the VRV – mandatory account groups and cost approach subsections[60]. This data is provided for the last full fiscal year every October. So, 2016's statements of account for all unlocked municipalities were made available on "OffenerHaushalt.at" in October 2017. Additional data releases by the local governments themselves – under the demand applied by the ÖStP 2012 and similar regulations – may be more detailed[61], provide figures of annual budgets and thereby, complete the view on municipal budgets as a whole. All this data on "OffenerHaushalt.at" published so far consists out of CSV files and is released generally in terms of CREATIVE COMMONS' attribution licence (CC-BY). Its coherence and consistency is ensured as only data compliant to the Austrian GHD interface (see chapter 2.2.6) can be uploaded by municipalities or is provided by STATISTIK AUSTRIA.

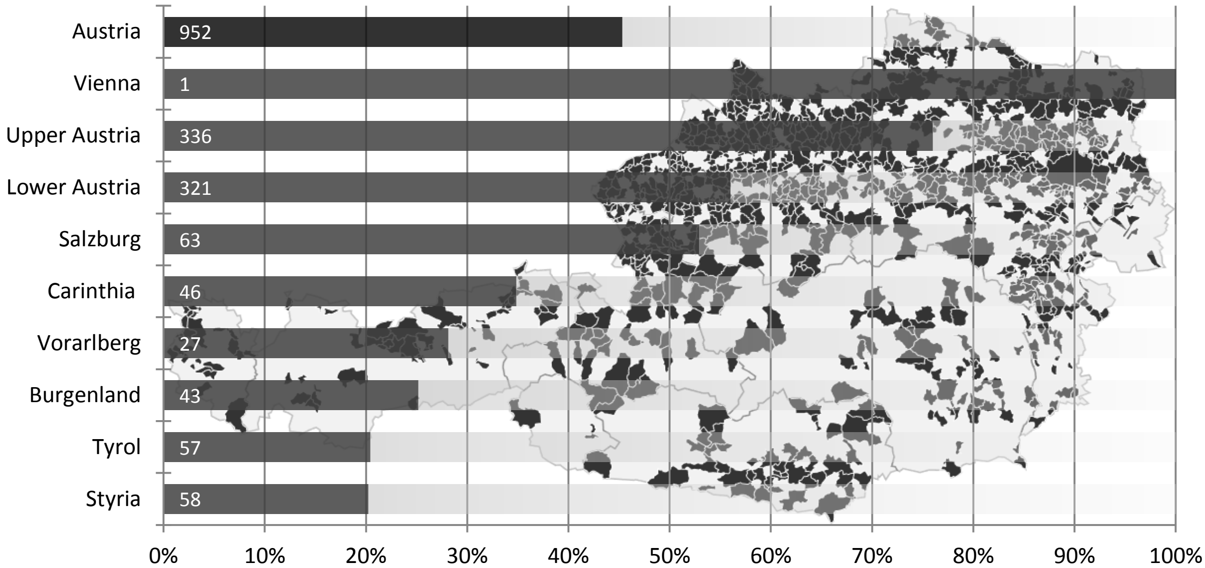
By the end of March 2017, "OffenerHaushalt.at" released a dedicated data portal with respective webserver endpoints, enabling automatic bulk access to the platform's database, as well as a download interface for manual bulk access – also licenced under CC-BY. Although a LOBD approach was obviously investigated during its development [130, p. 17], this new data portal confines itself with providing interfaces for metadata and bulk access to the mentioned CSVs instead of providing the data itself in a more sophisticated way. However, in its current form at least automatic parsing of the platform's data pool is possible and, hence, its enrolment on meta-platforms like "data.gv.at". The permission for the latter was indeed given by the KDZ to the Federal Ministry of Finance [130, p. 23], but to this date the datasets from "OffenerHaushalt.at" are still not available on "data.gv.at".

With respect to the overall coverage of released data on public finances of Austrian municipalities, "OffenerHaushalt.at" is leading (by far – compared to the spare available publications of public authorities via "data.gv.at"), but is still in its development and not all-encompassing throughout Austria: By August 2017, 952 municipalities (of 2,100) have released their budgetary data through "OffenerHaushalt.at" – with 45% of all municipalities, this is not even their majority in Austria. This as well as the respective figures per Federal State is depicted in **FIGURE 25**. Regionally seen, the local governments in the states of Upper Austria, Lower Austria and Salzburg (excluding Vienna as a city state, which has published its finances as OBD here too) are leading in releasing their budgetary data.

[60] These are the first three decades of each part of VRV's budgetary classification – see also chapter 2.2.4.

[61] Approach and account aggregations on level of all six – according to the VRV – possible decades can be provided hereby.

FIGURE 25 – Municipalities which released OBD on "OffenerHaushalt.at" in absolute and relative figures per Federal State



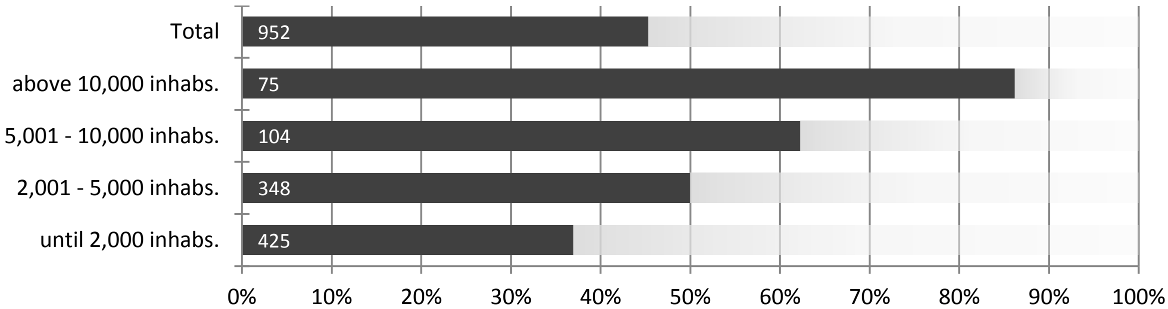
Source: Krabina et al. [129] & [130, p. 5], own calculations & illustration – data as per 15th August 2017

The reason for this somewhat unbalanced ranking may be (even internal) regulations and decisions on the implementation of ÖStP 2012 and similar legislation. However, as also shown by Krabina [130, p. 6], the population size of the affected community has obviously a significant effect on its OBD publication – on "OffenerHaushalt.at", at least: here, more than 86% of all municipalities with a population over 10,000 citizens have published their budgetary data, but only 44% from the rest. A declining trend – the smaller the municipality, the less likely is its local government to release its budgetary data – can be clearly seen through more detailed numbers depicted in **FIGURE 26**.

Provided that the residual municipalities do not publish their budgets and statements of accounts – according to the ÖStP 2012 and similar regulations – on their own (possibly still inexistent) website or directly on "data.gv.at", they should perform (at least) the release process on "OffenerHaushalt.at" as an easy mean to accomplish their publication obligations. Currently, about 60 municipals are approving the publication of their budgetary data on "OffenerHaushalt.at" each year – given the numbers from the last two years. Thereby, a full coverage of the platform's database for each and every municipality would be reached as late as in nearly two decades – provided that nothing regarding the present publication habits of Austrian municipalities will change.

Nevertheless, "OffenerHaushalt.at" has pioneered the field of nationwide, coherent and open distribution of municipal budget data in Austria – including state-of-the-art visualisations like those ones from the project "Where does my money go?" (see chapter 3.1.2) – and still does. Its future, growth and further developments remain to be seen.

FIGURE 26 – Municipalities which released OBD on "OffenerHaushalt.at" per population class



Source: Krabina et al. [129] & [130, p. 6], own calculations & illustration – data as per 15th August 2017

3.2 ANALYSIS & COMPARISON

After the survey of current implementations on publication of governmental OBD around the world, in the following, final part of this chapter these implementations should be assessed and evaluated according to common standards and guidelines in this field of information science. In the end, possible best practices are worked out for further usage in the last major part of this Diploma Thesis.

3.2.1 Assessment of current implementations

As seen in the preceding chapter 3.1, there already exist various actual implementations on real OBD and even LOBD. However, their quality and extend as well as the overall awareness for this (potentially very beneficial) topic varies tremendously across different countries and among all parties involved. On top of this, most of these implementations do not use existing, modern technologies for real and encompassing open and linked data distribution.

This perception is reflected by relevant surveys like OKI's OPEN DATA INDEX and its respective sub-categories on governmental budgets and spending as well as the more specific OPEN BUDGET SURVEY by the IBP and its accompanying OPEN BUDGET INDEX likewise. Both indices reveal that more than the majority of the world's central governments are not publishing (or, in some cases, even producing) sufficient information – let alone, machine-readable and semantic data – on public finances in a transparent way. Hence, the majority of the world's public (and if such information is not produced at all, also their government) is left behind unable to fully understand or monitor how much public finances are raised and how they are spent. This may allow undesirable developments, possible improvements and prospects to be unrecognized, both unintentionally and deliberately. As a consequence, effective use of public resources cannot be ensured (and monitored) this way.

Nevertheless, there are international examples of real OBD and even LOBD publications and such ones within the European Union by governmental authorities. In **TABLE 11** below, the discussed implementations are listed with key figures about their type and covered timespan including an assessment by the means of Tim Berners-Lee's 5-Star-Scheme (see chapter 2.4.4).

TABLE 11 – Overview and assessment of surveyed international Open Budget Data implementations

	United Kingdom	Germany	USA	European Union
ODI	2.	24.	11.	N/A
ODI (Budget)	1.	1.	33.	N/A
ODI (Spending)	7.	13.	8.	N/A
OBI	8.	13.	5.	N/A
Open Budget	(✓)	✓	✓	✗
Formats	PDF, CSV, RDF/SPARQL	PDF, XLS, CSV, XML	PDF, XLS, CSV	PDF, CSV, XML
Licence	Open Government Licence	Public Domain	Public Domain	proprietary
RDF Vocabulary	RDF Data Cube Vocabulary	N/A	N/A	N/A
Since	PDF: 1992 CSV, RDF & SPARQL: 2010	PDF: 2005 XLS: 2012 CSV & XML: 2013	PDF: 1996 XLS & CSV: 1998 with time series data since 1962	PDF: 1958 XML: 2013 CSV: 2014
Open Ranking	★★★★	★★★	★★★	-
Comments	spending as full & detailed (L)OBD in CSV & RDF format, revenues only as PDF	no open data on public spending, proprietary XML encoding for budget data	data for governmental spending on transactional level also available	no common licence used, proprietary XML encoding, revenues only as PDF

ODI = Open Data Index (with respective numbers for sub-ranking in the categories of governmental budget and spending)
OBI = Open Budget Index

Source: Own illustration & survey, Open Knowledge International [97], International Budget Partnership [99]

With that, it becomes obvious that the UK is one of the paragons in the field of OGD in general and OBD in particular. Their LOBD implementation is the first real one published by a governmental authority so far and can be considered demonstrating current state-of-the-art. This accomplishment is also represented by UK's ranking at the upper end in respective presented indices.

And even outside the European Union, budgetary transparency of US public finances made relatively good progresses in comparison to other countries in recent years and its today's level can be considered as widely corresponding to general standards and guidelines on openness.

However, the surveyed status on OBD showed also examples of how not to do it: the European Union itself failed to publish its own budget in terms of a generally accepted, open licence. And in Austria, municipal budgetary data is published by public authorities themselves rarely – and if so, partially in insufficient ways – although recent legislation demands it differently. The Austrian central government itself makes no exception, failing in publishing OBD at all, with no real implementation regarding this matter down to the present day. Still, Austria's three biggest cities – according to their population – have accomplished to release OBD on their budgets and even their statements of account. These national, governmental releases are also summarized – like in **TABLE 11** – with an assessment through the 5-Star-Scheme in **TABLE 12** below.

TABLE 12 – Overview and assessment of surveyed national Open Budget Data implementations in Austria

	Central Government	Vienna	Graz	Linz
Open Budget	✗	✓	✓	✓
Formats	PDF, XLS, HTML	PDF, CSV	PDF, CSV	PDF, CSV
Licence	All rights reserved	CC-BY	CC-BY	CC-BY
Budget since	PDF & XLS: 2012	PDF: 1999 CSV[1]: 2001 (until 2015)	PDF: 2006 CSV: 2009	PDF: 2007 CSV: 2013
Statement of Account since	HTML: 2012 PDF: 2015	PDF: 1998 CSV[2]: 2002	PDF: 2004 CSV: 2009	PDF: 2005 CSV: 2012
Open Ranking	-	★★★	★★★	★★★
Comments	No real OBD publications whatsoever until today	[1] ca. 2,700 records per year [2] only highly aggregated (17 records per year)	ca. 8,000 records per year and dataset	ca. 5,000 records per year and dataset
Source	bmf.gv.at	wien.gv.at & data.gv.at	graz.at & data.graz.gv.at	linz.at & data.gv.at

Source: Own illustration & survey

Where public authorities fail, private initiatives try to compensate. Internationally seen, the OPENSPENDING project and its LOBD derivate LINKEDSPENDING are most notable. In Austria, "OffenerHaushalt.at" is one of the major accomplishments of this kind, when it comes to municipal budgets and statements of account. However, both their coverage leave much to be desired.

It should be mentioned at this point, that not only the availability and quantity of OBD is important, but also the actual *quality* of the respective OBD implementation is of the essence. In general, budgetary data is highly complex information and deduced knowledge can only obtained with already existing knowledge on budgetary processes and their implementation in order to make any sense of the raw data. For the sake of this Diploma Thesis, this should be achieved through the detailed discussion in chapter 2.2. On the other hand, Linked Data – implemented in a correct and carefully considered way – can be used (like in the UK or with LINKEDSPENDING) to overcome this issue in automated way. However, today a lot of implementations in this field fail at this point. For example: Vienna, as Austria's capital city, is publishing only 17 records on data relating to statements of account each year, whereas Graz, as the second biggest city, publishes over 8,000 records of OBD on the very same matter. In both cases, possibly more knowledge can be obtained via already existing and commonly used PDF releases on public finances, as Vienna does not provide enough information in a machine-readable way, while the data provided by the city of Graz can become overwhelming for those ones, who are not familiar with the subject. If both municipalities would refrain from publishing their indeed quite structured, explanatory and informative PDF documents on public finances, transparency would decrease as a consequence. This should make oneself clear, that the pure existence of OBD does *not* state *anything* about its quality and a subsequent information gain. Beyond that, transparency could even be hampered by wrong implementations and insufficient budgetary data.

All in all, the current status of Austrian implementations on OBD in particular is somewhat devastating. However, this perception can be received throughout the world – as seen with the mentioned indices and surveys like the OPEN BUDGET SURVEY, which brings it full circle: Current implementations on the publication of budgetary data in a reasonable, comprehensive, transparent and open way are existing, but still in their early stages of development. The overall coverage and quality of the found implementations reveal that developments in applied OBD of governmental bodies are in an initial phase, awaiting more momentum and even more quality through recent developments in legislation, information technology and overall requirements by current society.

3.2.2 Recapitulation of possible best practices

Although, most of the existing implementations on OBD are – as seen in the past chapter of this work – not rudimentarily flawless, and examples of real LOBD are seen rarely up until now, they already provide in parts a state-of-the-art as well as possible best practices that can be used in order to provide more transparent, coherent, reasonable, comprehensive and comparable budgetary data for Austrian municipalities – and with respect to the scope of this Diploma Thesis.

These best practices can be narrowed down to three main subjects for an actual implementation:

- → data format, storage and distribution,
- → data vocabulary and ontology, and
- → data licencing.

Of course, all of these subjects should be done in an open and transparent way, as described in chapter 2.4. However, for each of them, respective actual implementations have been found in the course of the survey taken in this very chapter and the respective concepts were discussed in the previous ones. These insights should be used now for this work.

As described multiple times up until now, the concept of *Linked Data* can provide the means to distribute information with self-defined semantics and meaning, in order to overcome issues of highly complex topics like public finances. Both LINKEDSPENDING as well as the conversion of the released budgetary data by UK's government used the so-called "*RDF Data Cube Vocabulary*" to model such LOBD. Considering this as state-of-the-art in this field, this ontology will be discussed for further usage in the following last major chapter of this Diploma Thesis (see chapter 4.1).

And last but not least, as the release of budgetary data dedicated to public domain is not possible in Austria – as seen within the chapters 2.4.1 & 2.4.2, the *CC-BY licence* by CREATIVE COMMONS – as the minimal legally possible licencing in Austria – can be considered as national, common best practise, as it is picked up by the Austrian governmental framework for OGD platforms as well.

A detailed discussion about these best practices for use in an application for an exemplary approach to LOBD for Austrian municipalities will follow in the next, final chapter of this Diploma Thesis.

CHAPTER 4

AN EXEMPLARY APPROACH TO LINKED OPEN BUDGET DATA FOR AUSTRIA

In the following final part of this Diploma Thesis the combined knowledge laid out in the preceding chapters 2 and 3 as well as all the assessment done so far will be used to find an answer to the question already raised at the very beginning of this work in its first chapter:

→ How can the principles of Linked Open Data be applied on local government finances in Austria?

The respective answer will be tripartite: First of all, an exemplary approach – which will be discussed in detail in the upcoming chapter and, in short, combines the discussed principles of Open Data and Linked Data with local government finances in Austria – is created, in order to provide the means for future applications on a new, more transparent, coherent, reasonable, comprehensive and comparable level of implementation to better fulfil overall requirements by current society, governmental institutions and law. The mentioned approach will concentrate on Austrian municipalities and a respective data model and format, the creation of such data and its retrieval – not on its actual end-user-friendly representation. This is made on purpose, in order to limit the scope of this Diploma Thesis in an appropriate way – as already stated at the beginning in chapter 1.

The second major outcome of this very chapter will be an assessment of this approach according to the common standards and guidelines found in the course of the previous parts of this work. This is made in order to verify whether the chosen approach is really an appropriate solution for the given problem.

And last but not least, in the very end of this work (in sub-chapter 5.3) open questions and issues, which could not be addressed (on purpose) in this Diploma Thesis and its exemplary approach, will be discussed in a nutshell. In this way some helpful foundation for possible, future works and applications on the very same matter should be provided after all.

With that said, the final major chapter of this Diploma Thesis will structure the first and the second part of this answer to the formerly given question in three segments: In the **first** part the general concept and all given technology, that will be used for the described exemplary approach, will be discussed. In this way, all applied components for real LOBD should be presented, which are needed to address the given problem in the given field of public finances in Austria. The **second** part of this chapter will document the implemented exemplary approach itself, the used data model and format as well as the constructed software architecture for transforming, storing, handling and publishing such data. And in the **third** and final part of this chapter the chosen approach, its underlying concept as well as the prototype implementation itself will be assessed with predefined goals to measure its accuracy of fit with respect to the given problem as well as common standards and guidelines.

So, before starting off with this very chapter, the mentioned goals and aims, which should be reached with the to-be chosen approach, shall be laid out in advance. Eventually, this will serve as a guideline and golden thread on its own for the upcoming, final part of this Diploma Thesis.

General goals for an exemplary approach on LOBD for Austrian municipalities

During the discussion of all the technologies and standards seen so far, also a lot of principles and guidelines were presented – or, at least, touched – *how* they *should be* actually used in practise. Naturally, we won't go into very much detail on these matters (again); however, as the evaluation of the results of this work is crucial to its overall consistency and completeness, a predefined set of goals and aims for the upcoming development of an exemplary approach on LOBD for Austrian municipalities is inevitable.

Of course, the variety of possible aims in such a complex topic as public finances combined with state-of-the-art technology could be sheer endless. Nevertheless, the basic requirements for the to-be chosen approach need to be narrowed down to some meaningful and significant aspects that the result should feature, eventually – although, much more extensive considerations will take place during the implementation of the chosen approach, probably integrating other, more detailed requirements on certain parts. However, in short, these aspects can be condensed into the following, thematically separated points:

→ *Openness*
The implementation, its data model as well as all of its results should be open in the sense of OKI's OPEN DEFINITION (see chapter 2.4.1) and in full compliance with it. So, the resulting work itself as well as its incorporated licence and data format should be open according to that.

→ *Linked Data*
As announced earlier, the chosen approach will use Linked Data and its various technologies (see chapter 2.3) as such for data modelling, representation and distribution. The chosen approach – of course – should not use Linked Data just for its present's sake. Instead a full-blown approach to Linked Open Data (see chapter 2.4.4) is intended. In order to verify, whether this goal is reached in the very end, the resulting data is scored with Tim Berners-Lee's 5-Star-Scheme and a rating with all five stars shall be reached (see **TABLE 10** on p.58).

→ *Correct usage of Vocabulary*
The to-be chosen vocabulary used for modelling and representing the budgetary data as LOBD induces significant challenges on its own. It shall be guaranteed that the vocabulary is used as intended and the data is consistent in itself. The necessary formal constraints that need to be imposed on the resulting dataset and its application will heavily depend on the vocabulary and the created data model itself. They can and will derive from the specification of the to-be chosen vocabulary itself, of course.

→ *Accuracy of fit of the solution for the given subject*
Finally and most importantly, the result must be an adequate solution and properly implemented application in the field of public finances or municipal budgets, to be more precise. For the matter of this Diploma Thesis, the aim is, to verify this via a fully automated generation of budget profiles from exemplary municipalities in accordance to the VRV, which model the budgetary data on an aggregated, but encompassing level (see **TABLE 7** on p.24 or chapter 2.2.4 for a general description). The figures and numbers in these generated budget profiles based on LOBD of the chosen approach need to correspond – exactly and in every detail – with the ones from the officially published ones.

With these four criteria, the assessment, that will take place in the last part of this very chapter, shall ensure a comprehensive, but also reasonable evaluation of the practical portion of this Diploma Thesis. Additionally, these criteria can be used in a generalised way also by possible, similar works and applications on the very same matter in the future, by the way.

4.1 CONCEPT

In order to establish an exemplary approach for LOBD for Austrian municipalities in an orderly manner, a technical implementation concept should be laid out in the following sub-chapter of this Diploma Thesis. This concept will be separated in five different parts – all addressing primarily the eponymous data in some way or another, as a matter of fact:

→ data *integration*
→ data *format* & *structure*
→ data *identification*
→ data *storage* & *handling*
→ data *(re-)publication*

In this way, the overall applied foundations for the chosen approach based on all the theoretical aspects of this work (see chapter 2) as well as current state-of-the-art shall be covered in an appropriate way.

Data integration

As the data for the given approach will not be produced within the conceptualized application itself, but rather given budgetary data from some open source will be used. Of course, this data needs to be incorporated and integrated by this application somehow.

For this matter, a typical method used widely as well as commonly in Informatics will be applied: a so-called "ETL process", where "ETL" stands for "Extract", "Load" and "Transform". This process was already mentioned during the discussion of given examples of real LOBD or LINKEDSPENDING, respectively (see chapter 3.1.1). A similar process as in this practical example will be also used by the exemplary approach of this very work. But what is actually happening in this process? To answer this question, it is necessary to describe this ETL process in general.

First of all, such ETL processes exist in fact since the beginning of database applications altogether, although, separate, explicit discussions about this topic emerged not long ago in the first decade of the 21st century [131, p. 1095]. As data processing became more and more important, and an even more challenging and complex task in the recent past – especially, when considering the growing amount of available, digital data – more and more attention was attracted by this ETL process.

With respect to the process itself, it consists – as its abbreviation might suggest – of three separate steps. Very briefly and concentrating on the relevant parts for this work, these steps are responsible for [131, pp. 1097-98]:

→ *Extraction of data from a given data source*
In technical terms of this Diploma Thesis, this is the simplest task of all, as budgetary data in nearly all cases is already existing – in some form or another – as extracted OBD (mostly, CSV files) from the original governmental accounting or reporting systems. This work will confine itself to choosing an appropriate data source, which provides coherent and consistent OBD for its application's sake.

→ *Transformation of the given data into a target data format*
The acquired data needs then to be transformed into the target format – in case of this work: Linked Data or, to be more precise, RDF data using *some* existing vocabulary. During this operation a wide variety of things can and will happen; among other things these are: cleaning, filtering, splitting, joining, or even, enrichment of data. These steps can be implemented on various levels of a given dataset: on schema-, record- as well as value-level.

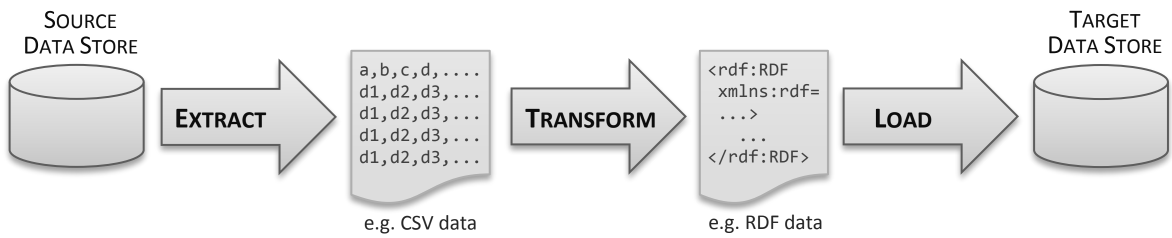
The transformation during an ETL process is the most complex part of the overall topic and various specific works in the field of Informatics are tackling it in more detail. However, in order to remain within the scope of this Diploma Thesis, there will be no more detailed, theoretical discussion at this point – interested readers are referred to relevant literature.

→ *Loading of the transformed data into the target data store*

In the last and final step of the overall ETL process the transformed data is loaded into the target data store. With respect to this very Diploma Thesis, this will be done via an SPARQL endpoint with *some* triplestore (see chapter 2.3.2). Often, in this step various problems regarding performance need to considered. However, this affects in most cases only (really) big amounts of data and, therefore, should *not* be of major interest for the upcoming approach of annually aggregated, budgetary data.

Below, **FIGURE 27** gives a general overview of this abstract process at a glance. The overall ETL process can be indeed much more complex as described at this point. However, for the matter at hand, this discussion will prove to be sufficient.

FIGURE 27 – A general overview of the abstract ETL process in common usage



Source: Own illustration

The specifics of the actual implemented ETL process for given Austrian municipal budget data will be discussed subsequently in chapter 4.2.

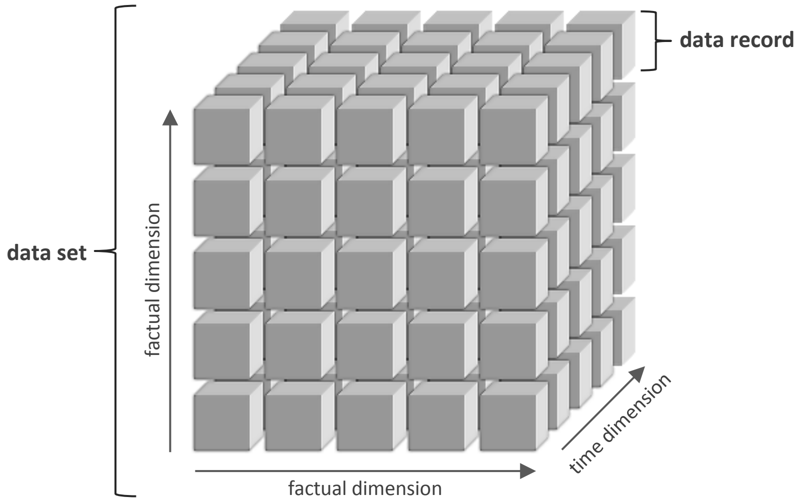
Data format & structure

As overall concept for data representation, this work relies on the presented principles and concepts of Linked Data with its technical low-level implementation of RDF (see chapter 2.3.2). As a result, municipal budgets will be represented in a reasonable, comprehensive, transparent and open way. In order to model the respective data accordingly – with self-defined semantics and meaning as well as to comply with overall requirements by budgetary data and Linked Data itself – the RDF ontology used for data modelling subsequently is chosen carefully and should not reinvent the wheel altogether. Indeed, there are already suitable and standardised RDF vocabularies and ontologies available and in use for this very matter as similar international exemplary implementations show (see chapter 3). But before diving into such standards, the overall used data structure should be explained briefly in advance.

Most of the statistical as well as budgetary data is nowadays organized digitally in multi-dimensional so-called “data cubes”. These data cubes are also often known as “OLAP cubes”, where “OLAP” is standing for “Online Analytical Processing” – a term originating from the field of data analysis. The primary features of such a (three dimensional and simple) data cube are illustrated in **FIGURE 28**.

As depicted, an arbitrary *data set* originating the same thematic area composes a *data cube* with a virtually unlimited number of individual *data records* comprising its elements. Each and every data record itself consists of various attributes and one or more measurands. The record can be located or, to be more precise, is identified by an explicit number of its attributes or key indices. Each combination of these indices is unique for every data record in the cube (e.g. the time and place the

FIGURE 28 – Structure and elements of an illustrated data cube with three dimensions



Source: Own illustration

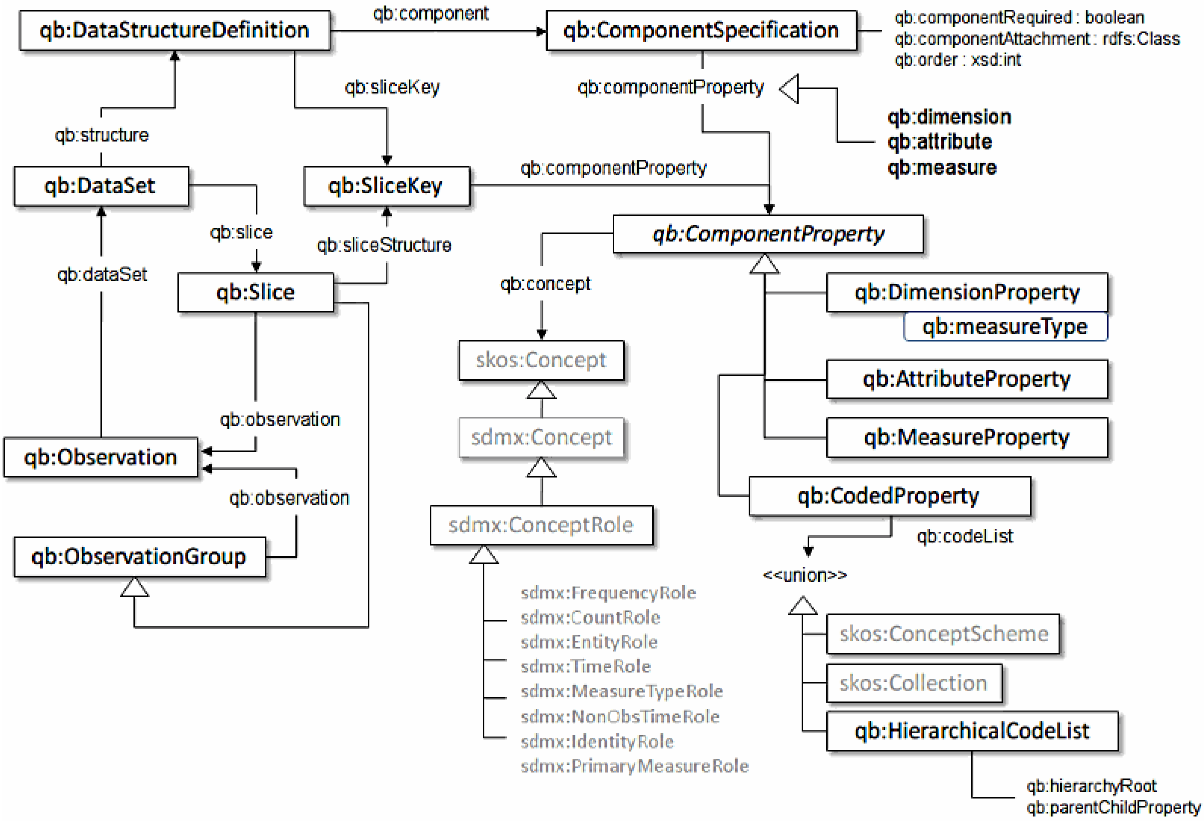
data was recorded and so on). Like the three variables x, y and z are normally used to identify any point in a three-dimensional space, the mentioned indices identify every data record in the cube unambiguously and are called likewise *dimensions*. Normally, one dimension is dedicated as time axis incorporating the changes of the recorded data over time. If a data set comprises exactly three such dimensions, the resulting data cube looks like the illustrated one in **FIGURE 28**. If more dimensions are used, the corresponding structure is often also referred to as a so-called *hypercube*, since it features more than the three spatial dimensions, which we normally experience.

The advantage of organizing data in this – of course – more complex way (in comparison e.g. with classical two dimensional spreadsheets) is – among other things – that reality can be modelled and, as a consequence, described in a more correct and fine-grained way. The data itself as well as analytics based on it are able to be much more meaningful, informative and, if nothing else, more detailed. Of course, this more complex method needs more attention during modelling in order to work as intended.

For this purpose, an initiative called "Statistical Data and Metadata eXchange" (SDMX) emerged in 2002 as a joint venture that should maintain and, if necessary, develop new standards for digital data exchange via such means [132]. The SDMX initiative and its technological input is significant because it is made up by seven important, world-wide organisations: the Bank for International Settlements, the European Central Bank, EUROSTAT, the International Monetary Fund, the Organisation for Economic Cooperation and Development, the United Nations Statistical Division and the World Bank. The first version of the SDMX's set of technical standards was published in 2004, revised multiple times and was finally approved in 2013 as an ISO standard (ISO 17369). The United Nations Statistical Division recommended SDMX as the preferred standard for statistics in 2007 [133].

Based on the data cube model underlying SDMX's information model and after initial, individually published versions beginning in 2010, an official W3C recommendation for a finalized respective RDF ontology was published in January 2014 [133] – the so-called "RDF Data Cube Vocabulary". Although it is based upon concepts for statistical data by SDMX, this ontology is actually not limited to such data. In fact, the incorporated cube model is a more generalized one and, hence, the Data Cube vocabulary can be used for all kinds of other datasets comprising any sort of regular data. But first and foremost, the Data Cube vocabulary is primarily focused on the publication of multi-dimensional data as Linked Data on the World Wide Web (as *Web of Data*, see also chapter 2.3) – and therefore, it is of primary interest for this Diploma Thesis.

FIGURE 29 – Abstract overview of all conceptual elements and their relationship in the "RDF Data Cube Vocabulary"



Source: Cyganiak & Reynolds [133]

The actual, overall structure of the "RDF Data Cube Vocabulary" is depicted in **FIGURE 29** above. At the heart of this ontology are two vital concepts – illustrated in the upper left corner – which every data cube, represented by the means of this vocabulary, should provide:

→ (possibly multiple) *data sets*, and
→ their respective *data structure definition*.

Thereby, these RDF data cubes are self-describing with respect to their meaning as well as their semantics: from every part of the actual represented data, one can find the enclosing data set and the corresponding definition of its structure and meaning. The individual data records themselves are organized as so-called *observations*. Each observation in turn can have one or even more associated *measurements* (e.g. population size, ozone levels or, simply and obviously, the recorded amount of cash flow). In order to be able to reliably interpret such measurements, other information can be provided with each observation or, generalized, over the whole data set via respective *attributes* (e.g. the unit of measurement, or even, virtually any more complex information). All of these possible information structures in the actual implementation are specified in the contained data structure definition as individual *component specifications*, which detail the individual parts an implemented data set (e.g. whether individual attributes are mandatory for each observation or not). With that, actually hierarchical organized, complete lists of codes can be incorporated in such a RDF data cube.

Moreover, this ontology provides the means to organize data explicitly in (arbitrary) so-called *observation groups*. With that, the data can be grouped in beforehand for presentation or analytical purposes or, simply, for ease of access and navigation through the respective data. A sub-class of such observation groups are *slices* of the respective incorporated data cube. Observations grouped together this way have one or multiple fixed dimensions. This is done for the same reason as composing observation groups. If all dimensions except the time dimension or axis are fixed, the resulting slice is called a *time series*; if the free one is a non-time dimension, the respective slice is called a *section*. However, the various types of slices are handled within the “RDF Data Cube Vocabulary” the same way. Explicit definitions of such sub-classes are left to derived vocabularies in the future [133].

In addition to that, the W3C standard is defining certain terms with respect to this ontology for the characteristics of the modelled data cubes themselves, which can be met by them, if certain, explicitly defined criteria are fulfilled:

→ *Normalized* data cubes associate all data (dimensions, measurands & attributes) for each and every observation to the respective RDF subject. So, every observation holds all data connected to it, even if this leads to redundancy within the data structure.
→ *Abbreviated* data cubes associate common parts of data for a set of observations in other, higher levels of the data structure (an observation group, slice or in the definition of the data set itself). For example, if a data slice has fixed the time dimension, the actual time value need not be defined and associated to each observation of this slice; instead, the slice itself can be annotated with this value. If this method is applied to the whole data cube, possible overheads in transmission and storage of such data cubes can be minimized in certain environments. The W3C standard for this ontology provides a normalisation algorithm implemented as SPARQL queries for converting abbreviated data cubes into normalized ones in a standardized way.
→ *Well-formed* data cubes comply with certain, explicitly defined integrity constrains, which should be met by every RDF graph implementing the “RDF Data Cube Vocabulary” correctly. For example, there is checked whether there exists a corresponding and appropriate data structure definition to each existent data set. All 21 integrity constraints are also given as SPARQL ASK queries by the W3C standard.

The overall “RDF Data Cube Vocabulary” and its standardisation had been referenced already in chapter 3.1, although anonymously. Anyhow, this ontology is used explicitly by the surveyed project LINKEDSPENDING since 2014 and – even before it was standardised by the W3C – in UK’s LOBD publication of COINS data in 2010. As it can be considered current state-of-the-art, it will be also used for the implementation done in the course of this very Diploma Thesis. The actual implemented structure of Austrian municipal budget data will be discussed subsequently in chapter 4.2.

Data identification

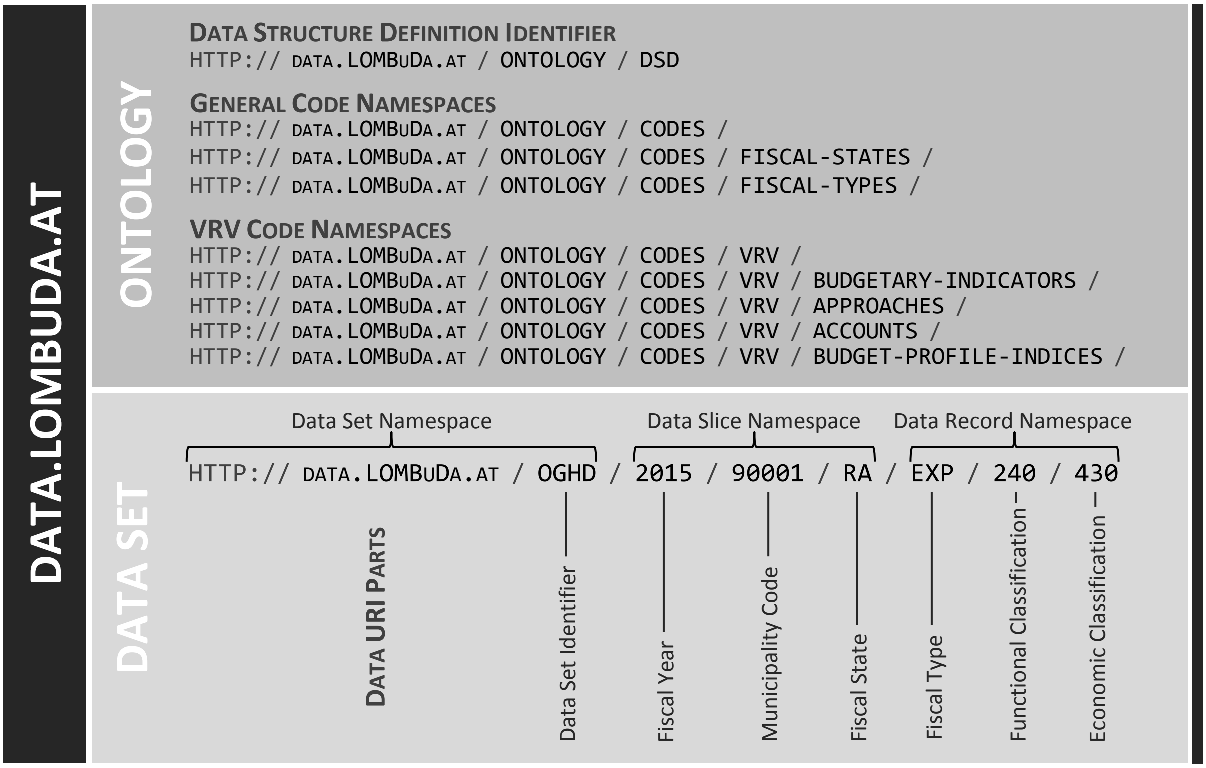
As seen in chapter 2.3 and, especially, its sub-chapter 2.3.3, while using Linked Data and respective ontologies like the “RDF Data Cube Vocabulary” URI’s are used to uniquely identify real world things as well as abstract concepts – or, speaking more specifically, RDF subjects, properties or objects – so, in general, so-called resources. So, for the given approach to model LOBD for Austrian municipalities, naturally also URIs are used for identification of the different data records and their underlying ontology elements – all based on the discussed “RDF Data Cube Vocabulary”.

As these URIs are a basic foundation of Linked Data (see FIGURE 8 on p.31) as well as the to-be conceptualized application itself, already at this point the used URIs or, to be more precise, a respective so-called “URI scheme” should be discussed.

Considering Linked Data’s overall structure as well as the to-be-used “RDF Data Cube Vocabulary” in particular, two major groups of resources need to be considered for a set of URIs or a respective scheme for use in conjunction with Linked Data:

→ The *ontology* and its elements for Austrian municipal LOBD, or its *data structure definition*, and
→ the respective *data set* with its *data slices* and respective *records*.

FIGURE 30 – The URI scheme of the exemplary approach for Austrian municipal LOBD at a glance



Source: Own illustration

Above, **FIGURE 30** shows all the necessary, basic namespaces for the given, exemplary approach. The base URI – or the "authority" as it is called – "DATA.LOMBUDA.AT" stands for <u>L</u>inked <u>O</u>pen <u>M</u>unicipal <u>Bu</u>dget <u>Da</u>ta. This URI (sub-)domain is reserved for the vocabulary and data of the given approach *only* and is controlled by the author of this work directly. Additionally, it is separated completely from the rest of the implementation of technical prototype at "LOMBUDA.AT".

However, in the upper half of **FIGURE 30**, respective namespaces of the ontology elements or the data set's data structure definition are listed: First of all, the identifier of the data structure definition of the vocabulary itself, followed by all namespaces for respective codes within the implemented LOBD – separated in those codes, which are explicitly defined by the VRV, and those, which are defined by the application itself. In the lower half of **FIGURE 30**, the URI scheme for the actual data records is given (or, in fact, one example). The parts of these URIs' paths resemble all the used dimensions in the data set. In their respective order, these are:

→ The *fiscal year*, from which the respective data record origins,
→ the *municipality code*[62] from the municipality the data record is associated to,
→ the *fiscal state* or budgetary document[63] from which the data originates,
→ the *fiscal type*[64] of the cash flow recorded,
→ the *functional classification* or, to be more precise, all three mandatory decades of VRV's approach subsections[65], and
→ the *economic classification* or all three also mandatory decades of VRV's account groups[65].

While the first three dimensions or parts of each data record's URI form the URI of the respective data slice, which corresponds with the original budgetary document, where the data record was

[62] This code – five digits in length & assigned by STATISTIK AUSTRIA – identifies every Austrian municipality in a unique way.
[63] Budgets & statements of account are encoded as follows: VA (for "Voranschlag") and RA (for "Rechnungsabschluss").
[64] Revenues and expenditures – as fiscal types – which are modelled as data records are encoded as follows: REV or EXP.
[65] VRV's classifications of municipal cash flows were discussed in detail in chapter 2.2.4.

published in the first place (e.g. Vienna's statement of account for 2015 – as given in the URI in FIGURE 30), the whole set of all six dimensions identifies each individual data record itself in this approach for Austrian municipal LOBD (e.g. Vienna's expenditures for food in kindergartens according to 2015's statement of account – again as given in the URI in FIGURE 30). Apart from this, each data slice's and data record's URI begins with the identifier of the dataset itself: HTTP://DATA.LOMBUDA.AT/OGHD, where "OGHD" stands for OFFENE GEMEINDEHAUSHALTSDATEN or municipal OBD – based on the official GHD interface (see chapter 2.2.6). Yet, more on these dimensions and the actual structure of the implemented dataset later in chapter 4.2.

It should be noted at this point, that the given URI scheme as-it-is is obviously free from implementation-specific parts (like file extensions, e.g. `.rdf`, `.html` or `.php`) – except for the general domain-related specifics of the VRV[66]. Also, versioning within the URIs is omitted altogether. Thereby, the defined URIs conform to the definition of "Neutral URIs" as given in chapter 2.3.3.

With that said, this structure should guarantee a consistent URI scheme or design for the exemplary approach given by this Diploma Thesis that correspond to the three overall characteristics of URIs RDF's standardization by the W3C had in mind: simplicity, stability and manageability (see chapter 2.3.3 or [71]). All in all, overall usability and convenience as well as general requirements by Linked Data can be ensured easily by the implemented URI scheme as it has been described so far.

Data storage & handling

As the actual technical prototype for the given approach is written in Java[67], an appropriate Open Source framework for handling of RDF data as well as building a Semantic Web application in general and which at the same time supports Java, needed to be chosen in advance. As there are only a few such frameworks available that also cover all the technical requirements for the described approach on LOBD for Austrian municipalities, the choice[68] fell on the popular framework called "Jena" by the APACHE SOFTWARE FOUNDATION [134].

This framework was originally developed by researchers of HEWLETT PACKARD LABS, an exploratory research and development group at the eponymous company, in the United Kingdom already by the year 2000 – one year after the final standardisation of RDF. After HEWLETT PACKARD LABS abandoned Jena's development in 2009, the famous APACHE SOFTWARE FOUNDATION adopted the framework's development in November 2010. Jena is still maintained this way down to the present day.

The framework itself comprises various technical components needed for full-blown Linked Data applications – nearly all of them are also required for the exemplary approach worked out in this very chapter. In a nutshell, these components are:

→ an *RDF API* (Jena's core API for parsing, processing and serialization of RDF graphs)
→ *ARQ* (a SPARQL query engine compliant with its most recent specification version 1.1)
→ *TDB* (a native, high performance and persistent triplestore for RDF graphs)
→ *Fuseki* (depending on its configuration, a standalone SPARQL server or an independent web application – incorporating TDB as data store and providing a full-blown SPARQL endpoint)

---

[66] These specifics were generalised on their own: With the knowledge of state-of-the-art implementations of chapter 3 and the common characteristics of the VRV 1997 and 2015 (see chapter 2.2.5), the given URI scheme *should* be persistent for municipal budgets in the future (of course, not unconditionally, but as best as it can get). For this reason, e.g. VRV's budgetary indicator as such does not appear as a dimension on purpose – as it will be dropped with the VRV 2015.

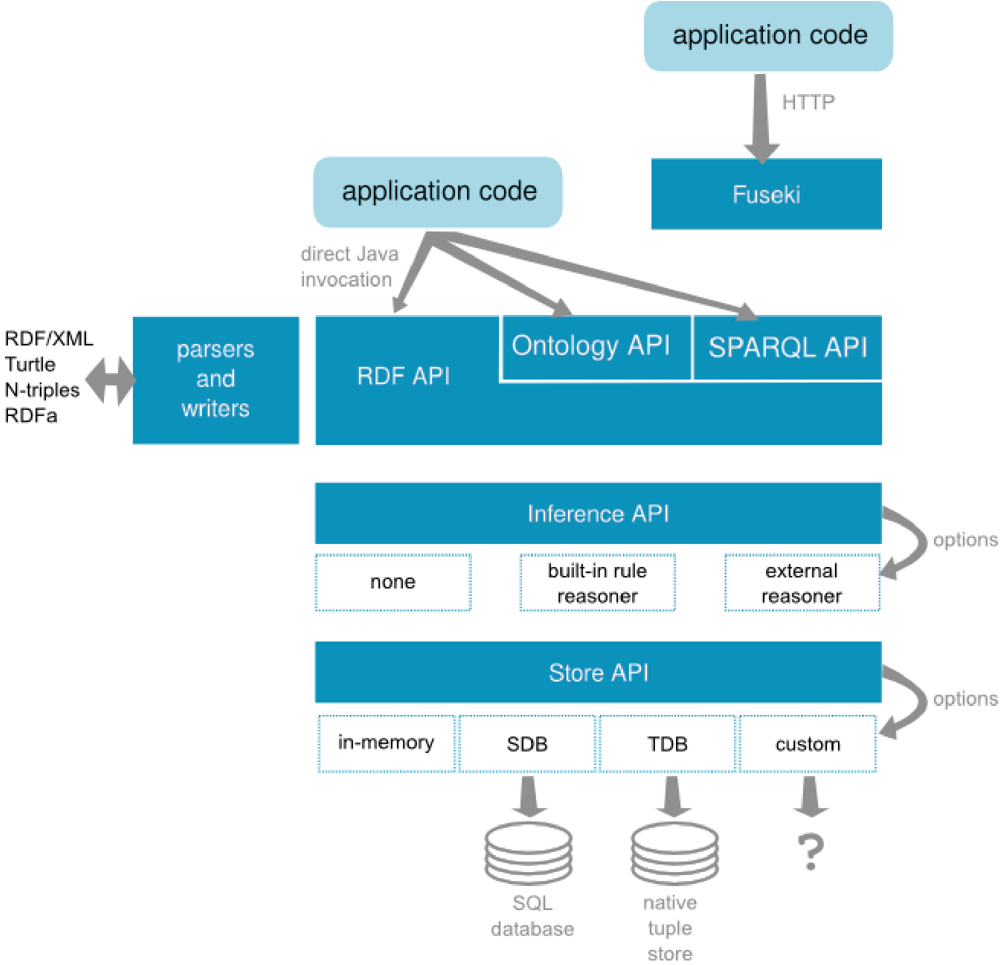
[67] Java is the name of a popular object-orientated, platform-independent programming language.

[68] Obviously, at this point there is *no* detailed assessment for this choice. However, as long as it deems fit (and the given choice provides all technologic possibilities needed for the exemplary approach laid out by this Diploma Thesis), this choice shall be sufficient for this work's sake. In the end, this work is *not* about choosing the right Linked Data framework, but about creating an appropriate data model for Austrian municipal LOBD and verifying this (framework-independent) model.

→ A *ontology API* as well as an *inference API* including a rule based Semantic Web *reasoner*

All listed components except the last one, which – as a Semantic Web technology – goes beyond the scope of this Diploma Thesis, will be used in this very chapter as well as its accompanying technical implementation. These components are also depicted in the overall framework architecture of Jena in **FIGURE 31** below, which outlines the possibilities with this framework graphically.

FIGURE 31 – Jena's framework architecture including possible interactions between its different components



Source: The Apache Software Foundation [134]

What should be not forgotten is that Jena – with its RDF API – is able to read and write a wide variety of RDF serialization syntaxes – like the ones already presented in chapter 2.3.2, but also others:

→ Turtle
→ RDF/XML
→ N-Triples
→ JSON-LD
→ RDF/JSON
→ TriG
→ N-Quads
→ TriX
→ RDF Binary

In addition, other syntaxes can be integrated dynamically by the individual applications using Jena for both parsing and serialization. As the ones listed above as well other RDF syntaxes are not of primary interest for this Diploma Thesis, no further discussion on this topic will take place at this point.

With respect to the actual implementation itself, Jena shall be used with its RDF API and ARQ in the technical implementation for the chosen approach of LOBD for Austrian municipalities itself and its component Fuseki for providing a public SPARQL endpoint as well as a persistent storage of the generated data in the end – however, more on that later in chapter 4.2.

Data (re-)publication

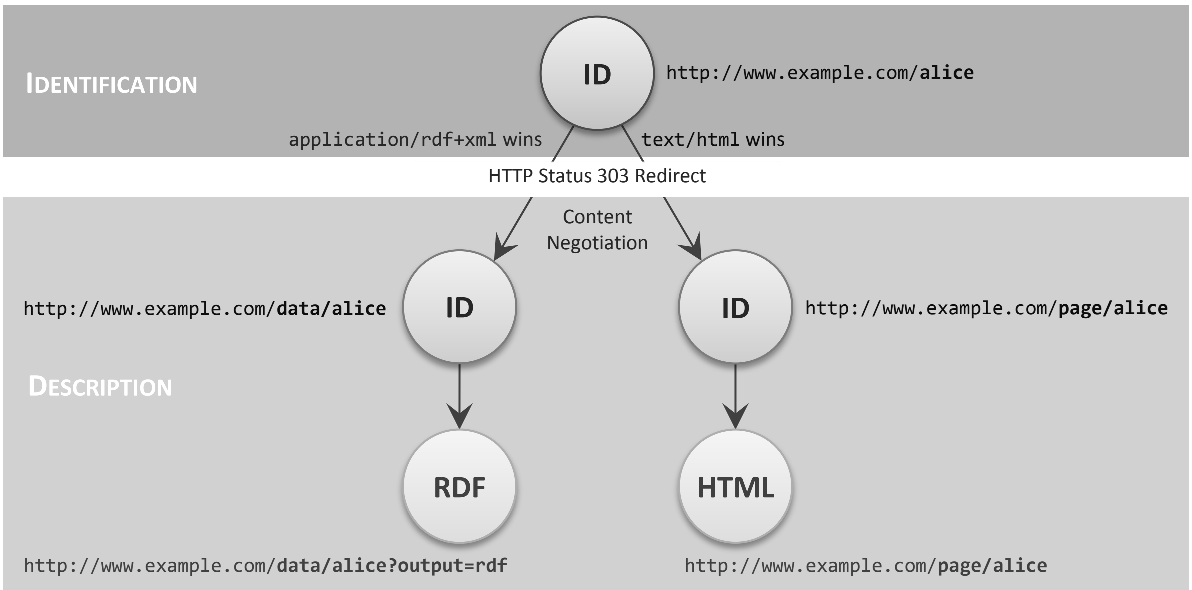
While Jena's Fuseki-component provides a public SPARQL endpoint for the chosen approach, further pieces of the puzzle are needed to achieve real Linked Data as well as LOBD in particular – especially, when seen in the sense presented in chapters 2.3.3 or 2.4.4, respectively. To be more precise, starting points and an appropriate foundation for open licensing as well as dereferenceable, useful URIs are still missing. Both parts for this Thesis' concept shall be tackled in the following last part of this very sub-chapter.

First of all, there are already implemented solutions available to address the issue of separation of RDF's URIs *identifying* resources from those *describing* the latter – including a respective technical concept and implementation for redirecting requests from the former to the latter with sophisticated content negotiation (see chapter 2.3.3). To limit the scope of this Diploma Thesis (again) in an appropriate way, such an existing implementation has been chosen[69] for the sake of providing encompassing and real LOBD: the Linked Data frontend for SPARQL endpoints called "Pubby".

Pubby was developed by Richard Cyganiak[70] (with the involvement of Chris Bizer) at the Free University of Berlin starting in 2007 [135]. It was used by the famous "DBpedia" project, which we already got to know while discussing the Linked Open Data cloud (see **FIGURE 21** on p.61), until the architectural as well as overall demands of this platform increased and Pubby was replaced by conceptually differing software [136].

Below, Pubby's implemented URI and redirect design based on the concept described in chapter 2.3.3 as *303 URIs* is depicted in **FIGURE 32**. As suggested right at the end of the respective discussion, in this case – as human- and machine-readable description of a given URI can quite differ – the content negotiation is situated right at the identifying URI – immediately redirecting the client to separate, descriptive URIs through responding with HTTP Status Code "303" to the original request.

FIGURE 32 – The separation of identification and description URIs of a resource as implemented by "Pubby"



Source: Own illustration

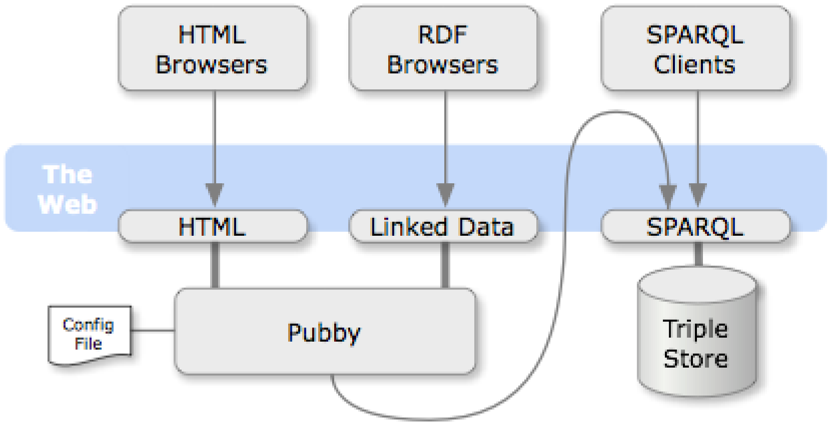
[69] Again, *no* detailed discussion about this choice is made at this point. Like for data handling & storage, the given technology – as long as it deems fit for the requirements of the exemplary approach created in the course of this Diploma Thesis (which the given choice does) – should prove to be sufficient for this work's sake.

[70] As interested readers might have already noticed, Richard Cyganiak was already mentioned at various points of this very Diploma Thesis: Like as co-author of the "Linked Open Data cloud" as well as the "RDF Data Cube Vocabulary". He has contributed in a lot of works in the field of Linked Data and can be considered as one of the respective experts nowadays.

With respect to the depicted example in **FIGURE 32**, it defines “`http://www.example.com/alice`” as URI identifying a person called Alice with “`http://www.example.com/data/alice`” providing machine-readable data as well as with “`http://www.example.com/page/alice`” a human-readable web page including the respective information about her – while the latter *descriptive* URLs are automatically negotiated between Pubby and the client’s web browser (or any other application) after requesting the first *identifying* URI.

With that, Pubby’s overall data provisioning architecture – as depicted in **FIGURE 33** below – is very easy to explain: this Linked Data frontend makes URIs within a given namespace of Linked Data, which is provided by a third-party, local or remote SPARQL endpoint, dereferenceable. Pubby itself is implemented as a Java web application that passes client request towards a given webserver domain or URI to the mentioned SPARQL endpoint, and responses – via respective content negotiation – appropriately to the original request directly, while other SPARQL clients can still freely access the very same SPARQL endpoint Pubby uses as well. This enables transparent publication of Linked Data of a given domain incorporating all the requirements discussed previously in this work.

FIGURE 33 – The data provision architecture as implemented by “Pubby”



Source: Cyganiak & Bizer [135]

Nevertheless, by today, Pubby seems to be somehow abandoned as no development or new releases are made since 2014 [135]. However, as the source code is open source and publicly available for anyone, the application was heavily customized in the course of this work to fulfil its purpose to its full extend for the sake of real, state-of-the-art LOBD in this given approach. These customizations or adoptions of the original code affect primarily following two parts of Pubby:

→ *Frontend representation*
The frontend is probably the part of Pubby that was reworked the most. On the one hand, the HTML template was replaced by a responsive and more contemporary one. Additionally, it is equal to the one of the rest of the technical prototype – so, the presented LOBD integrates seamlessly to the rest of the website. On the other hand, the presentation of the data itself was revised to be more clear and comprehensible – among other things, this includes that properties are now listed in a more logical order (e.g. the type definition of a given resource including its label and textual description is always at the top) and all the URLs themselves are parsed and split into logical, navigable parts. Furthermore, a huge number of similar statements with the same property for a given resource are displayed in a more usable and condensed manner by now, while still preventing any information loss.
It is obvious, that these changes affect the human-readable (HTML) interface only – the RDF data interface remained (in most parts) unchanged.

FIGURE 34 – Screenshot of customized “Pubby” version after delivering a descriptive HTML document for a given URI

HTTP :// DATA.LOMBUDA.AT / OGHD

Linked Open Municipal Budget Data

This is an exemplary data set containing information on Austrian municipal budgets as Linked Open Data. It is represented in a respective Linked Data ontology, which is adapted to Austrian specifics and defined in the same way as this very dataset on this web server as Data Structure Definition in the RDF Data Cube Vocabulary.

Property	Value
rdf:type	▪ qb:DataSet
rdfs:label	▪ Linked Open Municipal Budget Data (en)
dct:title	▪ Linked Open Municipal Budget Data (en)
rdfs:comment	▪ This is an exemplary data set containing information on Austrian municipal budgets as Linked Open Data. It is represented in a respective Linked Data ontology, which is adapted to Austrian specifics and defined in the same way as this very dataset on ... »more» (en)
dct:description	▪ This is an exemplary data set containing information on Austrian municipal budgets as Linked Open Data. It is represented in a respective Linked Data ontology, which is adapted to Austrian specifics and defined in the same way as this very dataset on ... »more» (en)
lombuda:currency	▪ dbr:Euro
dct:issued	▪ 2017-10-31 (xsd:date)
cc:licence	▪ <http://opendatacommons.org/licenses/by/1.0/>
qb:slice	▪ <http://data.lombuda.at/OGHD/2002/40101/RA> ▪ <http://data.lombuda.at/OGHD/2003/40101/RA> ▪ <http://data.lombuda.at/OGHD/2004/40101/RA> ▪ »28 more»
qb:structure	▪ lombuda:dsd
dct:subject	▪ sdmx-subject:2.5
sdmx-attribute:unitMult	▪ sdmx-code:unitMult-0

This page displays information obtained from the SPARQL endpoint at http://lombuda.at/sparql.

Source: Own creation

→ *SPARQL Query building*
Some queries implemented originally by Richard Cyganiak needed to be altered or corrected[71], respectively, in order to work properly with the given approach’s vocabulary and RDF data.

The result of those changes – especially, the ones made regarding the first part – can be observed in **FIGURE 34** above, which represents a screenshot of a part of the implemented technical prototype, for which Pubby provides the respective web services, data provisioning and presentation. The resource description or the respective, processed RDF data displayed in **FIGURE 34** represents the declaration of a dataset for Austrian municipal LOBD using the “RDF Data Cube Vocabulary”.

In addition to that, the second part for appropriate data (re-)publication is still missing in this discussion: respective open licensing. As most of this topic in all its aspects was already covered as part of chapter 2.4.1, the discussion regarding this matter can be kept deliberately short at this point:

For the implemented prototype and all its components (including its source code) the CC-BY license (see p.52) shall be used; for the respective vocabulary for municipal LOBD as well as the generated data based on it the more appropriate, but equivalent ODC-by license (p.51) will be utilized.

[71] There will be *no* further discussion at this point about the exact changes made concerning this matter as the actual implementation (and its source code) is online available for the reader and there, the changes are much more apparent.

4.2 IMPLEMENTATION

With the discussed concept in the previous chapter 4.1, already individual parts for the exemplary approach on LOBD for Austrian municipalities given by this Diploma Thesis were presented in a – more or less – detailed way. However, a discussion as well as documentation of the interaction and integration of these parts with each other in a full-blown implementation and, if nothing else, a description of the vocabulary for Austrian municipal LOBD itself has been omitted up to this point. This should be caught up now in the following, upcoming sub-chapter of this work. The approach for this discussion will be rather feature-driven – in contrast to the discussion of the previous chapter 4.1, which was based on the different technical components.

Integration of existing municipal OBD

First and foremost, existing OBD for Austria municipalities is obviously needed in order to provide a respective implementation in the first place. As discussed in chapter 3.1.3, there already exist (at least *some*) options or open data sources, respectively, for Austrian municipal OBD. Although the given approach as well as its implementation shall be relatively flexible regarding the actual given data structure for to-be integrated data itself, it relies on the commonly available CSV format for such data *only*, and in order to prevent any side effects or consistency errors from the very beginning, one single, consistent data source was preferred. Considering this, the choice was obvious (as there are no true alternatives present so far): the openly licensed budgetary data on "OffenerHaushalt.at" was used from the very beginning of the work on the respective prototype implementation.

The data itself is integrated as Linked Data via an ETL process as described in chapter 4.1. Within this process, following steps are made to provide the respective results:

- → *Data cleaning* (dropping of unnecessary data within the CSV files, which are not modelled)
- → *Data validation* (verifying given data according to VRV and other, similar rules)
- → *Data mapping* (transformation of e.g. given classification codes into RDF compliant resources)
- → *Data merging* (aggregation of maybe more detailed data as modelled to each 3-digit approach and account classifications as well as neglecting the to-be abandoned budgetary indicator except the differentiation between revenues and expenditures)
- → *Data enrichment* (adding – also externally hosted – information and resources with respect to e.g. municipalities to the generated data)
- → *Data clustering* (linking data of the same budgetary document to a respective data slice)

Since the given data from "OffenerHaushalt.at" is available in an extremely clean and consistent form, the integration or ETL process, respectively, can be automated in a straightforward way. Therefore, in the course of this Diploma Thesis, a specific application (as part of the overall web-based technical prototype) was written in Java, which takes care of all the points mentioned above.

The actual "*extraction*" part of the overall ETL process is – of course – not of major interest for this work, as budgetary data is normally available in an already processed, extracted form by the original governmental accounting or reporting systems anyway, and bulk data processing was never the aim of this Diploma Thesis, instead a technical prototype concentrating on chosen budgetary data from Austrian municipalities. As a result, the core "*transformation and loading*"-part is triggered manually via uploading given CSV data into the application. However, the implemented approach is still conceptualized and hence, in principle capable of handling relatively big amount of data – considering the annually aggregated budgetary data for Austrian budgets and statements of account.

With respect to the chosen budgetary data itself, a representative and large enough set of valid[72] respective data records were selected from the available documents on “OffenerHaushalt.at”. In order to use and test the given approach extensively, the biggest and diversified datasets per document available as time series data for statements of account from the three biggest municipalities in Austria was chosen: Vienna, Graz and Linz. Additionally, the appropriate data source attribution (stating “OffenerHaushalt.at” with its CC-BY licensing as the origin) is made in each respective data slice, which – in each case – corresponds to the original budgetary document the data derives from.

An exemplary data structure for municipal LOBD in Austria

At the true core of this Diploma Thesis now finally a vocabulary or, to be more precise, an exemplary data structure for municipal budgets in Austria building upon the eponymous Linked Data shall be discussed. This data structure, which was implemented and verified with the already mentioned prototype, can be considered being one of the major contributions of this work for future ones as well as the overall field of LOBD in Austria. Accordingly, the following part should receive respective attention by the interested reader.

Below, in LISTING 6 an exemplary, transformed data record including its respective data slice is shown as Linked Data using this created vocabulary. In this example, the slice – using the URI scheme discussed in chapter 4.1 – represents a statement of account for the city of Graz from the fiscal year 2015. The following data record is associated with it through the predicate “`observation`” from the “RDF Data Cube Vocabulary” and features the actual aggregated cash flow as well as its budgetary

LISTING 6 – An example of a data record for municipal budget data and its respective data slice (in “Turtle” syntax)

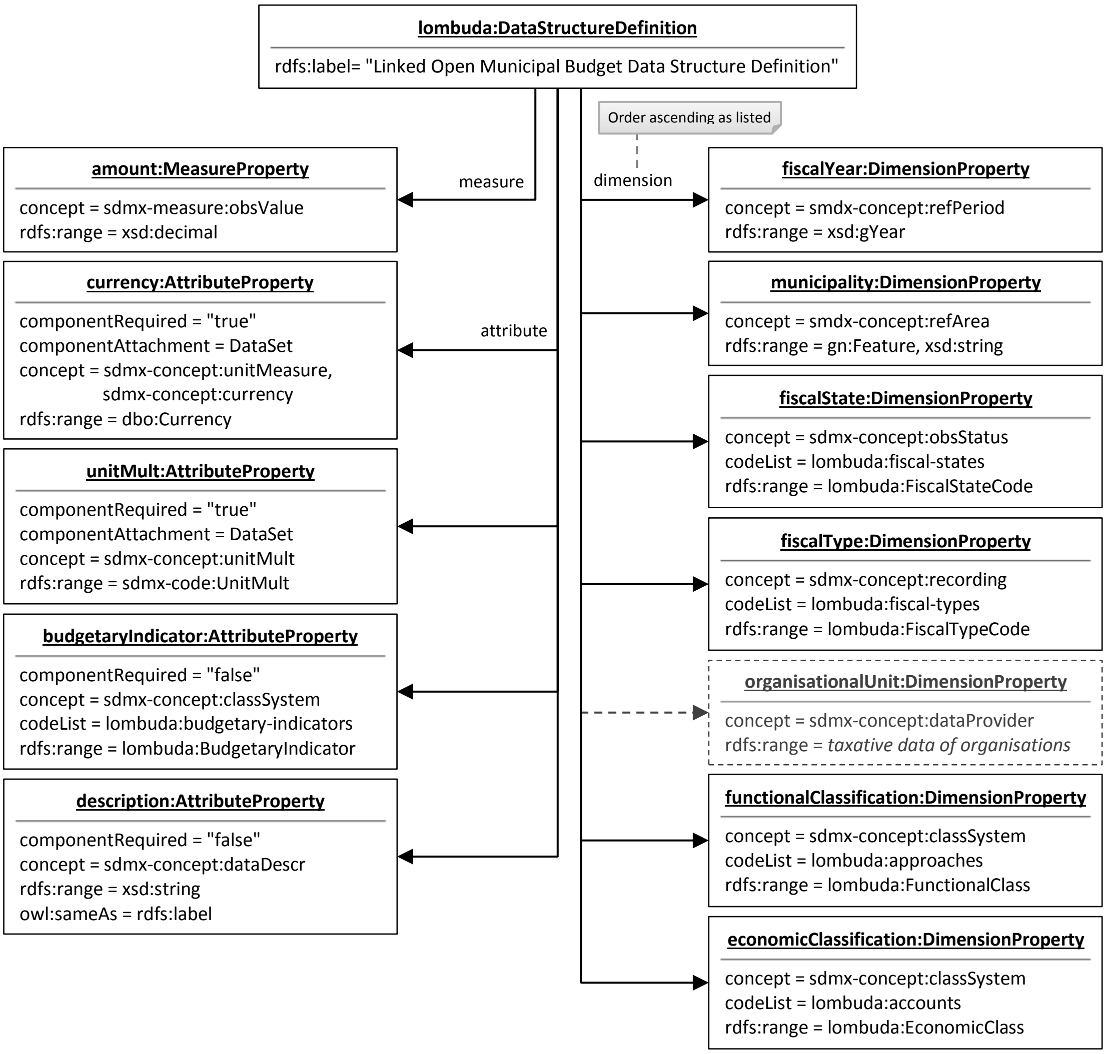
```
<http://data.lombuda.at/OGHD/2015/60101/RA> a qb:Slice ;
    rdfs:label             "A Data Slice"@en ;
    rdfs:comment           "An example slice for municipal budget data from Graz's
                            statement of account in the fiscal year 2015"@en ;
    dct:source             <http://offenerhaushalt.at> ,
                           "This derived data originates from 'OffenerHaushalt.at'
                            by 'KDZ - Zentrum für Verwaltungsforschung' and is used
                            under CC-BY-3.0-AT."@en ;
    qb:sliceStructure      lombuda-slice:fiscalDoc ;
    lombuda:fiscalYear     "2015"^^xsd:gYear ;
    lombuda:municipality   <http://sws.geonames.org/2761333> ,
                           "Graz (60101)" ;
    lombuda:fiscalState    lombuda-fs-code:RA ;
    qb:observation         <http://data.lombuda.at/OGHD/2015/60101/RA/EXP/420/700>
    .

<http://data.lombuda.at/OGHD/2015/60101/RA/EXP/420/700> a qb:Observation;
    rdfs:label                        "A Data Record"@en ;
    qb:dataSet                        lombudata:OGHD ;
    lombuda:fiscalType                lombuda-ft-code:EXP ;
    lombuda:functionalClassification  lombuda-fc-code:420 ;
    lombuda:economicClassification    lombuda-ec-code:700 ;
    lombuda:amount                    "257105.95"^^xsd:decimal ;
    lombuda:budgetaryIndicator        lombuda-bi-code:1 ;
    lombuda:description               "Test Data Record #1"@en
    .
```

A declaration of any namespace prefixes has been omitted in this example for the sake of clarity and briefness (see p.115 for all them).
Source: Own creation

[72] It is important to note that valid and consistent data sets were chosen from the original data source. Unsurprisingly, available budgetary data published by local governments is *not* consistent necessarily and the available data sometimes does *not* exactly correspond with all its numbers and figures to the officially published documents. For example, data for some statements of account for the city of Vienna show this in a way, that the overall administrative annual result (№99 of VRV’s budget profile, see **TABLE 7**) does *not* equal zero – as it should be, also according to the official documents.

FIGURE 35 – The abstract data structure definition for the given Linked Data model for Austrian municipal budget data



The full data structure definition with all its RDF data is listed in the Annex (see p.115).
Source: Own illustration

origin or classification, respectively, including:

- → the *fiscal type* (whether the concerned data record represents revenues or expenditures)
- → the *functional classification* (the first three mandatory decades of VRV's approach code)
- → the *economic classification* (the first three mandatory decades of VRV's account code)
- → VRV's *budgetary indicator* (which can be optionally given to prevent information loss)
- → and optional, descriptive fields with a textual statement relevant for the data record.

The actual data composition used in this example as well as all other similar data records is defined in a formal way through the "RDF Data Cube Vocabulary" as a data structure definition in RDF. This definition is listed to its full extend within the Annex (see p. 115). Above in **FIGURE 35**, it is depicted as an UML object diagram in an abstract, simplified way.

As it can be seen easily, this data model defines all different kinds of properties also used in the example given in **LISTING 6** – and more (like a currency and measurement unit definition for the overall dataset). Furthermore, it includes value ranges (in form of simple data types, like "`xsd:string`" or even internal as well as external resource classes like "`gn:Feature`"), if needed, references to separately defined code lists, as well as a declaration of the used statistical concept according to SDMX's standardization and, last but not least, whether an attribute is actually required for each and every instance of a data record or not. These attributes (e.g. the currency) can also be defined to be situated at specific aggregating resources, like a data slice or the overall dataset through the predicate "`componentAttachment`", resulting in an *abbreviated* data cube as defined in chapter 4.1.

Additionally, within in the overall concept – depicted in **FIGURE 35** as fifth dimension property – a separate dimension for organizational units can be found, with which (eventually) certain parts of budgetary data could be separately categorized completely – including also associated corporations of local governments, which were neglected so far. Although this dimension was not implemented in the technical prototype – as it is not (yet) required for municipal budgets in Austria, this was (at least) conceptualized following the model used for publication of UK's COINS dataset and foreseeing a likely more dedicated separation or – depending on the individual case – inclusion of associated institutional units from municipalities within Austrian public finances in the future.

With respect to the used code lists, the implemented technical prototype provides the means to automatically transform the respective definitions for approach, account and budget profile indices into hierarchical organized concepts, so they can be also modelled and subsequently referenced as full-blown Linked Data. This provides the means to actually give those codes self-defined meaning through using references to the respective code resources (like shown in the example given in **LISTING 6**). These resources in turn hold respective textual descriptions of the actual codes as well as information about their hierarchical relationships, which can be easily obtained by respective application or the user himself this way, as all code URIs are indeed dereferenceable.

Considering all features of municipal budgets in Austria as well as respective legislation like the VRV (see chapter 2.2), this vocabulary tries to model the respective data structure of public finances in an appropriate and – considering its aggregated nature in public use – streamlined way. However, as Linked Data in its definition is – obviously – not finite, future additions to the given data structure definition or a completely new one, which builds upon it, can be easily implemented. Hence, specifics for other applications can be added, and future changes in respective legislation can be incorporated without reinventing the wheel altogether.

Nevertheless, an encompassing and exhaustive discussion of all possibilities and technical details of this exemplary data structure of LOBD for Austrian municipalities is nearly impossible at this point. If needed, a more detailed discussion of all the ontological elements used for the created vocabulary can be found in the specification of the "RDF Data Cube Vocabulary" in [133] as well as its underlying ontologies' specifications. And finally, it is left to the interested reader to have a look at the actual implemented full vocabulary and generated, exemplary data based on it as well as its underlying technical implementation: As already mentioned, the full data structure definition can be found in the Annex of this work (on p. 115). Additionally, all parts of the respective technical prototype are available online at HTTP://LOMBUDA.AT[73].

A transparent SPARQL endpoint for Linked Data

In order to implement encompassing and real LOBD as well as to allow bulk access, it is necessary to publish the transformed and carefully modelled data and its vocabulary in an appropriate way. This is done in the technical prototype application for the given approach of this Diploma Thesis via a dedicated SPARQL endpoint (by "Fuseki") and a Linked Data frontend based on it ("Pubby"). While the latter facet of this approach has already been discussed in detail – including this application's

[73] This domain is controlled and maintained by the author of this work himself. In this way, it should be guaranteed that this vocabulary as well as the overall results of this Diploma Thesis will be available for interested readers as well as other works in this field of Informatics and public finances also in the future.

customizations to it (see chapter 4.1), the former's actual implementation shall be addressed briefly in the following part of this chapter. But, just to make things clear: The application uses Jena's component "Fuseki" for all SPARQL services, anyway – however, *only* in its backend and not *directly* available for any potential end-users.

What does this mean? And why it is done this way at all? As Fuseki or its web interface, respectively, allows extensive operations on the underlying dataset (including its complete deletion), while not offering any sophisticated, built-in security measures [74] to protect the application against unauthorized access, the technical prototype encapsulates Fuseki's SPARQL endpoint as a whole. In this way, it is hidden, or – depending on the point of view – seamlessly integrated within the technical prototype's application itself, which – in turn – takes care of forwarding all valid SPARQL queries from the public endpoint URL to Fuseki and, at the same time, providing an appropriate human-usable interface for it, if necessary.

This means, that a human-readable HTML page resides at the very same URL as the public SPARQL endpoint itself. This is also done for reasons of convenience and usability. The application responses with a page, which shows a human-friendly and -usable interface at first – instead of trying to querying the underlying dataset with a potentially non-existing query. As alternative, a sophisticated Query Editor is available to create and transmit SPARQL queries directly on this very endpoint including an appropriate display of respective result sets and the possibility to download them as CSV files. This Query Editor as well as its result parser – some really outstanding pieces of technology – is one of the used third-party libraries for implementation of the technical prototype application for this work: the so-called "Yet Another Sparql GUI" (YASGUI)[75].

In order to perform actual queries on the application's SPARQL endpoint – for example, via a third-party application – it is necessary to pass a (URL-encoded) SPARQL query as a so-called URL query parameter (see **FIGURE 9** on p.31) in a HTTP "`GET`" or "`POST`" request to the URL of the endpoint.[76] For large queries, it is possible to pass the (uncoded) SPARQL query via an HTTP "`POST`" request body, while declaring the HTTP "`Content-Type`" header as "`application/sparql-query`". Furthermore, the actual output by Fuseki can be controlled this way as well by using the HTTP "`Accept`" header field (e.g. "`text/ttl`" will result in a response in "Turtle" syntax).

In conjunction with the dereferenceable URIs of this approach's dataset, provided by "Pubby", the resulting publicly available Linked Data conforms extensively to the standard and guidelines discussed in chapters 2.3.3 and 2.4.4.

Automated generation of VRV's budget profiles

Finally, for verifying the overall results of the discussed ETL process as well as to ensure consistency and accuracy of the created model for municipal budget data in the end, the technical prototype provides the means to generate full budget profiles for all saved data slices[77] of the underlying data records according the VRV (see **TABLE 7** on p.24 or chapter 2.2.4 for a general description). These budget profiles are generated for each user request on the website of the given application individually, through respective SPARQL queries.

---

[74] Indeed, Fuseki provides some basic security by using "Apache Shiro" [134]. However, this measures only apply to the level of URLs – *not* individual requests, as it is necessary for the given application (e.g. normal queries are allowed, as long as they do *not* try to alter the underlying dataset, which is possible as Fuseki supports SPARQL 1.1 – see also chapter 2.3.2).

[75] For more information on this JavaScript-based web component, the interested reader is referred to [147].

[76] A respective request that performs the valid SPARQL query "`DESCRIBE <http:data.lombuda.at/OGHD>`" would look like this: HTTP://LOMBUDA.AT/SPARQL?QUERY=DESCRIBE+%3CHTTP%3A%2F%2FDATA.LOMBUDA.AT%2FOGHD%3E

[77] As mentioned earlier, each data slice corresponds to the respective budgetary document from where its data originates; so, it fixes the dimensions for the fiscal year, affected municipality and the fiscal type (*budget* versus *statement of account*).

LISTING 7 – The implemented SPARQL query for aggregating saved datasets for a budget profile according to the VRV

```
SELECT ?budgetProfileIndex (SUM(?amount) as ?sum)
WHERE {
    ?bpi    a lombuda-code:BudgetProfileIndex ;
            skos:notation          ?budgetProfileIndex ;
            skos:narrower          ?account ;
            lombuda:fiscalType     ?fiscalType .
    ?slice  a qb:Slice;
            lombuda:fiscalYear     "2015"^^xsd:gYear ;
            lombuda:municipality   "Vienna" ;
            lombuda:fiscalState    lombuda-fs-code:RA ;
            qb:observation         ?obs .
    ?obs    a qb:Observation ;
            lombuda:amount         ?amount ;
            lombuda:fiscalType     ?fiscalType ;
            lombuda:economicClassification ?account .
} GROUP BY ?bpiNo
```

A declaration of any namespace prefixes has been omitted in this query for the sake of clarity and briefness (see p.115 for all them).
Source: Own creation

These SPARQL queries in turn take care of the automated aggregation of all affected data records to the respective, regulated indices of a budget profile. Therefore, a complete list of all these indices including a proper association to the related VRV account groups has been incorporated as Linked Data by the prototype application as well. Such an example query, which sums up the complete cash flows per index – including those from public corporations (approach sections 85 to 89), is shown **LISTING 7** above. Through the means of Linked Data and SPARQL itself, this query is written in a straightforward way without joining datasets or similar. This would likely be impossible when using a conventional relational database and a respective query engine.

The resulting budget profile from this query is shown in **FIGURE 36** below – as processed and presented by the implemented technical prototype on its website. The included separate data for

FIGURE 36 – A screenshot of a generated budget profile (Vienna's statement of account from 2015) according to the VRV

№	Description	Sum OB + EB	thereof A85-89	Sum w/o A85-89
I.	PROFILE			
19	SUM #1: Operating Revenues	11.177.099.289,59 EUR	559.943.110,63 EUR	10.617.156.178,96 EUR
29	SUM #2: Operating Expenditures	10.270.855.480,08 EUR	395.150.288,91 EUR	9.875.705.191,17 EUR
91	**BALANCE #1: Operating Result**	**906.243.809,51 EUR**	**164.792.821,72 EUR**	**741.450.987,79 EUR**
39	SUM #3: Revenues from Asset Accounting	196.346.443,47 EUR	13.987.198,50 EUR	182.359.244,97 EUR
49	SUM #4: Expenditures from Asset Accounting	1.326.531.927,87 EUR	74.169.375,86 EUR	1.252.362.552,01 EUR
92	**BALANCE #2: Asset Account Result w/o Financial Transactions**	**-1.130.185.484,40 EUR**	**-60.182.177,36 EUR**	**-1.070.003.307,04 EUR**
59	SUM #5: Revenues from Financial Transactions	1.707.035.162,37 EUR	30.684.699,00 EUR	1.676.350.463,37 EUR
69	SUM #6: Expenditures from Financial Transactions	1.483.093.487,48 EUR	18.856.224,02 EUR	1.464.237.263,46 EUR
93	**BALANCE #3: Financial Account Result**	**223.941.674,89 EUR**	**11.828.474,98 EUR**	**212.113.199,91 EUR**
94	**BALANCE #4: ANNUAL RESULT W/O FINAL SETTLEMENT**	**0,00 EUR**	**116.439.119,34 EUR**	**-116.439.119,34 EUR**
II.	DEDUCTION OF THE FISCAL BALANCE			
70	Annual Result w/o A85-89 & Financial Transactions	-328.552.319,25 EUR		
71	Clearing of Annual Result of A85-89	116.439.119,34 EUR		
95	**FISCAL BALANCE ("MAASTRICHT RESULT")**	**-212.113.199,91 EUR**		
III.	TOTAL BUDGET OVERVIEW			
79	SUM #7: Total Revenue	13.080.480.895,43 EUR		
89	SUM #8: Total Expenditure	13.080.480.895,43 EUR		
99	**ADMINISTRATIVE ANNUAL RESULT**	**0,00 EUR**		

Source: Own creation

approach sections 85 to 89 are aggregated – by the way – through a similar query link the one in LISTING 7 and subsequent procedures resulting in the depicted second column of figures. Eventually, the last column is simply the difference between the first and the second column of figures.

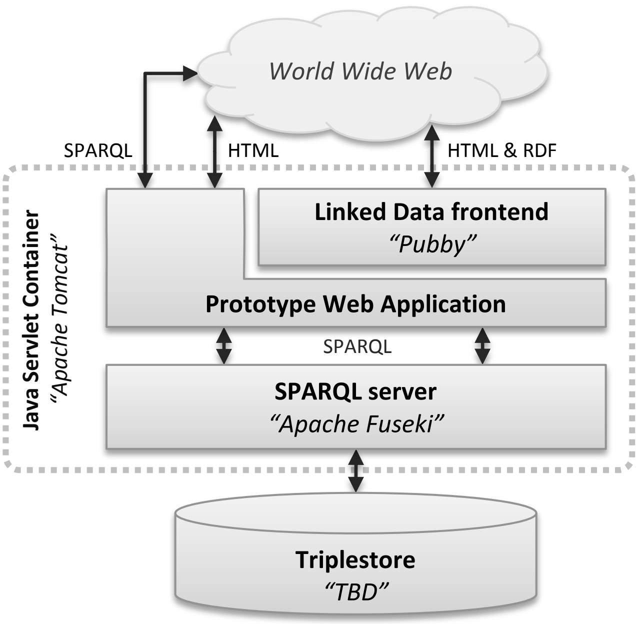
This application of implemented LOBD for Austrian municipalities shows the advantages of the approach, discussed and presented by this Diploma Thesis, although only with one major example. Of course, in the field of public finances various other applications for such processed data would be easily conceivable. Furthermore, in this way also federated queries – so, aggregating data from various, distributed datasets, which use the same vocabulary – would be possible.

The overall software architecture

Last but not least, an overview of all the different technical components responsible for the actual technical implementation discussed in this chapter shall be given. This way, the different matters should be combined in comprehensible way; so that, it is possible to understand the overlying technical relationships for the given approach after all.

Before diving into its details, the actual implementation – as mentioned already in a more or less detailed way – is based on Java 8 or, to be more precise, a respective web application. The needed runtime environment is provided by APACHE SOFTWARE FOUNDATION's Tomcat in its Version 8, where all the discussed components are hosted. This runtime environment acts as a web server and is operating from a virtual machine in the AWS cloud – for the time being.

FIGURE 37 – Overall software architecture of the given approach for LOBD of Austrian municipalities



Source: Own illustration

Above in FIGURE 37, the overall software architecture with all important components is shown in an abstract way within this environment: "Fuseki" as SPARQL server with an integrated, persistent triplestore is connected with the discussed prototype application via its full-blown SPARQL endpoint. "Pubby" in turn serves as Linked Data frontend of this application, while the application itself encapsulates the public SPARQL endpoint as well as the means to automatically generate authentic budget profiles with the underlying, provided data on the prototype's website.

Hence, the architecture of the chosen approach is designed in a streamlined way, while using existing state-of-the-art software components – eliminating the need to reinvent the wheel altogether.

4.3 ASSESSMENT OF THE CHOSEN APPROACH

After the chosen approach on LOBD for Austrian municipalities was discussed exhaustively in the previous sub-chapters, at this point the overall results shall be matched with the goals set at the very beginning of this chapter. In this final analysis of this Diploma Thesis, the achievement of objectives for the given approach shall be reviewed in a formal, qualitative way.

In the following, all four predefined criteria from page 81 will be worked through again for evaluation. The respective titles of the list below are given in the same order and wording, whereby the goals' definition can be matched with the respective results in a convenient way.

→ *Openness*
The overall result or, to be more precise, the created vocabulary as well as the data generated with it – including the technical implementation – is publicly available for free – without any restriction – under the terms of OKI's Open Data Commons Attribution Licence (ODC-by) or CC's one (CC-BY), respectively. The data format itself is open in the sense as it is with Linked Data machine-readable in a state-of-the-art kind of way. With that the overall result can be considered to *conform in its entirety* with OKI's OPEN DEFINITION (see chapter 2.4.1).

→ *Linked Data*
The implemented application of Linked Data is available online (at HTTP://DATA.LOMBUDA.AT) as non-proprietary, machine-readable and structured RDF data, where dereferenceable URIs are used to identify given concepts as well as the data records themselves. These URIs conform to overall requirements and "expectations of behaviour" of Linked Data as discussed in chapter 2.3.3. Furthermore, even Linked Data outside the dataset's namespace is referenced (e.g. for identification of municipalities). Thereby, it is also *fully compliant* to Tim Berners-Lee's 5-Star-Scheme (see **TABLE 10** on p.58) and can be rated with a total of five of its stars comprising real LOD and an unprecedented contribution to the evolving *Web of Data* in Austria.

→ *Correct usage of Vocabulary*
In order to verify whether the used "RDF Data Cube Vocabulary" is applied by the LOBD for the given approach in a correct and appropriate way, all 21 integrity constraints – defined in its specification [133] verifying a so-called *well-formed* data cube – are reviewed automatically[78] by the implemented application after every modification of the underlying dataset. And, indeed, this review *confirms* the consistent usage of this ontology to this day.

→ *Accuracy of fit of the solution for the given subject*
Finally and most importantly, the results of the created vocabulary and the respectively generated data needed to be verified, whether they are an adequate solution and properly implemented model in the field of public finances or municipal budgets, to be more precise. For the matter of this Diploma Thesis, a fully automated generation of budget profiles from the exemplary chosen municipalities[79] was implemented in accordance to the VRV. Thereby, the figures and numbers in these generated budget profiles based on LOBD were matched with the ones from the officially published ones. All in all, more than 30 datasets from statement of accounts were transformed into the given approach on LOBD and, indeed, the numbers of the generated budget profiles *correspond – exactly and in every detail –* with those in the officially published ones.

To sum it up, the chosen approach proves to have achieved in fact *all* of the predefined goals for reaching full-blown, encompassing and comprehensive LOBD for Austrian municipalities in the end.

---

[78] The results of this automated, regularly made verification are online available at HTTP://LOMBUDA.AT/DATA/IC.
[79] These are the three biggest Austrian cities: Vienna (with datasets from 12 fiscal years), Graz (with 5) and Linz (with 14).

CHAPTER 5
CONLUSIONS

Increasing overall requirements by current society, governmental institutions and legislation related to governmental transparency – and especially, public finances in particular – require a change within public administration and used technologies on an unprecedented scale. New approaches for data integration and interconnection – applying current state-of-the-art and best practice – must be implemented in order to provide actual real-life solutions to the currently existing issues – especially, concerning transparency and openness.

For local government finances as discussed by this Diploma Thesis in its chapter 2.2, one of the major driving forces in this field is the concept of Open Budget Data, which becomes increasingly important in recent history – not only in public perception, but also in respective national laws. The definition of such Open Data is already quite clear and commonly accepted, and respective implementations including legal aspects like appropriate licensing can be found in a broad variety of areas. The benefits of this concept with respect to Open Government Data can be hardly measured yet; neither quantitatively nor qualitatively. However, relevant studies and projections – as made in chapter 2.4.2 of this work – show that overall direct and indirect effects for the economy of the European Union can yield in a potential economic benefit of more than 300 Billion or, in case of Austria, up to 7 Billion Euro. Furthermore, qualitatively spoken, other benefits are obvious: increased transparency, better governance, reuse of governmental data, which in turn induces participation and collaboration, and, if nothing else, an engine of innovation are created through opening up information by the public administration. Therefore, the specific benefits of Open Budget Data are evident and should not go unnoticed.

However, the technical distribution of such data in a digital and open way induces challenges not only with respect to the data itself but also to the embedded budgetary processes and systems as well as respective underlying concepts. During the discussion seen in chapter 2.3, Linked Data – as a concept derived from the Semantic Web vision – was introduced to address exactly these challenges related to information technology. Although it is commonly perceived as a relatively new concept, Linked Data's origins can be traced back to the beginnings of the World Wide Web itself with its core technologies specified already in the late 90s. A real breakthrough and spreading of Linked Data as applied technology in large scale remains to be seen.

In order to provide a new, more transparent, coherent, reasonable, comprehensive and comparable approach for opening up local government finances in Austria – considering relevant requirements, the combination of these principles of Open Data and Linked Data with budgetary data of Austrian municipalities has been discussed – and even practically tested – in the course of this Diploma Thesis. With that, a way to reach more budgetary transparency as well as a new level of best practice focused on the peculiarities of Austria's governmental budget structure has been shown for future applications in this field of public finances.

In the following final part of this work, the overall results of the assessments and the implementation made in the preceding chapters regarding Open Budget Data – especially in Austria – will be recapped in a nutshell. Additionally, open questions and issues related to the discussed topic shall be addressed as well. Thereby, one final overview about this work on Linked Open Budget Data in Austria is given in an appropriate way.

5.1 STATUS & PERSPECTIVE OF LINKED OPEN BUDGET DATA IN AUSTRIA

The current status of Austrian implementations on Open Budget Data in general and Linked Open Budget Data in particular is somewhat devastating as seen in chapter 3.1.3 of this Diploma Thesis. However, this perception also resembles the current state of respective governmental releases on public finances throughout the world. Relevant rankings and underlying studies like the OPEN DATA INDEX by OPEN KNOWLEDGE INTERNATIONAL or the OPEN BUDGET INDEX by the INTERNATIONAL BUDGET PARTNERSHIP – as discussed in chapter 3.1.1 – reveal that governments *typically fail* to publish Open Budget Data and the majority of the world's population can*not* access adequate budgetary information about public finances of the country they live in. Both indices reveal that more than the half of the world's central governments are *not publishing* (or, in some cases, even *producing*) sufficient information on public finances in a transparent way – let alone, machine-readable or semantic data. Hence, the majority of the world's public (and if such information is *not produced* at all, also their government) is left behind unable to fully understand or monitor how much public finances are raised or how they are spent. This may allow undesirable developments, possible improvements and prospects to be unrecognized, both unintentionally and deliberately. As a consequence, effective use of public resources cannot be ensured (and monitored) this way.

Austrian governmental authorities make *no* exception to this perception nowadays. While the central government fails – with missing respective open licensing and usage of proprietary data formats – to publish actual Open Budget Data altogether, local government finances are published by responsible public authorities themselves rarely – and if so, partially in insufficient ways – although recent legislation like the Austrian Stability Pact or ÖSTERREICHISCHER STABILITÄTSPAKT demands it differently since more than five years.

Nevertheless, there are exemplary exceptions from this rule. Although only *eight* datasets on each budgets and statements of account for municipal finances are available on the Austria's Open Government Data platform "data.gv.at" at the time of this survey in August 2017, they include respective datasets for at least the three biggest cities in Austria: Vienna, Graz and Linz. All of them conform to the definition of Open Data and provide time series data for at least five years. Private initiative like "OffenerHaushalt.at" can be seen as reinforcement to the efforts in this field, providing budgetary data for nearly the half of all Austrian municipalities by now. Additionally, private as well as governmental *international* pioneering implementations of Linked Open Budget Data show, where the technological journey will lead us in the future, eventually – through providing more than simple Open Budget Data or *Documents on the Web*: to an imminent *Web of Data*.

As seen so far, the quality and extend of implementations as well as the overall awareness for this (potentially very beneficial) topic varies tremendously across different countries and among all parties involved. All in all, current implementations on reasonable, comprehensive, transparent and open publication of budgetary data are existing, but still in their early stages of development. The overall coverage and quality of the found implementations reveal that developments in applied (Linked) Open Budget Data of governmental bodies are in an initial phase, awaiting more momentum and even more quality through recent developments in legislation, information technology and overall requirements by current society.

5.2 AN EXEMPLARY APPROACH & GOALS FOR TRANSPARENT LINKED OPEN BUDGET DATA

For complex topics like local government finances as discussed by this Diploma Thesis in its chapter 2.2, sophisticated technical solutions are needed in order to ensure an actual transparency gain through the release of respective Open Data that is normally heterogeneously organized, in the first place. In general, budgetary data in particular is highly complex information and deduced knowledge can only be obtained with already existing knowledge on the financial background, the budgetary processes and their implementation in order to make any sense of the raw data at all. With recent technical developments and progresses with respect to information technologies, the possibilities to automate or – at least – facilitate this process are greater than ever. With the incorporation of Linked Data into a holistic technological approach for Linked Open Budget Data, it is possible to create budgetary information with self-defined semantics and meaning, so that machines as well as, ultimately, users are able to deduce information about known or potential relationships and dependencies in public finances that can be easily shown and understood by everybody. Thereby, a new level of transparency can be reached in this field of information science.

FIGURE 38 – The combination of thematic concepts shown within this Diploma Thesis

Source: Own illustration

Hence, the concepts of three thematically independent topics – as shown in **FIGURE 38** above – are combined in an innovative way to reach this specific goal in the very end. A possible approach to this combination related to local government finances specifically in Austria is given by this Diploma Thesis in its fourth chapter and the accompanying, respective prototype implementation. Therefore, also relevant requirements and peculiarities of Austria's governmental budget structure were incorporated. Of course, this approach was not chosen arbitrarily. Instead, the adoption and elaboration of this concept runs like a golden threat from the very beginning to the final end of the present work. Additionally, a four-stage, qualitative assessment is given in chapter 4.3 to ensure that the given implementation is really an appropriate application of Linked Open Budget Data that fulfils its purpose in an adequate way.

With that, an unprecedented exemplary approach for Austria on Open Budget Data in general and Linked Open Budget Data in particular was developed in the course of this work, providing a new level of best-practise and a state-of-the-art concept for future applications in this part of the world.

5.3 OPEN QUESTIONS & ISSUES RELATED TO LINKED OPEN BUDGET DATA

Naturally, as it can be expected with complex and diverse topics as Linked Open Budget Data shown within this work, an appropriate abstraction and concentration on certain parts has need to be made in order to condense it within the scope of this Diploma Thesis. Hence, there are of course certain questions and issues that could *not* be addressed – directly or indirectly – over the course of the past five chapters; although, some were explicitly pointed out as a matter of fact. In the following, these topics will be addressed collectively, but briefly as a final conclusion of this work.

First of all, the given approach obviously concentrates on eponymous local government finances in Austria. Other governmental levels like the ones of Federal States or the National State have been omitted on purpose; although, the principles of the given approach may be applied to them with ease in future works. Additionally, only the most important and major parts of municipal budgets were discussed, analysed and implemented as Linked Open Budget Data in the very end. Statements on assets, their deprecation or overall liabilities, which are also part of overall municipal documents on budgets, are still untouched – although, they are not less important. And, last but not least, future implementations of the Budgeting and Accounts regulation or VORANSCHLAGS- UND RECHNUNGSABSCHLUSSVERORDNUNG (like the VRV 2015) and succeeding regulations of the current Austrian Stability Pact or ÖSTERREICHISCHER STABILITÄTSPAKT will impose probably their own requirements on the underlying budgetary data; although, the data structure of the given approach was conceptualized accordingly, and allows respective adaptions in a straightforward way for future applications building upon it.

On another level, also the actual *quality* of Open Budget Data released by governmental authorities is crucial for the overall concept presented by this Diploma Thesis. To be precise, the amount, accuracy and structure of such data need to be consistent and available across states, or even countries – in the best case. As shown in the course of this work, this is *not* necessarily the case down to the present day. In fact, today in most cases more knowledge can be obtained via already existing and commonly used PDF releases on public finances, which prove to be more consistent through respective legislation and standing practice, in contrast to corresponding Open Budget Data. If municipalities would refrain from publishing their indeed quite structured, explanatory and informative PDF documents on public finances right now, transparency would *decrease* as a consequence. This should make oneself clear, that the pure existence of Open Budget Data does *not* state *anything* about its quality and a subsequent information gain. Beyond that, transparency could even be *hampered* by wrong implementations and insufficient budgetary data.

What is left out completely by this discussion – as well as the rest of this work – is the actual end-user-friendly *presentation* of such public finances. Of course, this would give sufficient content for one or multiple Diploma Theses. However, on the one hand, there exist already various such papers, which focus on several aspects related to this point, and furthermore, it was – in the author's view – out of scope for this Diploma Thesis as well as business informatics' primary applications in general.

In the end, it is the appropriate publication of budgetary data by the public administration, which enables and facilitates developments like Linked Open Budget Data in the first place. Currently, still *no* encompassing release of such data – especially, in a timely manner – has happened *quantitatively* and the small amount of already published data sets on governmental platforms is obviously varying in structure, consistence and even quality itself. An improvement through such approaches as given by this Diploma Thesis remains to be seen.

BIBLIOGRAPHY

[1] European Parliament & Council of the European Union, "Directive 2003/98/EC on the re-use of public sector information", *Official Journal of the European Union*, Volume 46, 17th November 2003.

[2] Republic of Austria, "Vereinbarung zwischen dem Bund, den Ländern und den Gemeinden über einen Österreichischen Stabilitätspakt 2012 (ÖStP 2012)", *BGBl. I Nr. 30/2013*, 23rd January 2013.

[3] United Nations Statistics Division, "Classification of the Functions of Government (COFOG)", 1999. Online available: http://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=4&Lg=1. Accessed on 8th December 2016.

[4] Rechnungshof, "Bericht des Rechnungshofes - Der abgestufte Bevölkerungsschlüssel", Publication series *Bund 2016/4*, Vienna, 13th April 2016.

[5] European Commission, "European system of accounts — ESA 2010", Luxembourg: Publications Office of the European Union, 2013.

[6] European Parliament & Council of the European Union, "Regulation (EU) No 549/2013 on the European system of national and regional accounts in the European Union", *Official Journal of the European Union*, Volume 56, 21st May 2013.

[7] Council of the European Union, "Regulation (EC) No 2223/96 on the European system of national and regional accounts in the Community", *Official Journal of the European Communities*, Volume 39, 25th June 1996.

[8] OECD, "OECD Glossary of Statistical Terms", Paris: OECD Publications, 2008.

[9] Republic of Austria, "Verordnung der Bundesministerin für Finanzen über die Statistik der Gebarung im öffentlichen Sektor (Gebarungsstatistik-VO 2014 - GebStat-VO 2014)", *BGBl. II Nr. 345/2013*, 14th November 2013.

[10] Statistik Austria, "Gemeindeeinheiten in Österreich - Bestandsaufnahme 2012", Direktion Volkswirtschaft, 2013.

[11] Statistik Austria, "Overview of general government and government controlled units according to ESA", March 2016. Online available: http://www.statistik.at/web_en/statistics/Economy/Public_finance_taxes/public_finance/public_sector/076255.html. Accessed on 14th February 2017.

[12] Republic of Austria, "Bundes-Verfassungsgesetz (B-VG)", StF: *BGBl. Nr. 1/1930* idF *BGBl. I Nr. 194/1999*, 1930.

[13] D. Wastl-Walter, "Gemeinden in Österreich: im Spannungsfeld von staatlichem System und lokaler Lebenswelt", Vienna: Böhlau Verlag, 2000.

[14] A. Parrer, "Finanzen managen in österreichischen Gemeinden: Handbuch für Politik & Verwaltung", Vienna: planetVERLAG, 2012.

[15] Republic of Austria, "Bundesverfassungsgesetz über die Regelung der finanziellen Beziehungen zwischen dem Bund und den übrigen Gebietskörperschaften (Finanz-Verfassungsgesetz 1948 - F-VG 1948)", StF: *BGBl. Nr. 45/1948* idF *BGBl. I Nr. 194/1999*, 1948.

[16] F. Klug, "Einführung in das kommunale Haushalts- und Rechnungswesen", *RFG - Recht und Finanzen für Gemeinden,* no. 4, 2011.

[17] Republic of Austria, "Voranschlags- und Rechnungsabschlussverordnung 2015 (VRV 2015)", *BGBl. II Nr. 313/2015*, 19th October 2015.

[18] J. Hiebler, "Analyse der finanziellen Situation in den steirischen Gemeinden: eine wirtschaftshistorische Studie", Graz: Uni-Press Graz Verlag, 2011.

[19] J. v. Lucke, C. P. Geiger, A. Hoose and M. Schreiner, "Open Budget 2.0 & Open Budget Data Öffnung von Haushaltswesen und Haushaltsdaten", zeppelin universität, Friedrichshafen, 2011.

[20] R. Holzer, "Budget und Rechnungsabschluss in meiner Gemeinde", Schwanenstadt: Verlag der Grünen Bildungswerkstatt OÖ, 2004.

[21] A. Enzinger and M. Papst, “Mittelfristige Finanzplanung in Gemeinden”, *RFG - Recht und Finanzen für Gemeinden,* no. 6, 2009.

[22] Republic of Austria, “98. Bundesgesetz, mit dem das Bundes-Verfassungsgesetz, das Rechnungshofgesetz 1948, das Verfassungsgerichtshofgesetz 1953 und das Verwaltungsgerichtshofgesetz 1985 geändert werden”*, BGBl. I Nr. 98/2010*, 14th December 2010.

[23] Rechnungshof, “Themen der öffentlichen Finanzkontrolle - Prüfkompetenz bei Gemeinden ab 10.000 Einwohnern”, Publication series *Bund 2010/14*, Vienna, 23rd December 2010.

[24] Verwaltungsgerichtshof, “Erkenntnis 93/01/0670”*,* 17th May 1995.

[25] Amt der niederösterreichischen Landesregierung, “Bescheid über die Auflösung des Gemeinderates von Deutsch-Wagram”*,* St. Pölten, 29th April 2009.

[26] K. Nedyalkov, P. Gartner and M. Mayer, “Gemeindefinanzbericht 2016”, Kommunalkredit Austria, Vienna, 5th December 2016.

[27] Republic of Austria, “Bundesgesetz mit dem ein Finanzausgleichsgesetz 2017 erlassen wird (Finanzausgleichsgesetz 2017 - FAG 2017)”*, BGBl. I Nr. 116/2016*, 30th December 2016.

[28] E. Hauth, “Aktuelle Regeln und Anwendungen des Haushaltswesens in den Bundesländern: Eine kritische Analyse”, Studie im Auftrag des Staatsschuldenausschusses, March 2009.

[29] M. Köfel, “Kennzahlen zum Voranschlags- und Rechnungsquerschnitt”, *Stabile Finanzwirtschaft in Gemeinden,* vol. 125, pp. 91-104, 2013.

[30] H.-J. Hörmann, “Das neue Haushaltsrecht der Gemeinden”, 2015. Online available: http://www.flgoe.at/content/wp-content/uploads/2015/11/VRV-2015_Ausblick.pdf. Accessed on 1st May 2017.

[31] Rechnungshof, “Themen der öffentlichen Finanzkontrolle - Die Voranschlags- und Rechnungsabschlussverordnung 2015”, Publication series *Bund 2015/18*, Vienna, 28th December 2015.

[32] V. Farré Capdevila, “Mehr Transparenz und Rechenschaftspflicht durch die VRV 2015”, March 2016. Online available: https://www.fiskalrat.at/workshops/fiskalregeln-fuer-laender-und-gemeinden.html. Accessed on 1st May 2017.

[33] J. Bröthaler, “Open Budget Data - Offenheit und Transparenz öffentlicher Haushalte in Österreich”, *Der öffentliche Sektor – The Public Sector,* no. 1/2015, pp. 29-46, 2015.

[34] Republic of Austria, “Bundesgesetz über die Bundesstatistik (Bundesstatistikgesetz 2000 - BStatG)”, StF: *BGBl. I Nr. 163/1999*, 2000.

[35] C. Rumpold, “Was bringt das neue Jahr?”, Kommunalnet, 20th December 2013. Online available: https://www.kommunalnet.at/news/einzelansicht/was-bringt-das-neue-jahr.html. Accessed on 5th May 2017.

[36] Statistik Austria, “Satzaufbau für die Lieferung der Gemeindehaushaltsdaten”*,* Direktion Volkswirtschaft: Version 3.7, 14th May 2014.

[37] Statistik Austria, “Satzaufbau für die Lieferung der Landeshaushaltsdaten”*,* Direktion Volkswirtschaft: Version 3.7, 14th May 2014.

[38] F. Bauer and M. Kaltenböck, “Linked Open Data: The Essentials”, Vienna: edition mono/monochrom, 2011.

[39] R. Schulmeister, “Hypertext - Geschichte, Systeme, Strukturmerkmale und Werkzeuge”, *Lehrbuch für Lernen und Lehren mit Technologien*, L3T, 2011.

[40] I. Herman, “W3C Semantic Web Frequently Asked Questions”, World Wide Web Consortium, 12th November 2009. Online available: https://www.w3.org/RDF/FAQ. Accessed on 14th May 2017.

[41] T. Berners-Lee, “Linked Data”, Design Issues, 27th July 2006. Online available: https://www.w3.org/DesignIssues/LinkedData.html. Accessed on 14th May 2017.

[42] C. Bizer, T. Heath and T. Berners-Lee, “Linked Data - The Story So Far”, *International Journal on Semantic Web and Information Systems,* vol. 5, no. 3, 2009.

[43] T. Berners-Lee, “A roadmap to the Semantic Web”, September 1998. Online available: https://www.w3.org/DesignIssues/Semantic.html. Accessed on 14th May 2017.

[44] T. Berners-Lee, J. Hendler and O. Lassila, “The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities”, *Scientific American,* vol. 284, no. 5, pp. 34-43, 2001.

[45] T. Berners-Lee, “Weaving the Web: The Past, Present and Future of the World Wide Web by its Inventor”, London: Orion Business Books, 1999.

[46] M. Hausenblas, “Exploiting Linked Data For Building Web Applications”, *IEEE Internet Computing,* vol. 13, no. 4, pp. 68-73, 2009.

[47] World Wide Web Consortium, “Facts About W3C”, Online available: https://www.w3.org/Consortium/facts. Accessed on 14th May 2017.

[48] T. Berners-Lee, “Semantic Web on XML - Architecture”, 6th December 2000. Online available: https://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html. Accessed on 13th May 2017.

[49] S. Bratt, “Semantic Web, and Other Technologies to Watch”, January 2007. Online available: https://www.w3.org/2007/Talks/0130-sb-W3CTechSemWeb/#(24). Accessed on 13th May 2017.

[50] M. Obitko, “Semantic Web Architecture”, 2007. Online available: http://obitko.com/tutorials/ontologies-semantic-web/semantic-web-architecture.html. Accessed on 13th May 2017.

[51] T. Berners-Lee, R. Fielding and L. Masinter, “RFC 3986 - Uniform Resource Identifier (URI): Generic Syntax”, Internet Engineering Task Force, 2005.

[52] The Unicode Consortium, “The Unicode Standard”, Version 9.0, Mountain View, CA, 2016.

[53] T. Bray, J. Paoli, M. Sperberg-McQueen, E. Maler and F. Yergeau, “Extensible Markup Language (XML) 1.0”, 5th ed., World Wide Web Consortium, 10th February 1998.

[54] M. Kifer and H. Boley, “RIF Overview”, World Wide Web Consortium (W3C), 22th June 2010. Online available: https://www.w3.org/TR/2010/NOTE-rif-overview-20100622/. Accessed on 25th May 2017.

[55] O. Lassila and R. R. Swick, “Resource Description Framework (RDF) Model and Syntax Specification”, World Wide Web Consortium (W3C), 22th February 1999. Online available: https://www.w3.org/TR/1999/REC-rdf-syntax-19990222/. Accessed on 26th May 2017.

[56] G. Klyne, J. J. Carroll and B. McBride, “RDF 1.1 Concepts and Abstract Syntax”, World Wide Web Consortium (W3C), 25th February 2014. Online available: https://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/. Accessed on 26th May 2017.

[57] D. L. McGuinness and F. van Harmelen, “OWL Web Ontology Language - Overview”, World Wide Web Consortium (W3C), 10th February 2004. Online available: https://www.w3.org/TR/owl-features/. Accessed on 26th May 2017.

[58] V. Miličić, “Problems of the RDF model: Blank Nodes”, Bew Citnames, 14th July 2011. Online available: http://milicicvuk.com/blog/2011/07/14/problems-of-the-rdf-model-blank-nodes/. Accessed on 26th May 2017.

[59] B. DuCharme, “Learnig SPARQL - Querying and Updating with SPARQL 1.1”, 1st ed., Sebastopol, CA: O'Reilly Media, Inc., 2011.

[60] T. Berners-Lee, “Primer: Getting into RDF & Semantic Web using N3”, 16th August 2005. Online available: https://www.w3.org/2000/10/swap/Primer.html. Accessed on 28th May 2017.

[61] D. Beckett and I. Herman, “RDF Primer — Turtle version”, World Wide Web Consortium (W3C), February 2007. Online available: https://www.w3.org/2007/02/turtle/primer/#L1995. Accessed on 28th May 2017.

[62] D. Beckett, T. Berners-Lee, E. Prud'hommeaux and G. Carothers, “RDF 1.1 Turtle - Terse RDF Triple Language”, World Wide Web Consortium (W3C), 25th February 2014. Online available: https://www.w3.org/TR/turtle/. Accessed on 27th May 2017.

[63] D. Brickley, R. Guha and A. Layman, “Resource Description Framework (RDF) Schemas”, World Wide Web Consortium (W3C), 9th April 1998. Online available: https://www.w3.org/TR/1998/WD-rdf-schema-19980409/. Accessed on 28th May 2017.

[64] D. Brickley and R. Guha, “RDF Schema 1.1”, World Wide Web Consortium (W3C), 25th February 2014. Online available: https://www.w3.org/TR/rdf-schema/. Accessed on 28th May 2017.

[65] D. Connolly, J. Hendler and G. Schreiber, “Web-Ontology (WebOnt) Working Group (Closed) - Schedule/Milestones”, World Wide Web Consortium (W3C), 15th June 2004. Online available: https://www.w3.org/2001/sw/WebOnt/#L151. Accessed on 28th May 2017.

[66] W3C OWL Working Group, “OWL 2 Web Ontology Language - Document Overview (Second Edition)”, World Wide Web Consortium (W3C), 11th December 2012. Online available: https://www.w3.org/TR/owl2-overview/. Accessed on 28th May 2017.

[67] E. Prud'hommeaux and A. Seaborne, “SPARQL Query Language for RDF”, World Wide Web Consortium (W3C), 12th October 2004. Online available: https://www.w3.org/TR/2004/WD-rdf-sparql-query-20041012/. Accessed on 31st May 2017.

[68] E. Prud'hommeaux and A. Seaborne, “SPARQL Query Language for RDF”, World Wide Web Consortium (W3C), 15th January 2008. Online available: https://www.w3.org/TR/rdf-sparql-query/. Accessed on 31st May 2017.

[69] The W3C SPARQL Working Group, “SPARQL 1.1 Overview”, World Wide Web Consortium (W3C), 21st March 2013. Online available: https://www.w3.org/TR/sparql11-overview/. Accessed on 31st May 2017.

[70] B. Hyland and B. Villazón Terrazas, “Linked Data Cookbook”, World Wide Web Consortium (W3C), December 2011. Online available: https://www.w3.org/2011/gld/wiki/Linked_Data_Cookbook. Accessed on 3rd June 2017.

[71] D. Ayers and M. Völkel, “Cool URIs for the Semantic Web”, World Wide Web Consortium (W3C), 3rd December 2008. Online available: https://www.w3.org/TR/cooluris/. Accessed on 3rd June 2017.

[72] R. Fielding and J. Reschke, “RFC 7231 - Hypertext Transfer Protocol (HTTP/1.1): Semantics and Content”, Internet Engineering Task Force, 2014.

[73] T. Heath and C. Bizer, “Linked Data: Evolving the Web into a Global Data Space”, 1st ed., Morgan & Claypool, 2011.

[74] P. Klimpel, “Freies Wissen dank Creative-Commons-Lizenzen”, Berlin: Wikimedia Deutschland , 2012.

[75] Open Knowledge International, “The Open Definition”, Online available: http://opendefinition.org/. Accessed on 17th Juni 2017.

[76] S. Amini, “Der freie Zugang zu wissenschaftlicher Information - Urheberrecht”, Open Access, Online available: https://open-access.net/AT-DE/informationen-zu-open-access/rechtsfragen/urheberrecht/. Accessed on 24th June 2017.

[77] J. Rosén, “Individualism and Collectiveness in Intellectual Property Law”, Cheltenham: Edward Elgar Publishing, 2012.

[78] Republic of Austria, “Bundesgesetz über das Urheberrecht an Werken der Literatur und der Kunst und über verwandte Schutzrechte (Urheberrechtsgesetz - UrhG)”, StF: *BGBl. Nr. 111/1936*, 1936.

[79] Republic of Austria, “Bundesgesetz über den Schutz personenbezogener Daten (Datenschutzgesetz 2000 - DSG 2000) ”, StF: *BGBl. I Nr. 165/1999*, 1999.

[80] European Parliament & Council of the European Union, “Regulation (EU) 2016/679 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)”, *Official Journal of the European Union*, Volume 59, 27th April 2016.

[81] Open Knowledge International, “Open Data Commons - Licences”, Online available: https://opendatacommons.org/licenses/. Accessed on 24th June 2017.

[82] Creative Commons, “State of Creative Commons - Data Notes & Sources”, 2017. Online available: https://stateof.creativecommons.org/data-and-acknowledgements-page/. Accessed on 24th June 2014.

[83] Creative Commons, “Downloads - Icons”, Online available: https://creativecommons.org/about/downloads/. Accessed on 16th June 2017.

[84] J. v. Lucke and C. P. Geiger, “Open Government Data – Frei verfügbare Daten des öffentlichen Sektors”, zeppelin universität, Friedrichshafen, 2010.

[85] B. Obama, "Memorandum on Transparency and Open Government", *Federal Register,* pp. 4685-4686, 26th January 2009.

[86] G. Eibl, J. Höchtl, B. Lutz, P. Parycek, S. Pawel and H. Pirker, "Framework for Open Government Data platforms", Project Group Open Government Data Austria, 23rd August 2013.

[87] E. Hödl, T. Rohrer and M. Zechner, "Open Data und Open Innovation in Gemeinden", *RFG - Recht und Finanzen für Gemeinden,* no. 5, 2015.

[88] European Commission, "Digital Agenda: Turning government data into gold", Press Release, 12th December 2011.

[89] Republic of Austria, "Bundesgesetz über die Weiterverwendung von Informationen öffentlicher Stellen (Informationsweiterverwendungsgesetz – IWG)", *BGBl. I Nr. 135/2005* & *BGBl. I Nr. 76/2015*, 18th November 2005 & 9th July 2015.

[90] Land Salzburg, "Gesetz mit dem das Gesetz über Auskunftspflicht, Datenschutz und Landesstatistik geändert wird", *LGBl Nr 69/2007*, 7th September 2007.

[91] B. Huber, A. Kurnikowski, S. Müller and S. Pozar, "Die wirtschaftliche und politische Dimension von Open Government Data in Österreich", Institute for Entrepreneurship & Innovation, Vienna University of Economics and Business, 2013.

[92] G. Vickery, "Review of Recent Studies on PSI-Re-use and Related Market Developments", Information Economics, Paris, 2011.

[93] R. Cyganiak, "Linked Open Data star badges", 7th December 2010. Online available: http://lab.linkeddata.deri.ie/2010/lod-badges/. Accessed on 28th June 2017.

[94] T. Davies, "What's in the Linked Open Data Stack?", 5th May 2011. Online available: http://www.opendataimpacts.net/2011/05/whats-in-the-linked-open-data-stack/. Accessed on 28th June 2017.

[95] M. Schmachtenberg, C. Bizer and H. Paulheim, "Adoption of the Linked Data Best Practices in Different Topical Domains", University of Mannheim, Research Group Data and Web Science, 2014.

[96] S. Auer, I. Ermilov, J. Lehmann and M. Martin, "LODStats", Universität Leipzig, 2016. Online available: http://stats.lod2.eu/. Accessed on 29th June 2017.

[97] Open Knowledge International, "Global Open Data Index", Online available: https://index.okfn.org/. Accessed on 29 July 2017.

[98] International Budget Partnership, "International Budget Partnership - History", Online available: http://www.internationalbudget.org/who-we-are/history/. Accessed on 29th July 2017.

[99] International Budget Partnership, "Open Budget Survey 2015", Washington D.C., 2015.

[100] International Budget Partnership, "Open Budget Survey: 2017 Update", Online available: http://www.internationalbudget.org/opening-budgets/open-budget-initiative/open-budget-survey/2017-news/. Accessed on 29 July 2017.

[101] Open Knowledge International, "OpenSpending", 2017. Online available: https://openspending.org/. Accessed on 9th August 2017.

[102] R. Pollock, P. Walsh and A. Kariv, "OpenSpending Next", Open Knowledge International, 2017. Online available: http://community.openspending.org/next/. Accessed on 9th August 2017.

[103] K. Höffner, M. Martin and J. Lehmann, "LinkedSpending: OpenSpending becomes Linked Open Data", *Semantic Web,* vol. 7, no. 1, pp. 95-104, 2016.

[104] OpenSpending Tech Team, "OpenSpending Next", Open Knowledge International, 2017. Online available: http://community.openspending.org/next/. Accessed on 9th August 2017.

[105] K. Höffner, M. Martin and J. Lehmann, "LinkedSpending", 2014. Online available: http://linkedspending.aksw.org. Accessed on 9th August 2017.

[106] M. F. Gerli, "Tracking Federal Funds: USAspending.gov and Other Data Sources", Congressional Research Service, Washington D.C., 13th May 2015.

[107] Office of Management & Budget, "Budget", 2017. Online available: https://www.whitehouse.gov/omb/budget/. Accessed on 10th August 2017.

[108] Office of Management & Budget, Budget Analysis Branch, “Public Budget Database User’s Guide”, May 2017.

[109] Office of Management and Budget, “USAspending.gov”, Online available: https://www.USAspending.gov. Accessed on 10th August 2017.

[110] Texas Comptroller of Public Accounts, “Texas Transparency”, Online available: http://www.texastransparency.org. Accessed on 10th August 2017.

[111] Publications Office of the European Union, “Budget on line”, Online available: http://eur-lex.europa.eu/budget/www/index-en.htm. Accessed on 11th August 2017.

[112] European Commission, “Commission Decision on the reuse of Commission documents”, *Official Journal of the European Union*, Volume 54, 12th December 2011.

[113] Publications Office of the European Union, “European Union Open Data Portal”, Online available: https://data.europa.eu/euodp/. Accessed on 11th August 2017.

[114] Interoperability Solutions for European Public Adminstrations (ISA) Programme of the European Commission, “Controlled Vocabularies for EU Budget”, 8th August 2016. Online available: http://joinup.ec.europa.eu/site/core_vocabularies/registry/eubudget/. Accessed on 11th August 2017.

[115] E. Druel, P. Chrzanowski, R. Pollock and J. Gray, “Where Does Europe’s Money Go? - A Guide to EU Budget Data Sources”, Open Knowledge International, Cambridge, July 2017.

[116] The National Archives, “Open Government Licence - Version 3”, 31st October 2014. Online available: http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/. Accessed on 12th August 2017.

[117] HM Treasury, “Combined Online Information System”, 2010-2013. Online available: https://data.gov.uk/dataset/coins. Accessed on 12th August 2017.

[118] HM Treasury, “HMT OSCAR: publishing from the database”, 2013-2017. Online available: https://www.gov.uk/government/collections/hmt-oscar-publishing-from-the-database. Accessed on 12th August 2017.

[119] HM Treasury, “Online System for Central Accounting and Reporting - OSCAR”, 2013. Online available: https://data.gov.uk/dataset/oscar. Accessed on 12th August 2017.

[120] I. Dickinson, “COINS as Linked Data”, HM Treasury, 11th November 2010. Online available: https://data.gov.uk/resources/coins. Accessed on 12th August 2017.

[121] HM Treasury, “Understanding the OSCAR Data Release 2016”, 17th November 2016.

[122] HM Treasury, “HM Treasury”, Online available: https://www.gov.uk/government/organisations/hm-treasury. Accessed on 12th August 2017.

[123] Open Knowledge International, “Where Does My Money Go?”, Online available: http://app.wheredoesmymoneygo.org/. Accessed on 9th November 2016.

[124] Bundesministerium der Finanzen, “Bundeshaushalt-Info.de”, 2012-2017. Online available: https://www.bundeshaushalt-info.de/. Accessed on 14th August 2017.

[125] Finanzbehörde Hamburg, “GovData - Das Datenportal für Deutschland”, 2016. Online available: https://www.govdata.de/. Accessed on 14th August 2017.

[126] Austrian Ministry of Finance & City of Vienna, “data.gv.at - offene Daten Österreichs”, 2012-2017. Online available: https://www.data.gv.at/. Accessed on 16th August 2017.

[127] M. Kaltenböck, S. Härtig, J. Jakobitsch and T. Thurner, “LODpilot”, Semantic Web Company GmbH, 16th October 2014. Online available: https://lodpilot.at/. Accessed on 16th August 2017.

[128] Austrian Ministry of Finance, “Das Budget”, Online available: https://www.bmf.gv.at/budget/das-budget/das-budget.html. Accessed on 16th August 2017.

[129] B. Krabina, C. Hödl, M. Köfel, P. Biwald, R. Titl and M. Wallisch, “OffenerHaushalt.at”, KDZ - Zentrum für Verwaltungsforschung & Interactives, 2013-2017. Online available: https://www.offenerhaushalt.at. Accessed on 15th August 2017.

[130] B. Krabina, “data.offenerhaushalt.at - Endbericht”, KDZ - Zentrum für Verwaltungsforschung, Vienna, 2017.

[131] P. Vassiliadis and A. Simitsis, “Extraction, transformation, and loading”, *Encyclopedia of Database Systems*, New York, Springer US, 2009, pp. 1095-1101.

[132] SDMX, “The official site for the SDMX community”, Online available: https://sdmx.org/. Accessed on 5th September 2017.

[133] R. Cyganiak and D. Reynolds, “The RDF Data Cube Vocabulary”, World Wide Web Consortium (W3C), 16th January 2014. Online available: https://www.w3.org/TR/vocab-data-cube/. Accessed on 5th September 2017.

[134] The Apache Software Foundation, “Apache Jena”, 2017. Online available: http://jena.apache.org/. Accessed on 3rd November 2017.

[135] R. Cyganiak and C. Bizer, “Pubby - A Linked Data Frontend for SPARQL Endpoints”, 2007-2014. Online available: http://wifo5-03.informatik.uni-mannheim.de/pubby. Accessed on 4th November 2017.

[136] DBpedia, “Learn more about DBpedia - The DBpedia Data Provision Architecture”, 2017. Online available: http://wiki.dbpedia.org/about/architecture. Accessed on 4th November 2017.

[137] EUROSTAT, “GDP and main components (output, expenditure and income)”, 26th June 2017. Online available: http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=nama_10_gdp&lang=en. Accessed on 27th June 2017.

[138] Republic of Austria, “Inkrafttreten der Vereinbarung zwischen dem Bund, den Ländern und den Gemeinden über einen Österreichischen Stabilitätspakt 2012 für das Land Salzburg”, *BGBl. I Nr. 45/2013*, 8th March 2013.

[139] Dublin Core Metadata Initiative (DCMI), “Dublin Core Metadata Element Set - Version 1.1”, 14th June 2012. Online available: http://dublincore.org/documents/dces/. Accessed on 26th May 2017.

[140] Open Knowledge International, “Open Knowledge International Blog – our new name!”, 23rd May 2016. Online available: https://blog.okfn.org/2016/05/23/open-knowledge-international-our-new-name/. Accessed on 17th June 2017.

[141] European Parliament & Council of the European Union, “Directive on the protection of individuals with regard to the processing of personal data and on the free movement of such data”, *Official Journal of the European Union*, Volume 38, 24th October 1995.

[142] S. Fuchs, C. Jünger, G. Kasper, C. Linhart and T. Walch, “Open Government Data – Offene Daten für Österreich”, Wirtschaftsuniversität Wien, Vienna, 14th May 2013.

[143] R. Cyganiak and A. Jentzsch, “The Linking Open Data cloud diagram”, 30th August 2014. Online available: http://lod-cloud.net/. Accessed on 3rd February 2017.

[144] Microsoft, “Excel specifications and limits”, Online available: https://support.office.com/en-us/article/Excel-specifications-and-limits-1672b34d-7043-467e-8e27-269d656771c3. Accessed on 13th August 2017.

[145] State government of Lower Austria, “NÖ Gemeindeordnung 1973”, StF: *LGBl. 1000-0* idF *LGBl. 1000-22*, 18th November 2013.

[146] State government of Carinthia, “Kärntner Allgemeine Gemeindeordnung”, StF: *LGBl Nr 66/1998* idF *LGBl. Nr. 3/2015*, 14th January 2015.

[147] L. Rietveld, “Yet Another Sparql GUI”, 2017. Online available: http://doc.yasgui.org/. Accessed on 5th November 2017.

APPENDIX

Glossary for technical vocabulary in the fields of municipal budgets

In particular, the terminology for chapter 2.2 (and subsequent ones basing on it) contains technical vocabulary, wherefore – in most parts – no common or even official English translation exist. For continuous usage within this Diploma Thesis, respective ones were made and listed at this point with their respective German counterparts, as they may not be clear even for experts immediately.

English	German
advancing funds	Vorschussgelder
association of local authorities	Gemeindeverband
Association of Towns and Municipalities	Städte- und Gemeindebund
Austrian Stability Pact	Österreichischer Stabilitätspakt
autonomous economic body	selbstständiger Wirtschaftskörper
budget	Rechnungsvoranschlag
budget affecting accounting	voranschlagswirksame Gebarung
budget ineffective accounting	voranschlagsunwirksame Gebarung
budget profile	Rechnungsquerschnitt
Budgeting and Accounts Regulation	Voranschlags- und Rechnungsabschlussverordnung
Constitutional Court	Verfassungsgerichtshof
contingent liabilities	Eventualverbindlichkeiten
Court of Audit	Rechnungshof
double-entry bookkeeping (system)	Doppik / doppelte Buchführung
extraordinary budget	außerordentlicher Haushalt
Federal Chancellery	Bundeskanzleramt
Federal Constitutional Law	Bundesverfassungsgesetz
federal state	Bundesland
fees and charges	Gebühren und Abgaben
finance statistic	Gebarungsstatistik
fiscal equalization	Finanzausgleich
fiscal management	Haushaltsführung
funding allocation	Bedarfszuweisung
generally accepted accounting principles	Grundsätze ordnungsmäßiger Buchführung
government commissioner	Regierungskommissär
grants to the State Government	Landesumlagen
housing subsidies	Wohnbauförderung
local council	Gemeinderat
local government (level)	Kommunalebene
local tax	Ortstaxe
municipal	kommunal
municipal code	Gemeindeordnung
municipal tax	Kommunalsteuer
municipality	Gemeinde
municipality code	Gemeindekennziffer
non-profit institutions serving households	Organisationen ohne Erwerbszweck
ordinary budget	ordentlicher Haushalt
primary residence	Hauptwohnsitz
property tax	Grundsteuer
public budget	Öffentlicher Haushalt
public sector	Öffentlicher Sektor
reserves	Rücklagen
service revenues	Leistungserlöse
shares of common taxes	Ertragsanteile
single-entry bookkeeping (system)	Kameralistik
statement of account	Rechnungsabschluss
supplementary budget	Nachtragsvoranschlag
Supreme Administrative Court	Verwaltungsgerichtshof
territorial authority	Gebietskörperschaft
transitory funds	Verwahrgelder
uniform municipality	Einheitsgemeinde

Economic benefit of PSI within the European Union & Austria incl. projections

Basic figures

Year / Data	GDP (EU-28)	%Δ GDP (EU-28)	GDP (Austria)	%Δ GDP (Austria)	est. PSI market growth
2008	13.055		0.292		7%
2009	12.298	-5.80%	0.286	-1.97%	7%
2010	12.818	4.23%	0.295	2.95%	7%
2011	13.193	2.93%	0.309	4.75%	7%
2012	13.449	1.94%	0.317	2.75%	7%
2013	13.559	0.82%	0.323	1.71%	7%
2014	14.011	3.33%	0.330	2.44%	7%
2015	14.728	5.12%	0.340	2.87%	7%
2016	14.825	0.65%	0.349	2.78%	7%

Projection of the overall PSI potential

Year / Data	overall PSI potential (EU) GDP projection	overall PSI potential (AT) GDP projection	overall PSI potential (EU) PSI market growth projection	overall PSI potential (AT) PSI market growth projection
2008	**200.000**	4.472	**200.000**	4.472
2009	188.398	4.384	214.000	4.384
2010	196.362	4.514	228.980	4.691
2011	202.111	4.728	245.009	5.020
2012	206.033	4.858	262.159	5.371
2013	207.718	4.941	280.510	5.747
2014	214.644	5.062	300.146	6.149
2015	225.636	5.207	321.156	6.580
2016	**227.112**	**5.352**	**343.637**	**7.040**

Projection of direct PSI potentials

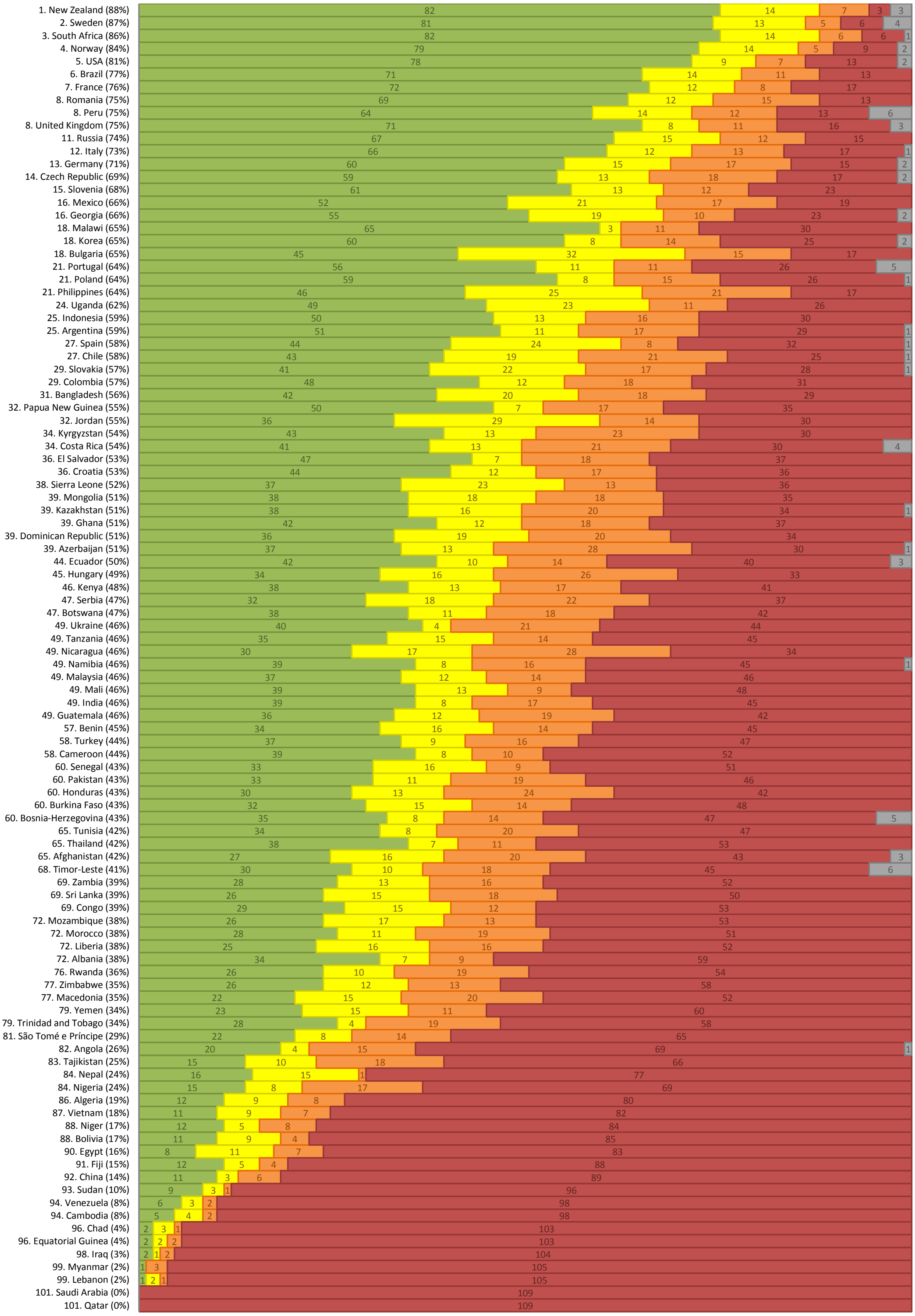
Year / Data	min direct PSI potential (EU) GDP projection	min direct PSI potential (AT) GDP projection	min direct PSI potential (EU) PSI market growth projection	min direct PSI potential (AT) PSI market growth projection	avg direct PSI potential (EU) GDP projection	avg direct PSI potential (AT) GDP projection	avg direct PSI potential (EU) PSI market growth projection	avg direct PSI potential (AT) PSI market growth projection	max direct PSI potential (EU) GDP projection	max direct PSI potential (AT) GDP projection	max direct PSI potential (EU) PSI market growth projection	max direct PSI potential (AT) PSI market growth projection
2009	**35.300**	0.821	**35.300**	0.821	**40.000**	0.931	**40.000**	0.931	**47.100**	1.096	**47.100**	1.096
2010	36.792	0.846	37.771	0.879	41.691	0.958	42.800	0.996	49.091	1.128	50.397	1.173
2011	37.869	0.886	40.415	0.941	42.912	1.004	45.796	1.066	50.528	1.182	53.925	1.255
2012	38.604	0.910	43.244	1.006	43.744	1.031	49.002	1.140	51.509	1.215	57.700	1.343
2013	38.920	0.926	46.271	1.077	44.102	1.049	52.432	1.220	51.930	1.235	61.738	1.437
2014	40.218	0.948	49.510	1.152	45.572	1.075	56.102	1.306	53.662	1.265	66.060	1.537
2015	42.277	0.976	52.976	1.233	47.906	1.106	60.029	1.397	56.410	1.302	70.684	1.645
2016	**42.554**	**1.003**	**56.684**	**1.319**	**48.220**	**1.136**	**64.231**	**1.495**	**56.778**	**1.338**	**75.632**	**1.760**

All figures are in Billions of Euros (except explicitly declared percent values).
All values for GDP of the EU-28 and Austria in particular are originating from EUROSTAT; all values for est. PSI Market growth and direct & overall potential at the level of 2008/2009 from Vickery's study for the European Commission.
The most important figures, some of which are referred by this work in chapter 2.4.2, are printed in bold type.

Sources: EUROSTAT [137], Vickery [92, pp. 37-38], own illustrations & calculations

OPEN BUDGET INDEX 2015 – Overall Ranking

Below the ranking of 2015's OPEN BUDGET INDEX for all 102 surveyed countries is shown incl. a classification of all of the 109 answers which lead to this very ranking (green = full standard met or exceeded, yellow = standard met basically, orange = standard met minimally, red = standard not met at all, grey = not applicable).



Source: International Budget Partnership [99]

Data Structure Definition for Austrian municipal LOBD

In the following, all the basic RDF data for the exemplary approach for real LOBD on public finances for Austrian municipalities as discussed in chapter 4 is given in RDF's "Turtle" syntax (see chapter 2.3.2). Please note, that all code definitions within the `http://data.lombuda.at/ontology/code/`-namespace are skipped *on purpose*. That's simply due to the lack of available space within this Annex – e.g. the full RDF approach code list definition according to the VRV with the mandatory first three decades measures more than 3000 lines of text! The interested reader is referred to the website of the technical prototype or its respective code repository for the full RDF definitions, their data source as well as the technical prototype itself, which are available at:

HTTP://LOMBUDA.AT

&

HTTP://GITLAB.COM/AXIO/LOMBUDA

However, in the following, all commonly used namespace prefixes, the data set definition, its data structure definition itself as well as all data slice and property definitions are listed, nevertheless. This should give an appropriate overview of the created vocabulary anyway.

Common Namespace Prefixes

```
@prefix rdf:        <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs:       <http://www.w3.org/2000/01/rdf-schema#> .
@prefix owl:        <http://www.w3.org/2002/07/owl#> .
@prefix xsd:        <http://www.w3.org/2001/XMLSchema#> .
@prefix skos:       <http://www.w3.org/2004/02/skos/core#> .
@prefix dct:        <http://purl.org/dc/terms/> .
@prefix cc:         <http://creativecommons.org/ns#> .

@prefix qb:         <http://purl.org/linked-data/cube#> .

@prefix sdmx-concept:     <http://purl.org/linked-data/sdmx/2009/concept#> .
@prefix sdmx-dimension:   <http://purl.org/linked-data/sdmx/2009/dimension#> .
@prefix sdmx-attribute:   <http://purl.org/linked-data/sdmx/2009/attribute#> .

@prefix dbo:              <http://dbpedia.org/ontology/> .
@prefix dbr:              <http://dbpedia.org/resource/> .
@prefix gn:               <http://www.geonames.org/ontology#> .

@prefix lombuda:          <http://data.lombuda.at/ontology/> .
@prefix lombuda-slice:    <http://data.lombuda.at/ontology/slice/> .
@prefix lombuda-code:     <http://data.lombuda.at/ontology/code/> .
@prefix lombuda-fs-code:  <http://data.lombuda.at/ontology/code/fiscal-states/> .
@prefix lombuda-ft-code:  <http://data.lombuda.at/ontology/code/fiscal-types/> .
@prefix lombuda-vrv-code: <http://data.lombuda.at/ontology/code/VRV/> .
@prefix lombuda-bi-code:  <http://data.lombuda.at/ontology/code/VRV/budgetary-indicators/> .
@prefix lombuda-fc-code:  <http://data.lombuda.at/ontology/code/VRV/approaches/> .
@prefix lombuda-ec-code:  <http://data.lombuda.at/ontology/code/VRV/accounts/> .
@prefix lombuda-bp-index: <http://data.lombuda.at/ontology/code/VRV/budget-profile-indices/> .
@prefix lombudata:        <http://data.lombuda.at/> .
```

Data Set Definition

```
lombudata:OGHD      a       qb:DataSet;
    rdfs:label              "Linked Open Municipal Budget Data"@en ;
    rdfs:comment            "This is an exemplary data set containing information on Austrian
                             municipal budgets as Linked Open Data. It is represented in a respective
                             Linked Data ontology, which is adapted to Austrian specifics and defined
                             in the same way as this very dataset on this web server as Data Structure
                             Definition in the RDF Data Cube Vocabulary."@en ;
    dct:creator             "Paul Blasl"^^xsd:string ;
    dct:issued              "2017-10-31"^^xsd:date ;
    cc:licence              <http://opendatacommons.org/licenses/by/1.0/> ;
    cc:attributionName      "LOMBuDa.at"^^xsd:string ;
    cc:attributionURL       <http://lombuda.at> ;
    dct:subject             sdmx-subject:2.5 ; # Government finance statistics
    qb:structure            lombuda:dsd ;
    lombuda:currency        dbr:Euro ;
    sdmx-attribute:unitMult sdmx-code:unitMult-0
    .
```

Data Structure Definition

```
lombuda:dsd      a       qb:DataStructureDefinition ;
   rdfs:label            "Linked Open Municipal Budget Data Structure Definition"@en ;
   rdfs:comment          "The data structure definition (DSD) for Austrian municipal budget data
                          as Linked Open Data."@en ;
   dct:creator           "Paul Blasl"^^xsd:string ;
   dct:issued            "2017-09-19"^^xsd:date ;
   dct:modified          "2017-10-26"^^xsd:date ;
   cc:licence            <http://opendatacommons.org/licenses/by/1.0/> ;
   cc:attributionName    "LOMBuDa.at"^^xsd:string ;
   cc:attributionURL     <http://lombuda.at> ;

   qb:component

   # The dimensions
      [ qb:dimension     lombuda:fiscalYear;    qb:componentAttachment qb:Slice; qb:order 1;
        rdfs:label       "Time Dimension"@en ],
      [ qb:dimension     lombuda:municipality; qb:componentAttachment qb:Slice; qb:order 2;
        rdfs:label       "Spatial Dimension"@en ],
      [ qb:dimension     lombuda:fiscalState;   qb:componentAttachment qb:Slice; qb:order 3;
        rdfs:label       "Fiscal State Dimension"@en ],
      [ qb:dimension     lombuda:fiscalType;                          qb:order 4;
        rdfs:label       "Fiscal Type Dimension"@en ],
      [ qb:dimension     lombuda:functionalClassification; qb:order 5;
        rdfs:label       "Functional Dimension"@en ],
      [ qb:dimension     lombuda:economicClassification;   qb:order 6;
        rdfs:label       "Economic Dimension"@en ];

   # The measurand
   qb:component [ qb:measure lombuda:amount ; rdfs:label "The measurement: amount of money"@en ];

   # The attributes
   qb:component [ qb:attribute lombuda:currency ; rdfs:label "Currency attribute"@en ;
                   qb:componentRequired "true"^^xsd:boolean ; qb:componentAttachment qb:DataSet ] ,
                 [ qb:attribute sdmx-attribute:unitMult ; rdfs:label "Unit Multiplier attribute"@en ;
                   qb:componentRequired "true"^^xsd:boolean ; qb:componentAttachment qb:DataSet ] ,
                 [ qb:attribute cc:license ; rdfs:label "Licence attribute"@en ;
                   qb:componentRequired "true"^^xsd:boolean ; qb:componentAttachment qb:DataSet ] ,
                 [ qb:attribute cc:attributionName ; rdfs:label "Attribution (Name) attribute"@en ;
                   qb:componentRequired "true"^^xsd:boolean ; qb:componentAttachment qb:DataSet ] ,
                 [ qb:attribute cc:attributionURL ; rdfs:label "Attribution (URL) attribute"@en ;
                   qb:componentRequired "true"^^xsd:boolean ; qb:componentAttachment qb:DataSet ] ,
                 [ qb:attribute dct:source ; rdfs:label "Data source attribute"@en ;
                   qb:componentRequired "false"^^xsd:boolean ; qb:componentAttachment qb:Slice ] ,
                 [ qb:attribute lombuda:budgetaryIndicator ; rdfs:label "Budgetary Indicator
                   attribute"@en ; qb:componentRequired "false"^^xsd:boolean ] ,
                 [ qb:attribute lombuda:description ; rdfs:label "Description attribute"@en ;
                   qb:componentRequired "false"^^xsd:boolean ] ;

   # The slices
   qb:sliceKey lombuda-slice:fiscalDoc ;
   .
```

Data Slice Definition

```
lombuda-slice:fiscalDoc a qb:SliceKey ;
   rdfs:label              "Data Slice by Fiscal Documents"@en ;
   rdfs:comment            "Slice by grouping fiscal values of one specific budgetary document together,
                            fixing municipality and time value"@en ;
   qb:componentProperty lombuda:fiscalYear, lombuda:municipality, lombuda:fiscalState ;
   .
```

Definition of all dimension properties

```
lombuda:fiscalYear  a   rdf:Property, qb:DimensionProperty ;
    rdfs:label          "Reference fiscal year"@en ;
    rdfs:range          xsd:gYear ;
    rdfs:subPropertyOf  sdmx-dimension:refPeriod ;
    qb:concept          sdmx-concept:refPeriod
    .

lombuda:municipality a  rdf:Property, qb:DimensionProperty ;
    rdfs:label          "Reference municipality"@en ;
    rdfs:comment        "A reference to the municipality at GeoNames to which the referenced data
                         applies to incl. some textual representation of its name and municipality
                         code"@en ;
    rdfs:range          gn:Feature, xsd:String ;
    rdfs:subPropertyOf  sdmx-dimension:refArea ;
    qb:concept          sdmx-concept:refArea
    .

lombuda:fiscalState  a  rdf:Property, qb:CodedProperty, qb:DimensionProperty ;
    rdfs:label          "Respective fiscal state of the measurement"@en ;
    rdfs:range          lombuda-code:FiscalStateCode ;
    qb:codeList         lombuda-code:fiscal-states ;
    qb:concept          sdmx-concept:obsStatus
    .

lombuda:fiscalType  a   rdf:Property, qb:CodedProperty, qb:DimensionProperty ;
    rdfs:label          "Respective fiscal type of the measurement"@en ;
    rdfs:range          lombuda-code:FiscalTypeCode ;
    qb:codeList         lombuda-code:fiscal-types ;
    qb:concept          sdmx-concept:recording
    .

lombuda:functionalClassification a rdf:Property, qb:CodedProperty, qb:DimensionProperty ;
    rdfs:label          "Respective functional classification of the measurement"@en ;
    rdfs:range          lombuda-code:FunctionalClass ;
    qb:codeList         lombuda-vrv-code:approaches ;
    qb:concept          sdmx-concept:classSystem
    .

lombuda:economicClassification a rdf:Property, qb:CodedProperty, qb:DimensionProperty ;
    rdfs:label          "Respective economic classification of the measurement"@en ;
    rdfs:range          lombuda-code:EconomicClass ;
    qb:codeList         lombuda-vrv-code:accounts ;
    qb:concept          sdmx-concept:classSystem
    .
```

Definition of the measurand

```
lombuda:amount    a     rdf:Property, qb:MeasureProperty ;
    rdfs:label          "Recorded amount of cash flow"@en ;
    rdfs:subPropertyOf  sdmx-concept:obsValue, rdf:value ;
    rdfs:range          xsd:decimal
    .
```

Definition of all other attributes

```
lombuda:currency   a    rdf:Property, qb:AttributeProperty ;
    rdfs:label          "Currency of the measurement"@en ;
    rdfs:range          dbo:Currency ;
    qb:concept          sdmx-concept:unitMeasure, sdmx-concept:currency
    .

lombuda:budgetaryIndicator a rdf:Property, qb:AttributeProperty ;
    rdfs:label          "Budgetary indicator of the measurement"@en ;
    rdfs:range          lombuda-code:BugdetaryIndicator ;
    qb:codeList         lombuda-vrv-code:budgetary-indicators ;
    qb:concept          sdmx-concept:classSystem
    .

lombuda:description  a  rdf:Property, qb:AttributeProperty, rdfs:label ;
    rdfs:label          "Description of the measurement"@en ;
    rdfs:comment        "A human readable description of the measurement
                         (e.g. transaction descriptor or posting text)"@en ;
    rdfs:range          xsd:string ;
    qb:concept          sdmx-concept:dataDescr ;
    owl:sameAs          rdfs:label
    .
```